JSEALS
Journal of the Southeast Asian Linguistics Society

Managing Editor:
Paul Sidwell (Pacific Linguistics, Canberra)

Editorial Advisory Board:
Mark Alves (USA)
George Bedell (Thailand)
Marc Brunelle (Canada)
Gerard Diffloth (Cambodia)
Marlys Macken (USA)
Brian Migliazza (USA)
Keralapura Nagaraja (India)
Peter Norquest (USA)
Amara Prasithrathsint (Thailand)
Martha Ratliff (USA)
Sophana Srichampa (Thailand)
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 (jseals.org).

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 this publication. As an additional service we also admit data papers, reports and notes, subject to an internal review process.

JSEALS is published twice a year. Papers can be submitted to the Managing Editor, electronically (paul.sidwell@anu.edu.au or paulsidwell@yahoo.com) at any time.
Contents

Editorial iv

Papers

A Cognitive Linguistics Storyline Analysis of Iu-Mien Narrative Discourse
_T. Daniel Arisawa_ 1

Fixed Autosegmentism in Thai Emphatic Reduplication.
_Lezlie Lee_ 41

Sorbung, an Undocumented Language of Manipur: its Phonology and Place in
Tibeto-Burman
_David Mortenson and Jennifer Keogh_ 64

Floating Quantifiers in Burmese and Thai
_Andrew Simpson_ 115

The Evolution of Polyfunctionality of dây Construction in Thai: Split Patterns of
Possibility-Related Modal Concepts
_Kiyoko Takahashi_ 147
Welcome to JSEALS Volume 4.1. We are now into our forth year and I can only report that our journal is humming along nicely; the quality of papers is excellent, and I am pleased to relate that there is quite a queue of papers shaping up for 4.2.

This volume also sees abstracts and keywords for all papers, reflecting the ongoing process of improving standards and, importantly, accessibility. And in this regard, I urge you please to check the updated Guide for Authors on the last page - things run so much smoother when authors strictly follow the submission guidelines.

As issue 4.1 is being released, the 2011 SEALs meeting is about to get underway. The host institution - Kasetsart University in Bangkok - is offering beautiful modern meeting rooms in a great campus setting. Acharn Kitima Indrambarya, Chair of the Organizing Committee, is doing a superb job, and we look forward to a very memorable meeting.

Finally, I would like to pay homage to Uri Tadmor, who recently stepped down from the SEALs International Committee and the JSEALS Editorial Advisory Board. Uri, now working as an Editorial Director for De Gruyter Mouton, has been a supporter of SEALs over many years, and we especially acknowledge the great job he did taking on and running the 16th SEALs meeting at Atma Jaya University, Jakarta, Indonesia, in 2006. Thank you Uri.

Paul Sidwell (Managing Editor)
May 2011
A COGNITIVE LINGUISTICS STORYLINE ANALYSIS OF IU-MIEN NARRATIVE DISCOURSE

T. Daniel Arisawa
Chiangrai Rajabhat University, Thailand
<danielta928@hotmail.com>

Abstract
A storyline is a foregrounded mainline of development in narrative discourse. Taking the narrative as a mental picture viewed by a hearer/reader (i.e. conceptualizer), the Cognitive Linguistics posits two groups of constructions which make storyline emerge to the foreground: progression and sequential structures.

The storyline progression in Iu-Mien narrative is encoded with the unmarked action/motion verbs, the development adverb aengx ‘furthermore’, Serial Verb Construction, topic chains, and the multiclausal constructions. The sequential structures include the prospective conjunction ziouc ‘and then (soon)’, the retrospective conjunction cinx_daaih ‘therefore’, the topic marker aeqv ‘as for’, and the ad-clause containing V-liuz ‘after V-ing’.

Keywords: discourse, cognitive linguistics, text analysis

0. Introduction
A storyline is a foregrounded mainline of development in narrative discourse. This study seeks to investigate a storyline in Iu-Mien narrative discourse. The purpose is to show what it means that the storyline is foregrounded. The approach taken here is Cognitive Linguistics (CL), particularly Cognitive Grammar (CG) theorized by Ronald W. Langacker (e.g. 1987, 1991a, 1991b, 1991c, 2000), rather than a traditional and popular textlinguistics developed by Robert E. Longacre (e.g. 1981, 1983, 2003a, 2003b). Though Longacre’s theory was much influenced by Hopper and Thompson’s correlation between high/low transitivity and the foreground/background in discourse, he has never elaborated the relation.

Longacre emphasizes (to simplify various the argume nts) that a discovery the particular verb which has a morphosyntactic marking for the past tense or perfective aspect inevitably leads a researcher to identify the storyline. Iu-Mien, however, totally lacks such a marking in the verb; hence requires an alternative method. Following Somsonge’s non-verb-morphology dependent storyline identification method (i.e. temporal movement approach) in Thai (1990, 1992) and Hlai (2002), it turns out that the storyline in Iu-Mien can be analyzed in terms of two major factors: the chronological forward movement of action/event (e.g. topic chains and serial verb constructions, etc), and a bundle of the

---

1 I would like to thank the anonymous reviewers for their remarks, and also to thank Paul Sidwell the managing editor for his invaluable comments for improvement of my paper, editorial advice and his patience and kindness. I take full responsibility for any flaws found in this paper.
linguistic expressions indicating sequentiality (e.g. a developmental adverb and an inchoative conjunction, etc.) in the narrative. In this study, the former is referred to as “progression”, and the latter “sequential structure”, owing Talmy (2000).

The method in terms of progression and sequential structure explains cognitive linguistically why the storyline is foregrounded. From the CL perspective, a narrative is taken as a mental picture unfolded in front of the hearer/reader’s eyes. The narrative’s storyline is viewed as a thick, noticeable line that stands out (i.e. foregrounding) to the visual perception in the hearer/reader’s mental space as a result of combining two major factors: progression and sequential structure of the story.

Though seven story texts of Iu-Mien (over 700 sentences) were analyzed in search of storyline, among other discourse features, only one sample story is appended in this paper. Some examples are drawn from other sources.

1. Departure from Longacre’s theory
Longacre’s storyline theory within the general framework of textlinguistics is based on the discourse analyses of some eighty languages of the world. It is significantly influenced by his over fifty year study of Biblical Hebrew and Hopper and Thompson’s (1980) research on the correlation between high/low transitivity parameters and the foreground/background.

1.1. Longacre’s Premise
Longacre’s strong emphasis is on the virtual equation between the preterite verb which furthers the story and the foregrounded events or the storyline (e.g. 1981:337-47). For him, discovery of the preterite tense or completive aspect almost automatically eliminates the off-the-line materials as the background, i.e. non-storyline, then arrives at identification of the storyline.

1.2 Beyond Longacre: Temporal Movement
Gradually in the course of development of Longacre’s theory, some questions regarding the rigid connection between the morphosyntactic marking (e.g. preterite verb) and the foregrounded events or the storyline began to be raised even from among his colleagues and former students. One example is Somsonge who states:

---

2 I appreciate the reviewers’ comments for improvement, particularly their recommendation that the term ‘sequentiality’ should be more general, non-particularized and straightforward. Upon consideration, however, I kept it in some places as in my original draft following Givon (1993, 1984), whose influence can be seen in Endo (1996, 2003). It is their notion of ‘topic-continuity’, ‘sequential-action’, ‘theme continuity’ (Givon 1984:296-97) and the materials which ‘control the flow of the story’ (Endo 1996:232, 244-97) that my ‘sequentiality’ means. Talmy’s ‘sequential structure’ is also comprehensive enough to signify the sense by Givon and Endo; it refers to ‘the patterns in which a number of the elements of some category combine in a sequence through time’ (Talmy 2000:463). Thus, in this paper both ‘sequentiality’ and ‘sequential structure’ can include verbs with the sequential form or adverb, and conjunctions. The use of ‘sequentiality’ in this study is restricted in the storyline. By contrast, the common textlinguistic term ‘coherence’ (i.e. a text’s property that enables a hearer/reader to ‘fit its different elements into a single overall mental representation’ (Dooley and Levinsohn 2001:23)) is broader and ‘cohesion’ (i.e. ‘the use of linguistic means to signal coherence’ (ibid: 27)) is narrower than the ‘sequentiality’.

3 Other examples are Somsonge (1992) and Dry (1992) in the anthology for Longacre.
The study of bipartite structure [i.e. storyline vs. non-storyline materials] of discourse information in Thai reveals that in language without verbal inflection as Thai, the bipartite structure is not expressed solely by the verb system as in English but by a conspiracy of non-systemic ways which include types of verbs, adverbs, time phrases, sequential signals, temporal clauses/phrases/words, auxiliaries, pre-serial verbs, and post-serial verbs. (Somsonge 1990:76) [Underline added]

She has developed, through the studies of Thai (1990, 1992) and Hlai (2002) and others, the way to identify the storyline by finding the temporal movement of narrative which may be indicated by the various constructions mentioned in the above quotation. It should be noted that such constructions as “adverbs, time phrases, sequential signals, temporal clauses/phrases/words, auxiliaries”, etc. were all (but punctiliar adverbs) categorized as the non-storyline materials by Longacre. However, they play an important role in identifying the storyline in Iu-Mien. Just as Thai is a language without verbal inflection, so is Iu-Mien. Inspired by Somsonge, the materials indicating temporal movement and sequentiality of narrative will be investigated.

1.3 Application to Iu-Mien: Progression and Sequential Structure
The materials contributing to the storyline in Iu-Mien can be categorized into two major groups. The first are those which indicate chronological forward movement, increment of new information, transition of actions/events in the flow of discourse: they include the development adverb aengx ‘furthermore’, Serial Verb Constructions (SVCs), and multiclausal constructions containing SVCs and topic chains. The property found in them can be summarized in the term “progression” (Talmy 2000:425-26).

The second involves those which indicate sequentiality within and across sentences: i.e. the prospective (or inchoative) conjunction ziouc ‘then’, the retrospective conjunction cingx_daaih “therefore”, the ad-clause V + liuz “after V-ing”, and the topic marker aeqv “as for”. These constructions may be treated under the term “sequential structure” (Talmy 2000:463-76) or “sequentiality”.  

2. A Cognitive Grammar Outlook on Narrative: Current Discourse Space (CDS)
While textlinguistics tends to view narrative text outside a mind, e.g., as a written text on pages of a book or some sorts of recording, Cognitive Linguistics (CL) including CG considers it as a mental entity conceptualized in a mind. For example, Talmy expresses his fundamental assumption in approaching narrative as “the existence of a mind that has produced the narrative as well as of a mind that is cognizing the narrative” (2000:417-8). The schematization of cognizing or perceiving of the narrative is proposed in the framework of Current Discourse Space (CDS) by Langacker (2001:144-50; 2008:281, 464-67), an extension from his “stage model” (Langacker 1991a:211) at the clause level.

---

4 Earlier than Somsonge, Dry (1981, 1983) studied this aspect.
5 In questioning Hopper and Thompson’s (1980) absence of mentioning “sequentiality”, Longacre stated, “Sequentiality must in fact be taken as prime characteristic of the storyline, i.e., of foregrounding in narrative” (1996:25) but has not developed this point further.
6 Earlier, in the 1960s and 70s, Longare and his colleagues’ works were largely on the field-recorded written narrative discourses; hence their theory was usually called “discourse analysis”. However, when the term began to be applied to wider usage in such disciplines as conversation analysis or sociolinguistic study on people’s interaction as a discourse, Longare adopted the term “textlinguistics”.
7 Other example is Gärdenfors (1999:21), who says, “Cognitive models are mainly perceptually determined (meaning is not independent of perception)”.
CDS as defined by Langacker is “the mental space comprising those elements and relations constructed as being shared by the speaker and hearer as a basis for communication at a given moment in the flow of discourse” (2001:144). The Figure 1 adapted from Langacker (2001, 2008) shows that the speaker (S) and the hearer (H) are interacting (indicated by the dashed two-way arrow) in mentally viewing (indicated by the upward dashed arrows) the focused part of a narrative which is profiled in the heavy line. “A discourse comprises a succession of frames each representing the scene being ‘viewed’ and acted on by the speaker and hearer at a given instant” (Langacker 2001:151).

![Figure 1. Current Discourse Space in the Cognitive Grammar outlook on narrative](image)

Relevant parts of the Figure 1 will be used to explain the conceptualization of storyline materials (e.g. briefly mentioned in 1.3) in the following pages. The important thing at this point is that a narrative discourse is viewed as a cognitive entity in a CL perspective.

3. Progression
The first component of the storyline is progression and it is analyzed based on a fundamental nature of human’s perceptive ability, called cognitive psychological principle I in this study. In analyzing some constructions contributing to progression, the principle I plays a role to explain how the storyline is foregrounded.

3.1 Cognitive Psychological Principle I
The analysis of progression in the storyline is based on a general nature of visual perception. Figure 2 shows that a moving object (b) is more conspicuous than a static one (a) to the perception of the viewer (V). This is a fundamental tendency of human cognitive

---

8 For the detail explanation of each part of the diagram, see Langacker (2001:145) and (2008:466).
ability. In the analysis of storyline, the viewer (V)\(^9\) can be represented by a speaker and writer or a hearer and reader of narrative, i.e. S and H in Figure 1.

![Figure 2. Static and moving objects](image)

3.2 Unmarked action/motion verbs

The first linguistic construction on the storyline is the unmarked action/motion (as opposed to static) verb, which is the simplest and most common, as exemplified in (1).

(1) AS.070  
\[ Ninh \ \textit{mingh} \ \textit{wuov} \ \textit{ndiev} \]  
3sg go there down_there  
\[ \text{pn.p} \ \text{vi} \ \text{pn.dem} \ \text{adv.loc} \]  
‘He went down there.’

An intransitive verb is used in (1), whereas a transitive verb in (2).

(2) AS.015  
\[ \textit{Nzopv} \ \textit{jiepv} \]  
pierce bear  
\[ \text{vt} \ \text{n} \]  
‘(She) pierced into the bear('s mouth).’

The verb \textit{mingh} ‘go’ in (1) is depicted in Figure 3, where the mover (M) (c.f. Langacker 1991c:288) in the original location (L\(_1\)) moves to the new location (L\(_2\)) through an action of going. On the other hand, the conceptualization of the verb \textit{nzopv} ‘pierce’ in (2) which has a patient (PAT) is described in Figure 4, following Langacker’s “stage model” (1991a:211), also called “Canonical Event Model” (1991c:285). The transmission of energy is depicted in these models and the energy transmission is a core element of events in a discourse. In this way the conceptualization of events is analyzed in a way analogous to a theater-goer viewing a play on the stage.

---

Besides the above quoted single unmarked verb in the storyline\textsuperscript{10}, the main verb can be accompanied by an aspectual verb which always comes in a post verb position. Example (3) shows the construction ‘unmarked verb + post verb aspectual verb.’

\textit{(3) AS.097}

\begin{verbatim}
97.1 Mingh wuov ndiev! gorngv ziangx
97.2 go there down_there speak finish
vi pn.dem adv.loc v vi
\end{verbatim}

\textsuperscript{97.1} (The governor’s people) went far down there \textsuperscript{97.2} (and) finished talking (with the brave man).

\textbf{3.3 Development adverb aengx ‘further’}

Secondly, another construction that has a force to push the story forward is ‘aengx + verb’, where aengx can be (i) most basically translated as ‘again’, ‘also’ meaning repetition; and (ii) ‘and then’, ‘further’, ‘furthermore’, or ‘moreover’ at the discourse level. This adverb often introduces a new development of a story, hence called the development adverb. In (4) the first aengx (65.1) introduces a chronological development; the second (65.2) presents one more happening on top of the previous ones in a story.

\textit{(4) AS.065}

\begin{verbatim}
65.1 Aengx taux da’nyeic hnoi,
adv v n n
65.2 aengx maaih nda’maauh faanv.
adv vt n v.st
\end{verbatim}

\textsuperscript{65.1} And then another day, \textsuperscript{65.2} furthermore there was a disturbing tiger.’

Figure 4(a) represents the conceptualization of aengx ‘furthermore’, where the CDS holds three viewing frames: minus, zero and plus frames (c.f. Langacker 2001:151), each

\textsuperscript{10} Besides (a) the unmarked action/motion verb, (b) the unmarked verb + post verb aspectual verb, another expression that pushes the story forward that shold come under section 3.2 is (c) the repetition of the same unmarked verb, e.g. \textit{he ran, ran, ran}. 
of which represents a usage event (corresponding to the previous, current, and anticipated usage events of Figure 1). The development adverb *aengx* ‘furthermore’ in the zero frame (i.e. the frame getting the present attention) introduces and profiles (described by the heavy line) a relationship in which an additional event \( E_2 \) follows from another \( E_1 \). Further, \( E_1 \) is identified (connected by the dashed line) with the preceding event (E) in the minus frame (i.e. the prior frame in discourse). The relationship of the extended \( E_2 \) and the present \( E_1 \), treated as “anchor-increment”, profiles an evocation of or expectation toward a further event that is inherent in the construction, i.e. *aengx* ‘furthermore’. Inevitably, the incremental nature of *aengx* increases another frame, the plus frame, containing \( E_2 \) identical with (depicted by the dashed line) the one in the zero frame.\(^\text{11}\) The basic notion of “anchor-increment” is sketched in Figure 4(b) (Langacker 2000:265).

![Figure 4](image)

**Figure 4.** *aengx* ‘furthermore’ and incremental discourse space

The incremental characteristic of certain linguistic constructions discussed by Langacker under the notion “discourse expectations” (2001:151-53) is an important element that pushes the story forward. In the case of *aengx*, it has dual functions: prospective in that it evokes the subsequent discourse and retrospective in that it makes a relation to the prior discourse.

### 3.4 Serial Verb Constructions (SVCs) and Billiard-Ball Model

Thirdly, progression is most evidently expressed by a combination of Serial Verb Construction (SVC) and topic chain. Six defining factors of SVC are proposed by Aikhenvald (2005): 1) component verbs in SVC act as a single predicate, 2) monoclausality, 3) no intonation breaking between components, 4) shared tense, aspect, mood, modality and polarity value, 5) describes one event, 6) has one argument but may share arguments and obliques. An example of the simplest kind is (5).

(5) AS.008

\[
\begin{array}{llll}
\text{Nyörng} & \text{bieq} & \text{amz} & \text{mingh.} \\
\text{creep} & \text{enter} & \text{granary} & \text{go} \\
\text{vi} & \text{vi} & \text{n} & \text{vi}
\end{array}
\]

‘They crept into the granary.’

\(^{11}\) The translation value ‘again’ of *aengx* or the repetition of the same verb, rather than ‘furthermore’, is retrospect only. Hence, the retrospect use of *aengx* has a different conceptual structure than the dual functions; a possible diagram should neither have an increment arm of \( E_2 \) in the zero frame nor a dashed line toward the plus frame.
In (6), SVC is in the clause 23.2, in which the three verbs share one elliptical subject, i.e. the hero and his wife.

(6) AS.023
23.1
\begin{tabular}{l}
\textit{Gengh} & wuov_{\text{nzunc}_hnoi} & paaix & orv & daath \\
adv & adv.tmp & vt & n & v.asp \\
\end{tabular}

\begin{tabular}{l}
\textit{dorh} & mingh & bun \\
take & go & give Ampe \\
vt & vi & vt & n.prp \\
\end{tabular}

Really this time (they) divided up the meat \(^{23.2}\) (and) took it (to) give to the governor.’

Example (7) shows a case where an SVC has different subjects in it, which is referred to as “switch-subject SVC” (Aikhenvald 2005).

(7) AS.040
\begin{tabular}{l}
buonv & daic & mingh. \\
shoot & die & go \\
vt & vi & v.asp \\
\end{tabular}

‘(Those who were inside the boat) shot (the crocodile and it) died.’

In order to account for conceptual structure of progression comprised of SVC and topic chains in the storyline, the notions of “action chain” and “billiard-ball model” are drawn upon from Langacker (1991c:282-329). Figure 5 shows a transition of energy (indicated by the double arrows representing verbs) from one object (depicted by the circles representing nouns) to others in sequence. The energy transition starts from the first object named the Head and terminates in the last one termed the Tail.

\begin{center}
\textit{Figure 5. Action chain in billiard-ball model}
\end{center}

As is easily observed, the action chain in the billiard-ball model is an extended chain of the Canonical Event Model in Figure 4. Whereas the term transitivity by Hopper and Thompson (1990) was restricted to lexical semantics and energy transmission within a clause, Langacker’s billiard-ball model can be applied to analyze SVCs, topic chains, and their combination in the form of multiclausal constructions (3.6).

\footnote{The notion of “action chain” can be said that it is the Langacker’s version of explaining transitivity in view of revealing its conceptual structure, which was discussed by Hopper and Thompson (1980) in relation to the foreground in discourse.}
3.5 Topic Chains

Used in conjunction with SVCs, fourthly, topic chains are also substantially used in the storyline of Iu-Mien narratives. Topic chains are “chains of clauses which share a common topical participant” (Van Valin 2005:103). In (8), clauses 40.3 and 40.4 are the topic chain, followed by a switch-subject SVC with an aspectual verb.

(8) AS.040
40.1 Wuov_deix, yiem nzangv gu'nyuoz wuov_deix
those_people be_in boat inside those_people
tn.dem vi n n pn.dem
40.2 , zorqv caang nzopy.
take spear pierce vt n vt
40.3 , zorqv congx buonv, buonv daic mingh.
take gun shoot shoot die go vn vt vt vi v.asp
40.4 , 40.1 Those people, 40.2 those who were inside the boat, 40.3 took spears (and) pierced (crocodile), 40.4 took guns (and) shot, 40.5 shot (it and it) died.’

While SVCs give a story the sense of terseness and swiftness, the topic chains can elaborate the action and motions. The combined use of them forwards the story dynamically.

3.6 Multiclausal Constructions Containing SVCs and Topic Chains

The SVCs and topic chains can be combined to compose a long string of multiclausal construction as in (9a). An intonation break shows that it is composed of two sentences, but connected tightly describing one event comprising sub events. Clause 30.1 and Clause 30.2 are SVCs presenting a setting for Clause 30.3. There is no intonation break between Clauses 30.4-30.6, which comprise the three-member switch-subject SVCs. Clauses 30.3-30.6 is a topic chain.

(9a) AS.030
30.1 Mienh geh jakv jiex, nzaeng nzangv jiex,
person ride boat pass n vi v
30.2 ninh cuotv daaih
3sg come_out come pn.p vi
30.3 ning czangv
30.4 [St1] nzhngv, mbienv ndortv
press_down_on_hard boat overturn fall vn vi

[St2] ngatv nzangv,
30.5 press_down_on_hard boat overturn fall
v vi
People rode a boat to pass (the river), paddled to pass, it [the crocodile] came out pressed down on the boat; (the boat) turned over (and people) fell off, (crocodile) got people (to) eat.’

(9b) is a schematic presentation of action chains mentioned in section 3.4. In (9b), “billiard-ball model” (3.4) is superimposed on the multiclausal construction (9a), in which the circles represent NPs and the arrows Vts.

(9b) AS.030

This schematization is a simple technique to make the dynamism of action/event progression in the multiclausal construction explicit. The longer the verbs are strung with interspersed NPs, the more vivid the actions become; the perception of the storyline thus becomes obvious.

An extended application of the billiard-ball model based on the cognitive psychological principle I (3.1) to both SVCs and topic chain is schematized in Figure 6. Circles represent nouns, double arrows verbs. Each time the verb is uttered by a narrator, an action occurs. As the action increases as indicated by a black arrow in each line along the axis Y, the mind of audience (i.e. hearer or reader) conceptualizes the motion of the black arrow as a progression of actions/events along the axis X.

Figure 6. The conceptual schematization of action increase in SVCs and topic chains
3.7 Summary
Progression of a story is expressed in such constructions as unmarked action/motion verbs, the development adverb *aengx* ‘furthermore’, SVCs, topic chains, and the multicausal construction containing SVCs and topic chains. The bundle of these constructions constitutes progression, i.e. advancement of action, energy, motion, and events, which forms the storyline. The progression in the storyline has been analyzed on the basis of the cognitive psychological principle I, i.e. a moving object is more conspicuous than a static one to the perception of the viewer. Taking a narrative as analogous to a theater play viewed by the audience, the storyline stands out as the foreground in the mental space of the hearer or reader through the progression energized by the above analyzed constructions. The clear and dynamic movement in the storyline is foregrounded in the visual perception in the mind of hearer/reader. In CL outlook on the narrative discourse, the storyline is an outcome of the energy, action, event progression.

4. Sequential Structures
Besides the progression, there is yet another factor by which storyline is foregrounded to the perception of the hearer or reader, that is, sequential structures.

4.1 Cognitive Psychological Principle II
A principle on which sequentiality is deemed as important to the foreground is schematized in Figure 7. That is, a condensed cluster of objects in a linear order (c) is more conspicuous to the perception of the viewer than scattered objects with no connection between them (a). This may be termed as cognitive psychological principle II.

![Figure 7. Scattered and condensed objects](image)

Sequentiality is expressed by *ziouc* “then”, *cingx_daaih* “therefore”, *aeqe* “as for”, ad-clause containing *liuz* “after V-ing”, and independent clause marker *gau* “and then”. Four of them are selected to show the schematizations of their conceptual structures.

4.2 Sequential Marker (Prospective Conjunction) *ziouc* “then”
First, a very frequently used construction for connecting two clauses is S + *ziouc* + V, where *ziouc* can be translated as ‘and then, then soon, after which’. As in (10), *ziouc* always comes in the second position after S in the subsequent clause (126.2), which is connected to the preceding clause (126.1). With this grammatical behavior, *ziouc* can be
simply called a conjunction. At the same time, it gives a story a sense of quick transition of action/event across the clauses; hence its appropriate term may be the sequential marker. Example (10) shows a series of quick actions of a hero the main figure (126.2) after all soldiers fled (126.1).

<table>
<thead>
<tr>
<th>(10) AS.126</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>126.1</strong></td>
</tr>
<tr>
<td><em>Uv! biaux ninh ziouc zorqv baeng nyei congx daath</em></td>
</tr>
<tr>
<td>Oh! flee 3sg so_then take soldier POSS gun come</td>
</tr>
<tr>
<td>intj v pn.p seq.mk vt n poss n v.dr</td>
</tr>
<tr>
<td><strong>126.2</strong></td>
</tr>
<tr>
<td><em>buonv nyei buonv</em></td>
</tr>
<tr>
<td>shoot ADVBLZR shoot</td>
</tr>
<tr>
<td>vt part vt</td>
</tr>
<tr>
<td><strong>126.3</strong></td>
</tr>
<tr>
<td><em>nzuangv nyei nzuangv zunc jienv mingh.</em></td>
</tr>
<tr>
<td>aimless ADVBLZR aimless chase SML.ACT go</td>
</tr>
<tr>
<td>v.st part v.st vt asp vi</td>
</tr>
<tr>
<td><strong>126.4</strong></td>
</tr>
</tbody>
</table>
| *Oh! (they) ran away, 126.2 so he (immediately) took the soldier’s guns 126.3 (and) shot repeatedly, 126.4 (and) chased them away furiously.*’

In addition to expressing an immediate action following the first one, this construction also introduces an additional event in the course of happenings in succession as in (11).

<table>
<thead>
<tr>
<th>(11) AS.071</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Wuov hnoi lungh_aanx</em></td>
</tr>
<tr>
<td>that day after_noon</td>
</tr>
<tr>
<td>pn.dem n adv.tmp</td>
</tr>
<tr>
<td><em>ziouc haiz wuov_jiex laangz mienh.</em></td>
</tr>
<tr>
<td>so_then hear upper_side village people</td>
</tr>
<tr>
<td>seq.mk vt n n n</td>
</tr>
</tbody>
</table>
| ‘In the afternoon of that day, then soon (they) heard (a sound of) the upper village people.’

*Ziouc* is more than a simple conjunction; it has a force to let a story advance. In parallel to los in the White Hmong, which means “(and) then/so, thus, therefore, yet, and it happens, and it turns out, with the result” (Clark 1988:93), *ziouc* in Iu-Mien seems to have the function of “an inchoative conjunction” (ibid). On this assumption, *ziouc* is considered to be an important element which belongs to the storyline. The “inchoative” semantic value is also testified by some native speakers of Iu-Mien, as saying that *ziouc* means “future”.13

The conceptual structure of *ziouc* has much in common with that of *aengx* ‘furthermore’ (Figure 4). Figure 8 profiles the increment of E2 upon the utterance of *ziouc*.

---

13 Zanh Gueix-Fongc 2005: personal communication
in E₁ in the zero frame, which in turn evokes (indicated by the dashed line) the plus frame. What is different from aengx ‘furthermore’ in Figure 4 is that ziouc ‘and then’ lacks the anaphoric reference, hence Figure 8 does not have a dashed line which connects E and E₁.

![Figure 8. Prospective conjunction ziouc ‘and then’, ‘then soon’](image)

While keeping connection with the previous clause, ziouc also pushes an event forward in CDS. Thus, the sequential marker ziouc can also be referred to as a prospective conjunction.

### 4.3 Retrospective Conjunction cing_daaih “therefore”

The second material that ensures sequentiality is cing_daaih ‘therefore’. Whereas ziouc is analyzed as the prospective conjunction, cing_daaih ‘therefore’ is deemed as retrospective conjunction, thus, two are in the contrastive positions in their semantic value. (12a) and (12b) are connected discourse, where cing_daaih bases (12b) on (12a). The clause after cing_daaih presents a result of the previous sentence. (In fact, in this story cing_daaih ‘therefore’, ‘as a result’ here is used humorously, since the hero is only accidentally honored by a success of others.)

(12a) AS.060

60.1 60.2 60.3 60.4

_Those people took spears pierced (and) pierced repeatedly (the snake) (and it) died._

(12b) AS.061

_Cingx_daaih aengx orn yietc nzunc therefore again appoint one time_

_The conceptual structure of cing_daaih ‘therefore’ is depicted in Figure 9._
A parallelism with *therefore* in English analyzed by Langacker can be recognized to explain Figure 9. The following quotation is applicable to *cingx_daaih* : “*therefore* [and *cingx_daaih*] introduces and profiles a relationship in which one proposition, labeled $P_2$, follows from another, $P_1$. Moreover, $P_1$ is identified as the proposition expressed in the previous usage event, just earlier in the flow of discourse” (Langacker 2001:149).

It should be noted that *cingx_daaih* often emphasizes a logical consequence of the proceeding preposition, in contrasted to the construction $V + liuz$ ‘after $V$-ing’ which signifies temporal transition (4.4). As such, $P_1$ and $P_2$ are confined in the zero frame and there is no increment toward the plus frame.

### 4.4 Ad-Clause $V + liuz$ “after $V$-ing”

The third construction to contribute to the storyline is an ad-clause containing the phrase $V + liuz$ ‘after $V$-ing’, where *liuz* itself originally means ‘to finish’. The ad-clause always comes in the sentence initial position followed by the main clause. Compare (13a-d). They are not from the story in the appendix but all from actual situations except for (13a) which is unnatural.

(13a) *nyanc* liuz hnaangx

<table>
<thead>
<tr>
<th>eat</th>
<th>finish</th>
<th>rice</th>
</tr>
</thead>
<tbody>
<tr>
<td>vt</td>
<td>aux</td>
<td>n</td>
</tr>
</tbody>
</table>

(unnatural)

(13b) Nyanc liuz hnaangx aqv

<table>
<thead>
<tr>
<th>eat</th>
<th>finish</th>
<th>rice</th>
<th>PFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>vt</td>
<td>aux</td>
<td>n</td>
<td>part.asp</td>
</tr>
</tbody>
</table>

‘I have finished having a meal.’

(13c) Nyanc liuz nhaangx, (yie) cingx_daaih daaih

<table>
<thead>
<tr>
<th>eat</th>
<th>finish</th>
<th>rice</th>
<th>(1sg) therefore</th>
<th>come</th>
</tr>
</thead>
<tbody>
<tr>
<td>vt</td>
<td>aux</td>
<td>n</td>
<td>pn.p conj</td>
<td>vi</td>
</tr>
</tbody>
</table>

‘I have finished having a meal, so I came.’

(13d) nyanc liuz nhaangx, (yie) ziouc daaih

<table>
<thead>
<tr>
<th>eat</th>
<th>finish</th>
<th>rice</th>
<th>(1sg) consequently</th>
<th>come</th>
</tr>
</thead>
<tbody>
<tr>
<td>vt</td>
<td>aux</td>
<td>n</td>
<td>pn.p conj</td>
<td>vi</td>
</tr>
</tbody>
</table>
‘Having finished a meal, I came right away.’

The native speakers’ response to (13a) would be “and then what happened?” This means that *nyanc liuz hjangx* is just a fragment until either it is completed by an aspectual particle *aqv* as in (13b) or followed by a main clause as in (13c-d). Both the narrator and hearer cannot stop after the *liuz*-clause. This ad-clause inevitably evokes an expectation in hearer’s mind.

Reutilizing the discourse space increment diagram from Figure 4(b), the conceptual structure of *liuz*-clause can be explained as in Figure 10. In the ad-clause *V + liuz* ‘after *V*-*ing*’, the *V* carries the old information and the whole construction projects a new discourse space for more information.

![Figure 10. Discourse space increment through ad-clause *V + liuz* ‘after *V*-*ing’](image)

Thus, the construction is a prospective element in a similar vein with *ziouc* ‘and then (soon)’ in Figure 8. The effect on the audience in hearing and reading the ad-clause that contains *liuz* is to cause an expectation or a whetting of the appetite to listen for more. Naturally, it has an impact on the audience in creating a mental space even before hearing what comes next. As a result, the use of *liuz* pushes the story forward.

4.5 Topic marker *aeqv* “as for”

In the fourth place, a topic marker *aeqv* ‘as for’ is also an important element in the sequential structure. This is true when the above discussed retrospective and prospective elements (i.e. *cingx daaih* ‘therefore’ and *ziouc* ‘and then’) are taken in consideration to understand sequentiality of storyline. The main function of topicalization, according to Levinsohn, is to indicate discontinuities while maintaining “the overall unity and continuity” (Levinsohn 1994:7). He says: “This is because topicalized constituents have a bidirectional function: (a) they serve as a point of departure for what follows and (b) they indicate the primary basis for linking what follows to its context”. In Examples (14)-(16), the underlined parts are topicalized by the topic mark *aeqv*; the clause after them (i.e. (113.2), (36.2), and (96.3)) are all main clauses.

The first example (14) has a personal pronoun topicalized; (15) exemplifies a verb phrase topicalized.

(14) AS.113

113.1 113.2

*Ninh aeqv, sin zinx haic aav lov!*

3sg TOP body tremble very SURPRISE

113.1 As for him, 113.2 his body trembled terribly.
As (they) reach there, (they) again passed (the river) paddling a boat.

Not only topicalizing a noun phrase and a verb phrase, but also multiple phrases can be topicalized as in (16). Each time aeqv is uttered, the floor of narration is held until the whole sentence is concluded with the main clause.

If we don’t, (we) can’t win and then we do not have magic here.

The held floor of narration forces the narrator to continue to tell story and the hearer to expect more to listen. Evidently, the use of the topic marker aeqv let the story go on.

The conceptual structure of the bi-directional function of topic marker aeqv ‘as for’, which evokes an upcoming proposition creating a new discourse space, is illustrated in Figure 11 following Langacker 2001.

A lengthy quotation from Langacker well explains Figure 11:

A topic marker refers schematically to the thing profiled by the noun phrase with which it combines; the nominal referent is represented as a circle in the focus frame. Such marker is prospective by virtue of signaling that the profiled entity will function as a conceptual reference point for purposes of interpreting a subsequent proposition, as shown in the plus frame. But it is also retrospective, in the sense that the topic needs to be an entity already accessible in the prior discourse. Thus the profile is shown as corresponding to a thing present in the minus frame. (Langacker 2001:152)
4.6 Summary
The sequential structure premises the analyses on the cognitive psychological principle II: a condensed cluster of objects in a linear order is more conspicuous to the perception of the viewer (i.e. hearer/reader of narrative) than scattered objects with no connection between them. On this premise, the linguistic constructions that have functions of combining and sequential nature have been analyzed. The prospective conjunction ziouc ‘and then (soon)’ and the retrospective conjunction cingx_daaih ‘therefore’ ensure the flow of discourse in sequence. Besides them, the topic marker aeqv ‘as for’ has the bi-directional function, which provides the sequentiality for the story as well as preventing the narration from stopping by its floor-holding function. The ad-clause containing V-liuz ‘after V-ing’ has the incremental nature of discourse space. This construction not only keeps sequentiality in the storyline but causes the hearer to expect to listen for more. Thus, the sequential structure makes the storyline obvious as the foreground to the visual perception in the mind of the hearer/reader.

5. Conclusion
The term “foreground” is itself pertains to an act of viewing: it is the locus of a view which is the nearest to an observer, hence the most salient or noticeable. Though the term has been used to describe the event-line/mainline/storyline of the narrative discourse in numerous literatures, the optical notion of the foreground has never been highlighted in the study of storyline. CL including CG and cognitive semantics, however, assumes that the narrative discourse is a cognitive entity viewed in the mind of the conceptualizer, i.e. viewer/hearer/reader. In the course of hearing/reading the narrative, by analogy with an act of viewing a play on the stage, the hearer/reader perceives the storyline as the foreground in the mental space. The storyline is optically foregrounded as an outcome of two main components of the narrative: progression and sequentiality.

The progression of the storyline comprises such constructions as the unmarked action/motion verbs, the development adverb aengx ‘furthermore’, SVCs, topic chains, and the multiclausal construction containing SVCs and topic chains. They advance the story. The sequential structures include such elements as the prospective conjunction ziouc ‘and then (soon)’, the retrospective conjunction cingx_daaih ‘therefore’, the topic marker aeqv ‘as for’ with the bi-directional function, and the ad-clause containing V-liuz ‘after V-ing’. The storyline encoded with these constructions of the forward-moving and the tightly sequenced becomes easily observable in front of the mental eyes of the hearer/reader, i.e. the viewer. In other words, progression and sequentiality in the narrative make the storyline emerge to the foreground in the perception of the viewer.

This approach works well for Iu-Mien which totally lacks morphological markers in the verb, as well as revealing semantics of storyline. It is also expected that a CL approach can assist in discourse analysis in various languages in Eastern and Southeast Asia, such as within Chinese, Tai, Hmong-Mien, and Vietnamese languages. As these languages are areally similar in their structure (e.g. little morphological inflection), adopting a similar cognitive-semantic approach can be illuminating.

---

14 Purnell 2009: personal communication
References


Appendix: Text of “A Story of Aahan” (AS)

Aa^han Gouv
Yunh Zosh gorngv
Yunzoi speak
n.prp v

Mr. Yunzoi narrated.

AS.001
1.1
Maaih i_gox   mienh  za’gengh  jomc  nyei.
have married_couple person really be_poor PRS.ST
v n n adv v.st asp

1.1 There was a couple  1.2 (who were) very poor.

AS.002
2.1    2.2    2.3
Jomc  nyei,  ninh  mingh  maaic  gong  lorz  hnaangx  nyanc.
be_poor  PRS.ST  3sg go  sell  work look_for  rice  eat
v.st asp pn.p vi vt n vt n vt

2.1 Poor they were, 2.2 they went (and) did labor for other people 2.3 looked for rice (to) eat [i.e. made their living.]

AS.003
3.1
I_gox   ndaam  jienv  ndaamx
married_couple carry SML.ACT carrying_pole
n vt asp n

3.1 The couple would carry baskets on their shoulders with poles 3.2 just like Northern Thai would do, 3.3 (they) went dangling, dangling and dangling.

AS.004
4.1 4.2
Mingh  gau  hmuangx  aqv.
go DPCL dark CHG-O-ST
vi part adj part.asp

4.1 As they went, 4.2 (it) became dark.

AS.005
5.1    5.2    5.3
Hmuangx  aqv.  taux  wuov  ndaamv-jauv,
dark CHG-O-ST reach there half-way
adj part.asp v pn.dem n

5.1 It became dark 5.2 (when they) were on the way, 5.3 (then there was) a granary just like mine, 5.4 it has been built for long time.
(When they) arrive there  
(they) became dark.

Ah, wherever (they) went  
(they) have not gone as far as to enter a granary (to) sleep.

They crept into the granary.

A bear had given birth to a cub (there).

The bear really ran out  
(it) wanted (to) bite the couple.

Her husband ran away.

There was nothing his wife could do,  
(she) was not strong.

(She) was carrying things.
(She) swung (it) dangling, dangling, then (she) took her carrying pole like this, that is, her pole was long, (she) released (it and) let (it) go down (and it) came off; (she) stamped on this end (let it drop (on the bear) (and it) pierced the bear.

(She) pierced into the bear’s mouth as (she) pierced, (she) shoved the pole really hard (and the bear) died.

He wanted to divide the bear (to) take it (and give to the district township official).
They said, "Ah, what shall we do? We don't have a knife or a gun."

"If you say…" (if) his wife beat (the bear to death)… her husband would be ashamed more.

"(If) say her husband beat… that's not (that) the husband beat; (the truth is) his wife beat.

But his wife said, “Ah, I am a woman; I’m sure I don’t want an honor. I give it to you, you tell them you beat it.”
23.1 Really this time (they) divided up the meat 23.2 (and) took it (to) give to the governor.

24.1 (They) reached there, 24.2 the governor asked them, “You don’t have a gun, don’t have a knife. How could you get the bear?”

25. “Oh, we beat (it to death). (We) caught it (and) beat and beat. (We) took a stick and shoved (it) into its mouth repeatedly (and it) died.”

26. “Wow! You are really so strong!”

27. (The governor) entitled that man (to) be Bear the Brave.

28. At this occasion (he) got one considerably big name.
Then another day, like the river down there (where) we ride a boat, there was a kind of crocodile, (which) was big.

**AS.030**

<table>
<thead>
<tr>
<th>Person</th>
<th>Ride</th>
<th>Boat</th>
<th>Pass</th>
<th>Paddle</th>
<th>Boat</th>
<th>Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mienh</td>
<td>Geh</td>
<td>Jakv</td>
<td>Jex,</td>
<td>Nzaeng</td>
<td>Nzangv</td>
<td>Jex.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>VT</td>
<td>N</td>
<td>VI</td>
<td>V</td>
<td>N</td>
<td>VI</td>
</tr>
</tbody>
</table>

30.1 People rode a boat to pass (the river), paddled to pass, it [the crocodile] came out pressed down on the boat; (the boat) turned over (and people) fell off, (crocodile) got people (to) eat.

**AS.031**

There was nothing the Northern Thai people could do about it again.

**AS.032**

Ah, (we) can't manage to do anything.

**AS.033**

(We) must go find that Bear the Brave and bring him here to let him catch (the crocodile).

**AS.034**

(They) really went to look for (him).

**AS.035**

(They) found the Bear the Brave.
As (they) reach there, (they) again passed (the river) paddling a boat.

As soon as (the boat) passed (to) reach the middle of the river, (which was) really absolutely black and wide, came out.

He was very afraid, (his) body trembled, jerked intensely.

Those people, those who were inside the boat, took spears (and) pierced (it), took guns (and) shot, shot (it and it) died.

“Don’t you yet pierce it.”

“I’m just releasing feces to let it eat.”
I want to catch it alive.

I want to take it bound to show gratitude to the provincial officer.

How come you guys pierced it to death?

He said like that.

Yes, this time (he)’s got to go again.

(He) taking this crocodile (and) went to the provincial office.

(The governor) appointed him again (to) be Suthi the Brave.

(He) has got two names, Bear the Brave (and) Suthi the Brave; (he) got (them).

(He) has got two names, Bear the Brave (and) Suthi the Brave; (he) got (them).
Daniel Arisawa

Then another day, there was a gigantic snake, (which) was big fell into a city water source for the residents of the whole town, the whole city.

Indeed, (people) became unable to get the drinking water. (i.e. the water became undrinkable.)

The really very big snake fell into (the well).

And again the people of that village, of the whole capital discussed.

Well, there was nothing (they) could do about it.

There is no other way than (we) must go look for Bear the Brave, Suthi the Brave (and ask him to) catch (the snake).

(They) went (to) look for him, went (and) reached that village, found him (there), (he was) like help(ing) you (to) make a fence.
Being there, (he was) building in a fence sticking a scabbard like mine (in his back part of the belt); (and they) called him. (He answered) “Ah, I will go and see that kind (of thing).”

The land was steep; which means the well was on the lower side, he was on the upper side.

(They) went (and) reached (the well), as soon as (he) squatted down his scabbard hit this [i.e. his back], he just fell down there inside (the well).

Wow! The Northern Thai really praised him very much.

Look at him, (he) arrived here (he) immediately jumped into the well.
AS.059
Zorqv naang-jaang nanv jienv.
take snake-neck grasp DUR
vt n-n vt asp
(He) took the snake by the neck, grabbing (it)

AS.060
Wuov_deix zorqv caang nzopv nzopv nzopv nzopv daic m’aqv.
those_people take spear pierce pierce pierce pierce die RSLT
pn.dem vt n vt vt vt vi asp
60.1 Those people 60.2 took spears 60.3 pierced (and) pierced repeatedly (the snake) 60.4 (and it) died.

AS.061
Cingx_daaih aengx om yietc nzunc ninh zoux Ar’Han_Nguy.
therefore again appoint one time 3sg be Snake_the_Brave
conj adv vt numb n pn.p vi n.prp
Therefore, (the governer) again gave him another title (to) be Snake the Brave.

AS.062
Za’gengh longx haic aqv wuov nzunc.
really be_good very CH-O-ST that time
adv v.st adv asp pn.dem n
This time it was really good.

AS.063
Za’gengh duqv nyaanh duqv lui-houx.
really get money get clothes
adv vt n vt n
(He) actually got money and clothes.

AS.064
Gwyen (Jien) za’gengh a’hneiv haic ninh.
officer official really be_happy very 3sg
n n adv v.st adv pn.p
The governer was really happy about him.

AS.065
Aengx taux de’nyeic hnoi, aengx maaih nda’mauh faanv.
and_then reach second day further have tiger be_di sturbing
adv v n n adv vt n v.st
65.1 And then another day, 65.2 furthermore there was a disturbing tiger.

AS.066
Maaih laanh mienh nda’mauh daaih ngaatc gau, jai ngaatc gau, dungz ngaatc gau,
have CLF person tiger come bite DPCL chicken bite DPCL pig bite as
vt clf n n vi vt part n vt part n vt top
66.1 There was a person 66.2 the tiger came (and) bit (him); 66.3 then (it) bit chickens, 66.4 then (it) bit pigs, 66.5 (it)
wanted all people!

(They) discussed (saying) Ah, there’s nothing (we) can do (about it).

(We) must go look for Bear the Brave, Snake the Brave, Suthi the Brave (and have him) catch (the tiger), there is no other way.

And then (they) went (and) found him.

He went down there, like in this village, he went down there at the stream.

In the afternoon of that day, then soon (they) heard (a sound of) the upper village people.

(They) made a lot of noise chasing the tiger (and they) came.
When they came chasing the tiger (and) reached there, he climbed up a tree (and) stayed above.

Underneath the tree was hollow.

The tiger arrived there slinking along but did not come in (the hole).

(Its) tail was outside.

The tail was long (and) it stayed outside.

Those people came chasing (it) with a loud noise, came shooting guns.

(As) they came a little closer, he came down a little.
AS.081
81.1  Wuov_deix  daaih  fatv  deix  ninh  njiec  aiv  deix.
  those_people  come  near  some  3sg  descend  be_low  some
pn.dem  vi  v.st  adv  pn.p  vi  v.st  adv
(As) they came a little closer, he came down a little.

AS.082
82.1  Hnangv  wuov_deix  taux  wuov  ndau  ninh  liux  njiec  daaih
  as_if those_people  reach  that  place  3sg  run  go_down  come
adv  pn.dem  v  pn.dem  n  pn.p  vi  vi  vi
82.2  It looked like (that) when those people arrived there he ran down pulling the tiger’s tail.

AS.083
Nauc  wuov_deix,
  yell_at  those_people
vt  pn.dem
("Nqongh  daaih  nqongh  daaih  yie  zorqv  duqv  dauh  nda’maa uh  aqv.
  hurry  come  hurry  come  1sg  take  can  CLF  tiger  PFT
  v  vi  v  vi  pn.p  vt  v.aux  clf  n  part.asp").
(He) shouted to them, “Come quickly! I’ve just managed to catch the tiger.”

AS.084
Wuov_deix  daaih  taux  buonv  n yei  buonv.
  those_people  come  reach  shoot  ADVBLZR  shoot
pn.dem  vi  v  vt  part  vt
Those people arrived, shooting repeatedly.

AS.085
85.1  Baqv  nyei  baqv  daic  mingh.
  pierce  ADVBLZR  pierce  die  go
vt  part  vt  vt  vi  v.asp
85.2  (They) pierced and pierced (the tiger) pierced (it) died.

AS.086
Aengx  dorh  mingh  bun  gwyen  mangc.
  and_then  take  go  let  officer  look
adv  vt  vi  vt  n  vt
(He) took it back (to) let the governor look at (it).

AS.087
Gwyen  aengx
  officer  further_more  name  3sg  be  Tiger_the_Brave  Tiger_the_Brave
n  adv  vt  pn.p  vi  n.prp  n.prp
The governor furthermore called him Tiger the Brave, Tiger the Brave.
Tiger the Brave got (awarded) four times.

(He) was really strong.

Later, the country became unsettled again, like people here fight each other nowadays.

There were many fighters, many soldiers; (they) destroyed everywhere.

There was this officer living in that city.

Further those soldiers wanted to come (to) fight.

There was nothing (he) could do about it.
Ah, (we only) order Bear the Brave, Suthi the Brave, Snake the Brave, Tiger the Brave (to) go (and) catch (the soldiers), there is no other way.

If we don’t, (we) can’t win and then we do not have magic here.

(The governor’s people) went far down there (and) finished talking (with the brave man); on the slope on the opposite side of the great mountain.

From there as he went up there, there was one kind of fern.

Among the trees here a kind of fern traversed completely.

He climbed up sitting there.

The numerous soldiers really came.
Many thousand, (or we) don’t know if it was how many thousand, how many ten-thousand.

Didn’t know where they stayed either.

(They) went (and) were resting at the tree base, numerously resting there.

They were there (and they) discussed (i.e. they were discussing there).

“Oh, it’s scary! We are not afraid of the village people, we are not afraid of the people of the city. We only fear Snake the Brave, Bear the Brave.”

He could hear (the soldiers’ conversation) staying above.

“Oh! If (we) get Bear the Brave, Snake the Brave killed, the people of the whole village are not enough for us to eat for breakfast.”
He was terrified very much, terrified so much (that his) body trembled, clinging to a tree.

(His) body trembled much continuously.

(Other) came driving horses (they) were red.

As for him, his body trembled terribly.

As for this place, we are here, (we) suspect (that) they won’t come here.

If (they) really come here, (we) don’t have guns, don’t have swords.
AS.116
Fungc  zoux.
adv  v

Can't do anything.

AS.117

117.1
Aav_dangh  ninh_mbuo  ciepv  buatc  ndortv  aqv.
in_a_short_while  3_PL  look_at  see  fall  PFT
adv.tm  pn.p  vt  v  vi  part.asp

117.2
za'gengh  gamh_nziex  haic.
really  fear  very
adv  vt  adv

117.3
Suddenly they saw 117.2 (the brave man) fall down, 117.3 (they) were really terrified very much.

AS.118

Ninh  oix  daic  haic.
3sg  want_to  die  very
pn.p  v.mod  vi  adv

He was scared very much.

AS.119

Sin  zinx  nyanh  gau.
body  tremble  jerk  very
n  vi  vi  adv

His body trembled very much.

AS.120

120.1
Nyaalir-weih  nquatv
fern  crack_down
n  vi

120.2
ziouc  sus!  sus!  nyei
so_then  sound_of_spurting_air  sound_of_spurting_air  ADVBZR
seq.mk  onoma  onoma  part
fortc  wuov  ndau  daaih.
cry  that  ground  come
vi  pn.dem  n  vi

The fern cracked down, 120.2 so (he) zoomed down to the ground with a sound of spurting air.

AS.121

Maiv_haif_fungc_zoux  aqv.
can’t_do_anything  CHG-O-ST
idm  part.asp

He couldn’t help.

AS.122

122.1
Nyaayv  sin  daaih  mbaix  mbaix  mbaix
get_up_quickly  body  come  clap  hand  slap  foot
vt  n  v.dr  vt  n  vt  n

122.1 (He) got up quickly 122.2 clapped (his) hands 122.3 beat (his) thighs.
"Bear the Brave, Snake the Brave, Suthi the Brave, that’s me myself!"

(Even though) today (he) is alive, (or) dies today, (since he) frightened them today, let’s see, we only discuss here, there is no way to describe where he flew from.

(He) didn’t remember to look at the fern.

Oh! (they) ran away, so he took the soldier’s guns (and) shot repeatedly, (and) chased them away furiously.

Then he further carried their guns.

They all ran away.
Daniel Arisawa

AS.129

129.1
Ndaam jenv congx nzungx mingh bun gwen mingc.
carry SML.ACT gun return go let officer look
vt asp n v vi vt n vt

129.1
(He) went back carrying the guns on his shoulders, 129.2 (and) let the governor look (at them).

AS.130

130.1
Buonv baaic mi'aqv ninh_mbuo biaux nzoungc mi'aqv.
Shoot defeat RSLT 3.PL flee consumed RSLT
vt vt asp pn.p v vi asp

130.1
(He) shot (and) defeated (them) all; 130.2 they all completely ran away.

AS.131

Gwen aengx bun minh mingh
Officer further let person go
n adv vt n vi
zorqv congx lo_haax nzoung x daaih.
take gun and_so_forth return come
vt n n vi vi

The officer further let people go (to) get guns and so forth back.

AS.132

132.1
Wuuv_nzung_hnoi paaiv yielt gongv dei-bung yieti fouv-zingh
this_time appoint one corner region one CLF capital_city
adv.tmp vt numb n n numb clf n

132.3
bun ninh nyanc.
let 3sg eat
vt pn.p vt

132.3
Then this time, (the governor) assigned one corner of a region 132.2 (with) one capital city 132.3 (to) let him make a living.

AS.133

Za'gengh! ninh yaac duqv zoux hio.
really 3sg also get be big
adv pn.p conj vt vi v.st

Really he also became big.

AS.134

I_gox loz-hnoi za'gengh jomc haic.
marrried_couple old-days really poor very
n n adv v.st adv

Before that the couple was really very poor.

AS.135

Nae!
See
intj

I have told you.
FIXED AUTOSEGMENTISM IN
THAI EMPHATIC REDUPLICATION

Leslie Lee
University of California at San Diego
<leslie@ling.ucsd.edu>

Abstract
The classical approach to dealing with fixed material in reduplication is to assume that the reduplicant is prespecified with the fixed content (Marantz 1982, Yip 1982b). However, researchers working in OT have argued that fixed segmentism in reduplication can be explained without resorting to prespecification, and that the emergence of fixed segmentism in the reduplicant is either phonologically or morphologically determined (Alderete et al. 1999, Walker 2000). This paper extends this argument to Thai emphatic reduplication, where, instead of segmental content, it is the autosegmental content of tone that is fixed. I argue that the analysis of Thai emphatic reduplication does not require prespecification in the reduplicant, while accounting for the occurrence of high tone on CVVO syllables in emphatic reduplication - a phenomenon that is generally prohibited in the language.

Keywords: phonology, tones, reduplication

1. Introduction
It has been argued that fixed segmentism in reduplication can be explained without having to resort to prespecification in the reduplicant, and that the emergence of fixed segmentism in the reduplicant is either phonologically determined or morphologically determined (Alderete et al. 1999, Walker 2000). In this paper, I explore a related phenomenon in Thai emphatic reduplication, where the reduplicant appears with an invariant high tone regardless of the tone on the base, i.e. it is not segmental content that is fixed, but autosegmental content. In addition to the issue of fixed autosegmentism, the occurrence of a high tone on the reduplicant also highlights a corresponding puzzle in the tonal phonology of the language: high tones are generally prohibited from occurring on CVVO syllables in Thai, yet, this combination does occur on the CVVO reduplicant in emphatic reduplication. Therefore, the aim of this paper is twofold: (i) to provide an account of fixed autosegmentism in reduplication without prespecification, and (ii) to account for the presence of high tone on CVVO reduplicants.

---

15 I would like to thank Sharon Rose and an anonymous reviewer for detailed comments that have led to the improvement of this paper. Any remaining errors are my own.

Journal of the Southeast Asian Linguistics Society 4.1:41-63
Copyright vested in the author
Received 5/2/2010, revised text accepted 16/3/2011

41
1.1 Thai tones
Thai has five lexical tones. The five tones are traditionally (since Abramson 1962) described and transcribed as mid, low, high, falling, and rising. These are illustrated in (1):

(1) Mid  naa ‘rice field’
Low  nàa ‘custard apple’
High  náa ‘aunt’
Falling  nàa ‘face’
Rising  nàa ‘thick’  

(Morén and Zsiga 2006: 114)

Not all syllables can bear all five tones. Only open syllables with long vowels (CVV) or syllables closed by a sonorant coda (CVS, CVVS) can bear all five tones:

(2) CVV     CVS     CVVS
naa ‘rice field’  laŋ ‘crate’  laaŋ ‘omen’
nàa ‘custard apple’  láŋ ‘to flow’  lâaŋ ‘chime of a bell’
náa ‘aunt’  lám ‘to go beyond’  lâaŋ ‘to wash’
nàa ‘face’  lámk ‘sturdy’  lâaŋ ‘below’
nâa ‘thick’  lâŋ ‘back’  lâaŋ16 ‘grandchild’

(Morén and Zsiga 2006: 115)

In syllables closed by obstruents, tonal distribution is restricted. In CVO syllables, only low and high tones may appear, while in CVVO syllables, only low and falling tones are permitted:

(3)  CVO  CVVO
mid  ---  ---
low  lák ‘stake’  láak ‘various’
high  lák ‘to steal’  ---
falling  ---  láak ‘to tow’
rising  ---  ---

(Morén and Zsiga 2006: 116)

The table in (4) summarises the general distribution of tones in Thai with respect to syllable type:

(4) General distribution of tones in Thai

<table>
<thead>
<tr>
<th>σ Type/Tone</th>
<th>Mid</th>
<th>Low</th>
<th>High</th>
<th>Falling</th>
<th>Rising</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVV, CVS, CVVS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CVVO</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>CVO</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

16 The original transcription is laaŋ, which is surely a typographical error.
1.2 Reduplication in Thai

There are two types of reduplication in Thai that copy the base segments faithfully: simple reduplication and emphatic reduplication. In simple reduplication, the reduplicant is an exact copy of the base, including its tone, as illustrated in (5):

(5) dii ‘good’  dii-dii ‘really good, better, (do) well’
cháa ‘slow’  cháa-cháa ‘really slow, slower, (do) slowly’
rew ‘fast’  rew-rew ‘really fast, faster, (do) fast’
sǎaw ‘young’  sǎaw-sǎaw ‘really young, younger (female)’
nùm ‘young’  nùm-nùm ‘really young, younger (male)’
kɛ̀ɛ ‘old’  kɛ̀ɛ-kɛ̀ɛ ‘really old (person)’

(Iwasaki and Ingkaphirom 2005: 34)

Since the reduplicant is simply an exact copy of the base, simple reduplication obeys the general distribution pattern of tones in the language.

On the other hand, emphatic reduplication copies only the base segments, and assigns an emphatic high tone to the reduplicant irrespective of the tone on the base. (6) illustrates:

(6) dii ‘good’  díi-dii ‘very good’
ʔím ‘full’  ʔím-ʔím ‘very full’
yâak ‘difficult’  yâak-yâak ‘very difficult’
rɔ́ɔn ‘hot’  rɔ́ɔn-rɔ́ɔn ‘very hot’
wǎan ‘sweet’  wǎan-wǎan ‘very sweet’

(Iwasaki and Ingkaphirom 2005: 35)

Unlike the simple reduplicant which fills its autosegmental content through copy from the base, the emphatic reduplicant displays fixed autosegmental material. Recall that high tones do not occur on CVVO syllables in the language (see (3-4)). The emphatic reduplication of a CVVO syllable, as in the case of yâak-yâak, then presents a puzzling situation where a high tone is permitted on a CVVO syllable in spite of the fact that this is generally prohibited in the tonal system of the language. This suggests that the requirement on the reduplicant to bear a high tone overrides the general phonotactic constraints in the language.

In light of the data concerning emphatic reduplication, two questions beg to be answered:

(i) Why does the reduplicant surface with a high tone in emphatic reduplication? In other words, what is the basis for the fixed autosegmentism?
(ii) What is the explanation for the high tone occurring on CVVO syllables?

17 According to Haas, phonemically, the emphatic high tone can be treated as a modification of the high tone. Phonetically, however, it is not identical with the high tone, for the emphatic high is always higher in pitch and usually more protracted in length than the normal high (Haas 1946). I assume that the emphatic high is underlyingly the same as the regular high tone.

18 Iwasaki and Ingkaphirom omitted the onset glottal stops in order to represent casual speech. In this paper, I assume the underlying phonological representations.
2. Analyses of Thai tonal distribution

Before proceeding to answer these questions, the gaps in the Thai tonal system must first be accounted for. More specifically for present purposes, the occurrence of high tone on CVVO syllables must be ruled out in the general phonology of the language.

Prior to Morén and Zsiga (2006), no complete explanation has been proposed for the distributional gaps. As they note, two different partial explanations have been proposed among previous analyses, each with their own shortcomings. Gandour (1974, 1977) and Zhang (2002) account for the lack of rising tone on obstruent final syllables by appealing to duration. Citing phonetic studies showing that rising contours tend to be durationally longer than falling contours (Ohala and Ewan 1972; Sundberg 1973, 1979; Xu 1999a, 1999b), these authors argue that the relatively shorter duration of the sonorant portion of the rhyme in obstruent-final syllables does not allow sufficient time for the phonetic realisation of a rising tone. However, an account based on duration cannot account for the prohibition against mid tones, nor can it explain why high tones are prohibited on CVVO but allowed on CVO, the shortest syllable of all. Instead of duration, Yip (1982a, 2002) grounds the prohibition against rising contours in glottalisation. Noting that syllable-final obstruents in Thai are both voiceless and glottalised, she formulates a negative constraint “*LH[+glottal]” which prevents rising contours from associating with obstruent-final syllables. She does not, however, extend the analysis to the other gaps in the system (Morén and Zsiga 2006: 117).

Since Morén and Zsiga (2006) is the only account that successfully accounts for all the gaps in the Thai tonal system, I assume that this is the correct analysis. On this analysis, the lack of mid, high, and rising tones on CVVO syllables is due to a positive constraint that requires obstruent coda segments to be associated with a low tone. This analysis is outlined in the following section. As the distributional gaps in CVO syllables are not crucial to our discussion of emphatic reduplication, I shall ignore this issue here and refer the interested reader to the original article.

2.1 Bimoraicity in Thai

Morén and Zsiga argue that all stressed words in Thai are bimoraic (2006: 127-129). First, long and short vowels contrast, but there are no extra long vowels in the language. That is, vowels may be monomoraic or bimoraic, but not trimoraic. Second, short open syllables cannot occur in stressed positions, and underlyingly CV syllables that occur in stressed positions are made heavy by the addition of a glottal stop. Thus, all stressed syllables are (minimally) bimoraic. Third, since CVC syllables can bear stress, the final coda in a CVC syllable is moraic. Finally, based on acoustic measurements, they note that phonetic shortening of the long vowel and coda occurs in CVVC syllables, suggesting that Thai words are maximally bimoraic, and that the second mora is shared between the vowel and the coda in a CVVC syllable. Thus, they propose the moraic structure in (7):

(7) µ µ µ µ
    | | | |
    CVV CVC CVVC

2.2 Phonetic realisation of Thai tones in citation form

Based on data recorded from two native female speakers, Morén and Zsiga provide a detailed description of the shapes of the five tones in citation form by dividing the
frequency scale into three subranges: high (240-280 Hz), mid (200-240 Hz), low (160-200 Hz):

- The mid tone stays level or falls gradually through the mid range.
- The low tone begins in the mid range and falls, reaching its lowest point at the bottom of the range at the end of the syllable.
- The high tone stays level or falls slightly in the mid range through the first mora, then rises to reach the high range at the end of the syllable.
- The falling tone begins in the high range, starting with a rising high during the first mora, and then falls to the low range during the second mora.
- The rising tone begins in the mid range, falling and staying low during the first mora before rising during the second mora, ending in the mid range.¹⁹

(Morén and Zsiga 2006: 130)

The tonal shapes described here are generally consistent with the findings of previous acoustic analyses of the tonal shapes in citation form (Abramson 1962, Gandour et al. 1991).

2.3 Representation of Thai tones

Morén and Zsiga noticed that, across syllable types and between their two speakers, the inflection points in tonal contours correspond closely to the right edges of morae: mid syllable inflections occurred nearly exactly at the end of the first mora. This led them to argue that the tone-bearing unit (TBU) in Thai is the mora, and that the specified tones are aligned to the right of morae. On the assumption that phonologically toneless morae are phonetically mapped onto the neutral pitch range, mid tone syllables have no phonological tonal specification and remain at neutral pitch throughout; high tones, which remain in the mid range through the first half of the syllable and reach a high point only at the end of the rhyme, begin with no tonal specification on the first mora and end with a high tone on the second; low tones, which begin in the mid range and fall to a low point at the end, begin with a tonally unspecified first mora and end with a low tone on the second; falling tones, which first exhibit a rise in the high range followed by a fall to the low range, begin with a high tone on the first mora and end with a low tone on the second; rising tones, which fall and stay low before rising, begin with a low tone on the first mora and end with a high tone on the second. Their proposed phonological representations of Thai tones are shown below:

(8) Mid High Low Falling Rising
H L H L L H
µ µ µ µ µ µ µ µ

¹⁹ Morén and Zsiga (p. 130) note that the rising tone ends in what they classify as the mid range, and does not reach the target range for high. Nonetheless, the rising tone consistently ends at a higher frequency than the mid tone.
2.4 Analysis of CVVO syllables

Given the representation of tones in (8), the restriction against mid, high, and rising tones on CVVO syllables can be generalised as a requirement for CVVO syllables to end low. In order to rule out CVVO syllables that do not end low, Morén and Zsiga (p. 143) propose \(C.G.CODA \rightarrow L\), a positive constraint that requires obstruent coda segments to be associated with a low tone:

\[
C.G.CODA \rightarrow L \quad \text{Constricted glottis coda segments must be associated with L tone.}
\]

Since the second mora of a CVVO syllable is assumed to be shared between the vowel and the coda, as shown in (7), \(C.G.CODA \rightarrow L\) is only satisfied when the second mora of a CVVO syllable is associated with a L tone:

(9)                                  
\[
\begin{array}{c}
| \scalebox{0.7}{\text{L}} \\
\mid \\
\mu \mu \\
\mid \mid \\
\text{CVVO} \\
\end{array}
\]

On this analysis, the lack of mid tone on CVVO syllables is due to the ranking of \(C.G.CODA \rightarrow L\) above \(*L\), the markedness constraint against L tones:

(10) Neutralisation of mid tone to low tone in CVVO

<table>
<thead>
<tr>
<th>/laak/ ‘various’</th>
<th>(C.G.CODA \rightarrow L)</th>
<th>(*L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (\mu \mu)</td>
<td>(\mu \mu)</td>
<td></td>
</tr>
<tr>
<td>b. (\mu \mu)</td>
<td>(\mu \mu)</td>
<td>!</td>
</tr>
</tbody>
</table>

Given a CVVO input with no tonal specification, \(C.G.CODA \rightarrow L\) must outrank \(*L\) in order for candidate (10a), which surfaces with a L tone, to be preferred over the faithful candidate in (10b). In other words, it is less costly to have an extra L tone than to risk having an obstruent coda segment that is not associated with a L tone.20

The avoidance of high tones on CVVO syllables is achieved by ranking \(C.G.CODA \rightarrow L\) above both \(*[[TT]_\sigma]\) and \(ALIGN-R(T,\sigma)\), as shown in tableau (11).

---

20 One might ask why the markedness constraint \(*L\) is used instead of the faithfulness constraint \(\text{DEP-IO}(L)\), since candidate (10a) involves an unfaithful correspondence between the input and the output. This is certainly an option, but crucially, both \(*L\) and \(\text{DEP-IO}(L)\) must be ranked below \(C.G.CODA \rightarrow L\): ranking \(*L\) above \(C.G.CODA \rightarrow L\) incorrectly predicts (10b) to be the optimal candidate.
**[TT]_σ** Two tones within the same syllable domain are prohibited.

**ALIGN-R(T, σ)** Align the tone at the right edge of the syllable.

Given an input specified with a high tone, it is more harmonic to insert a L tone to satisfy C.G.CODA→L, as in the optimal candidate (11a), than to violate C.G.CODA→L, as in (11b), even though (11a) incurs violations of both *[TT]_σ* and ALIGN-R(T, σ). Finally, rising tones on CVVO syllables are ruled out by ranking both C.G.CODA→L and MAX-IO(H) above LINEARITY, shown in tableau (12).

**MAX-IO(H)** Every H tone in the input has a correspondent in the output.

**LINEARITY** Preserve the linear order of features and segments.

(11) Neutralisation of high tone to falling tone in CVVO

<table>
<thead>
<tr>
<th>/laak-H/ ‘to tow’</th>
<th>C.G.CODA→L</th>
<th><em>[TT]_σ</em></th>
<th>ALIGN-R(T, σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. H L</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>μμ</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>laak</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. H

| μμ          |          | !         |
|            |          |           |
| laak       |           |           |

(12) Neutralisation of rising tone to falling tone in CVVO

<table>
<thead>
<tr>
<th>/laak-LH/ ‘to tow’</th>
<th>C.G.CODA→L</th>
<th>MAX-IO(H)</th>
<th>LINEARITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. H L</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>μμ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>laak</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. LH

| μμ          |           | !         |
|            |           |           |
| laak       |           |           |

c. L

| μμ          |           | !         |
|            |           |           |
| laak       |           |           |
Given an input specified with a rising tone, it is preferable to inverse the linearity of the tones in order to satisfy C.G.Coda→L, as in the optimal candidate (12a), than to have the faithful candidate in (12b), which violates C.G.Coda→L, or to delete the H tone as in (12c), which violates Max-Io(H).

To summarise, the ranking of constraints shown in (13) ensures that mid, high, and rising tones do not occur on CVVO syllables, so that these syllables always end low and surface with either low or falling tone:

(13) Ranking of constraints required for CVVO syllables to end low

C.G.Coda→L => MAX-Io(H)

* L ALIGN-R(T,σ) *[TT]σ LINEARITY

3. Reduplication

Having shown how the distributional gaps in CVVO syllables are accounted for in Morén and Zsiga’s analysis, I now turn my focus to reduplication. I begin by outlining key theoretical assumptions.

3.1. Theoretical assumptions

The analysis is formalised in the constraint-based framework of OT (Prince and Smolensky 1993). The standard OT model of reduplication is shown in (14):


Input: /AfRED + Stem/

Output [Red ↔ Base]

B-R identity

In this model, input-output (IO) correspondence is evaluated with respect to the stem and the base, while base-reduplicant (B-R) correspondence is evaluated with respect to the base and the reduplicant. This model assumes the definition of the reduplicant, RED, in (15):

(15) Definition of RED (McCarthy and Prince 1994 (Part 1): 2)

REDk is a morpheme lexically unspecified for segmentism, but requiring a correspondence relation with its base, the phonological structure to which it attaches.

However, Walker (2000) observes that this definition is at odds with the principle of Richness of the Base (Prince and Smolensky 1993: 191), which hypothesises that all inputs are possible. Since OT constraints do not apply to inputs and merely evaluate output candidates, the null hypothesis is that prespecification could actually occur in the inputs of reduplicative affixes: otherwise, we would be stipulating restrictions on the input (Walker 2000: 90). Accordingly, she revises the definition of RED:
Revised definition of RED (Walker 2000: 90)

REDₖ is a morpheme requiring a correspondence relation with its base, the phonological structure to which it attaches.

Having eliminated the assumption that the reduplicative affix is lexically unspecified, there is now an input form of the reduplicative affix to which the output can correspond. Walker then proposes an elaborated model of reduplication, with correspondence between the input and output forms of the reduplicative affix:

**Elaborated model of reduplication (Walker 2000: 90)**

**Input**

/AFₖ + Stem/

**Affix RED -IO faithfulness**    ↑↓    ↑↓

**Stem-IO faithfulness**

**Output**

[AFₖ ↔ Base]

B-R identity

This is the model that I will assume in my analysis. Unlike the model in (14), this model evaluates IO correspondences not merely with respect to the stem and base, but also the input and output forms of the reduplicative affix.

The three core families of faithfulness constraints on correspondence relations are given in (18), where S₁ and S₂ refer to strings in a correspondence relation (McCarthy and Prince 1995: 264).

**a. MAX**

Every segment of S₁ has a correspondent in S₂.

**b. DEP**

Every segment of S₂ has a correspondent in S₁.

**c. IDENT(F)**

Let α be a segment in S₁ and β be any correspondent of in S₂. If α is [γF], then β is [γF]. (Correspondent segments are identical in feature F).

McCarthy and Prince cast these constraints in terms of correspondence relations on segments. (19) interprets these in terms of tone:

**a. MAX(T)**

Every tone of S₁ has a correspondent in S₂.

**b. DEP(T)**

Every tone of S₂ has a correspondent in S₁.

**c. IDENT(T)**

Let α be a mora in S₁ and β be any correspondent of in S₂. If α is [γT], then β is [γT]. (Correspondent morae are identical in tonal affiliations).

Since the concern of the present paper is with the form of the reduplicant, it will be unnecessary to focus on stem-IO correspondence, which is covered in the scope of Morén et al.'s analysis.
and Zsiga (2006). I shall thus ignore candidates that involve bases ruled out by Morén and Zsiga’s analysis.

Given the model in (17), if a reduplicative affix input came with prespecified input material, it has to be subject to both AffixRED-JIO faithfulness constraints and B-R identity constraints. The two general configurations that may arise from the competition between these sets of constraints are shown in (20):

(20)  a. Faith-BR >> AffixRED-Faith-IO

The ranking in (20a) yields a pattern in which the reduplicant is a maximally faithful copy of the base, i.e. prespecified material in the reduplicant does not overwrite the copied material from the base. This outcome corresponds to one in which there is no apparent prespecification. On the other hand, the ranking in (20b), which places some or all of AffixRED-Faith-IO above Faith-BR, allows prespecified material to appear in the output of the reduplicative affix at the cost of maximizing copied material from the base. The outcome yields prespecified material as the source of fixed segmentism in reduplication. Walker observes that if rankings such as that in (20b) could be eliminated, this would rule out the possibility of the emergence of prespecified material in the reduplicant. Thus, she proposes the following metaconstraint on reduplication correspondence:

(21) Reduplication correspondence metaconstraint (Walker 2000: 92)
     Faith-BR >> AffixRED-Faith-IO

Given this metaconstraint, Faith-BR must be ranked above AffixRED-Faith-IO, and prespecification is not an option for the source of fixed segmentism.

3.2. Simple reduplication

Having laid out the theoretical assumptions, I will now first develop an analysis of simple reduplication in Thai. As seen in section 1.2, tones from the base appear faithfully on the reduplicant in simple reduplication. In an OT analysis, this entails that the Faith-BR constraints outrank the AffixRED-Faith-IO constraints as well as the general markedness constraints in the language. I illustrate with the CVV syllable ƙɛ̀ɛ ‘old’ but the same basic analysis applies for all other syllable types as well, since surfacing as a faithful copy of the base entails that the reduplicant obeys the general distribution of tones in the language. I assume that the tonal alignment in reduplication is no different from that in other parts of the language. On this assumption, tonal alignment on the reduplicant falls out from Morén and Zsiga’s analysis, in which ALIGN-R(T,σ) outranks ALIGN-L(T,σ).

The tableau in (22) illustrates the faithful copying of the L tone from the base onto the reduplicant when a tonally unspecified reduplicative affix is attached to ƙɛ̀ɛ ‘old':
(22) Simple reduplication

<table>
<thead>
<tr>
<th>/RED+kɛɛ-L/</th>
<th>MAX-BR(T)</th>
<th>DEP-AF_{RED}(T)</th>
<th>*L</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>L</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td></td>
<td>µ</td>
<td>µ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>kɛɛ – kɛɛ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>µ</td>
<td>µ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>kɛɛ – kɛɛ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The optimal candidate in (22a), which copies the L tone from the base onto the reduplicant, violates DEP-AF_{RED}(T) since the L tone on reduplicant does not have a correspondent in its input. In faithfully reduplicating the base, the candidate in (22a) also incurs an additional violation of *L. In contrast, the candidate in (22b) only violates *L once and does not incur any Affix_{RED}-Faith-IO violations. However, it violates MAX-BR(T) as the L tone on the base does not have a correspondent in the reduplicant. Therefore, in order for the reduplicant to surface as a faithful copy of the base as in (22a), MAX-BR(T) must be ranked above DEP-AF_{RED}(T) and *L. The ranking order MAX-BR(T) >> \{DEP-AF_{RED}(T), *L\} ensures that a L tone from the base is copied faithfully onto the reduplicant.

The definition of RED in (16) allows prespecification in the reduplicative affix, so it is possible for the reduplicative affix to be tonally specified. This creates three possible scenarios: (i) overwriting, where the prespecified tone on the reduplicative affix replaces the tone that we would otherwise expect to be copied from the base, (ii) the creation of a contour tone, where the prespecified tone on reduplicative affix and the copied tone from the base each occupy one mora on the reduplicant, and (iii) backcopy, where the prespecified tone on the reduplicative affix is copied onto the base. Since simple reduplication simply involves the faithful copying of tones from the base to the reduplicant, these possibilities must be ruled out. For this to happen, Faith-BR must outrank Affix_{RED}-Faith-IO, as predicted by Walker’s metacontrast in (21).

First consider overwriting, illustrated in tableau (23) with a reduplicative affix specified with a H tone. The faithful candidate in (23b) violates MAX-BR(T), since the L tone on the base does not have a correspondent in the reduplicant. However, it does not violate any Affix_{RED}-Faith-IO constraints. On the other hand, the optimal candidate in (23a), which surfaces without the prespecified H tone and instead copies the L tone from the base, violates both DEP-AF_{RED}(T) and MAX-AF_{RED}(T), but not MAX-BR(T). It has been shown in (22) that MAX-BR(T) must be ranked above DEP-AF_{RED}(T). (23) now shows that for (23a) to be the optimal candidate, MAX-BR(T) must also outrank MAX-AF_{RED}(T). Thus overwriting in simple reduplication is ruled out by the ranking order MAX-BR(T) >> \{DEP-AF_{RED}(T), MAX-AF_{RED}(T)\}.
(23) No overwriting in simple reduplication

<table>
<thead>
<tr>
<th>/RED-H+κɛɛ-L/</th>
<th>MAX-BR(T)</th>
<th>DEP-AF_{RED}(T)</th>
<th>MAX-AF_{RED}(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textsubscript{a}. L L</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kɛɛ – kɛɛ</td>
<td></td>
</tr>
<tr>
<td>b. H L</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>kɛɛ – kɛɛ</td>
<td></td>
</tr>
</tbody>
</table>

In order to avoid violating MAX-BR(T) like the candidate in (23b), the L tone from the base must be copied onto the reduplicant. However, it is also possible for this to happen without violating MAX-AF_{RED}(T): if the prespecified H tone on the reduplicative affix and the copied L tone from the base each occupy one mora on the reduplicant, forming a derived contour tone, the H tone on the reduplicant input would have a correspondent in its output and MAX-AF_{RED}(T) would be satisfied. This is illustrated in tableau (24). Since both candidates in (24) copy the L tone from the base onto the reduplicant, both violate DEP-AF_{RED}(T) equally. However, in allowing the prespecified H tone to surface on the RED, the candidate in (24b) violates DEP-BR(T), as the prespecified H tone does not have a correspondent in the base. Therefore, in order for (24a) to be the optimal candidate, DEP-BR(T) must be ranked above MAX-AF_{RED}(T). The ranking order DEP-BR(T) >> MAX-AF_{RED}(T) ensures that a derived contour tone does not result on the reduplicant as a result of the concatenation of a prespecified tone and a copied tone.

(24) No derived contour tone in simple reduplication

<table>
<thead>
<tr>
<th>/RED-H+κɛɛ-L/</th>
<th>DEP-BR(T)</th>
<th>DEP-AF_{RED}(T)</th>
<th>MAX-AF_{RED}(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textsubscript{a}. L L</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kɛɛ – kɛɛ</td>
<td></td>
</tr>
<tr>
<td>b. HL L</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>kɛɛ – kɛɛ</td>
<td></td>
</tr>
</tbody>
</table>

Backcopy can be avoided by ranking Stem-Faith-IO above Affix_{RED}-Faith-IO. This is illustrated in (25). Unlike the candidate in (24b), the candidate in (25b) does not violate DEP-BR(T) as each tone on the reduplicant has a correspondent in the base. Therefore, in order to rule out this candidate and prevent a tone that was prespecified on the reduplicative affix from appearing on the base, DEP-IO(T) must be ranked above MAX-
AF_{RED}(T). The ranking order Dep-IO(T) \gg Max-AF_{RED}(T) thus prevents backcopy in simple reduplication.

(25) No backcopy in simple reduplication

<table>
<thead>
<tr>
<th>/RED-H+kɛɛ-L/</th>
<th>Dep-IO(T) \gg Dep-AF_{RED}(T)</th>
<th>Max-AF_{RED}(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{a.} L L</td>
<td>\text{*}</td>
<td>\text{*}</td>
</tr>
<tr>
<td>\text{\mu\mu \mu\mu}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\text{\kɛɛ \kɛɛ}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\text{b. HL HL}</td>
<td>\text{*!}</td>
<td>\text{*}</td>
</tr>
<tr>
<td>\text{\mu\mu \mu\mu}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\text{\kɛɛ \kɛɛ}</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The ranking of constraints required for simple reduplication is summarised in (26). Generally speaking, in order for the reduplicant to be an exact copy of the base, the Faith-BR constraints must be ranked above the Affix_{RED}-Faith-IO constraints, in accordance with the general configuration in the metaconstraint in (21).

(26) Ranking of constraints for simple reduplication

\text{Dep-IO(T) \gg Dep-BR(T) \gg Max-BR(T) \gg Max-AF_{RED}(T) \gg Dep-AF_{RED}(T) \gg *L}

3.3. Emphatic reduplication

Having presented an analysis of simple reduplication, I turn now to emphatic reduplication. The first question with respect to emphatic reduplication was why the reduplicant always surfaces with a high tone; in other words, why does the reduplicant appear with invariant autosegmental material? The classical approach to dealing with fixed material in the reduplicant is to assume that the reduplicant is prespecified with the fixed content (Marantz 1982, Yip 1982b). On this analysis, the reduplicative affix would simply be prespecified with a H tone when used in emphatic reduplication.

However, recent work within OT has explored the notion that fixed segmentism in reduplication can be explained without prespecification, and that the emergence of any fixed content in the reduplicant is either phonologically determined or morphologically determined (Alderete et al. 1999, Walker 2000). Moreover, as shown in (23-25), a prespecified tone on the reduplicative affix cannot be allowed to surface on the reduplicant for simple reduplication to take place. Now, in order to have a unified analysis of both simple and emphatic reduplication, the grammar in (26) presented for simple reduplication must also hold for emphatic reduplication. Therefore, not only is the prespecification of a H tone on the reduplicative affix in emphatic reduplication unnecessary, assuming it would contradict the rest of the grammar of the language. Thus, in line with the spirit of recent work, I will eschew the prespecification approach and explore instead the option of
deriving the fixed autosegmental content in the reduplicant from the premises of OT; reduplicative affixes prespecified with autosegmental content will not be considered from this point on.

Fixed segmentism that has a phonological basis falls under the OT rubric of the emergence of the unmarked (TETU), whereby only unmarked structure is allowed on the reduplicant while the corresponding marked structure is allowed elsewhere in the language (Alderete et al. 1999: 328). On Morén and Zsiga’s analysis of Thai tones, high tones are marked. If emphatic reduplication were indeed an instance of TETU, we would expect the emphatic reduplicant to surface instead with a mid tone, which is treated as the neutral pitch. This means that if Morén and Zsiga’s analysis is correct, the invariant H tone on the reduplicant in emphatic reduplication cannot have a phonological basis.

Morphological fixed segmentism, on the other hand, is a kind of affixation, whereby the fixed content is treated as an affix that is realised simultaneously with the reduplicant, overwriting part of it (Alderete et al. 1999: 328). Thus, the identity of the fixed content is morphologically determined. In the case of Thai emphatic reduplication, it is not segmental content that is fixed, but tone, i.e. autosegmental content. In order to pursue the line of phonological reasoning that the invariant high tone that occurs in Thai emphatic reduplication has a morphological basis, I assume that the emphatic high tone is an affix that is realised simultaneously with the reduplicant. As Yip notes, “Although we usually begin by thinking of a morpheme as something with segments, purely tonal morphemes abound” (Yip 2002: 106). On this conception, the emphatic high tone can be treated as a tonal morpheme sans segments that affixes to the reduplicative compound. The concatenation of the emphatic tonal morpheme with the reduplicative compound results in the following model:

\[
\text{(27) Elaborated model of Thai emphatic reduplication} \\
\text{Input} /Af_H + Af_{RED} + \text{Stem/} \\
\text{Affix}_{RED}-\text{IO faithfulness} \uparrow \downarrow \uparrow \downarrow \text{Stem-IO faithfulness} \\
\text{Output} [Af_{RED} \leftrightarrow \text{Base}] \\
B-R \text{ identity}
\]

IO faithfulness constraints govern the correspondence relations between the stem and base as well as the input and output forms of the reduplicative affix, while B-R faithfulness constraints regulate the correspondence between the reduplicant and the base. However, these constraints do not apply to the tonal morpheme, since it is not part of the stem, base, or reduplicative affix. Hence, in addition to the correspondence relations between related phonological structures, it is necessary to assume a mapping between morphology and phonology, in order to guarantee the appearance of the tonal morpheme in the phonological output. I adopt a constraint requiring a morpheme to be realised in the phonological content of the output (Samek-Lodovici 1993; Akinlabi 1996; Gnandesikan 1997; Rose 1997a,b; Walker 1998, 2000):

\[
\text{REALISE-}\mu \quad \text{A morpheme must have some phonological exponent in the output.}
\]

In essence, when applied to emphatic reduplication in Thai, \text{REALISE-}\mu ensures that the tonal morpheme will be realised in the output.
3.3.1. Fixed Autosegmentism
I now proceed to account for the fixed autosegmentism in Thai emphatic reduplication. Since CVV/CVS/CVVS syllables are the least marked in terms of tonal distribution, it would be useful to begin the analysis with these syllables before moving on to the CVVO syllables.

First consider why the emphatic high tone appears, and why it appears on the reduplicant. Given an input stem that is not specified for tone, e.g. dii ‘good’, the reduplicative compound consists of a reduplicant that appears with a high tone and a base that remains phonologically toneless, as in dii-dii ‘very good’. In order for the emphatic high tone to appear only on the first half of the reduplicative compound, both REALISE-µ and Stem-Faith-IO must outrank Faith-BR.

(28) compares the optimal candidate with one in which the emphatic tonal morpheme does not surface. The optimal candidate in (28a) violates DEP-BR(T) as the H tone on the reduplicant does not have a correspondent in the base. It also violates DEP-AF<sub>RED</sub>(T), since the H tone does not have a correspondent in the input of the reduplicant. Candidate (28b) does not incur any faithfulness violations, since the stem, base, and reduplicant stand in perfect correspondence. However, it incurs a violation of REALISE-µ, as the tonal morpheme does not have a phonological exponent in the output. Thus, in order for candidate (28a) to be more harmonic than candidate (28b), REALISE-µ must be ranked above DEP-BR(T) and DEP-AF<sub>RED</sub>(T). The ranking order REALISE-µ >> {DEP-BR(T), DEP-AF<sub>RED</sub>(T)} ensures that the emphatic tonal morpheme is realised in the phonological output.

(28) Emphatic high tone appears on the RED

<table>
<thead>
<tr>
<th></th>
<th>/H+RED+dii/</th>
<th>REALISE-µ</th>
<th>DEP-BR(T)</th>
<th>DEP-AF&lt;sub&gt;RED&lt;/sub&gt;(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>µµ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>µµ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d i i –d i i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>dii – dii</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now compare the optimal candidate with one in which the tonal morpheme appears not on the reduplicant, but on the base, as shown in (29). Both candidates satisfy REALISE-µ. However, the output candidate in (29b) contains a tone in the base that does not have a correspondent in the stem and thus violates DEP-IO(T). (29b) also violates MAX-BR(T), since the H tone on the base does not have a correspondent in the reduplicant. Thus, in order for (29a) to be more harmonic than (29b), either DEP-IO(H) or MAX-BR(T) must outrank DEP-BR(T), so that it is preferable to have the tonal morpheme realised on the reduplicant than on the base.
(29) Emphatic high tone appears on the red instead of the Base

<table>
<thead>
<tr>
<th>/H+RED+dii/</th>
<th>DEP-IO(T)</th>
<th>MAX-BR(T)</th>
<th>DEP-BR(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>µµ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d i i – d i i</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>µµ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d i i – d i i</td>
<td>*(!)</td>
<td></td>
</tr>
</tbody>
</table>

The ranking ranking order DEP-IO(H) >> DEP-BR(T) is also relevant in ensuring that there is no backcopy, and that the emphatic high tone does not appear on the base to match the tone on the reduplicant, as shown in (30). Even though candidate (30b) satisfies Faith-BR, it violates Stem-Faith-IO and is less harmonic than the optimal candidate in (30a) which violates DEP-BR(T). Thus, DEP-IO(H) >> DEP-BR(T) ensures that backcopy does not occur in emphatic reduplication. Since this ranking is independently required, I assume, for reasons of parsimony, that the violation of DEP-IO(H) is the fatal violation in (29), and leave open the possibility of MAX-BR(T) and DEP-BR(T) being unranked with respect to each other.

(30) No backcopy

<table>
<thead>
<tr>
<th>/H+RED+dii/</th>
<th>DEP-IO(T)</th>
<th>DEP-BR(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. H</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>µµ</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d i i – d i i</td>
<td>*</td>
</tr>
<tr>
<td>b. H</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>µµ</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d i i – d i i</td>
<td>*(!)</td>
</tr>
</tbody>
</table>

To summarise thus far, fixed autosegmentism in the emphatic reduplicant is accounted for by the ranking of constraints in (31):

(31) Ranking of constraints for fixed autosegmentism in emphatic reduplication

\[
\text{DEP-IO(T)} \rightarrow \text{REALISE-µ} \rightarrow \text{DEP-BR(T)} \rightarrow \text{DEP-AF}_{RED}(T)
\]
Combining this with the grammar in (26), the interim ranking of constraints required for reduplication in Thai is shown in (32):

\[ \text{(32) Interim ranking of constraints for Thai reduplication} \]

\[
\begin{array}{c}
\text{DEP-IO(T)} \\
\text{DEP-BR(T)} \\
\text{MAX-AF}_{\text{RED}}(T) \\
\text{REALISE-} \mu \\
\text{MAX-BR(T)}
\end{array}
\]

(28-30) illustrated fixed autosegmentism with a stem that is unspecified for tone. Now consider emphatic reduplication that involves a stem that is specified for tone. Given an input stem specified for low tone, such as \( \tilde{\text{i}}\text{m} \) ‘full’, the reduplicative compound consists of a reduplicant that appears with a high tone and a base that retains the low tone, as in \( \tilde{\text{i}}\text{m}-\tilde{\text{i}}\text{m} \) ‘very full’. Unlike the case of the toneless base in (28-30), where the emphatic high tone is merely added onto the reduplicant, the emphatic high tone now overwrites the tone on the reduplicant that we would have otherwise expected to be copied from the base when the base is specified for tone. Since high and low tones are, on Morèn and Zsiga’s analysis, specified on only one mora, one wonders why the concatenation of the high tonal morpheme and the low tone from the base does not result in a contour tone on the reduplicant, e.g. *\( \tilde{\text{im}}-\tilde{\text{im}} \). Such a candidate, which copies the base tone onto the reduplicant, would better satisfy MAX-BR(T) than one for which overwriting occurs, while satisfying REALISE-\( \mu \) at the same time. (33) illustrates.

\[ \text{(33) Derived contour in emphatic reduplication?} \]

<table>
<thead>
<tr>
<th>/H+RED+\tilde{\text{i}}\text{m}-L/</th>
<th>IDENT-BR(T)</th>
<th>DEP-BR(T)</th>
<th>MAX-BR(T)</th>
<th>DEP-AF\text{RED}(T)</th>
<th>*L</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ( \text{H L} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \mu \mu \mu \mu )</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ? i m - ? i m )</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>---b. ( \text{H L L} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \mu \mu \mu \mu )</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ? i m - ? i m )</td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
</tr>
</tbody>
</table>

Both candidates in (33) satisfy REALISE-\( \mu \), since the emphatic tonal morpheme surfaces on the reduplicant. Both candidates also violate IDENT-BR(T) and DEP-BR(T) equally: in each case, (i) there is one set of corresponding morae that do not have identical tonal affiliations, and (ii) the emphatic tonal morpheme that surfaces on the reduplicant does not have a correspondent in the base. The candidate in (33a) violates MAX-BR(T) as the L tone on the base has no correspondent in the reduplicant. In contrast, the candidate in
(33b) satisfies MAX-BR(T), at the expense of incurring additional penalty for violating DEP-AF\text{RED}(T) and \*L. However, as the latter two constraints have been established in (22) to be ranked below MAX-BR(T), these extra violations do not count against the candidate in (33b). Therefore, while the markedness constraint \*L penalises the candidate in (33b), it is ranked too low to affect the choice of the optimal candidate. On the other hand, the faithfulness constraint MAX-BR(T) penalises candidate (33a) and not (33b). This leads to the incorrect predictions that the candidate in (33b) is the optimal candidate, and that overwriting does not occur in emphatic reduplication.

3.3.2. Feature geometry

It appears that neither markedness nor faithfulness constraints can account for the overwriting phenomenon in emphatic reduplication. What then is the explanation for the overwriting? I propose that we need to refer to the internal feature geometry of the tonal features. Bao (1990) proposes the model of internal feature geometry of tonal features in (34):

\[
\sigma \\
| \\
o \quad \text{Tonal Node} \\
\wedge \\
H \quad \text{Contour} \\
\wedge \\
l \quad h 
\]

(Yip 2002: 53)

In this model, Tone features are dominated by a node of their own, called a Contour, which is a sister of the Register feature. The Contour node and Register feature are in turn dominated by a Tonal node. For present purposes, it is not necessary to consider the Register feature or the Tonal node; what is relevant is the idea that individual tones can form a constituent, as represented in the part of the model highlighted in bold.

Assuming that the individual tones within a syllable form a constituent, the constituent as a whole can be assessed for constraint violation. This entails a modification of Morén and Zsiga’s proposed representation of Thai tones, illustrated in (35):

\[
\begin{array}{cccc}
\text{Mid} & \text{High} & \text{Low} & \text{Falling} & \text{Rising} \\
C & C & C & C(ontour) \\
\mid & \mid & \wedge & \wedge \\
H & L & H & L & H \\
\mid & \mid & \mid & \mid & \mid \\
\mu & \mu & \mu & \mu & \mu \\
\end{array}
\]

On this analysis, both the emphatic tonal morpheme and the tone on the base project their own Contour nodes. Since only the emphatic high tone can occur on the reduplicant, overwriting can be understood as the result of a prohibition against multiple Contour nodes within a syllable:

\*\[CC]_\sigma \\
Two Contour nodes within the same syllable domain are prohibited.
By assuming the elaborated representation of tones in (35) and the markedness constraint \[^{\text{[CC]}_\sigma}\] overwriting in emphatic reduplication can now be accounted for, as shown in (36):

\[
(36) \text{ Overwriting in emphatic reduplication}
\]

<table>
<thead>
<tr>
<th></th>
<th>(\text{[CC]}_\sigma)</th>
<th>Max-BR(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>C C</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H L</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\mu \mu \mu \mu)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>? im - ? im</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>C C C</td>
<td>*!</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H L L</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\mu \mu \mu \mu)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>? im - ? im</td>
<td></td>
</tr>
</tbody>
</table>

As we saw earlier, the optimal candidate in (36a) violates Max-BR(T). This constraint is not violated in candidate (36b), which copies the L tone from the Base. However, candidate (36b) incurs a violation of \[^{\text{[CC]}_\sigma}\]. This is because the entire Contour node projected by the L tone is copied from the base, while the emphatic high tone projects its own Contour node, resulting in two Contour nodes within the single syllable domain of the reduplicant. By ranking \(^{\text{[CC]}_\sigma}\) above Max-BR(T), the analysis correctly chooses (36a) as the optimal candidate.

Thus, feature geometry seems to provide a plausible explanation for the overwriting phenomenon. Notice that postulating the elaborated representation in (35) does not necessarily affect Morén and Zsiga’s analysis in any way, since one could still make reference to the individual tone features in the assessment of constraint violation, i.e. we could simply retain Morén and Zsiga’s analysis while ranking the faithfulness constraints on Contour nodes low enough so that their impact is not felt in the grammar.

3.3.3. \(H\) tone on CVVO syllables

The second question that this paper seeks to answer is why the emphatic high tone is permitted on CVVO syllables when such a combination is otherwise prohibited. That the emphatic reduplicant can surface with a high tone must mean that the overwriting process does not only override Max-BR(T), but also the requirement for these syllables to end low. On Morén and Zsiga’s analysis, the positive constraint C.G.Coda→L rules out CVVO syllables that do not end low. Therefore, in order for CVVO syllables to surface with a H tone in emphatic reduplication, the two constraints Realise-\(\mu\) and \(^{\text{[CC]}_\sigma}\) must be ranked higher than C.G.Coda→L. Consider (37), which illustrates overwriting in CVVO syllables in emphatic reduplication:
(37) Overwriting in CVVO in emphatic reduplication

<table>
<thead>
<tr>
<th>/H+RED+CVVO/</th>
<th>REALISE-µ</th>
<th>C.G.CODA→L</th>
<th>DEP-BR(T)</th>
<th>MAX-BR(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>L</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CVVO - CVVO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>L</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVVO - CVVO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By not copying the L tone from the base onto the reduplicant, the candidate in (37a) violates both MAX-BR(T) and C.G.CODA→L. Additionally, it also incurs a violation of DEP-BR(T) since the emphatic tonal morpheme has no correspondent in the base. The candidate in (37b) on the other hand, does not violate any of these constraints: it ends on a L tone, and the tones on the base and the reduplicant are in correspondence with each other. However, it fails to realise the emphatic tonal morpheme and thus violates REALISE-µ. Therefore, in order for (37a) to be chosen as the optimal candidate, REALISE-µ must be ranked above C.G.CODA→L, DEP-BR(T), and MAX-BR(T). The ranking order REALISE-µ >> {C.G.CODA→L, DEP-BR(T), MAX-BR(T)} guarantees that overwriting takes place in CVVO syllables during emphatic reduplication so that the emphatic high tone surfaces on the CVVO reduplicant.

Besides having to ensure that the emphatic tonal morpheme appears on the reduplicant, we also need to prevent a contour tone from occurring on the reduplicant due to the concatenation of the emphatic tonal morpheme and the tone copied from the base. For this to happen, *[CC]₀ must be ranked above C.G.CODA→L, so as to resist the tendency of CVVO syllables to end low. This is illustrated in tableau (38):

(38) No derived contour tone in emphatic reduplication

<table>
<thead>
<tr>
<th>/H+RED+CVVO/</th>
<th>*[CC]₀</th>
<th>C.G.CODA→L</th>
<th>MAX-BR(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>H</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVVO - CVVO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVVO - CVVO</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The candidate in (38a) violates both Max-BR(T) and C.G.CODA→L by not copying the L tone from the base onto the reduplicant. On the other hand, neither of these constraints is violated by the candidate in (38b), which copies the L tone. However, by copying the L tone along with its Contour node, while having the emphatic tonal morpheme and its associated Contour node appear within the same syllable domain, (38b) violates *[CC]ο. Thus, to rule out (38b) in favour of (38a), *[CC]ο must be ranked above Max-BR(T) and C.G.CODA→L. The ranking order *[CC]ο >> {Max-BR(T), C.G.CODA→L} prevents a derived contour tone from occurring on a CVVO syllable during emphatic reduplication.

(39) Ranking of constraints for Thai reduplication

\[
\begin{align*}
\text{DEP-JIO(T)} & \quad \text{REALISE-µ} & \quad *[CC]ο \\
\text{DEP-BR(T)} & \quad \text{MAX-BR(T)} & \quad \text{C.G.CODA→L} \\
\text{DEP-AF}_{\text{RED}}(T) & \quad \text{MAX-AF}_{\text{RED}}(T) & \quad *L
\end{align*}
\]

Therefore, the occurrence of H tone on CVVO syllables in emphatic reduplication falls out naturally from the existing tools of the analysis, providing additional evidence for the utility of the modified representation in (35). This completes the rankings that sanction the reduplication patterns in Thai, summarised in (39).

4. Conclusion

This paper presented an analysis of reduplication in Thai, with special emphasis on two issues related to emphatic reduplication: fixed autosegmentism and the occurrence of high tone on CVVO syllables. In the spirit of work in OT showing that fixed segmentism can be accounted for without appealing to prespecification, the first goal of this paper was to provide an account of fixed autosegmentism in Thai emphatic reduplication without prespecification. Indeed, in order to have a coherent account of both simple and emphatic reduplication, one cannot appeal to tonal prespecification in the reduplicative affix. This is because in order for simple reduplication to take place successfully, the faithfulness constraints on base-reduplicant identity must be ranked higher than the faithfulness constraints on the reduplicative affix. In turn, this ranking configuration rules out prespecification as the source of fixed autosegmentism in emphatic reduplication. I argued that the emergence of the invariant high tone is morphologically determined and proposed that the emphatic high tone that occurs on the reduplicant has its origins in a tonal morpheme that attaches to the reduplicative compound. As long as REALISE-µ, which ensures that a morpheme has a phonological exponent in the output, remains undominated, the emphatic tonal morpheme is guaranteed to be realised in the phonological output. In order to ensure that overwriting takes place in emphatic reduplication so that the reduplicant surfaces with an invariant high tone, I proposed a modification to Morén and Zsiga’s representations of Thai tones by allowing individual tones within a syllable domain to form a constituent, or Contour node. On this analysis, both the emphatic tonal morpheme and the tone on the base project their own Contour nodes, and overwriting in emphatic reduplication is analysed as the result of a prohibition against multiple Contour
nodes occurring within the same syllable domain, *[CC]₀. With regard to the puzzling phenomenon of high tone bearing CVVO syllables that results from emphatic reduplication, I showed that the occurrence of the emphatic high tone on CVVO syllables falls out naturally from the interaction of *[CC]₀ and C.G.CODA→L, thereby providing additional motivation for and utility to the modified representation of Thai tones.

References
Thai Emphatic Reduplication


SORBUNG,
AN UNDOCUMENTED LANGUAGE OF MANIPUR:
ITS PHONOLOGY AND PLACE IN TIBETO-BURMAN

David Mortensen
University of Pittsburgh
<drm31@pitt.edu>

Jennifer Keogh
Independent Scholar
<jennifer.r.keogh@gmail.com>

Abstract
Sorbung is a Tibeto-Burman language of Manipur, India. This study outlines the phonology of the language based on data elicited from a speaker of residing temporarily in the United States, where he was studying theology. Sorbung shares characteristics with at least two branches of Tibeto-Burman—Tangkhulic and Kuki-Chin—without showing unambiguous evidence of belonging to either group. A word list based upon this collection of data is provided as an appendix.

Keywords: phonology, classification, Tibeto-Burman

1. Introduction
Sorbung is a previously undocumented Tibeto-Burman language of Manipur, India. The current study outlines the major features of the phonology of the language, describing the segment inventories, prosodic structures, phonotactic constraints, and phonological alternations that characterize Sorbung. This sketch is based on lexical and other data elicited from a single native speaker of Sorbung (a male in his thirties) residing temporarily in the United States, where he was studying theology. Elicitation was conducted

The original field research on Sorbung was supported by a UC Berkeley Summer Research grant (2004) to the first author.

Although we have no reason to doubt that our consultant’s speech was typical of the speech of the broader language community, caution should be used in interpreting these results. Like most other speakers of Sorbung in his age group, our consultant was proficient to some degree in Standard Tangkhul, Meithei, Hindi, and English. Some instances of language interference are therefore likely. Since his spouse was not from Sorbung, he spoke to her and his children primarily in Standard Tangkhul. He had resided in the United States for approximately three years at the time the first author worked with him and his use of Sorbung was primarily confined to telephone conversations with family members during that time. However, he expressed considerable confidence in his Sorbung language competence, and was both prompt and consistent in his responses to elicitation prompts. Unfortunately, since the current political situation in Manipur makes fieldwork difficult and dangerous, it is unlikely that work on Sorbung based on the current consultant will ever be possible.

Journal of the Southeast Asian Linguistics Society 4.1:64-114
Copyright vested in the author
Received 29/9/10, revised text accepted 26/3/11

64
primarily in English but Standard Tangkhul and visual illustrations were used for difficult forms. Audio recordings were made of most of the lexical material elicited from the consultant, and these were used to verify and correct the transcriptions made while the first author was working with him. A word list based upon this collection of data is provided as an appendix.

Sorbung is phonologically interesting for a number of reasons. From a comparative standpoint, it shares characteristics with at least two branches of Tibeto-Burman—Tangkhulic (Mortensen & Miller 2009a; Mortensen & Miller 2009b) and Kuki-Chin (VanBik 2009)—without showing unambiguous evidence of belonging to either group. From a synchronic point of view, it displays a variety of productive alternations (both segmental and tonal) that may be of theoretical interest. It also shows interesting patterns of synchronic variation that may provide evidence regarding diachronic developments that have occurred elsewhere in Tibeto-Burman, specifically the development of manner distinctions from the “collapse” of lexical prefixes into the onsets of roots.

1.1. Geographic and Demographic Information
Sorbung is spoken in Sorbung village, a community of about 300 located in the southeast corner of Ukhrul District, Manipur State, India (about 40 km east of the border with Burma). Despite some significant linguistic and cultural differences, the inhabitants of Sorbung village consider themselves to be ethnic Tangkhuls, the majority ethnic group in Ukhrul District. The immediately neighboring villages are also ethnic Tangkhuls, though there is significant Kuki (Thadou) settlement that separates Sorbung from the larger group of Tangkhul villages to the north. According to our consultant, many of these villages are homes to dialects closely related to Sorbung. Others, like Tusom village, are inhabited by speakers of languages belonging unambiguously to the Tangkhulic group. To the south and southwest is a network of Maring and Old Kuki villages. The linguistic and ethnic affiliation of the villages to the east is not clear.

It is not well established whether Tangkhulic (or Kuki-Chin) are separate top-level branches of Tibeto-Burman or whether they form a subgroup with one another or other Tibeto-Burman languages. Traditionally, they have been grouped together as part of a Kuki-Chin-Naga group. However, as Burling (2003) notes, this subgroup has been based more on intuition than argument. Mortensen and Miller (2009b) present some evidence for innovations shared among Kuki-Chin, Tangkhulic, and an additional subset of “Naga” languages and suggest that they share a common ancestor more recent than Proto-Tibeto-Burman. Since there is no consensus on this subject, we will treat Tangkhulic and Kuki-Chin only as distinct taxonomic units, without specifying the relationship between them.

Like other ethnic Tangkhul, the inhabitants of Sorbung village are multilingual. Most aspects of family and village life are negotiated in the local language (here called Sorbung). Church sermons and some schooling are carried out in the ethnic lingua franca we call Standard Tangkhul (a Tangkhulic language originally based on the speech of Ukhrul village, now the district headquarters). Other schooling is done in Meithei—the dominant regional language, Hindi, and English. English is important not only because of its international stature, but because the overwhelming majority are Baptist Christians with a

---

on the speech of a larger number of individuals will be available for some time unless it is carried out by local linguists.
persisting historical relationship to the American Baptist denomination. Loanwords from all of these languages appear in the Sorbung lexicon.

1.2. Previous Work

There is no previous work on Sorbung directly, but at least one published data set from a different dialect of the same language exists. Brown (1837) published word lists collected by Capt. George Gordon, a British political agent in Manipur, for a variety of languages in Manipur. These included “Luhuppa,” which is largely identical to the Tangkhul language of Ukhrul town, “Champung,” the language of another Tangkhul village, and three varieties explicitly labeled as Tangkhul: “Northern Tangkhul,” “Central Tangkhul,” and “Southern Tangkhul.” Of these, Southern Tangkhul is the least like the others but very similar to Sorbung in its lexical, morphological, and phonological properties. Unfortunately, Brown’s word lists were only 60 words long, and provide the only published information on this language up until the present.

Mortensen (2003) and Mortensen and Miller (2009a; 2009b) show that most of the languages spoken by members of the Tangkhul tribe belong to a single subgroup within Tibeto-Burman. They establish criteria for membership in this group, “Tangkhulic,” on the basis of phonological, morphological, and lexical innovations. All of the Tangkhul languages enumerated by Brown (1837) appear to belong to this group, with the exception of “Southern Tangkhul.” Our more complete corpus of data from Sorbung makes it clear that it does not descend from the Proto-Tangkhulic reconstructed by Mortensen and Miller. While Sorbung speakers belong to the same ethnic group as the speakers of Tangkhulic language, their language appears to come from a somewhat different branch of the Tibeto-Burman family.

1.3. The Place of Sorbung within Tibeto-Burman

It is well known that ethnic and linguistic boundaries do not align neatly. This is no less true in South and Southeast Asia than elsewhere. It appears likely that Tangkhuls, as an ethnicity, may be the result of the convergence of two or more populations already related in language and culture. One of these comprises, roughly, the forebears of Tangkhulic speakers; another comprises the ancestors of Sorbung and “Southern Tangkhul” speakers. We contend that this second group spoke a language closely related to the so-called “Old Kuki” languages that are spoken to the south and west of Sorbung. While the population around Sorbung came to identify ethnically with population of Tangkhulic speakers, and while their language was subject to some lexical and structural influence from Tangkhulic languages, they retain a linguistic tradition distinct from their northern neighbors. This was facilitated, no doubt, by the high degree of linguistic heterogeneity present even among Tangkhulic languages. It is apparent that Sorbung and “Southern Tangkhul” fall outside of Tangkhul as a group only when they are compared with Old Kuki languages like Kom (Kom Rem).

This is not to say that Sorbung does not resemble Tangkhul languages in some respects (only that its resemblance to Old Kuki languages is stronger). Like Tangkhul but unlike most Kuki-Chin languages, Sorbung shows no evidence of subject agreement in verbs. It also shows little or no evidence of verbal stem alternation, an important Kuki-
Chin feature. Phonologically, too, Sorbung resembles Tangkhul in some respects. For
example PTB *kr-, *kl- become c- in Sorbung, just as they did in Proto-Tangkhulic.
Compare the following Sorbung forms with related reconstructions from Proto-Tangkhulic
(PTk) and Proto-Tibeto-Burman (PTB):

**Table 1:** Palatal reflexes of velar clusters in Proto-Tangkhulic and Sorbung.

<table>
<thead>
<tr>
<th>PTB</th>
<th>PTk</th>
<th>SORBUNG</th>
</tr>
</thead>
<tbody>
<tr>
<td>*krap</td>
<td>*cap</td>
<td>caap</td>
</tr>
<tr>
<td>*kri(y)</td>
<td>*ci</td>
<td>-cii</td>
</tr>
<tr>
<td>*klaw</td>
<td>*cow</td>
<td>ciu</td>
</tr>
</tbody>
</table>

Unlike Tangkhul, it appears to be a general process of palatalization before
palatal vowels and glides, medial liquids having previously become palatal glides
(Mortensen & Miller 2009b). In Sorbung, it is limited to PTB *kr- and *kl- specifically;
*k- does not become c- before high front vowels, when has undergone secondary
aspiration, does not become c- even before medial *-l-.

**Table 2:** Forms displaying palatization in Tangkhulic but not Proto-Kuki-Chin or
Sorbung.

<table>
<thead>
<tr>
<th>PTB</th>
<th>PTk</th>
<th>PKC</th>
<th>SORBUNG</th>
</tr>
</thead>
<tbody>
<tr>
<td>*klum</td>
<td>*jim</td>
<td>*kh lum</td>
<td>kʰum</td>
</tr>
<tr>
<td>*g-loy</td>
<td>*-jí</td>
<td>*kh līí</td>
<td>kʰii</td>
</tr>
<tr>
<td>*g-la</td>
<td>—</td>
<td>*kh lāa</td>
<td>kʰaa</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>*kh l uak</td>
<td>kʰuak</td>
</tr>
<tr>
<td>—</td>
<td>*-ci</td>
<td>*kii</td>
<td>kii</td>
</tr>
<tr>
<td>*kəy</td>
<td>*sa-kií</td>
<td>saksií</td>
<td>‘deer’</td>
</tr>
<tr>
<td>*g(y)əy</td>
<td>*-ci</td>
<td>*kii (PNC)</td>
<td>bakíí</td>
</tr>
</tbody>
</table>

Furthermore, when lexical items with palatal stops are compared with their cognates in the
Old Kuki languages Kom and Moyon, and the lexically-similar language Puiron, it is
apparent that these languages share the same pattern of palatalization as Sorbung:

24 Many Kuki-Chin languages have a morphological process that derives an inflected stem form (Form II)
from the basic form (Form I). For a useful comparative study of this phenomenon, see Button
(2009:182ff). Note, however, that there are Old Kuki languages which display evidence of neither
subject agreement nor stem alternation (Grierson 1903:288-290).

25 Except where otherwise indicated, Proto-Tibeto-Burman reconstructions are from Matisoff (2003) and
Proto-Tangkhulic reconstructions are from Mortensen & Miller (2009b).

26 Except where otherwise indicated Proto-Kuki-Chin (PKC) reconstructions are from VanBik (2009).
Reconstructions labeled “DRM” are due to the first author.

27 Ethnically, the speakers of Puiron are a subset of the Nruangmei tribe, most of whom speak languages
belonging to the Zeme group (Burling 2003). Burling suggests that Puiron is a Zeme language with many
Kuki loans. Below we suggest that Puiron is, like Sorbung, a “Nagafied” Old Kuki language.
Table 3: Parallels in palatalization between Sorbung, Old Kuki (Kom and Moyon), and Puiron.

<table>
<thead>
<tr>
<th>PTB</th>
<th>PKC</th>
<th>SORBUNG</th>
<th>KOM</th>
<th>MOYON</th>
<th>PUIRON</th>
</tr>
</thead>
<tbody>
<tr>
<td>*krap</td>
<td>*krap</td>
<td>cùap</td>
<td>kə-çəp</td>
<td>—</td>
<td>cap</td>
</tr>
<tr>
<td>*kri(y)</td>
<td>*kriʔ</td>
<td>ʔənciʔ</td>
<td>—</td>
<td>—</td>
<td>kací</td>
</tr>
<tr>
<td>*klan̥</td>
<td>*kroŋ (DRM)</td>
<td>cóŋ</td>
<td>toŋ rih</td>
<td>lcoŋ</td>
<td>—</td>
</tr>
<tr>
<td>*tsywap</td>
<td>*tsuap</td>
<td>ʔəcúap</td>
<td>—</td>
<td>—</td>
<td>əcú</td>
</tr>
<tr>
<td>*m-t(s)i</td>
<td>*tsii</td>
<td>macíi</td>
<td>—</td>
<td>—</td>
<td>mací</td>
</tr>
</tbody>
</table>

In other words, palatal reflexes of PTB *krJ, *klJ, etc. cannot be used to associate Sorbung with Tangkhulic in preference to Kuki-Chin.

Another area of similarity between Tangkhulic and Sorbung is their personal pronouns. In this respect, Sorbung actually is quite different from most Kuki-Chin languages but similar to Tangkhulic languages (when the family is taken as an aggregate).

Table 4 compares the Sorbung person pronouns (both the bound-root forms and the free forms) with their cognates in Proto-Tangkhul:

Table 4: Comparison of Sorbung and Proto-Tangkhulic pronominal forms.

<table>
<thead>
<tr>
<th>FREE PRONOUNS</th>
<th>BOUND PRONOMINAL PROCLITICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SORBUNG</td>
<td>PTK</td>
</tr>
<tr>
<td>FIRST ʔoo (&lt;*ej)</td>
<td>*ʔej, *ʔi</td>
</tr>
<tr>
<td>SECOND naŋ</td>
<td>*naŋ</td>
</tr>
<tr>
<td>THIRD maa</td>
<td>*me, *ʔa</td>
</tr>
</tbody>
</table>

Apart from the third person proclitic, each of the Sorbung pronouns has a clear cognate in Tangkhulic. This is unsurprising for the second person pronouns, which are widespread in Tibeto-Burman (Matisoff 2003) and are also found in Kuki-Chin languages. For first and third person pronouns, however, Kuki-Chin languages usually have both free forms and politics from *kai ‘1’ and *ʔa ‘3’ (La Polla 2006). However, this match does not tie Sorbung to Tangkhulic exclusively. Meithei pronouns match those in Sorbung even more exactly:

Table 5: Comparison of Sorbung and Meithei personal pronouns.

<table>
<thead>
<tr>
<th>FREE PRONOUNS</th>
<th>BOUND PRONOMINAL PROCLITICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SORBUNG</td>
<td>MEITHEI</td>
</tr>
<tr>
<td>FIRST ʔoo (&lt;*ej)</td>
<td>ʔay</td>
</tr>
<tr>
<td>SECOND naŋ</td>
<td>naŋ</td>
</tr>
<tr>
<td>THIRD maa</td>
<td>má</td>
</tr>
</tbody>
</table>

The idea that the Sorbung pronouns (and possibly, those of Tangkhulic) have been borrowed from the politically and economically powerful Meithei-speaking community cannot be dismissed out of hand. It is also true that all of the Sorbung pronouns seem to have cognates among Kuki-Chin languages. Consider, for example, the pronouns of Kuki Thadou (Hyman 2007b):
Table 6: Comparison of Sorbung and Kuki Thadou personal pronouns.

<table>
<thead>
<tr>
<th></th>
<th>FREE PRONOUNS</th>
<th>BOUND PRONOMINAL PROCLITICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIRST EXCLUSIVE</td>
<td>ʔoo (&lt;*ej)</td>
<td>?i-, ?əN-</td>
</tr>
<tr>
<td>FIRST INCLUSIVE</td>
<td>eĩ</td>
<td>i-</td>
</tr>
<tr>
<td>SECOND</td>
<td>naŋ</td>
<td>na-</td>
</tr>
<tr>
<td>THIRD</td>
<td>maa</td>
<td>ma-</td>
</tr>
</tbody>
</table>

The proclitic i- marks ‘first person inclusive’ in some Kuki-Chin languages. The free first person pronoun in Puiron is ai, probably cognate with Sorbung ʔoo. In many Old Kuki languages, the third-person pronoun consists of *a- ‘3’ plus a formative like ma: Hallam a-ma, Aimol ə-ma, Kolren ə-ma. In Purum, this formative acts by itself as the marker of third person, as in mo-ju ‘3-NOM’. The same is true of Lamgang, where the 3sg prefix is ma-(Grierson 1903:282-284). In summary, the evidence supporting a special relationship between Sorbung and Tangkhulic breaks down under examination. It is unusual for Sorbung to share a characteristic with Tangkhulic unless the same characteristic is shared with Old Kuki languages.

The term “Old Kuki” is not without its complexities. As an ethnic term, it has been used to describe the Kuki-identified peoples who settled in Manipur at a relatively early time, prior to the arrival of the larger and more powerful Kuki groups like the Mizo (Lushai) and Thadou. These include the Anal, Aimol, Chothe, Purum, Koiren, Kom (Kom Rem), Lamkang, Moyon, Chiru, Tarao, and Vaiphei. Some of these groups, for political reasons, now identify as Naga (e.g. the Lamkang and the Tarao). We propose that certain other Naga groups, such as the Puiron (a ethnic subgroup of the Nruangmei) are the descendents of outlying Old Kuki groups that affiliated themselves ethnically with their Naga neighbors. It is not completely clear what groups should be characterized as Old Kuki and whether the languages of these people form a genetic subgroup (p.c., Kenneth VanBik). The internal coherence of Old Kuki was assumed in LSI, and this assumption has been carried on in part because of the lack of reliable data with which to test it.

VanBik, in his dissertation on Proto-Kuki-Chin, tentatively classified Old Kuki languages as part of his Northern (Zo) group, which is, in turn, part of his Peripheral group. This grouping is based on lexical resemblances between Northern Chin languages like Thadou and Paite and the Old Kuki language Purum. However, this speculation was removed from the published version of this work (VanBik 2009). We argue that the old Kuki languages cannot be part of VanBik’s Northern subgroup. The criterial innovation definition the Peripheral group is the fortition of PKC *r- > *g-. However, there is no evidence of this change in Old Kuki languages (or in Sorbung). Etyma with PKC *r- retain r- in Anal, Kom, Moyon, Puiron, and Sorbung, as they do in Mizo (a Central Kuki-Chin language) but develop g- Tedim, a Northern Kuki-Chin languages, as show in Table 7.
Table 7: Reflexes of Proto-Kuki-Chin *r- and *hr- in Tedim, Old Kuki (Anal, Kom, and Moyon), Puiron, and Sorbung.

<table>
<thead>
<tr>
<th>PKC</th>
<th>MIZO</th>
<th>TEDIM</th>
<th>ANAL</th>
<th>KOM</th>
<th>MOYON</th>
<th>PUIRON</th>
<th>SORBUNG</th>
</tr>
</thead>
<tbody>
<tr>
<td>*ruʔ</td>
<td>ruʔ</td>
<td>ruʔ</td>
<td>ru</td>
<td>ru</td>
<td>ru</td>
<td>?ʔ-ruʔ</td>
<td>‘bone’</td>
</tr>
<tr>
<td>*p-ruul</td>
<td>ruul</td>
<td>gu:lu</td>
<td>p-ruul</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>*rik</td>
<td>rik</td>
<td>gik’</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>rik</td>
<td>‘heavy’</td>
</tr>
<tr>
<td>*ril</td>
<td>ril</td>
<td>gi²</td>
<td>—</td>
<td>ʔ-ri</td>
<td>l-ʔr̥r̥</td>
<td>—</td>
<td>ʔ-rii</td>
</tr>
<tr>
<td>*raʔ</td>
<td>rəʔ</td>
<td>gəʔ</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>tak-ra</td>
<td>‘fruit’</td>
</tr>
<tr>
<td>*ruy ~ *hruy</td>
<td>hruy</td>
<td>gui</td>
<td>rui</td>
<td>rui</td>
<td>rui</td>
<td>rui</td>
<td>‘vine; tendon’</td>
</tr>
<tr>
<td></td>
<td>*hriŋ</td>
<td>hriŋ</td>
<td>—</td>
<td>—</td>
<td>l-ʔriŋ</td>
<td>riŋ</td>
<td>Riŋ</td>
</tr>
</tbody>
</table>

If all Northern Kuki-Chin languages are in the Peripheral group, and PKC *r- > g- in the common ancestor of all peripheral languages, the Old Kuki languages cannot be part of the Northern group. Given the morphological and lexical conservatism of Old Kuki languages, this introduces the possibility that they constitute a separate top-level branch of Kuki-Chin. However, they have been little-exploited in comparative reconstruction because relatively little data on them has been available. If it can be established that Sorbung belongs to this group, the body of data that we have assembled here could be of considerable value in reconstructing the history of Kuki-Chin as a family.

Aside from the evidence that has been presented thus far, there are three types of evidence that suggest Sorbung is a member of the Kuki-Chin family with specific affinities to the Old Kuki languages: shared lexical innovations, shared morphological innovations, and shared phonological innovations.

One striking aspect of the Sorbung lexicon is the presence of many lexical items that are widespread within Kuki-Chin but which are not widely found elsewhere in Tibeto-Burman:

Table 8: Lexical resemblances between Sorbung and Kuki-Chin.

<table>
<thead>
<tr>
<th>SORBUNG</th>
<th>PKC</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>som</td>
<td>*som</td>
<td>‘ten’</td>
</tr>
<tr>
<td>pan</td>
<td>*puan</td>
<td>‘blanket; cloth’</td>
</tr>
<tr>
<td>bùu</td>
<td>*ɓuʔ</td>
<td>‘rice (cooked)’</td>
</tr>
<tr>
<td>kèe</td>
<td>*kee</td>
<td>‘foot’</td>
</tr>
<tr>
<td>lu-kaŋ</td>
<td>*luu</td>
<td>‘head’</td>
</tr>
<tr>
<td>lu-siíp</td>
<td>*tship</td>
<td>‘crown (of head)’</td>
</tr>
<tr>
<td>ʔèk</td>
<td>*ʔeek</td>
<td>‘feces; shit’</td>
</tr>
<tr>
<td>ma-rái</td>
<td>*raay</td>
<td>‘be pregnant’</td>
</tr>
<tr>
<td>cèm</td>
<td>*tsem</td>
<td>‘knife’</td>
</tr>
<tr>
<td>ʔəták</td>
<td>*tak</td>
<td>‘flesh’</td>
</tr>
<tr>
<td>wàm</td>
<td>*wom</td>
<td>‘black’</td>
</tr>
<tr>
<td>dài</td>
<td>*daay</td>
<td>‘cool; quiet’</td>
</tr>
<tr>
<td>paasàa</td>
<td>*pa-sal</td>
<td>‘husband; man’</td>
</tr>
</tbody>
</table>
However, Old Kuki languages (and Puiron) have even more specific lexical resemblances to Sorbung:

Table 9: Innovative lexical items shared between Sorbung, Old Kuki (Kom, Aimol, Moyon, Chiru) and Puiron.

<table>
<thead>
<tr>
<th>SORBUNG</th>
<th>KOM</th>
<th>AIMOL</th>
<th>MOYON</th>
<th>CHIRU</th>
<th>PUIRON</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʔóo</td>
<td>ei</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>ai</td>
</tr>
<tr>
<td>̕hóon</td>
<td>ka-hoŋ</td>
<td>ə-joŋ</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>ʔiin</td>
<td>in</td>
<td>—</td>
<td>lń</td>
<td>—</td>
<td>in</td>
</tr>
<tr>
<td>kʰoɔp</td>
<td>—</td>
<td>Khop</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>ʔən-çáŋ-lăm</td>
<td>ēŋ</td>
<td>—</td>
<td>ti-cáŋ</td>
<td>—</td>
<td>caŋ-lam ‘right (side)’</td>
</tr>
<tr>
<td>ʔən-suŋ</td>
<td>in-suŋ</td>
<td>ə-suŋ</td>
<td>—</td>
<td>a-su</td>
<td>—</td>
</tr>
<tr>
<td>ʔìin</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>lée</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>lo</td>
</tr>
<tr>
<td>Kaa</td>
<td>kha-wa</td>
<td>Kha</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Wa</td>
<td>hi-wa</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>ham-buu</td>
<td>hum-pui</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>kóon</td>
<td>koŋ</td>
<td>—</td>
<td>kuŋ</td>
<td>—</td>
<td>koŋ   ‘waist; back’</td>
</tr>
</tbody>
</table>

The lexical similarities between Sorbung and Old Kuki languages are very strong and indicate, at the very least, a history of contact between southern Tangkhuls and Kuki-Chin speaking peoples. The morphological similarities (in both inflection and derivation) help to reinforce the idea that the relationship is a genetic one. Hartmann (2001) shows that there are a pair of valence-related prefixes in Daai Chin (a Southern Kuki-Chin language): ng-, occurring in low-valence verbs, and m-, occurring in transitive (especially causative) verbs:

(1) a. thei ‘be clever’ ng-thei ‘learn’ m-thei ‘teach’
    b. püi ‘be together’ ng-püi ‘be included’
    c. kyüh ‘be afraid’ m-kyüh ‘make afraid’

Cognates of these two prefixes appear in Sorbung as ʔəN- and mə-, respectively:

(2) a. ʔon-tʰee ‘awake; be awake’ mə-tʰee ‘get somebody up’
    b. ʔon-ci ‘fear; be afraid’ mə-ci ‘frighten’

Various innovative inflectional markers found in Sorbung are also found in Old Kuki languages. The Sorbung progressive suffix -om (< ʔóm ‘exist; be a place’) has a cognate at least in Aimol. Compare Aimol sekor a-chongaa-om ‘horse 3-ride-PROG (he is riding a horse)’ with Sorbung tʰiŋkoŋ-ua kiák-ôm-ôo ‘tree-DEM fall-PROG-DECL (the tree is falling)’. The Sorbung genitive suffix -ta also appears to have cognates in Langrong, Purum, and Kolren (all of which have a formally similar suffix in genitive pronouns).

Since comparative reconstruction of languages in the India-Burma borderlands area remains in its infancy, relatively speaking, it is difficult to speak of phonological innovations in terms as explicit as would be desirable. As noted above, Sorbung shares with Old Kuki languages a common development of PTB *kr-, *kl-, *ts(y)-, and *dzy-. It was also noted that a similar change has occurred in Tangkhulic and that it is not entirely
possible to establish the complete independence of these changes. The same may be said for another cluster of changes that are shared by Kuki-Chin, Sorbung, and Tangkhulic:

**Table 10:** Reflexes of Proto-Tibeto-Burman coronal fricatives and affricates in Proto-Kuki-Chin, Proto-Tangkhulic, and Sorbung.

<table>
<thead>
<tr>
<th>PTB</th>
<th>PKT</th>
<th>PKC</th>
<th>SORBUNG</th>
</tr>
</thead>
<tbody>
<tr>
<td>*s-, *sy₁-</td>
<td>*tʰ-</td>
<td>*th-</td>
<td>tʰ-</td>
</tr>
<tr>
<td>*sy₂-</td>
<td>*s-</td>
<td>*sʰ-</td>
<td>s-</td>
</tr>
<tr>
<td>*ts-</td>
<td>*s-</td>
<td>*s-, *θ-</td>
<td>s-</td>
</tr>
<tr>
<td>*tsy-</td>
<td>*ts-, *c-</td>
<td>*ts-, s-</td>
<td>c-, s-</td>
</tr>
<tr>
<td>*dz-</td>
<td>*ts-</td>
<td>*ts-</td>
<td>c-</td>
</tr>
<tr>
<td>*dzy-</td>
<td>*c-, *ts-</td>
<td>*ts-</td>
<td>c-</td>
</tr>
</tbody>
</table>

One piece of evidence that at least the Kuki-Chin and Sorbung innovations are shared is the fact that irregular and variable etyma tend to pattern the same way in PKC and Sorbung (given that PKC *ts- regularly corresponds to Sorbung c-):

**Table 11:** Comparison of regular and variable developments from PTB *ts-, *ts(y)-, and *dzy-. Note that Sorbung patterns like PKC rather than PTk.

<table>
<thead>
<tr>
<th>PTB</th>
<th>PKT</th>
<th>PKC</th>
<th>SORBUNG</th>
</tr>
</thead>
<tbody>
<tr>
<td>*tsam</td>
<td>*sam</td>
<td>*sam</td>
<td>sàam ‘hair of head’</td>
</tr>
<tr>
<td>*ts(y)a-t</td>
<td>*tsa</td>
<td>*saa</td>
<td>saa ‘hot; ill’</td>
</tr>
<tr>
<td>*ts(y)a</td>
<td>*tsa</td>
<td>*θaa</td>
<td>saa ‘child’</td>
</tr>
<tr>
<td>*dzya-k/t/n</td>
<td>*tsa</td>
<td>*θaʔ</td>
<td>saa ‘eat’; sak ‘feed’ ‘eat; food; feed’</td>
</tr>
<tr>
<td>*tsywap</td>
<td>*tsup</td>
<td>*tsuap</td>
<td>cúap ‘spleen’ ‘lung’</td>
</tr>
</tbody>
</table>

It is also the case that Sorbung and Proto-Kuki-Chin developed initial glottal stop from the same—reasonably diverse—set of onsets:

**Table 12:** Sources of Proto-Kuki-Chin and Sorbung initial glottal stop.

<table>
<thead>
<tr>
<th>PTB</th>
<th>PKT</th>
<th>PKC</th>
<th>SORBUNG</th>
</tr>
</thead>
<tbody>
<tr>
<td>*ʔaːr</td>
<td>*ar</td>
<td>*ʔaar</td>
<td>ʔaa chicken; fowl</td>
</tr>
<tr>
<td>*kʷəy</td>
<td>*hwɪ</td>
<td>*ʔuy</td>
<td>ʔʉ ́ʉ dog (Canis familiaris)</td>
</tr>
<tr>
<td>*hyen</td>
<td>—</td>
<td>*ʔen</td>
<td>ʔen look</td>
</tr>
<tr>
<td>*k-yim</td>
<td>*ʃim</td>
<td>*ʔim</td>
<td>ʔin house</td>
</tr>
</tbody>
</table>

Perhaps the most striking evidence of a close relationship between Old Kuki and Sorbung is that fact that at least one Old Kuki language, Kom, shares with Sorbung a productive voicing alternation that seems to be active in the same environments. This alternation is discussed at length in Section 0 below.

Based on the evidence that we have presented here, we argue that Sorbung is best seen as a Kuki-Chin language (closely related to Old Kuki languages like Kom, Aimol, and Moyon) that has come under influence from Tangkhulic languages. We further suggest that it is not the only “Nagafied” Kuki language and that Puiron might be another case of this type, a matter that cannot be discussed at length here.
1.4. Important Characteristics

The significance of Sorbung partly lies in that it demonstrates changes in progress that are important for accounting for historical developments elsewhere in Tibeto-Burman. These changes orbit around minor syllables (“presyllables” or “prefixes”), both their development from etymological major syllables and their collapse into following major syllables, yielding a new stop series.

Sorbung shows a pervasive pattern of reduction in pretonic syllables. Compounds consisting of two noun roots that would consist of heavy syllables in isolation are realized as a sequence of a light syllable and heavy syllable. This stress-conditioned pattern of reduction creates a kind of structure that is intermediate between full disyllables and sesquisyllables and provides a likely pathway for the development of new minor syllable prefixes from compounded roots.

In true minor syllables, though, another process is in progress. The vocalism in such syllables is now completely predictable; thus, the /ə/ vowels in minor syllables are analyzed here as epenthetic. However, in certain environments, epenthesis variably fails to occur. Thus, instead of [ʔəkái] ‘hips’, our consultant often produced [ʔkái]. Such preglottalized productions could easily develop into fortis [kˀái] or geminate stops [kːái], which could be reanalyzed by learners as aspirates [kʰái]. Since the prefix [ʔə ~ ʔ] has an unpredictable distribution, such a change would lead to what would appear to be, retrospectively, an unconditioned change in manner. Matisoff (2003:87ff) has suggested that many of the apparently sporadic manner correspondences among Tibeto-Burman onsets should be attributed to the effects of lexical prefixes that have collapsed into the following major syllable onset. Because the distribution of Tibeto-Burman lexical prefixes seems to be heavily determined by analogical processes, but seldom shows deterministic grammatical conditioning, the presence of a prefix often has to be inferred from the effect its existence is meant to explain. While this presents a methodological problem (invoking a hidden cause), cases like Sorbung give additional empirical support to the idea that semiproductive lexical prefixes can give rise to new laryngeal distinctions in onsets.

2. Syllable Structure and Prosody

Sorbung, like many other Tibeto-Burman languages (indeed, many Southeast Asian languages generally), tends to have sesquisyllabic word stems. That is, word stems often consist of a minor-syllable “prefix” (which is sometimes a morphological prefix) and a major syllable “root”. This pattern is consistent with a general tendency towards right-headed (iambic) prosodic constituents. It is also essential to understanding the distribution to tone (which is only contrastive on major syllables) and the general syllable-structure constraints in the language.

2.1. Stress

Stress in Sorbung has the following correlates:

1. **Duration**: Unstressed vowels are never long; underlying long vowels are shorted in unstressed syllables.
2. **Amplitude**: Stressed syllables are louder than unstressed syllables.
3. **Pitch**: Pitch excursions are more pronounced in stressed syllables.
All of these correlates point to a general iambic pattern of stress. This is not to say, however, that Sorbung stress is predictable without reference to morphological structure. The following general principles seem to hold:

1. Minor syllables are never stressed. This may be due to the fact that the vowels in these syllables are epenthetic.
2. The last root syllable in a compound stem, not counting derivational suffixes, is stressed.
3. Two consecutive unstressed syllables are not tolerated.
4. Stress assignment is cyclic. ‘Underarm hair,’ from the compound cubˈlaa (cuup ‘breast’ plus laa ‘crotch’) plus mii ‘hair’ is cub, laa mii rather than *ˌcuublaˈmii.
5. Suffixes do not affect the stress of the base to which they are attached: ˌlook ‘big’, ˈloog- oo ‘big-DECL’, ʔˈuu ‘dog’, ʔˈuu-e ‘dog-NOM’.
6. The members of one set of suffixes are always stressed (including the obligatory final suffixes marking mood). Another set of suffixes are always unstressed (including case suffixes).
7. Some non-lexicalized compounds, particularly verbs compounded with nominal objects, may not display stress subordination. Thus, while móo ‘fire’ is unstressed and reduced in mó-sèm ‘fire-blow; blow on fire’, cuup ‘breast’ is fully stressed and not reduced in cuup-ʔin-e ‘breast-drink-IMP; suckle!"

The following data illustrate these principles more fully. It can be seen in (3) that /ŋaa/ ‘fish’ is stressed when final or when preceded by a minor syllable, but is unstressed when it is initial:

(3) a. ˈŋaa ‘fish’
   b. ʔəJˈŋaa ‘0’ + ‘fish’
   c. ʔəJˌŋaaJˈmíi ‘0’ + ‘fish’ + ‘hair/feather’ ‘scale; fin’
   d. ɲaˈmíi ‘fish’ + ‘hair/feather’ ‘flying fish’

The examples in (4-6) show that /kee/ ‘foot’ is unstressed when it occurs penultimately (5) but is stressed before a minor syllable (6):

(4) ˈkèe ‘foot’
(5) a. ké-ˈlúuŋ ‘foot’ + ‘print’ ‘footprint’
   b. ké-ˈkíuk ‘foot’ + ‘crippled’ ‘lame’
   c. ké-ˈbůu ‘foot’ + ‘great one’ ‘big toe’
   d. ké-ˈsáa ‘foot’ + ‘child’ ‘little toe’
(6) a. ˌkèe-məˈtòo ‘foot’ + ‘thigh’ ‘crotch’
   b. ˌkèe-məˈrái ‘foot’ + ‘calf’ ‘calf’
   c. ˌkèe-məˈjáa ‘foot’ + ‘palm/sole’ ‘sole’
   d. ˌkèe-məˈjúum-ˈràa ‘foot’ + ‘digit’ + ‘fruit’ ‘toe’

In each word, one stress is more prominent than the others (specifically, it has a greater amplitude and longer duration). This is indicated here as primary stress [ˈ]. Other stressed syllables are marked as bearing secondary stress [ˌ]. By way of illustration, duration of ‘foot’ in (5a-d) is much shorter than the duration of ‘foot’ in (6a-d). This, in turn, is somewhat shorter than the duration of ‘foot’ in (4). We have not identified any
cases where the distinction between primary and secondary stress is contrastive. However, its patterning does suggest that both Sorbung feet and prosodic words are right-headed.

2.2. Word Structure

Barring loanwords, non-compound stems in Sorbung are always monosyllabic or sesquisyllabic. Monosyllabic stems consist of a single major syllable. Sesquisyllabic stems consist of a minor syllable followed by a major syllable. Underlyingly, we believe that all Sorbung roots have a single vowel and that the predictable schwa-vocalism in minor syllables is the result of epenthesis. However, the system is easier to describe if it is approached from the surface.

2.3. Minor syllables

With one exception, the minor syllables in Sorbung have the shape Cə, where C can be any of the onsets allowed in major syllables except for /p/, /t/, /w/, /j/, and /ʃ/. The exception to this pattern is /ʔəN/, where N is a nasal that shares the same place of articulation as the following consonant. In all instances of minor syllables, [ə] is the only vowel to appear (though it is possible that the quality of these vowel is affected by coarticulation with following vowels). The complete inventory is given in the following table:

<table>
<thead>
<tr>
<th>Table 13: Minor syllables in Sorbung.</th>
</tr>
</thead>
<tbody>
<tr>
<td>bə</td>
</tr>
<tr>
<td>bəkii ‘parrot’</td>
</tr>
<tr>
<td>pə</td>
</tr>
<tr>
<td>tə</td>
</tr>
<tr>
<td>tərət ‘tie’</td>
</tr>
<tr>
<td>cə</td>
</tr>
<tr>
<td>cəkòo ‘river’</td>
</tr>
<tr>
<td>kə</td>
</tr>
<tr>
<td>kəléen ‘loose’</td>
</tr>
<tr>
<td>ʔə</td>
</tr>
<tr>
<td>ʔəntʰée ‘awake’</td>
</tr>
<tr>
<td>ʔəŋkʰaaw ‘fish’ sp.’</td>
</tr>
<tr>
<td>pʰə</td>
</tr>
<tr>
<td>pʰəláa ‘wing’</td>
</tr>
<tr>
<td>mə</td>
</tr>
<tr>
<td>məmìt ‘extinguish’</td>
</tr>
<tr>
<td>sə</td>
</tr>
<tr>
<td>səkúu ‘porcupine’</td>
</tr>
<tr>
<td>rə</td>
</tr>
<tr>
<td>lə</td>
</tr>
<tr>
<td>ləʃùa ‘put’</td>
</tr>
<tr>
<td>ʔəN</td>
</tr>
<tr>
<td>ʔəŋkʰaaw ‘fish’ sp.’</td>
</tr>
</tbody>
</table>
In actual production, each of the Cə minor syllables has a variant without a schwa. Thus, /s-µuk/ ‘cattle’ can be realized either as [samûk] or [smûk] and /ʔ-kʰaaw/ ‘grasshopper’ can be realized either as [ʔkʰaaw] or [ʔkʰaaw]. The variants with initial consonant clusters occur more frequently when there is a sharp increase in sonority between the two consonants.

Problematically, our database contains minor syllables with /b/ but not with any other voiced stop (in spite of the fact that /d/ appears as the onset of major syllables). We believe this to be an accidental gap. Also, aspirated-stop minor syllables have a restricted distribution, occurring only before sonorant onsets. However, in this environment they contrast with their unaspirated counterparts, at least on the surface:

(7) a. kʰərâan ‘spider’ kəraak ‘voice’
   b. pʰəlâa ‘wash’ pəlái ‘snake’
   b. tʰərìi ‘boundary’ tərât ‘tie’

We propose that this pattern actually results from a morphological cause: the aspirated stop minor syllables occur only in monomorphemic stems and are the result of epenthesis into underlying stop-sonorant clusters. The minor syllables with unaspirated stops occur in both monomorphemic stems, as in (7) and bimorphemic stems where the minor syllable is a lexical prefix. Underlying, we posit that monomorphemic stems like those in (7) have forms as in (8) while bimorphic stems have forms like those in (9):

(8) a. /kʰrâã/ ‘spider’ /kraak/ ‘voice’
   b. /pʰlâã/ ‘wash’ /plâã/ ‘snake’
   b. /tʰrìã/ ‘boundary’ /trât/ ‘tie’

(9) a. /m-mît/ ‘extinguish’ /t-tîr/ ‘bar’
   b. /r-nâa/ ‘five’ /l-fîa/ ‘put’
   b. /s-loo/ ‘buffalo’ /ŋ-nâãr/ ‘tie’

This accounts for the freer distribution of unaspirated-stop minor syllables versus aspirated-stop minor syllables and for the greater frequencies of schwa-less variants stems in the aspirated-stop minor syllable group.

In some cases, there is compelling evidence that minor syllables are distinct morphological units; in other cases, the data suggest that minor syllables are simply part of a root. Evidence for morphological independence includes the following:

1. A root may occur with or without a minor syllable, depending on morphological or syntactic context.
2. The same root may occur with different minor syllables in related lexical items.
3. A cluster of stems containing the same minor syllable share a component or meaning or belong to the same semantic field.
4. The minor syllable is transparently related to an independent lexical item.
According to these criteria, at least the following minor syllables should be considered to be prefixes, rather than parts of the root:

(10) **ba-** fossilized animal prefix
    a. *bərīi* ‘snake’
    b. *bəkīi* ‘parrot’

(11) **ca-** ‘water’ prefix (*cūh* ‘water’)
    a. *cəkāā* ‘drought’
    b. *cəlōk* ‘flood’
    c. *cəluāŋ* ‘flow’

(12) **ka-** lexical prefix
    a. *kənīiŋ* ‘two’ (*semnīi* ‘twenty’)
    b. *kərūuk* ‘six’

(13) **ʔə-** lexical prefix
    a. *ʔərūu* ‘bone’
    b. *ʔəcūh* ‘egg’ (*ʔəcūh-ʔəcūh* ‘lay egg’)
    c. *ʔənāa* ‘fish’ (*ʔənāa* ‘flying fish’)
    d. *ʔəmīi* ‘hair; fur’ (*ʔəmīi* ‘facial hair’)

(14) **mə-** body-part prefix (< PTB *mi* ‘human’?)
    a. *məsāa* ‘body’
    b. *mətōo* ‘lap; thigh’
    c. *mətīn* ‘nail; claw’

(15) **mə-** ‘mouth’ prefix (< *mor* ‘mouth’?)
    a. *məlit* ‘vomit’
    b. *məjūup* ‘kiss’
    c. *məcūmīi* ‘saliva’
    d. *məkʰāa* ‘jaw; chin’

(16) **mə-** fossilized static prefix
    a. *mənēe* ‘be soft’
    b. *məkʰāi* ‘be crooked’
    c. *məsāaŋ* ‘be high’
    d. *mənēm* ‘be low’

(17) **mə-** productive causative prefix (< *mii* ‘give’?)
    a. *mətʰée* ‘wake up (v.t.)’ (*ʔənthée* ‘wake up (v.i.)’)
    b. *məcàap* ‘make cry’ (*càap* ‘hurt’)
    c. *məʔōo* ‘make bite’ (*ʔōo* ‘bite’)

(18) **ʔəN-** valence-decreasing prefix
    a. *ʔəntʰèe* ‘wake up (v.i.)’
    b. *ʔəncīi* ‘fear (v.i.)’
    c. *ʔəndīa* ‘fall (v.i.)’
(19) **lə-** ‘hand/arm’ prefix (*< PTB *l(y)ak ‘hand/arm’?)
   a. ləkʰàa ‘shoulder’
   b. ləʃùa ‘put’
   c. ləkium ‘bar’

(20) **sa-** ‘animal’ prefix (*< saa ‘animal’)
   a. səloo ‘buffalo’
   b. səriám ‘mithun’
   c. səkíi ‘deer’
   d. səkúu ‘porcupine’

None of the aspirated stop-minor syllables belong in this set, consistent with the hypothesis that the aspirated stops are morphologically part of the root.

Minor syllables are subject to an alternation that does not appear in major syllables. In compounds, nasal codas do not assimilate in place to following obstruents:

(21) a. [inpì] ‘host (house owner)’
   b. [puankʰóŋnaa] ‘loom’
   c. [somtʰúum] ‘thirty’
   d. [ciíntòon] ‘hill’

The nasal of the minor syllables [ʔəN] and [təN] assimilates in place to the following segment. The following examples show this for [ʔəN], which appears as a possessive prefix on kinship terms and as a stative/valency-reducing prefix on verbs:

(22) a. [ʔəmbíi] ‘paternal grandfather (of ego)’
   b. [ʔəmbúu] ‘maternal grandfather (of ego)’
   c. [ʔəntʰée] ‘awake’
   d. [ʔəntʰúum] ‘thirty’
   e. [ʔəndía] ‘fall’
   f. [ʔənnuu] ‘mother (of ego)’
   g. [ʔəngguu] ‘child’
   h. [ʔəŋkʰau] ‘fish species’

Similarly, the nasal in [təN], for which no general function is discernable, shares the place of articulation with the subsequent onset.

(23) a. [təmbuelàm] ‘left hand’
   b. [təmbàak] ‘valley’
   c. [təŋjük] ‘pestle’
   d. [təŋguáp] ‘headdress’
   e. [təŋgoon] ‘ant-eater’
   f. [təŋkʰai] ‘half’

---

This form is probably borrowed from Meithei (see Chelliah 1997).
Before palatals, as in (23c) ‘pebble’, the nasal becomes coronal. This may be due to a general restriction against palatal nasals in Sorbung. Another interpretation of the evidence is that nasal is underlingly /n/. However, we suggest that the nasal is underlingly unspecified for place, for the following reasons:

1. Throughout the corpus, nasal codas generally do not assimilate in place to following obstruents; the different, assimilating behavior of [ʔəN] and [təN] can be explained by the “need” for the final nasals to acquire place features.
2. Nowhere in the data does one of these formatives not occur before an obstruent from which it could acquire place features.

This position is subject to verification, pending the availability of further data.

2.4. **Major syllable types**

Major syllables in Sorbung may be either open or closed and may have either long or short nuclei. Length is only contrastive in closed syllables and only when syllables are stressed. Unstressed syllables always surface with short nuclei and underlying long vowels are shortened when they occur in weak positions. Stressed syllables, on the other hand, are always closed, have long nuclei, or both.

<table>
<thead>
<tr>
<th>Stressed Syllables</th>
<th>Unstressed Syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Syllables CVV</td>
<td>CV</td>
</tr>
<tr>
<td>Closed Syllables CVC, CVVC</td>
<td>CVC</td>
</tr>
</tbody>
</table>

Long nuclei may be either long monophthongs or diphthongs. There appears to be no quantitative (length) contrast among diphthongs. As discussed in section 0 below, monophthongs and rising-sonority diphthongs occur in closed syllables relatively freely. However, (phonological) falling sonority diphthongs do not occur in closed syllables. Interactions between constraints on stress assignment and quantity produce length alternations, which will be discussed at greater length in section 0 below.

2.5. **Tone**

Tone is contrastive in Sorbung, but the lexical load born by tonal contrasts is relatively low and the phonetic distance between tonal categories is relatively small. Tone interacts with intonation to such an extent that it was sometimes impossible, working from recordings, to determine the tones of a lexical item with complete certainty. Furthermore, all statements about the tonal inventory of Sorbung remain tentative until the tone sandhi system of Sorbung is worked out in greater detail.

2.5.1. **Inventory**

Sorbung has three contrasting tones, which we will refer to as H, L, and M. These are indicated in our transcriptions by diacritics on the first vowel of a rhyme: acute (́), grave (̀), or no diacritic (’), for H, L, and M, respectively. Minimal and near-minimal sets illustrating these contrasts are given in Table 15 below:
### Table 15: Example of the three contrastive tones of Sorbung.

<table>
<thead>
<tr>
<th>H</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>sāa</td>
<td>sāa</td>
<td>saa</td>
</tr>
<tr>
<td>mòo</td>
<td>mòo</td>
<td>moo</td>
</tr>
<tr>
<td>tʰòo</td>
<td>kʰòo</td>
<td>kʰoo</td>
</tr>
<tr>
<td>fí</td>
<td>kʰí</td>
<td>fí</td>
</tr>
<tr>
<td>wàŋ</td>
<td>wàm</td>
<td>wàn</td>
</tr>
<tr>
<td>rée</td>
<td>née</td>
<td>ree</td>
</tr>
<tr>
<td>náap</td>
<td>náap</td>
<td>nákʰúak</td>
</tr>
</tbody>
</table>

The labels given here should not be taken as literal phonetic realities: H is not simply a high-pitched tone and L is not simply a low-pitched tone. The tones have characteristic pitch contours and also differ from one another in voice quality. The H tone rises to a peak about two-thirds through its duration before dropping slightly. It often has a tense voice quality. The L tone falls through its duration and sometimes starts higher in pitch than the terminus of a preceding L or M tone. L tone syllables are usually produced with lax or breathy voice quality. The pitch of M tone syllables falls gently before rising again. The voice quality is usually modal.

The contrastive load of tone in Sorbung is relatively low. This may be responsible for some of the difficulty we encountered in producing a complete analysis of the tone system. Our consultant was aware that there were tonal contrasts in Sorbung, and produced a number of minimal sets with little prompting (e.g. ‘rain cover’, ‘sand’, ‘crab’, ‘star’), but found it difficult to say whether two items were similar or different in tone.
2.5.2. Distribution
Tonal contrasts in Sorbung are subject to two major restrictions. First, tonal contrasts only occur on major syllables. The pitch of minor syllables is predictable from the pitch of surrounding major syllables. Second, there is only a two-way tonal contrast in stopped syllables (syllables with a stop coda) but a three-way contrast in major syllables of other types. Neither of these restrictions is unusual in the area.

2.5.3. Alternations
Given the current state of knowledge about Sorbung, it is clear that the language has a robust system of tonal alternations. However, the data are not sufficient to paint a clear and complete picture of these alternations. What follows, then, is only an introduction to some of the most obvious alternations.

Sorbung displays dissimilation of high tones. In smooth syllables, underlying H become M after H:

---

Figure 1: *Averaged pitch plots for the three Sorbung tones.*
As can be seen from (24a-c) and (25a-c), the morphemes meaning ‘water; liquid’ and ‘bone’ are realized with a H tone word-initially and when preceded by M or L. However, as shown by (24d-e) and (25d-f), they are realized with a M tone when preceded by a H tone.

This process interacts with another alternation. Stopped syllables with an underlying H tone are realized as L except word-finally. Thus, /kʉ́t/ ‘hand; arm’ is realized as [kʉ́t] when it occurs finally but as [-kʉ́t] when it appears elsewhere:

(26) a. kʉ́t ‘hand’ ‘hand, arm’
    b. kʉ́t-sáa ‘hand’ + ‘child’ ‘finger’
    c. kʉ́d-biuli ‘hand’ + AUG ‘thumb’
    d. kʉ́d-məjáa ‘hand’ + ‘palm/sole’ ‘palm’

In isolation, the morphemes meaning ‘hand; arm’ and ‘breast’ are realized with H tones. However, when followed by other syllables, within a word, the tone changes to L. As shown by example (27c), this process counter-bleeds the H → M / H___ process described above. Were this not the case, we would expect *[cùb-cùu], with a H tone on the second syllable.

Our Sorbung database is not currently large enough, and does that have enough morphologically complex forms, to allow us to confidently characterize all of the tonal alternations that we have observed. These two patterns are described here because they were exemplified in a relatively large number of data. Based upon these relatively limited findings, though, we believe it is safe to characterize the tone sandhi system of Sorbung. On the one hand, Sorbung clearly has more tonal alternations than Tangkhulic languages like Ukhrul, where tone is morphophonologically very stable. On the other hand, Sorbung cannot claim as pervasive a system of tone rules as many Kuki-Chin languages (Hyman & VanBik 2002; Hyman & VanBik 2004; Hyman 2007a; Hyman 2007b; Yip 2004).
3. Onsets

3.1. Inventory

The onset inventory of Sorbung includes plosives, fricatives, nasals, trills, and approximates. Sorbung has five contrasting places of articulation: labial, coronal, palatal, velar, and glottal. Plosives appear at all of these places and display a three-way distinction among voiceless, aspirated, and voiced. In word-initial position, the /b/-/p/ and /d/-/t/ contrasts are retained but the /g/-/k/ contrasts is neutralized. Neither voicing nor aspiration is contrastive at the palatal place. There is no voiced-voiceless contrast among fricatives and the only voiced fricative is [v], which is a conditioned variant of /w/. There are labial, coronal, and velar nasals but no palatal nasal. The inventory also includes the trill /r/ ([ɾ] ~ [r]) and the approximates /l/, /w/, and /y/. The inventory is summarized in Table 16 below.

Table 16: Sorbung onsets. Onsets in parenthesis only occur as predictable variants of other underlying sounds.

<table>
<thead>
<tr>
<th>NASAL</th>
<th>LABIAL</th>
<th>CORONAL</th>
<th>PALATAL</th>
<th>VELAR</th>
<th>GLOTTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>n</td>
<td>k</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLOSIVE</td>
<td>p</td>
<td>t</td>
<td>c</td>
<td>kʰ</td>
<td>ʔ</td>
</tr>
<tr>
<td>pʰ</td>
<td>ʔʰ</td>
<td>b</td>
<td>d</td>
<td>g</td>
<td></td>
</tr>
<tr>
<td>FRICATIVE</td>
<td>(v)</td>
<td>s</td>
<td>f</td>
<td>h</td>
<td></td>
</tr>
<tr>
<td>TRILL</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APPROXIMATES</td>
<td>l</td>
<td>j</td>
<td>w</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The language variably allows a few consonant clusters, comprising of /pl/ /pʰl/ /tr/ /kʰr/ and /kl/, (see section 0 above).

3.1.1. Examples

Table 17 provides examples of each of the onsets.

Table 17: Examples of each onset.

<table>
<thead>
<tr>
<th>/m/</th>
<th>/n/</th>
<th>/ŋ/</th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
<th>/g/</th>
<th>/ʔ/</th>
<th>/pʰ/</th>
<th>/s/</th>
</tr>
</thead>
<tbody>
<tr>
<td>mai</td>
<td>náa</td>
<td>ŋaa</td>
<td>paasàa</td>
<td>taa</td>
<td>ka</td>
<td>ʔəŋgúu</td>
<td>ʔaa</td>
<td>pʰáaglàaŋ</td>
<td>sàa</td>
</tr>
<tr>
<td>‘face’</td>
<td>‘hurt’</td>
<td>‘fish’</td>
<td>‘man’</td>
<td>‘old’</td>
<td>‘that’</td>
<td>‘child’</td>
<td>‘fowl’</td>
<td>‘wall’</td>
<td>‘animal’</td>
</tr>
<tr>
<td>miit</td>
<td>niìŋ</td>
<td>ɲia</td>
<td>pii</td>
<td>tatir</td>
<td>ʔəki</td>
<td>cáap</td>
<td>?iin</td>
<td>pʰíit</td>
<td>lusìip</td>
</tr>
<tr>
<td>‘eye’</td>
<td>‘mind’</td>
<td>‘stand’</td>
<td>‘aunt’</td>
<td>‘crossbar’</td>
<td>‘corner’</td>
<td>‘cry’</td>
<td>‘sleep’</td>
<td>‘hit’</td>
<td>‘crown’</td>
</tr>
<tr>
<td>mìu</td>
<td>nèe</td>
<td>ɲée</td>
<td>pée</td>
<td>te</td>
<td>kèe</td>
<td>ʔìin</td>
<td>ʔök</td>
<td>ʔəpʰʉ̀t</td>
<td>ʔəsóo</td>
</tr>
<tr>
<td>‘see’</td>
<td>‘small’</td>
<td>‘white’</td>
<td>‘give’</td>
<td>‘sister’</td>
<td>‘leg’</td>
<td>‘sleep’</td>
<td>‘feces’</td>
<td>‘breath’</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Two of the onsets have visibly limited distributions.

- 

- /ʃ/- never occurs before /ee/, /e/, or /aa/.

While it would seem desirable to collapse these onsets with other phonemes, they are both in contrastive distribution with all of the possible candidates.

A more interesting case is the relationship between /g/ and /k/. Both of these occur in the onsets of major syllables following vowel-final minor syllables, establishing them as distinct phonemes:

\[
\begin{align*}
(28) & \quad a. \quad \text{tagók} & \quad \text{'pot'} \\
 & \quad b. \quad \text{cogóó} & \quad \text{'river'}
\end{align*}
\]

\[
\begin{align*}
(29) & \quad a. \quad \text{sökii} & \quad \text{'deer'} \\
 & \quad b. \quad \text{mkee} & \quad \text{'kidney'} \\
 & \quad c. \quad \text{kap kʰát} & \quad \text{'length from thumb to finger'} \\
 & \quad d. \quad \text{cákáap} & \quad \text{'tongs'} \\
 & \quad e. \quad \text{sökúan} & \quad \text{'horse'} \\
 & \quad f. \quad \text{ʔokó́r} & \quad \text{'peel'}
\end{align*}
\]

However, while /k/ can occur word-initially (30), and word-internally after another major syllable (31), /g/ does not:

\[
\begin{align*}
(30) & \quad a. \quad \text{kút} & \quad \text{'arm'} \\
 & \quad b. \quad \text{kée} & \quad \text{'foot'} \\
 & \quad c. \quad \text{kuum} & \quad \text{'year'} \\
 & \quad d. \quad \text{koo} & \quad \text{'do'}
\end{align*}
\]

\[
\begin{align*}
(31) & \quad a. \quad \text{pákùù} & \quad \text{'younger paternal uncle'} \\
 & \quad b. \quad \text{ké̂kiiuk} & \quad \text{'crippled, lame, handicapped'}
\end{align*}
\]

In this respect, /g/ is unlike the other voiced stops, /b/ and /d/. However, like then (and unlike the voiceless stops) /g/ can occur after nasal-final minor syllables:

\[
\begin{align*}
(32) & \quad a. \quad \text{ʔŋgguu} & \quad \text{'baby, child'} \\
 & \quad b. \quad \text{tŋguúp} & \quad \text{'hat'} \\
 & \quad c. \quad \text{tŋggoon} & \quad \text{'anteater'}
\end{align*}
\]
A reviewer suggests that this might be a case of free \([g \sim k]\) variation, as is found in Jingpho. This is possible. However, within our corpus there are no clear instances of \([g]\) varying with \([k]\) in the same lexical item without a clear conditioning environment (see 0 below).

### 3.2. Alternations

As discussed above, CəC sequences vary with CC sequences in the same lexical item. That is to say, the epenthetic process that breaks up underlying initial consonant clusters to produce minor syllables is variable. The \([^\text{ə}\text{]}\) appears to be the only vowel permitted in the minor syllables. We note three types of cases where epenthesis sometimes fails to occur, depending crucially on the nature of the two initial consonants. First, in /s/+sonorant clusters, epenthesis variably fails yielding, e.g. \([s^\text{ə}.m...]\) varying with \([s^\text{m}...]\), as in ‘cattle’:

\[(33) \quad [s^\text{múk}] \sim [s^\text{mũk}] \text{ ‘cattle’}\]

In this example, as in other cases with /s/+sonorant clusters, the /sm/ cluster is tautomorphemic (/s+mũk/ ‘animal prefix’ + ‘cattle’). This variation is apparently sensitive to word length and speech rate. For example, /s^mũk/ ‘cattle’ was usually realized with epenthesis, but /smukpaasaa/ ‘bull’ was usually realized without epenthesis.

Most of the obstruent-sonorant onset clusters in Sorbung contain stops. These clusters consist of voiceless labial or velar stops followed by /l/ or /r/:

\[(34) \quad \begin{align*}
\text{b. } [p^\text{ə}laaj] & \sim [p^\text{laaj}] \text{ ‘umbilical cord’} \\
\text{c. } [p^\text{ʰ}láa] & \sim [p^\text{ʰ}lãa] \text{ ‘wing (of a bird)’} \\
\text{d. } [k^\text{ʰ}rãa] & \sim [k^\text{ʰ}rãm] \text{ ‘spider’}
\end{align*}\]

In none of these cases is there convincing evidence that the minor syllable is an independent morphological unit. It is notable that while we have recorded instances of clusters with /p/ and /pʰ/, we have noted no clusters with /b/ even though the prefix /bJ/ occurs before the root /rió/. This leads us to believe that epenthesis between morphemes is more favored than epenthesis into morphemes.

The final, and probably most theoretically interesting, cluster of environments in which epenthesis sometimes fails to occur is between a glottal stop and a following consonant:

\[(35) \quad \begin{align*}
\text{a. } [ʔk^\text{ʰ}au] & \sim [ʔk^\text{ʰ}u] \text{ ‘grasshopper’} \\
\text{b. } [ʔk^\text{ʰ}úan] & \sim [ʔk^\text{ʰ}uän] \text{ ‘voice’} \\
\text{c. } [ʔkái] & \sim [ʔkái] \text{ ‘hips’} \\
\text{d. } [ʔk^\text{ʰ}úk] & \sim [ʔk^\text{ʰ}úk] \text{ ‘knee’} \\
\text{e. } [ʔɾík] & \sim [ʔɾík] \text{ ‘louse’} \\
\text{f. } [ʔpáa̯t] & \sim [ʔpá̯ät] \text{ ‘soft, mushy’} \\
\text{g. } [ʔŋaa] & \sim [ʔŋaa] \text{ ‘fish’}
\end{align*}\]

As with the /s/+sonorant clusters, the /ʔ/+consonant clusters are all tautomorphemic. There are a number of different prefixes in Sorbung which are underlyingly /ʔ-/∪, usually with a following epenthetic schwa. However, all of them seem to pattern similarly with regard to
epenthesis. When there is no epenthetic schwa, the glottal stop is realized as a faint glottal release (before stops), increased duration, or preglottalization (on a following sonorant). The variation does not seem to be constrained by the manner of the second consonant and seems to be a function mostly of speech rate and higher-level prosodic influences.

There are also cases, in Sorbung, where the first mora of a diphthong is devocalized and syllabified as an onset. The best examples of this involve the proximal demonstrative enclitic, /ʊá/. When the syllable to which /ʊá/ cliticizes ends in a consonant, the clitic is realized as [ʊā]:

(36) a. [cèmʊā] ‘this knife’
   b. [ʔɔjǐmʊā] ‘this needle’
   c. [tɔgɔguà] ‘this pot’
   d. [tɔŋguābuà] ‘this headdress’
   e. [tɛŋguàdʊà] ‘this bowl’

If the syllable to which it cliticizes ends in a vowel other than /a/, /ʊâ/ is usually realized as [wâ]:

(37) a. [tʰɛewâ] ‘this bow’
   b. [lɛewâ] ‘this medicine’
   c. [rɛewâ] ‘this war’
   d. [rɔowâ] ‘this axe’
   e. [ʔɔsɔowâ] ‘this spear’

After high vowels, though, /ʊâ/ varies between [wâ] and [vâ]:

(38) a. [məʃíiwâ ~ məʃíivâ] ‘this necklace’
   b. [rʉʉwâ ~ rʉʉvâ] ‘this rope’
   c. [məciïwâ ~ meciïvâ] ‘this necklace’

Finally, and most curiously, after /a/, /ʊâ/ varies between [vâ] and [βâ]:

(39) a. [bélcâavâ] ‘this spade’
   b. [iinkʰaavâ ~ iinkʰaaβâ] ‘this door’
   c. [laavâ] ‘this song’

It is not clear to us what the phonetic basis for this latter pattern is.29

29 A reviewer inquires about the fate of the initial low tone in /ʊâ/. When the /u/ is realized as a vowel, rather than a glide, it is produced with a distinct low tone and the following /a/ is produced with a high tone. When the /u/ is devocalized to [w], [v], etc., the low tone disappears so that in [məʃíiwâ], the last two syllables are realized as high tones with no intervening low. This suggests that the low tone may be a default inserted on toneless TBUs.
4. Rhymes

4.1. Inventory

We analyze Sorbung as having five short-long pairs of vowel phonemes in major syllables:

\[
\begin{array}{cccccccc}
/ii/ & /i/ & /ʉʉ/ & /u/ & /ʉʉ/ & /u/ \\
/ee/ & /e/ & /oo/ & /o/ & /a/ & /a/ \\
\end{array}
\]

Phonetically, the short-long distinction is realized through a number of different cues other than duration, including quality and dynamic quality. In general, long vowels have a more peripheral quality and short vowels a more centralized quality. For the high vowels, duration is the primary cue to length. The long mid-vowels /ee/ and /oo/ are phonetically the diphthongs [ej] and [ow] while their short counterparts are realized as the lower monophthongs [ɛ] and [ɔ]. The low vowels /aa/ and /a/ are phonetically [aː] and [ɐ]. A summary of the phonetic values of the vowel phonemes is given below:

\[
\begin{array}{cccccccc}
/aa/ & [aː] & /a/ & [ɐ] \\
\end{array}
\]

We readily acknowledge that it would be possible to reanalyze this system in terms other than quantity (vowel length). The similar vowel systems of Northern Chin languages, for example, were analyzed by Button (2009) purely in qualitative terms. Our analysis provides a number of advantages. It allows us to easily characterize the stress-conditioned vowel alternations as alternations in a single phonological parameter (length), even though different phonetic parameters are involved depending on the vowel. By treating /ee/ and /oo/ as long vowels rather than phonemic diphthongs, we save the generalization that codas do not appear after phonemic falling-sonority diphthongs (diphthongs with a phonemic off-glide occupying the coda position). It is not simply the case, though, that analyzing the Sorbung vowel system into short-long pairs allows us to better capture the phonological patterns in the language than a quality-based analysis. It is also conceptually simpler and requires us to invoke fewer phonological parameters in order to characterize all of the contrasts in the system.

Aside from vowel length, one other aspect of our analysis of the vowels calls for comment. Although they are contrastive on the surface, it is possible to eliminate /ʉʉ/ and /ʉ/ from the underlying vowel inventory if some abstractness in underlying representation is allowed. In stressed syllables, short [u] only occurs before coronal codas (/n/ and /t/) and thus occurs in complementary distributions with [u], which never occurs in these environments. In isolation from other facts, [u] could be reduced to an allophone of /u/*. We treat it as distinct because [u] can also surface as a result of stress-conditioned length alternations with [uu]. Compare, for example, cɨu ‘water’ and cu-rée ‘thirsty (water-thirst)’. Thus, the phonemic status of /u/ is dependent on the phonemic status of /uu/*. As noted in Section 0 (above), Sorbung -uu corresponds to PKC *-uy and probably reflects earlier *-ui. As will be seen below, it also behaves like the falling-sonority diphthongs /ai/ and /au/ in that it never occurs with a coda. For this reason, it is tempting to consider [uu]
to be the realization of underlying /ui/. This would not be problematic, apart from the degree of abstraction involved, if there were not a number of lexical items in Sorbung with surface [ui]. All of these appear to be loanwords from Meithei or Ukhrul Tangkhul:

(40)  **LOANS FROM UKHRUL TANGKHUL**
    a. ʃúi ‘tempt’
    b. lúi ‘finish’
    c. ɒkətʰu ‘world’

(41)  **LOAN FROM MEITHEI**
    cəkú ‘dance’

In principle, there is nothing to stop us from saying that there is a phonological process mapping /ui/ to [ʉʉ] that only affects the lexical stratum containing native vocabulary. Claims of the same type have been made for other languages (Itô & Mester 1995; Itô, Mester, & Padgett 1999). However, in the interest of descriptive neutrality, we have retained the more concrete analysis while highlighting some of its shortcomings.

Excluding a few loanwords which have final /l/, Sorbung allows major syllables to have one of seven coda consonants:

/p/ /t/ /k/
/m/ /n/ /ŋ/
/r/

With a few exceptions, these can occur after the long and short vowels and after the rising-sonority diphthongs /ia/, /ua/, and /iu/. They do not occur after the falling-sonority diphthongs, /ai/ and /au/, and after /ʉʉ/. **Table 18** shows which of the combinatorial possibilities between nuclei and rhymes are attested in our corpus. Examples are given in the subsequent tables.
### Table 18: Sorbung rhymes.

<table>
<thead>
<tr>
<th></th>
<th>–0</th>
<th>–r</th>
<th>–m</th>
<th>–n</th>
<th>–ŋ</th>
<th>–p</th>
<th>–t</th>
<th>–k</th>
</tr>
</thead>
<tbody>
<tr>
<td>[a]</td>
<td>ar</td>
<td>am</td>
<td>an</td>
<td>aŋ</td>
<td>ap</td>
<td>at</td>
<td>ak</td>
<td></td>
</tr>
<tr>
<td>aa</td>
<td>aar</td>
<td>aam</td>
<td>aan</td>
<td>aŋaap</td>
<td>aat</td>
<td>aak</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[e]</td>
<td>em</td>
<td>en</td>
<td>eŋ</td>
<td>et</td>
<td>ek</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ee</td>
<td>eem</td>
<td>een</td>
<td>eŋe</td>
<td>eek</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[o]</td>
<td>or</td>
<td>om</td>
<td>on</td>
<td>oŋ</td>
<td>[op]</td>
<td>ok</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oo</td>
<td>oom</td>
<td>oon</td>
<td>oonŋ</td>
<td>oop</td>
<td>ook</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[i]</td>
<td>ir</td>
<td>im</td>
<td>in</td>
<td>iŋ</td>
<td>[ip]</td>
<td>ik</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii</td>
<td>iin</td>
<td>iinŋ</td>
<td>ip</td>
<td>iit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[u]</td>
<td>ur</td>
<td>um</td>
<td>[un]</td>
<td>uŋ</td>
<td>[up]</td>
<td>[ut]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>uu</td>
<td>uur</td>
<td>uum</td>
<td>uuŋ</td>
<td>uup</td>
<td>uuk</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 19: Examples of Sorbung open and r-final rhymes.

<table>
<thead>
<tr>
<th></th>
<th>–0</th>
<th>–r</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>cinfiipá</td>
<td>‘ant’</td>
</tr>
<tr>
<td>aa</td>
<td>?aa</td>
<td>‘fowl’</td>
</tr>
<tr>
<td>e</td>
<td>këmiit</td>
<td>‘ankle’</td>
</tr>
<tr>
<td>ee</td>
<td>melëe</td>
<td>‘tongue’</td>
</tr>
<tr>
<td>o</td>
<td>tɔtrò</td>
<td>‘comb’</td>
</tr>
<tr>
<td>oo</td>
<td>móo</td>
<td>‘fire’</td>
</tr>
<tr>
<td>i</td>
<td>tʰìi</td>
<td>‘die’</td>
</tr>
<tr>
<td>ii</td>
<td>tʰìi</td>
<td>‘bleed’</td>
</tr>
<tr>
<td>uu</td>
<td>mëjùu</td>
<td>‘mouse’</td>
</tr>
<tr>
<td>uu</td>
<td>habhù</td>
<td>‘molar’</td>
</tr>
<tr>
<td>au</td>
<td>jau</td>
<td>‘sheep’</td>
</tr>
<tr>
<td>ai</td>
<td>mai</td>
<td>‘face’</td>
</tr>
<tr>
<td>ia</td>
<td>ñenfìà</td>
<td>‘enemy’</td>
</tr>
<tr>
<td>ua</td>
<td>kòfùà</td>
<td>‘rain’</td>
</tr>
<tr>
<td>iu</td>
<td>ciu</td>
<td>‘dig’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>–0</th>
<th>–r</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>?ətʰár</td>
<td>‘new’</td>
</tr>
<tr>
<td>aa</td>
<td>ŋənáar</td>
<td>‘snore’</td>
</tr>
<tr>
<td>e</td>
<td>cêm</td>
<td>‘knife’</td>
</tr>
<tr>
<td>ee</td>
<td>məlée</td>
<td>‘tongue’</td>
</tr>
<tr>
<td>o</td>
<td>mor</td>
<td>‘mouth’</td>
</tr>
<tr>
<td>oo</td>
<td>caŋkʰoom</td>
<td>‘milk’</td>
</tr>
<tr>
<td>i</td>
<td>mëtír</td>
<td>‘shrew’</td>
</tr>
<tr>
<td>ii</td>
<td>tʰìi</td>
<td>‘die’</td>
</tr>
<tr>
<td>uu</td>
<td>júur</td>
<td>‘nation’</td>
</tr>
<tr>
<td>uu</td>
<td>húar</td>
<td>‘bright’</td>
</tr>
</tbody>
</table>
Table 20: Examples of Sorbung nasal-final rhymes.

<table>
<thead>
<tr>
<th></th>
<th>–m</th>
<th>–n</th>
<th>–ŋ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>hambʉ́ʉ</td>
<td>‘tiger’</td>
<td>wàn</td>
</tr>
<tr>
<td>aa</td>
<td>səráam</td>
<td>‘otter’</td>
<td>laan</td>
</tr>
<tr>
<td>e</td>
<td>len</td>
<td>‘close’</td>
<td>sen</td>
</tr>
<tr>
<td>ee</td>
<td></td>
<td></td>
<td>kəléŋ</td>
</tr>
<tr>
<td>o</td>
<td>som</td>
<td>‘ten’</td>
<td>?itā kʰòn</td>
</tr>
<tr>
<td>oo</td>
<td>mohuŋoon</td>
<td>‘window’</td>
<td>jooŋ</td>
</tr>
<tr>
<td>i</td>
<td>ðənim</td>
<td>‘shadow’</td>
<td>mətín</td>
</tr>
<tr>
<td>ii</td>
<td>?iin</td>
<td>‘drink’</td>
<td>jiŋ</td>
</tr>
<tr>
<td>u</td>
<td>kʰumpʰék</td>
<td>‘duck’</td>
<td>kun</td>
</tr>
<tr>
<td>uu</td>
<td>kʰuam</td>
<td>‘sweet’</td>
<td></td>
</tr>
<tr>
<td>ia</td>
<td>səriám</td>
<td>‘mithun’</td>
<td></td>
</tr>
<tr>
<td>ua</td>
<td>múam</td>
<td>‘hold’</td>
<td>puan</td>
</tr>
<tr>
<td>ii</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 21: Examples of Sorbung stop-final rhymes.

<table>
<thead>
<tr>
<th></th>
<th>–p</th>
<th>–t</th>
<th>–k</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>nápkáaŋ</td>
<td>‘dry mucus’</td>
<td>tʰát</td>
</tr>
<tr>
<td>aa</td>
<td>naap</td>
<td>‘stick (v.)’</td>
<td>kʰáat</td>
</tr>
<tr>
<td>e</td>
<td>?ərét</td>
<td>‘eight’</td>
<td>mooʔék</td>
</tr>
<tr>
<td>ee</td>
<td></td>
<td></td>
<td>ŋəmek</td>
</tr>
<tr>
<td>o</td>
<td>kʰöop</td>
<td>‘be satisfied’</td>
<td>ŋək</td>
</tr>
<tr>
<td>oo</td>
<td>məcip</td>
<td>‘frighten’</td>
<td>mit</td>
</tr>
<tr>
<td>i</td>
<td>lusìip</td>
<td>‘top of head’</td>
<td>miit</td>
</tr>
<tr>
<td>ii</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>məjúp</td>
<td>‘suck (v.)’</td>
<td>kút</td>
</tr>
<tr>
<td>uu</td>
<td>məjúup</td>
<td>‘kiss (v.)’</td>
<td></td>
</tr>
<tr>
<td>ia</td>
<td></td>
<td></td>
<td>ŋəntiak</td>
</tr>
<tr>
<td>ua</td>
<td>ŋəcúap</td>
<td>‘spleen’</td>
<td>kʰuat</td>
</tr>
<tr>
<td>ii</td>
<td></td>
<td></td>
<td>kəkiuk</td>
</tr>
</tbody>
</table>

4.1.2 Gaps
Two major nucleus-coda co-occurrence restrictions are found in Sorbung. The first applies to the falling-sonority-coda diphthongs. As mentioned above, phonemic falling sonority diphthongs are never followed by coda consonants. Unlike /ia/, /ua/, and /iu/ that act as nuclei with codas, /ai/ and /au/ do not permit codas because the off-glide fills the coda position. In general, complex codas are not permitted in Sorbung. /ui/ follows the same pattern even though it is found only in loanwords from Ukhrul (Standard) Tangkhul and Meitheni. Similarly, no codas appear after long [uːi], which, as discussed above, is the modern reflex of historical **-ui.
(42)  a. ʔə-sái  ‘elephant’  săi-róoŋ  ‘slingshot’
     b. jau  ‘sheep’  ?okʰauríaŋ  ‘grasshopper’

(43)  a. ʔa-cíu  ‘chicken egg’  mə-nnu  ‘laugh’

(44)  a. ʔia  ‘stand’  ?on-tiàk  ‘green’
     b. fiu  ‘hit’  kè-kiuk  ‘crippled’
     c. kə-fuà  ‘rain’  fiuan  ‘weak’

(45)  a. ɲee  ‘white’  kə-léen  ‘loose’
     b. sóo  ‘run’  òok  ‘pig’

As mentioned above, the coda restriction does not apply to the long vowels /oo/ and /ee/, though phonetically they are falling-sonority [ow] and [ej]. Our analysis of [ow] and [ej] as underlying long vowels, /oo/ and /ee/, explains why they function as monophthongal nuclei instead of falling-sonority diphthongs with regard to syllable structure.

4.2. Alternations

The phonological alternations affecting Sorbung rhymes include voicing alternations in stop codas and vowel shortening in unstressed syllables.

The voicing of stop codas is predictable from the phonological environment. When the following segment in a word is a vowel or voiced consonant, the coda is voiced. Elsewhere, when word-finally and before voiceless consonants, stops are then voiceless. (46a) demonstrates the changing voicing of the coronal stop in /miit/ ‘eye’. When this stem is compounded with /mii/ to form ‘eyebrow’ (literally ‘eye hair’), the final stop is voiced as it precedes a voiced segment. However, when followed by the voiceless consonant-initial /kor/, the stop remains voiceless. Other examples in (46) further demonstrate the same pattern:

(46)  a. miit  ‘eye’  mid-mii  ‘eyebrow’  
     mit-kor  ‘eyelid’  mid-ék  ‘eye sand’
     b. kút  ‘hand’  kúd-matin  ‘finger nail’
     kút-kʰóom  ‘handspan’
     c. kʰòop  ‘full’  kʰóob-maa  ‘full-question’
     d. káap  ‘shoot’  káab-ôo  ‘shoot-ind’
     e. ʔèk  ‘feces’  ʔèg-len  ‘defecate’

This alternation cannot be attributed to a general process of voicing assimilation. Intervocalic voiceless stops occur frequently:

(47)  a. lukaanja  ‘head’
     b. kèkìuk  ‘crippled’
     c. pátáabàa  ‘elder paternal uncle’
     d. jáapee  ‘many’
     e. kʰūpú kʰuncaanja  ‘storm’
The examples in (47) show that voiceless stop onsets are not subject to intervocalic voicing and therefore dismiss a general voicing assimilation rule. However, it is not clear whether the voicing alternation in Sorbung stop codas should be treated as the voicing of underlying voiceless stops before voiced segments, the devoicing of underlying voiced segments word-finally and before voiceless segments, or the “filling” of laryngeal specifications into segments underlingly unspecified for voice. As both voiced and voiceless stops appear as onsets, intervocally or otherwise, we cannot assume all stops are either voiced or voiceless. In Kom, an Old Kuki language also spoken in Manipur which shares many traits with Sorbung, we find a similar pattern of voicing (Grierson 1903:244). Word final voiceless stops become voiced when a vocalic suffix is added (48):

(48) a. kut ‘hand’ a-kud-a ‘his hand on’
   b. kaap ‘shoot’ a-kaab-a ‘his shooting’

Like Sorbung, Kom also has examples of voiceless stops between vowels. While the available corpus of Kom data is not large enough to determine whether the alternations are identical, the similarity is suggestive.

In simple compounds consisting of two monosyllabic roots, the first syllable is consistently unstressed. If the first syllable’s vowel is short underlyingly, it will not be modified; if it is long, it will be shortened.

(49) a. kut ‘hand’ kudbiu ‘thumb’
   b. cóyn ‘words’ cóynʔáak ‘stutter’
   c. kʰoo ‘bee’ kʰociu ‘honey’
   d. miit ‘eye’ mitkor ‘eyelid’
   e. haa ‘tooth’ habwu ‘molar’
   f. cuup ‘breast’ cupcuu ‘milk’
   g. tʰi ‘blood’ tʰiʃuuk ‘bleed’
   h. jaŋ ‘penis’ jaŋhuun ‘foreskin’
   i. këe ‘leg, foot’ kemii ‘ankle’

However, the constituent structure of compound formation alters stress and thus reduction. As expected, the vowel reduces in the unstressed position in (51b), but remains a long vowel in the stressed position in (51c). The reduction appears to only apply to instances of endocentric compounds of two simple, one-syllable roots.

(51) a. ‘këe ‘leg, foot’
   b. ‘kë- mi ‘leg hair’
   c. ‘këe-mə jɑa ‘footprint’

(52) a. mə kʰáa ‘jaw’
   b. mə,kʰáa-ruu ‘jawbone’

In other complex compounds, especially those containing sesquisyllables, reduction does not apply in the same manner. (52b) displays a complex compound where the
nucleus of the head does not reduce. The length alternation follows almost entirely from
the alternations in stress discussed in Section 0 above.

5. Conclusion
Leaving aside Brown’s (1837) brief word list, this study is the first introduction of
Sorbung/Southern Tangkhul to the scholarly community. We have introduced some of the
significant properties of this language that are likely to be of interest to synchronic
theorists, typologists, and comparativists. At the same time, a great deal of work remains to
be done. Because our goals in this paper were descriptive, we have not ventured to provide
a theoretical rationale for the patterns and alternations that we have reported. We have only
ventured far enough into matters diachronic to promote our position that Sorbung is closer
to Kuki-Chin than to Tangkhulic and that its phonology should be seen in this perspective.
A detailed working-out of both the theoretical implications of the synchronic alternations
in Sorbung phonology—cyclic stress assignment, opaque tone sandhi, morphologically
conditioned place-assimilation, variable epenthesis, and problematic voicing alternations—
will have to be left for further work. So, too, will detailed study of the historical
development of Sorbung phonology. Likewise, we have scarcely touched upon the
implications that this work has for understanding Tangkhul ethnogenesis and for plotting
the historical relationship between the Tangkhuls of the Sorbung area and the rest of the
Tangkhul ethnicity. What we do hope to have established is that Sorbung presents
interesting data and problems to both the historical and synchronic linguist.

References
Brown, Nathan. 1837. Comparison of Indo-Chinese languages. Journal of the Asiatic
Society of Bengal 6(2). 1023-1038.
Button, Christopher Thomas James. 2009. A Reconstruction of Proto Northern Chin in Old
Burmese and Old Chinese Perspective. SOAS, University of London Phd dissertation.
Burling, Robbins. 2003, The Tibeto-Burman Languages of Northeast India. In G.
Thurgood and R. H. LaPolla (eds), The Sino-Tibetan Languages, 169–192.
17). Berlin: Mouton de Gruyter.
Grierson, George A. 1903. Tibeto-Burman Family: Specimens of the Kuki-Chin and Burma
of Government Printing.
Hartmann, Helga. 2001. Prenasalization and preglottalization in Daai Chin, with parallel
eamples from Mro and Maru. Linguistics of the Tibeto-Burman Area 24(2). 123–
142.
Experimental Approaches to Phonology. 7-24.
Berkeley Phonology Lab Annual Report.
Linguistics of the Tibeto-Burman Area 25(1). 113-120.


Appendix: Sorbung Word List

<table>
<thead>
<tr>
<th>SORBUNG</th>
<th>GLOSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1dl</td>
<td>abdomen (external bulge); belly</td>
</tr>
<tr>
<td>1pl</td>
<td>able</td>
</tr>
<tr>
<td>1sg</td>
<td>accept</td>
</tr>
<tr>
<td>1sg-gen</td>
<td>accustomed</td>
</tr>
</tbody>
</table>
| 1sg-refl | ad
| 2dl      | add together |
| 2pl      | adult |
| 2sg      | afterbirth; placenta |
| 2sg-gen  | alive; living, be |
| 2sg-refl | animal |
| 3dl      | animal |
| 3pl      | ankle |
| 3sg      | ant |
| 3sg-gen  | ant eater (pangolin); crocodile |
| 3sg-refl | antler (of deer) |
|          | anus; rectum |
|          | arm, upper |
|          | arm; hand |
|          | armpit; underarm |
|          | arrive |
|          | arrow |
|          | artery; blood vessel; vein |
|          | ascend; climb |
|          | ashamed; shy |
|          | asleep, be; sleep |
|          | aunt, elder paternal |
|          | aunt, maternal |
|          | aunt, wife of elder paternal uncle |

<table>
<thead>
<tr>
<th>SORBUNG</th>
<th>GLOSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʔərúkəniŋ</td>
<td>abdomen (external bulge); belly</td>
</tr>
<tr>
<td>ʔənəkʰuo</td>
<td>able</td>
</tr>
<tr>
<td>ʔoo</td>
<td>accept</td>
</tr>
<tr>
<td>ʔità</td>
<td>accustomed</td>
</tr>
</tbody>
</table>
| ʔōomasaa | ad
| nárúkəniŋ | adult |
| naanəkʰuo | afterbirth; placenta |
| naaŋ | alive; living, be |
| nanta | animal |
| naaŋməsaa | animal |
| mərúkəniŋ | ankle |
| maanəkʰuo | ant |
| mətiuk | ant eater (pangolin); crocodile |
| maanəkʰuo | antler (of deer) |
| mə́rúkəniŋ | anus; rectum |
| maaməsaa | arm, upper |
| maaməsaa | arm; hand |
| maaməsaa | armpit; underarm |
| maaməsaa | arrive |
| maaməsaa | arrow |
| mə́rúkəniŋ | artery; blood vessel; vein |
| maaməsaa | ascend; climb |
| maaməsaa | ashamed; shy |
| maaməsaa | asleep, be; sleep |
| maaməsaa | aunt, elder paternal |
| maaməsaa | aunt, maternal |
| maaməsaa | aunt, wife of elder paternal uncle |
aunt, wife of maternal uncle: pii
aunt, wife of younger paternal uncle: núgənùu
aunt, younger paternal: ʔənnìi
awake, be; wake up (v.i.): ʔəntʰèe
axe: rødùu
back: kòonq
back (of something): ʔənúuŋlàm
back basket: məkóonq
back, lower: kòonqcat
backbone; spine: kòonqruu
bad (morally): saaleè
bamboo spoon for curry: məkʰée
bar: lakium
bat: ʔəbák
bathe: cuʃii
beak; bill (of bird): mór
bear (Selenarctos thibetanus): ʔəwám
bear; birth, give: mədée
deard: mərmíi
beautiful: ʔəloo
become old: taalen
bee: kʰaa
behind: ʔənúuŋlàm
belly; abdomen (external bulge): wan
big: lòok
bile: məʃìŋkʰaa
bill (of bird); beak: mór
bird: ʔəwaa
birth, give; bear: mədée
bite (v.): ʔóo
bitter: kʰaa
black: wàm
blanket; cloth: púan
bleed (v.): tʰìʃúuk
blemish on skin: maihìŋ
blemish; mole: sə̀màamíi
blind: mìtcòo
blink (v.): mìtkʰəjùp
blood: tʰìi
blood vessel; vein; artery: maʃeetʰáaruu
blow (mouth): sèm
blow fire: mò-sèm
body: masàa
body, dead; corpse: tʰiîlenta masàa
boil: but
bone: ʔarúu
bone, jaw; jawbone; mandible  məkʰáaruu
bone, malar; cheekbone  mairúu
born, be  mədée
borrow  màmbúk
bosom  malúuŋ
boundary  tʰərìi
bow  tʰée
bowels; intestines  ?ərìi
bowl  ténguát
brain  ?əkʰúak
break  məkʰóo
break (glass, egg); split (a watermelon, other round object)  kʰoo
breast  cúup
breath  ?əpʰʉ̀t
breathe (v.)  ?əpʰʉ̀t-lée
bright  háar
brother of female, elder  moo
brother of female, younger  ?ináupa
brother of male, elder  moo
brother of male, younger  ?ináupa
brother-in-law (of ego)  ?inèeŋ
brow; forehead  məcèe
brush  ṭətáat
buffalo  səloo
bull; cattle, male  səmúk paasaa
burn off fields; slash and burn  lèe-hèe
burn; on fire, be  moo-còk
burnt  kaanlen
burst  məkʰóo
butterfly  ?əpàak
buttocks  ?əkái
buy  ?əlóo
calf; shin  kəemorài
call  kèe
carry  pʰùa
carry on the head  tooŋ
carry with hand; lift up  kʰaaŋ
cat (Felis domesticus)  kʰaaŋ
catch  ?əráak
catch; drive; chase  ?ajii
cattle  səmúk
cattle, female; cow  səmúk móosa
cattle, male; bull  səmúk paasaa
cave; hole (in rock)  luŋkʰua
charcoal  məhée
<table>
<thead>
<tr>
<th>English</th>
<th>Xam</th>
</tr>
</thead>
<tbody>
<tr>
<td>chase; catch; drive</td>
<td>?ajii</td>
</tr>
<tr>
<td>cheap</td>
<td>pai</td>
</tr>
<tr>
<td>cheek</td>
<td>mai</td>
</tr>
<tr>
<td>cheekbone; bone, malar</td>
<td>mairúu</td>
</tr>
<tr>
<td>chest (=liver)</td>
<td>maʔín</td>
</tr>
<tr>
<td>chew (v.)</td>
<td>?oo</td>
</tr>
<tr>
<td>chicken egg</td>
<td>?acʉ̀ʉ</td>
</tr>
<tr>
<td>chicken, jungle</td>
<td>?arip</td>
</tr>
<tr>
<td>chicken; fowl</td>
<td>?aa</td>
</tr>
<tr>
<td>chief</td>
<td>?əhuáŋbáa</td>
</tr>
<tr>
<td>child</td>
<td>?əŋgüú</td>
</tr>
<tr>
<td>chin; jaw</td>
<td>maʔk̡áa</td>
</tr>
<tr>
<td>choose</td>
<td>kəplaanj</td>
</tr>
<tr>
<td>chop (at tree)</td>
<td>dúu</td>
</tr>
<tr>
<td>clan; surname</td>
<td>sakái</td>
</tr>
<tr>
<td>claw (of animal)</td>
<td>maʔtín</td>
</tr>
<tr>
<td>clay</td>
<td>lámpʰu</td>
</tr>
<tr>
<td>clean</td>
<td>tʰàaar</td>
</tr>
<tr>
<td>clear</td>
<td>saŋ</td>
</tr>
<tr>
<td>clever; smart; wise</td>
<td>tʰàaanmee</td>
</tr>
<tr>
<td>climb; ascend</td>
<td>hánŋ</td>
</tr>
<tr>
<td>close; shut</td>
<td>lën</td>
</tr>
<tr>
<td>cloth; blanket</td>
<td>púan</td>
</tr>
<tr>
<td>cloud</td>
<td>kʰàmbuù</td>
</tr>
<tr>
<td>coccyx; tailbone</td>
<td>?skáiruu</td>
</tr>
<tr>
<td>cock (intact male)</td>
<td>?akʰóon</td>
</tr>
<tr>
<td>cold, very; quiet, very</td>
<td>dáiriʃì</td>
</tr>
<tr>
<td>cold; quiet</td>
<td>dái</td>
</tr>
<tr>
<td>comb (of rooster)</td>
<td>tətrò</td>
</tr>
<tr>
<td>come</td>
<td>hóonŋ</td>
</tr>
<tr>
<td>come back (imperative)</td>
<td>hoόŋledec</td>
</tr>
<tr>
<td>conceive</td>
<td>maʔrái</td>
</tr>
<tr>
<td>cook</td>
<td>jùuŋ</td>
</tr>
<tr>
<td>copulate (v.); have intercourse</td>
<td>?steŋná</td>
</tr>
<tr>
<td>cord, one; fathom, one</td>
<td>laam kʰåt</td>
</tr>
<tr>
<td>corner</td>
<td>?skii</td>
</tr>
<tr>
<td>corpse; body, dead</td>
<td>tʰ̱ilenta maʔsàa</td>
</tr>
<tr>
<td>correct</td>
<td>kún</td>
</tr>
<tr>
<td>cosmos; world</td>
<td>ókətʰui</td>
</tr>
<tr>
<td>cough</td>
<td>maʔk̡áa</td>
</tr>
<tr>
<td>cow; cattle, female</td>
<td>smúk móosa</td>
</tr>
<tr>
<td>crab</td>
<td>?ajii</td>
</tr>
<tr>
<td>crawl</td>
<td>?swák</td>
</tr>
<tr>
<td>crayfish</td>
<td>tukonŋ(g)lá</td>
</tr>
<tr>
<td>creator of the universe</td>
<td>semnùu sempáa</td>
</tr>
<tr>
<td>crippled; lame; handicapped</td>
<td>kékĩuk</td>
</tr>
</tbody>
</table>
(whether or not legs are involved)
crocodile; anteater (pangolin) təŋgoonŋ
crooked məkʰái
cross ?əkaan
crossbar tatir
crotch; fork of legs kèemətōo, kèemədóo
crow (v.); sound (v.i., as a musical instrument) ?akʰuán
crown; pate; head, top of lusìip
cry; weep cáap
cut (wood, vegetables) tán
dance cəkúi
dark jiŋ
darkness ?əjiŋ
daughter saa
daughter-in-law (of ego) ?iháa
day ?əʃʉn
dead, be; die tʰii
defaf nàkʰuák
decay məmǎaŋ
deceive mináam
deep tʰùuk
defer sakii
defecate (v.) ?èklen, ?èglen
descend júuŋ
desire; want pam
devil, ghost, demon, evil spirit rambúu
dew ?ədáicʉʉ
dhole rámʔʉ́ʉ
die; dead, be tʰii
difficult lūu
dig ciu
digit -majúumràa
dirty múàt
dish out; scoop out ʃuaglee
divide; split mətáan
do koo
dog (Canis familiaris) ?ʉʉ
door iinkʰaa
dove; pigeon ?əʃűu
dream máaŋ
drink (v.) ?iin
drive; chase; catch ?əjíi
drought cəkáaŋ
dry tʰaaŋ
dry land (vs. water); earth; ground ?əloo
duck
kùmpʰék

dull (not sharp)
ŋáimai

dumb; mute
conʃjúuk

ear
tshút

ear canal
bùumlík

eardrum
ʔaná

eye
ʔanáʔèk

early; fast
ʔèntʃià

earring
məmìt

earth; ground; dry land (vs. water)
ràmbʉʉ háaŋ

earthquake
ʔoloo

easy
pai

eat
saa

egg
ʔəwacuú

eggs, fish; roe
ʔənáʔèk

eight
ʔərēt

eighteen
ʔənáʔèk

elephant
ʔəsái

eleven
ʔənáʔèk

ember; hot coal
ʔənáʔèk

enemy
ʔənáʔèk

erase a chalkboard; extinguish
məmìt

err
ʔənáʔèk

evening
ʔənáʔèk

exist
ʔənáʔèk

exist, to
ʔənáʔèk

exit
ʔənáʔèk

exit, to
ʔənáʔèk

expect
ʔənáʔèk

expensive
ʔənáʔèk

extinguish; erase a chalkboard
məmìt

extinguish; put out fire
məmìt

eye
ʔənáʔèk

eye sand
ʔənáʔèk

eye, white of
ʔənáʔèk

eyeball
ʔənáʔèk

eyebrow
ʔənáʔèk

eye, white of
ʔənáʔèk

eyelash
ʔənáʔèk

eyelid
ʔənáʔèk

face
ʔənáʔèk

fall
ʔənáʔèk

family
ʔənáʔèk

far
ʔənáʔèk

fall
ʔənáʔèk

family
ʔənáʔèk

far
ʔənáʔèk
<table>
<thead>
<tr>
<th>English</th>
<th>Sorbung</th>
</tr>
</thead>
<tbody>
<tr>
<td>fart</td>
<td>məcipát</td>
</tr>
<tr>
<td>fast; early</td>
<td>tə̀ak</td>
</tr>
<tr>
<td>fat</td>
<td>?atóâu</td>
</tr>
<tr>
<td>fat</td>
<td>tə̀áu</td>
</tr>
<tr>
<td>father (of ego)</td>
<td>?əmbāa</td>
</tr>
<tr>
<td>father-in-law (of ego)</td>
<td>?əmbūu</td>
</tr>
<tr>
<td>fathom, half</td>
<td>laam ʔənkʰai</td>
</tr>
<tr>
<td>fathom, one; cord, one</td>
<td>laam kʰai</td>
</tr>
<tr>
<td>fear</td>
<td>ʔəncii</td>
</tr>
<tr>
<td>feather (of bird)</td>
<td>ʔəmiii</td>
</tr>
<tr>
<td>feces; shit</td>
<td>ʔèk</td>
</tr>
<tr>
<td>femur; thigh bone</td>
<td>matóóruu</td>
</tr>
<tr>
<td>few</td>
<td>kaajāŋ</td>
</tr>
<tr>
<td>field</td>
<td>lèe</td>
</tr>
<tr>
<td>fifteen</td>
<td>som rəŋāa</td>
</tr>
<tr>
<td>fifth born</td>
<td>tʰoombāa</td>
</tr>
<tr>
<td>fifty</td>
<td>sǒŋŋāa</td>
</tr>
<tr>
<td>fin; scale of fish</td>
<td>?əŋjammii</td>
</tr>
<tr>
<td>finger</td>
<td>kudmajjumráa, kutmjúumráa</td>
</tr>
<tr>
<td>finger width, one</td>
<td>kudmjumráa kʰát</td>
</tr>
<tr>
<td>finger, little</td>
<td>kúsáa</td>
</tr>
<tr>
<td>fingernail</td>
<td>kudmətín</td>
</tr>
<tr>
<td>finish</td>
<td>lůi</td>
</tr>
<tr>
<td>fire</td>
<td>móo</td>
</tr>
<tr>
<td>first born</td>
<td>moopaa</td>
</tr>
<tr>
<td>fish (general)</td>
<td>?ŋaa, ?ŋaa</td>
</tr>
<tr>
<td>fish species, light in color</td>
<td>?ŋkʰaau</td>
</tr>
<tr>
<td>fish, catfish</td>
<td>?ŋəŋ(j)lá</td>
</tr>
<tr>
<td>fish, flying</td>
<td>nəmii</td>
</tr>
<tr>
<td>fishing cat</td>
<td>sərəam</td>
</tr>
<tr>
<td>fit</td>
<td>cāa</td>
</tr>
<tr>
<td>five</td>
<td>rəŋāa</td>
</tr>
<tr>
<td>flea</td>
<td>ʔʉrík</td>
</tr>
<tr>
<td>flesh</td>
<td>ʔətáak</td>
</tr>
<tr>
<td>flood</td>
<td>cəlòk</td>
</tr>
<tr>
<td>flow (v.)</td>
<td>cəluáŋ</td>
</tr>
<tr>
<td>fly</td>
<td>tʰəwái</td>
</tr>
<tr>
<td>fly (v.)</td>
<td>wáŋ</td>
</tr>
<tr>
<td>follow</td>
<td>ʔənuŋ-aajii</td>
</tr>
<tr>
<td>foot</td>
<td>kée</td>
</tr>
<tr>
<td>footprint</td>
<td>kəluunj</td>
</tr>
<tr>
<td>forehead; brow</td>
<td>məcēe</td>
</tr>
<tr>
<td>foreskin</td>
<td>jaŋhűn</td>
</tr>
<tr>
<td>forget</td>
<td>mənii-ľén</td>
</tr>
<tr>
<td>fork of legs; crotch</td>
<td>kəemətőo, kəemədőo</td>
</tr>
<tr>
<td>four</td>
<td>məlii</td>
</tr>
</tbody>
</table>
fourteen  som məlíi
fourth born  miipáa
fowl, male; rooster ʔakʰ,oŋ
fowl; chicken ʔaa
fox ʔəmáʔu
freckle maihiŋ
friend ʔəjóok
frighten məcíp
frog ʔəcok
front (of something) məŋaalam
front of, in məŋaalam
frost ʔəhúr
fruit tʰiŋraa
fry tʰàu-jáu
full; satiated, be kʰ,òop
fur (of animal) ʔəmii
gape; open mouth (v.) ʔàañ
get somebody up matʰée
get up; rise ʔəntʰée
give pée
go jáu
goat màŋgəlée
goatee mòrmíi
gold sənáa
good ʔəloo
grandchild (of ego); son’s son ʔisàata ʔisàa
grandfather, maternal (of ego) ʔəmbúu
grandfather, paternal ʔəpúu
grandfather, paternal (of ego) ʔəmbúu
grandmother, maternal (of ego) ʔəmbíi
grandmother, paternal (of ego) ʔəmbíi
grease (for cooking); oil tʰàu
green ʔəntiàk
grasshopper ʔəkʰaurúaŋ
grasshopper ʔəkʰo, ʔkʰa
grind with pestle pʰənaat
grope (in the dark) matʰáp
ground; dry land (vs. water); earth ʔəloo
guest məłóoŋ
gums hákʰií
hail; sleet ʔəriáràa
hair (general) ʔəmii
hair of head sàam
hair, body ʔəmii
hair, facial mòrmíi
<table>
<thead>
<tr>
<th>English</th>
<th>Sorbung</th>
</tr>
</thead>
<tbody>
<tr>
<td>hair, female pubic</td>
<td>ŋirmii</td>
</tr>
<tr>
<td>hair, leg</td>
<td>kěmii</td>
</tr>
<tr>
<td>hair, male pubic</td>
<td>jāŋmii</td>
</tr>
<tr>
<td>hair, underarm</td>
<td>cūbláamii</td>
</tr>
<tr>
<td>hand</td>
<td>kút</td>
</tr>
<tr>
<td>hand span</td>
<td>kūtkʰōom</td>
</tr>
<tr>
<td>hand; arm</td>
<td>kút</td>
</tr>
<tr>
<td>handicapped (whether or not legs are involved); crippled; lame</td>
<td>kēkiuk</td>
</tr>
<tr>
<td>hard; tough</td>
<td>lūu</td>
</tr>
<tr>
<td>hat; headdress</td>
<td>təŋguáp</td>
</tr>
<tr>
<td>have intercourse; copulate (v.)</td>
<td>?atēŋná</td>
</tr>
<tr>
<td>head</td>
<td>lukaan</td>
</tr>
<tr>
<td>head, top of; crown; pate</td>
<td>lusii</td>
</tr>
<tr>
<td>headdress; hat</td>
<td>təŋguáp</td>
</tr>
<tr>
<td>hear (v.)</td>
<td>tʰoó</td>
</tr>
<tr>
<td>heart</td>
<td>məlúŋ</td>
</tr>
<tr>
<td>heart; locus of anger</td>
<td>məlúuŋ</td>
</tr>
<tr>
<td>heart; locus of emotions</td>
<td>niiŋ</td>
</tr>
<tr>
<td>heartbeat (=breath)</td>
<td>?apʰʉ̀t</td>
</tr>
<tr>
<td>heavy</td>
<td>riit</td>
</tr>
<tr>
<td>help</td>
<td>páan</td>
</tr>
<tr>
<td>hen</td>
<td>?aláa</td>
</tr>
<tr>
<td>hen, mother</td>
<td>?abuú</td>
</tr>
<tr>
<td>hiccup</td>
<td>sokáí</td>
</tr>
<tr>
<td>hide</td>
<td>mətʰuúp</td>
</tr>
<tr>
<td>hide; leather (dried animal skin)</td>
<td>?əmèkhùn</td>
</tr>
<tr>
<td>high</td>
<td>məsáan</td>
</tr>
<tr>
<td>hill</td>
<td>cíiŋtòoŋ</td>
</tr>
<tr>
<td>hips</td>
<td>?əkáí, ?kái</td>
</tr>
<tr>
<td>hit (with stick)</td>
<td>pʰiit</td>
</tr>
<tr>
<td>hit (with the fist)</td>
<td>jìu</td>
</tr>
<tr>
<td>hold in mouth (v.)</td>
<td>mùam</td>
</tr>
<tr>
<td>hole</td>
<td>?əkʰuá</td>
</tr>
<tr>
<td>hole (in rock); cave</td>
<td>luŋkʰua</td>
</tr>
<tr>
<td>honey</td>
<td>kʰōcúu</td>
</tr>
<tr>
<td>horn (of animal)</td>
<td>?əkii</td>
</tr>
<tr>
<td>horse</td>
<td>sakúan</td>
</tr>
<tr>
<td>host (house owner)</td>
<td>?inpù</td>
</tr>
<tr>
<td>hot</td>
<td>sáa</td>
</tr>
<tr>
<td>hot coal; ember</td>
<td>moʔék</td>
</tr>
<tr>
<td>house</td>
<td>?in</td>
</tr>
<tr>
<td>humble</td>
<td>tʰoónáa</td>
</tr>
<tr>
<td>hunchback</td>
<td>kəonjkʰuú</td>
</tr>
<tr>
<td>hundred</td>
<td>?ajaa kʰáat</td>
</tr>
<tr>
<td>hungry, be</td>
<td>búcám</td>
</tr>
</tbody>
</table>
hunt
hurt
hurt; sick, be; ill, be
husband
ice
ill, be; hurt; sick, be
insects
intestines; bowels
iron
itch (v.); itchy, be
itchy, be; itch (v.)
jaw; chin
jawbone; mandible; bone, jaw
joint
kick
kick; knead with the feet
kidney
kill
kiss (v.)
knead with the feet; kick
knee
knee cap; patella
knife
lac insect
ladder
lame; handicapped (whether or not legs are involved); crippled
langur; leaf-monkey
lap
large intestine
larynx (throat); adam’s apple
late; slow
laugh
lay egg
lead; start
leaf-monkey; langur
leather (dried animal skin); hide
leech (land)
left hand
leg
lend
length from thumb to forfinger
leprous
lick (v.)
lift up; carry with hand
ligament (bone to bone)
<table>
<thead>
<tr>
<th>English</th>
<th>Sorbung</th>
</tr>
</thead>
<tbody>
<tr>
<td>light (weight)</td>
<td>jaaŋ</td>
</tr>
<tr>
<td>lightning</td>
<td>kʰumléeláap</td>
</tr>
<tr>
<td>like</td>
<td>pam</td>
</tr>
<tr>
<td>like, be; resemble</td>
<td>máaan</td>
</tr>
<tr>
<td>lion</td>
<td>hambru</td>
</tr>
<tr>
<td>lip</td>
<td>mór</td>
</tr>
<tr>
<td>liver</td>
<td>matʰin</td>
</tr>
<tr>
<td>living, be; alive</td>
<td>riŋ</td>
</tr>
<tr>
<td>locus of anger; heart</td>
<td>məlúunŋ</td>
</tr>
<tr>
<td>locus of emotions; heart</td>
<td>niŋ</td>
</tr>
<tr>
<td>locus of thought; mind</td>
<td>niŋ</td>
</tr>
<tr>
<td>long</td>
<td>jōo</td>
</tr>
<tr>
<td>look</td>
<td>?en</td>
</tr>
<tr>
<td>look (imperative)</td>
<td>?enee</td>
</tr>
<tr>
<td>loom</td>
<td>puankʰóoŋnaa</td>
</tr>
<tr>
<td>loose</td>
<td>kaléenŋ</td>
</tr>
<tr>
<td>loose; untie</td>
<td>tərát-kalen</td>
</tr>
<tr>
<td>lost</td>
<td>máaan</td>
</tr>
<tr>
<td>louse</td>
<td>?ərik, ʔrik</td>
</tr>
<tr>
<td>low</td>
<td>mənèm</td>
</tr>
<tr>
<td>man</td>
<td>paasāa</td>
</tr>
<tr>
<td>man, blind</td>
<td>mətcóbāa</td>
</tr>
<tr>
<td>mandible; bone, jaw; jawbone</td>
<td>makʰáaru</td>
</tr>
<tr>
<td>mantis</td>
<td>ʔasóm</td>
</tr>
<tr>
<td>many; very</td>
<td>jáapee</td>
</tr>
<tr>
<td>marrow</td>
<td>ʔəlglák</td>
</tr>
<tr>
<td>meat</td>
<td>ʔəméek</td>
</tr>
<tr>
<td>medicine</td>
<td>lēe</td>
</tr>
<tr>
<td>milk</td>
<td>cúpcʉʉ</td>
</tr>
<tr>
<td>milk</td>
<td>caŋkʰoom</td>
</tr>
<tr>
<td>mind; locus of thought</td>
<td>niŋ</td>
</tr>
<tr>
<td>mithun</td>
<td>səriám</td>
</tr>
<tr>
<td>mock</td>
<td>mənufjā</td>
</tr>
<tr>
<td>molar</td>
<td>habuu</td>
</tr>
<tr>
<td>mole</td>
<td>ʔəbūu</td>
</tr>
<tr>
<td>mole; blemish</td>
<td>səmáamii</td>
</tr>
<tr>
<td>monkey</td>
<td>jooŋ</td>
</tr>
<tr>
<td>month</td>
<td>ʔəkʰáa</td>
</tr>
<tr>
<td>moon</td>
<td>ʔəkʰáa</td>
</tr>
<tr>
<td>morning</td>
<td>ʔəkʰuánlám</td>
</tr>
<tr>
<td>mosquito</td>
<td>tʰəwái</td>
</tr>
<tr>
<td>moth</td>
<td>ʔəpàa</td>
</tr>
<tr>
<td>mother (of ego)</td>
<td>ʔənnūu</td>
</tr>
<tr>
<td>mother-in-law (of ego)</td>
<td>ʔənnii</td>
</tr>
<tr>
<td>mountain</td>
<td>ciŋtōoŋ</td>
</tr>
<tr>
<td>mouse</td>
<td>majūu</td>
</tr>
</tbody>
</table>
moustache
mouth
mucus, dry
mucus, liquid
muscle (=flesh)
mushy (of rice, crushed banana); soft
mute; dumb
my village
nail
name
narrow
nation
nationality
navel
near
neck
necklace
needle
nephew (of ego)
nest
new
niece (of ego)
night
nine
nineteen
ninth
nipple
noon
nose
nose bridge
nipple
on fire, be; burn
one
open
open mouth (v.); gape
orphan
otter
owl
palm
parrot
pate; head, top of; crown

mòrmii
mor
nápkáŋ
nàap
ʔéták
ʔapáat, ʔpáat
conjúuk
ʔítà ʔkʰón
matín
ʔamín
nèe
júur
júur
p(o)láí
ʔánái
riiŋ
məʃíi
ʔajím
ʔitúsáa
ʔawaabúu
ʔatʰár
ʔitúsáa
ʔaján
ʔakúá
som ʔakúá
moolét
ʔúbúáa
ʔəjún
náráa
náráa
nákʰúur
ʔiín
ʔábú
ʔolúú
séndaa
ʔstaaríaa
moo-còk
kʰáat
mahóonj
ʔàanj
sàraasa
sàraám
ʔəwaahambúu
kudmájáa
bokíi
lusíip
patella; knee cap  kʰúk
paternal grandmother ʔəpii
path; road lambūu
path; this road lambūuwá
peel ?akó́r
penis jaaj
person, blind mitcòpà, mitconu
person, deaf nāakʰuák
person, old santáarià
perspiration; sweat məkʰáncʉʉ
pestle təŋjuk
pig ʔəkók
pigeon; dove ʔəʃúu
placenta; afterbirth ʔəlaam
pond; pool pukʰrii
pool; pond pukʰrii
porcupine sakúu
pot təgók
pot, clay lampʰu tokòo, dəgòo
pound rice ʔənpʰeeJnéŋ
powder; dust ʔəhʉ́t
practicioner of witchcraft léváaʃuán
pregnant, be mərái
price ʔəmáan
pus ʔənáí
put 1əʃuá
put out fire; extinguish məmìt
python bəríilòokpáa
queen (no such office) ʔəhuáŋnúu
quiet, very; cold, very dáiriji
cold; quiet dái
rain kâʃuà
rat majúu
rectum; anus ʔakái
red ʔaaŋ
release məkʰáan-lén
resemble; like, be máan
respect kʰəjáaʃá
rib cage wándà ʔərúu
right hand ʔəncáaŋlàm
rise; get up ʔəntʰée
car ʔəŋaacʉ́u, ʔŋaacʉ́u
roll (v.i.) laam
roll (v.t.)
rooster; fowl, male
rope
round
rub
run
rust
saliva
salt
sambar; sambhur
sambhur; sambar
sand
satiated, be; full
scab (hardened rice at bottom of pot)
scale of fish; fin
scar
scoop out; dish out
scrape; scratch
scratch
scratch; scrape
search; seek
second born
see (v.)
seek; search
sell
semen; sperm
seven
seventeen
seventh
sew
shadow
shadow
shallow
sharp
sharpen knife
sheep
shin; calf
shit; feces
shiver; shudder; tremble
shoot (v.)
short (height)
short (length)
shoulder
shout
shrew
shudder; tremble; shiver
shut; close lèn
shy; ashamed jāak
sick, be; ill, be; hurt nāa
taŋpùelām
side (left) ?ancāaŋlām
side (right) ?āliglāk
sinew; tendon (muscle to bone) laasāā
sing tee
sister of female, elder ?inàunū
sister of female, younger ?imēe
sister of male, elder tee
sister of male, younger ?isaanū
sister-in-law (of ego) ?ənjīŋ
sit buum
sit on eggs kōrūuk
six som kōrūuk
sixteen joombāā
skeleton ʔəurāaŋ
skin ʔohūn
skinny kōoŋ
skull lukāantlruu
skull ʔairūu
sky ʔωwaarām
slash and burn; burn off fields lēe-hēe
sleep; asleep, be ?iin
sleet; hail ʔərāraa
slingshot sairōṇ
slow; late jūum
small nēe
small intestine ʔəriinēepaa
smart; wise; clever tʰāaŋmēe
smell; sniff (v.) mənāam
smile (v.) mənūu
smoke mokʰūt
snail càbrùulā, càbrùurā
snake bərii
snake species, large boriis(ɔ)mūuk
sneeze hátii
sniff (v.); smell mənāam
snore ʔənāar
snow ʔshūr
soft mənēe
soft; mushy (of rice, crushed banana) ʔəpāat, ʔpāat
sole kēemajāā
son saa
son’s son; grandchild (of ego) ʔisāata ʔisāa
son-in-law (of ego) ʔihàa
song laa
soul, spirit (Meithei) t’owái
sound (v.i., as a musical instrument); crow (v.) ʔak’uán
sour t’uur
spade bèlcáa
speak (v.) cóŋ
spear ʔasóó
speech; words cóŋ
sperm; semen jaŋcùìu
spider k’oráan, k’ràan
spin məlám
spinal cord (=marrow) ʔolíglák
spine; backbone kòoŋrúu
spit (v.) məcamcùìu mət’úur
spit (v.) məcamcùìu masák
spleen ʔəcúap
split (a watermelon, other round object); break (glass, egg) k’òò
split (as with a candy bar) ʔatáanjèe
split; divide mətáan
spring cəlám
squeeze mit
squeeze; twist mit
squirrel ?əlóo
squirrel, giant ?əlóo
stammer; stutter cóŋʔàak
stand nia
stand up (imperative) nìaèe
star ʔəʃii
start; lead ʔənt’hée
stature ʔəráan
steal mərúù
stick (v.) náap
stink mənámʃoo
stomach wan
stone lùuŋ
storm k’upú k’uncaaj
straight kúŋ
strong məkàt
stupid pàaŋ, ʔəpàaŋ
stutter; stammer cóŋʔàak
suck (v.) məjúp
suckle (v.) (=drink); nurse ʔíin
sun ʔəníi
sunrise ʔəníi-ʃuuk
sunset ʔəníi-lút
surname; clan sakái
swallow (v.) majúu
sweat mækʰáan-cuú
sweat; perspiration mækʰáncuú
swell; swollen, be wór
swollen, be; swell wór
tail (of animal) ?əmóó
tailbone; coccyx ?akáiruu
take lée
tall ʃóo
talon (of bird) matín
tear (n.) mækʰácuú
tempt jűi
ten som
ten thousand lɛcjũ som
tendon (muscle to bone); sinew ʔəlíglák
termite; white ant lũuksáa
testicle jãŋráa
that kaa
the muscle comes out ʔəták ʃùugòo
the tail is long ʔəmóó sóoòo
the tail is short ʔəmóó tòoòo
thick lóok
thigh matóo
thigh bone; femur matóoruu
thin məmpáar
think ?ədúuk
third born teepáa
thirsty, be curée
d thirte en som ʔəntʰúum
thirty somtʰúum
this wa
this axe ròowá
this blanket; this cloth puanúá
this boundary tʰɔrìiwá, tʰɔrìivá
this bow tʰɛewá
this cloth; this blanket puanúá
this corner ʔəkiíwá
this crossbar tatirúá
this door iínkʰáavá
this earring ʔənáatʰíiŋùá
this knife cêmúá
this ladder ʔəláaguá
this loom
this medicine
this necklace
this needle
this pestle
this pot
this price
this road; path
this rope
this slingshot
this song
this spade
this spear
this wall
this war
this window
thousand
three
throat; uvula
thumb
thunder
tie
tiger
tight
to dry
toad
toad
thick
thin
toe, big
toe, little
toenail
tongs
tongue
tooth
torso; trunk
touch
tough; hard
trachea; windpipe
tremble; shiver; shudder
trunk; torso
twelve
twenty	
twist; squeeze
two
udder (of cow, goat)
ugly
puankʰóonjaavá
léewá
məjiiwá
ʔəjimúá
tanfugúá
tagóguá
ʔəmáanúá
lambúuwá
ruuwá
sàiròoŋúá
laavá
bélcaavá
ʔəsoówá
pʰáaglāanúú
réewá
mahutʔoonúú
leefjiŋ kʰáat
ʔəntʰúum
kəráak
kùdbuú
məcék kàa
tarât
hambúu
cín
məkaaŋ
ʔəcok
mahai
kèemajúumràa
kèbúu
kèsáa
kèematin
cəkáap
məlée
haa
məsàa
còo
lùu
kəráak
sái
məsàa
som kəníiŋ
semníi
mit
kəníiŋ
cúúp
càakʃa
umbilical cord  p(ə)lái
uncle, elder paternal  pátáabàa
uncle, husband of elder paternal aunt  pùu
uncle, husband of maternal aunt  pakùu
uncle, husband of younger paternal aunt  pùu
uncle, maternalUncle, younger paternal  pakùu
underarm; armpit  cúbláá
untie; loose  təràt-kəlen
urinate (v.)  məcée-len
urine  məcécʉ̀ʉ, məcécʉ̀ʉ
uvula; throat  kəráāk
vagina  jîr
valley (river valley)  təmbàak
vein; artery; blood vessel  məŋetʰáarʉʉ
vertebrae (of spine)  kənrūu
very; many  jáapec
village  ?akʰuɑn
voice  ?akʰúɑn, ?kʰúɑn
vomit (v.)  məlît
wait  nàak
wait (imperative)  nàagèe
wake up (v.i.); awake, be  ʔəntʰèe
walk  jáu
wall  pʰáaglàaŋ
want; desire  pam
war  rée
wart (= skin disease)  ʔəhʉ́nnáa
wash (clothes)  ʔəʃúu
wash (hands, vegetables, cars)  pʰərii
water  cʉ́ʉ
weak  juɑn
weep; cry  càap
well, be  ʔələo
wet  juu
whiskers (of animal)  mòrmíi
whistle  məhʉ́u
white  ŋée
white ant; termite  lűuksáa
whole  kəlúktaee
widow  nəmoomə̀kʰúu
wife  mənamóo
wind  kʰiį
window  məhʉtʰoon
windpipe; trachea  kərāāk
<table>
<thead>
<tr>
<th>English</th>
<th>Lush</th>
</tr>
</thead>
<tbody>
<tr>
<td>wing (of bird)</td>
<td>pʰəláa, pʰláa</td>
</tr>
<tr>
<td>wise; clever; smart</td>
<td>tʰàanmee</td>
</tr>
<tr>
<td>wolf</td>
<td>rámʔʉ́u</td>
</tr>
<tr>
<td>woman</td>
<td>moosáa</td>
</tr>
<tr>
<td>woman, blind</td>
<td>mitcònuu</td>
</tr>
<tr>
<td>woodpecker</td>
<td>?əwaatʰìŋhútpá</td>
</tr>
<tr>
<td>words; speech</td>
<td>cóŋ</td>
</tr>
<tr>
<td>world; cosmos</td>
<td>ókətʰui</td>
</tr>
<tr>
<td>yawn (v.)</td>
<td>mahám</td>
</tr>
<tr>
<td>year</td>
<td>kuum</td>
</tr>
<tr>
<td>yellow (color of curry)</td>
<td>jaiŋàanməcuu</td>
</tr>
<tr>
<td>young (age)</td>
<td>mənée</td>
</tr>
</tbody>
</table>
FLOATING QUANTIFIERS IN BURMESE AND THAI

Andrew Simpson  
*University of Southern California*  
<andrew.simpson@usc.edu>

**Abstract**

This paper considers patterns of floating quantifiers in Thai and Burmese and attempts to provide answers to three general questions. First, syntactically how are nouns and floating quantifiers displaced from each other? Second, what pragmatic/functional motivations underlie the optional use of floating quantifier structures? Third, are patterns of Q-float uniform across languages, or are there differences, and how might these be accounted for? It is argued that Q-float differences found in Thai and Burmese are due to the interaction of universal principals shaping linear word order (information structure) and language particular syntactic organisation (the head-initial/final parameter).

**Keywords:** quantifiers, topicalization, numerals

**1. Introduction**

In many languages certain elements which quantify over noun-phrases/NPs, such as numerals and quantifiers equivalent to English ‘all’ and ‘both’, may most frequently occur adjacent to the NP they modify, but at other times occur separated from the NP they relate to, when they are commonly referred to as ‘floating quantifiers’. Typical alternations between the NP-adjacent and floating occurrences of nominal quantifiers are illustrated in examples (2-4) from English, Thai and Burmese following the linear schematization in (1) (Q = quantifier):³⁰

(1)  
non-floated patterns: NP Q...... or Q NP....’  
common floated pattern: NP Q......  
less common floated pattern: Q NP....

**ENGLISH**

(2)  
a. *All the students* have arrived. *non-floated*  
b. *The students* have *all* arrived. *floated*  
NP = [the students]  
Q = [all]

³⁰The Thai and Burmese data presented in the paper was mostly gathered in a series of interviews carried out in 2004 with multiple native speakers of Thai and Burmese who were journalists regularly writing/broadcasting in Thai and Burmese, or teachers of one of the two languages. A preliminary version of the paper was subsequently presented at the SEALS conference in Chulalongkorn University in 2005. Sincere thanks are due to the language consultants who have helped with clarification of the patterns reported in the paper, and to two anonymous reviewers of the paper.
Floating quantifier patterns are linguistically interesting because they appear to show that it is possible to pull apart and separate two parts of a single grammatical unit (for example, a subject or object) into a discontinuous sequence of elements which are nevertheless still interpreted together, the quantifier being understood as modifying the reference of the NP. This gives rise to three important, general questions. First, how does this syntactic separation occur - what is the structure of non-floated NP/Q constituents, and how does it allow for the NP and the quantifier to be separated? Second, why are the NP and quantifier sometimes separated from each other? When elements are distanced from the position they would otherwise normally occur in, this may make sentences more difficult for hearers to parse and process. What functional purpose and benefit might there be in sometimes splitting apart NP and quantifier constituents? Third, there is a general typological question of whether floating quantifier patterns are indeed the same across different languages, and if not, what are the relevant differences and how might these be accounted for? Is it possible to identify any cross-linguistically shared properties in floating quantifier constructions?

This comparative study of floating quantifiers in Burmese and Thai sets out to investigate these issues and provide initial answers to these three questions which will hopefully be of use in the continued study of floating quantifiers. Concerning the particular choice of languages focused on in the present the study, it is interesting and potentially revealing to compare floating quantifier phenomena in Burmese and Thai because although the structure of NPs is similar in many ways in the two languages, the basic word of Burmese and Thai is fundamentally different, Burmese being an SOV head-final language, while Thai is a language with very typical SVO patterns. In the course of the paper, it will be suggested that differences in basic word order may indeed impact on the way that the phenomenon of floating quantifiers occurs in a language. In addition to Burmese and Thai, the paper will also make comparative reference to certain other
languages in which significant work on floating quantifiers has been carried out, in
particular English, and also Japanese and Korean.

2. Previous syntactic characterizations of floating quantifiers:
   the ‘movement’ analysis
Within generative grammar approaches to linguistics, it has been common to analyse
floating quantifiers as being transformationally derived from non-floating structures via an
operation of constituent displacement/movement. It is posited that quantifiers and their
associated NPs are regularly formed as single continuous syntactic units, and then in
certain instances separated from each other resulting in discontinuous sequences of
quantifier and NP. Such a hypothesis is argued to provide a principled account of various
properties of floating quantifier constructions, for example the observation that a quantifier
can frequently only occur in a floated position if it can also optionally be positioned
adjacent to its NP associate, as illustrated in (5) and (6)

(5)  [both] [the students] → [the students].....[both]...
    Both the students left  → The students have both left.
    [all]  [the students] → [the students]....[all]...
    All the students left.  → The students have all left.

Where a quantifier cannot be inserted directly adjacent to an associated NP (without other
supporting words), a floating form also seems to be unavailable, as shown in (6),
suggesting a systematic relation between floated and non-floated forms that can be
described in terms of a movement transformation converting non-floated sequences into
forms with NP and quantifier separated from each other.

(6)  *[few] [the students] → *[the students]....[few]...
    *Few the students left.  → *The students have few left.
    *[some] [the students] → *[the students]....[some]...
    *Some the students left.  → *The students have some left.

With regard to the question of which of the two elements NP/quantifier is understood to be
displaced and moved away from the other, there are two logical possibilities. First, it
might be hypothesized that the quantifier floats rightward away from an NP in subject
position, as schematized in (7) (with strike-through representing the original, underlying
posited position of the quantifier):

(7)  [All the students] have arrived.  → [All the students] have all arrived.

This possibility is commonly rejected as an analysis of floating quantifiers in English for
theory-internal reasons, as the hypothesized movement would have to be analyzed as
repositioning the quantifier in some structurally lower position in the syntactic
configuration. Movement transformations are otherwise uniformly assumed to reposition
elements in higher structural positions (which “c-command” the position moved from –
languages such as English is the quantifier itself is optionally left behind or “stranded”
when an NP moves from a lower VP-internal position to the regular, surface subject position preceding auxiliary verbs, as represented in (8):

(8)  

a. [**All the students**] have [all the students] arrived.

b. [**The students**] have [all [the students] arrived.

As there is other cross-linguistic evidence suggesting that subjects may originate in lower VP-internal positions (Ouhalla 1994), and the hypothesized movement in (8) is to a higher position in the syntactic structure, the analysis of quantifier floating in (8) has been widely adopted in transformational approaches to language. Such an analysis can be used to account for a range of phenomena, for example the observation that a quantifier such as ‘all’ can only occur floating between auxiliary verbs if there has been movement of the object to subject position, as in passive sentences such as (9), and ‘all’ may not simply be inserted in a floating position if no object-to-subject movement has occurred, as in parallel active transitive forms such as (10b). In (9), it is suggested that passive movement of the patient argument of the verb from object to subject position transits through an intermediate position between the auxiliary verbs, where the quantifier ‘all’ can be optionally stranded, giving rise to the floating pattern. This is schematized in representation (11):

(9) **The criminals** have **all** been arrested.

(10) a. ✔He has arrested **all the criminals**.
    b. *He has **all** arrested the criminals.

(11) [**The criminals**] have [**all the criminals**] been arrested [**all the criminals**].

The fuller documentation of floating quantifier patterns in English has also noted certain other regular properties constraining the distribution of NP/quantifier separation (Bobaljik 2003). First, although passive structures permit quantifier float, as seen in example (9), other constructions which are regularly analyzed as involving movement transformations in English such as topicalization and relative clause formation do not appear to permit the stranding of quantifiers associated with topicalized and relativized NPs, as shown in (12) and (13):

(12) a. *The students, I took **all** to the show.
    (cf. All the students, I took to the show).

b. *The reports, I recently **both** responded to.
    (cf. Both the reports, I recently responded to.)
Floating Quantifiers in Burmese and Thai

(13) *The students (who) I have all met today are very nice.
(cf. All the students (who) I have met today are very nice.)

Second, there is a definiteness restriction which applies to the NP in floating quantifier constructions in English – stranding and floating of a quantifier may only occur if its associated NP in subject position is definite (marked by ‘the’/’those’):

(14) a. All/Both participants have now arrived.
b. *Participants have both/all now arrived.
c. The/Those participants have now both/all arrived.

Such a restriction may account for the fact that numerals may not be stranded as floating quantifiers in English, as in cases such as (15) because the associated NP is indefinite:

(15) a. Three students have now arrived.
b. *Students have three now arrived.

In summary, floating quantifiers in English are assumed to be stranded by movement of an NP to the subject position of a sentence, and NPs raised to such a position must be definite in reference (preceded by ‘the’ or a demonstrative).

3. Burmese
Turning now to consider floating quantifier patterns in the southeast Asian language Burmese (Tibeto-Burman), it can be observed that there are two common differences between quantifierJfloat/QJfloat in English and Burmese, as well as other similarities. The first clear difference between the two languages is that floated quantifiers in Burmese are very frequently numerals (coupled with classifiers). As noted above (example 15b), numerals do not occur as floating quantifiers in English. Secondly, the NPs which occur separated from their modifying quantifiers in Burmese frequently do not occur in the subject position of the sentence as in English, and are instead often located in sentence-initial topic positions. Both such properties are illustrated in example (16):

(16) a. Daw Khin-Khin-ḵɛtʰan ngāa-caun wɛ-te non-floated
Daw Khin-Khin-NOM pencil 5-CL buy-REAL
‘Daw Khin-Khin bought 5 pencils.’

b. ḵɛtʰan canaw ngāa-se-daun wɛ-te floated
pencil 1 50-even buy-REAL
‘I bought as many as 50 pencils.’

Separation of the NP and quantifier elements in Burmese can be naturally analyzed as occurring via stranding of the quantifier when there is movement of the NP to a higher (topic-like) position. The quantifier element is naturally located in the position that the NP would occur in if there were to be no splitting and separation of the NP and quantifier. The alternation in (16) can be schematized as in (17). (16a) and (17a) are neutral S-O-V-Aux

---

31 Here and in subsequent examples, CL stands for classifier, and REAL for realis.
word orders. The sequences in (16b) and (17b) can be analysed as arising when the NP raises out of the neutral object position to the pre-subject topic position, stranding its associated quantifier Q in the object position:

(17)  a. \([\text{Subject Daw Khin-Khin-kə}] [\text{Object [NP } \text{khɛ̄tan}] [Q \text{ngāa-caun }] \text{ we-}\text{tɛ}\]

\[
\begin{array}{c}
\text{Subject} \\
\text{Daw Khin-Khin-kə} \\
\text{Object} \\
\text{[NP khɛ̄tan]} \\
\text{[Q ngāa-caun]} \\
\text{wɛ-tɛ}
\end{array}
\]

b. \([\text{NP } \text{khɛ̄tan}] [\text{Subject Daw Khin-Khin-kə}] [\text{Object [NP } \text{khɛ̄tan}] [Q \text{ngāa-caun }] \text{ we-}\text{tɛ}\]

An analysis of movement of the NP stranding the quantifier is supported by other patterns and restrictions in Burmese. A first restriction is that an NP separated from its associated quantifier must structurally “c-command” the quantifier. Example (18) below is ungrammatical because the NP meet8swee ‘friend’ inside the subject meet-swee ye seq-bein ‘friend’s bicycle’ does not c-command the quantifier it relates to. This indicates that floating quantifiers cannot simply be inserted into a sentence in a random way but are subject to clear syntactic restrictions. The ungrammaticality of examples such as (17) can be simply explained if floating quantifiers can only result from the movement of an NP to a position which c-commands the quantifier.

(17) *[měet-\text{\text{-swee-yɛ̄}} \text{seqbēin]-kə thǒun-yāuq akhoo-khan-yq-tɛ} \\
\text{friend-GEN bicycle-NOM 3-CL were-stolen}
\text{Intended interpretation: ‘Three friends’ bicycles were stolen.’}

A second, movement-related restriction is that an NP cannot be associated with a floating quantifier that occurs inside a syntactic “island” (constituents such as relative clauses, embedded questions, adjunct clauses – Ross 1967). As syntactic islands regularly disallow movement of a constituent from within the island to a position external to the island, this restriction on NP–quantifier relations in Burmese is again simply explained if the NP moves away from the position of the quantifier in sentences with floating quantifiers. Example (17) illustrates the ungrammaticality of a floating quantifier inside a relative clause island when its associated NP is external to the relative clause:

(18) *wiisakiihayɛ̄q canaw [manɛ-kə thǒun-palîn we tɛ meinmə-kə] thî-pa-tɛ \\
\text{whisky I yesterday 3-CL bought woman-ACC know}
\text{‘Whisky, I know the woman who bought three bottles yesterday.’}

A third restriction found in Burmese is that a subject NP cannot relate to a quantifier which follows the object in a sentence, as schematized in (19) and illustrated in example (20):

(19) *NP_{\text{subject-k}} \text{ NP_{object}} Q_{k} V

\[32\] The structural notion of c-command as it applies in syntactic tree structures is defined as follows: ‘A node X c-commands a node Y if the first branching node dominating X dominates Y, and X itself does not dominate Y.’ (adapted from Radford 1988 p.115).
Floating Quantifiers in Burmese and Thai

(20). *cāun-thāa(-ka) htaṃn-caw hna-vāuq hmāa-tē
student(-NOM) fried-rice 2-CL ordered
Intended meaning: ‘Two students ordered fried rice.’

The ungrammaticality of examples such as (20) can be accounted for if floating quantifiers arise when there is a movement of the associated NP away from the position of the quantifier. In (20) it can be suggested that the subject NP will never have occupied any position to the right of the object (given the SOV base word order of Burmese), and so will not be able to strand a quantifier in such a position. Such patterns therefore again support a movement and stranding analysis of floating quantifiers in Burmese. Note that other adverbs can occur to the right of the object, as shown in (21), which provides an argument against treating floating quantifiers as simple adverbial elements, and favors the movement-stranding account:

(21) canaw htaṃn-caw khana-khanq sāa-tē
 I fried-rice often eat
‘I often eat fried rice.’

A final restriction which can be noted here is that NPs cannot relate to floating quantifiers located inside postpositional phrases/PPs. Similarly, it is not possible for an NP contained within a PP to relate to a quantifier outside that PP. The generalization is therefore that if an NP and a quantifier occur together inside a PP, there is no way that such elements can be separated from each other. This follows from an analysis in which NPs relate to floating quantifiers via separation and movement of the NP from a position adjacent to the quantifier. Examples (22-24) illustrate such patterns with a range of different postpositions and PPs:

PP LOCATION
(23) a. turīt-twee-kā [pp[hote thōun-khu] hmāa] tē-ne-tē
tourist-PL-NOM hotel 3-CL in stay-ASP-REAL
‘The tourists were staying in 3 hotels.’
b. *hote turīt-twee-kā [ppthōun-khu hmāa] tē-ne-tē
hotel tourist-PL-NOM 3-CL in stay-ASP-REAL
c. *[pphote hmāa] turīt-twee-kā thōun-khu tē-ne-tē
hotel in tourist-PL-NOM 3-CL stay-ASP-REAL

PP COMITATIVE
 I Sulee-temple-ACC monk 3-CL with go-REAL
‘I went to Sulee Temple with 3 monks.’
b. *pounjii canaw Suulee-payāa-ko [ppthōun-paa]nē thwāa-tē
monk I Sulee-temple-ACC 3-CL with go-REAL
c. *pounjii nē canaw Suulee-payāa-ko thōun-paa thwāa-tē
monk with I Sulee-temple-ACC 3-CL go-REAL
As PPs are known to be constituents which disallow extraction in many languages, the restriction here can again be suggested to support an analysis of floating quantifiers as being derived by stranding and movement. Summarizing what has been presented in this section, it can be noted that the distribution of floating quantifiers in Burmese is highly compatible with an analysis of stranding as the result of movement of an NP, which is commonly to a topic-like position, and that floating quantifiers in Burmese are frequently comprised of a numeral and an appropriate classifier.

4. Thai
Standard Thai (Tai-Kadai) is the second southeast Asian language to be considered in detail in this paper. As noted earlier, the internal linear organisation of nominal expressions in Thai is similar to Burmese, with nouns/NPs being followed by numerals/other quantifiers and classifiers: [NP Numeral/Quantifier Classifier]. However, the neutral ordering of clause-level constituents in Thai is significantly different to Burmese, with Thai being an [S Aux V O] language and Burmese having neutral [S O V Aux] order. It is therefore interesting to consider whether issues of basic clausal word order might possibly affect the way floating quantifiers are distributed in the two languages.

4.1 A movement analysis of quantifier float in Thai?
In section 4 above, it was argued that floating quantifiers in Burmese arise via the stranding of a quantifier following movement of an NP to a higher position, as is frequently assumed for English. Shifting our focus to examine Thai, now, a natural comparative question is whether the basic mechanisms of movement and stranding are also responsible for creating floating quantifiers in this language? Consider first the occurrence of floating quantifiers in passive sentences in Thai, such as (26):

(26) rōt-Mercedes thūuk khəmɔy sīisiphàa-khan
    car-Mercedes PASS steal 45-CL
    ‘45 Mercedes were stolen.’

Such examples would appear to be compatible with the assumption that the numeral and classifier sequence seesiphaa-khan is stranded by movement of the theme NP rōt-Mercedes from object to subject position during the derivation of the passive sentence,
which in turn might support the conclusion that Thai floating quantifier constructions are fundamentally similar to those in Burmese and English. However, further data involving floating quantifiers in Thai show that there are clear differences between Thai, Burmese and English which indicate that floating quantifiers in Thai do not have the same syntactic derivation as those in Burmese and English.

A first difference with regard to English is that, despite sharing a similar SVO word order, Thai does not permit the occurrence of floating quantifiers in positions between auxiliary and main verbs, unlike patterns common in English:

(27) a. The children will **all** have arrived by now.
    b. The children will have **all** arrived by now.

(28) *phuak-dek aat-ca thuk-khon maa lǣw
    children may every-CL come ASP

(29) **The cars** were **all** stolen.

(30) *rōth-Mercedes thūk sīsiphàa-khan khămōy
    car-Mercedes PASS 45-CL steal

A second comparative difference is that in English floating quantifiers actually do not occur in post-verbal object position, unlike Thai (as illustrated in (26)):

(31) *The cars were stolen **all**.

Thirdly, and very importantly, when the position of floating quantifiers is considered more closely in Thai, it actually does not correspond to a potential “stranding” position which an associated NP could have moved from. This can be seen in a closer examination of passive sentences with floating quantifiers such as (32) below. Here the floating quantifier does not occur in the object-of-verb position following the verb ‘steal’ khamooi, and is instead located in sentence-final position following the adjunct of location ‘from a factory in Stuttgart’. The floating quantifier is therefore not located in a position from which the associated NP rot-Mercedes might have been moved, arguing against a simple movement-and-stranding analysis in Thai:

(32) rōth-Mercedes thūuk khămōy cāak rōong-ngaan
    car-Mercedes PASS steal from factory
    nay mung Stuttgart sīsiphàa-khan
    in town Stuttgart 45-CL

‘45 Mercedes were stolen from a factory in Stuttgart.’

Floating quantifiers in Thai also occur sentence-finally when relating to subject NPs in (active) sentences where the subject cannot be suggested to have raised from sentence-final position under any standard (transformational) analysis:
A woman came looking for you.'

‘Almost every student has read this chapter.’

Similarly, floating quantifiers associated with objects can occur clause-finally distanced from the latter in positions which could never before be occupied by direct object NPs:

‘He gave me 200 Baht.’

This occurrence of floating quantifiers in positions that are not possible positions for object NPs can also clearly be seen when the object is separated from the floating quantifier by aspect-marking elements such as yȕu, maa, pai, sɛ̆t and lēew. In none of the examples in (36-39) below could the object NP occur in the position occupied by the associated floating quantifier.

‘I only have one really good pair of trousers.’

‘He bought two books.’

‘Yesterday the police arrested two students in the market.’

‘He finished eating two of the dishes.’

The general observation resulting from such data is that floating quantifiers in Thai very frequently occur in positions which their associated NPs could not have previously occupied or have been moved from under any standard transformational analysis incorporating the notion of syntactic movement/displacement. This results in the conclusion that floating quantifiers in Thai, unlike English and Burmese, cannot be analysed as resulting from stranding following the movement/displacement of an associated NP.
This being so, the question remains as to how floating quantifiers in Thai may occur in different types of positions to those in English and Burmese? If it is supposed that some form of movement/displacement operation is still potentially involved in separating quantifiers from NPs in Thai, as assumed in other languages, the question then becomes what kind of rather different movement/displacement operation could apply to create the structures found in Thai? Arguably, the only obvious way to analyse the data in Thai in terms of movement is to assume that in Thai separation of NPs and their quantifiers is achieved by movement of the quantifier-element rather than movement of the NP, and that the quantifier-element is displaced to some rightward position in the clause, most commonly following other argument NPs and the aspect markers pay/maa/sêt/yûu and either preceding or following the aspect marker lêew, as schematized in (40):

(40) a. NP-Q V NP pay/maa/sêt Q lêew

b. NP-Q V NP pay/maa/sêt lêew Q

c. NP V NP-Q (NP) pay/maa/sêt Q lêew

d. NP V NP-Q (NP) pay/maa/sêt/lêew Q

The movement hypothesized above is quite different from the movement assumed in English and Burmese floating quantifier constructions. In English and Burmese the movement of the NP is leftward and can be suggested to occur for reasons of case (English) or topicalization (Burmese). In Thai the movement of the quantifier is to the right, and for (as yet) unclear reasons/motivations.

Such a working hypothesis naturally leads to the question of whether cross-linguistically it is possible to identify other instances of ‘rightward movement’ which might support a rightward movement analysis of floating quantifiers in Thai? The answer to this question is certainly ‘yes’, with instances of extraposition and ‘Heavy NP Shift’ being two common occurrences of the apparent displacement of syntactic constituents to the right of a clause, as illustrated with English (41) and (42) below:

(41) Someone [who would change our lives forever] then entered the room [who would change our lives forever].

(42) John recently sent [a book about Polish morphology] to me [a book about Polish morphology].

---

33 Heavy NP Shift characteristically involves the rightward displacement of heavy/long objects in double object constructions, as in (42). Instances of extraposition include (but are not limited to) the rightward displacement of PPs and relative clauses from subjects in English, as in (41).
Note that similar to NP/floating quantifier pairs, in cases of extraposition such as (41) there is a syntactic constituent (a relative clause) which would normally occur attached to an NP, but is here separated from that NP. Furthermore, as with Thai floating quantifier structures, the NP in sentences with extraposition such as (41) cannot itself occur in the position where the associated relative clause is found (as in: *Then entered the room someone.'), forcing the conclusion that the relative clause has not been stranded by movement of the NP from some clause-final position but has been moved rightwards away from the NP in subject position.

A further restriction found to characterize occurrences of extraposition is that the NP which relates to an extraposed relative clause must be interpreted as non-specific and indefinite. The NPs which are associated with floating quantifiers also commonly have the property of being non-specific, which suggests a further parallel between extraposition and Thai rightward floating quantifiers. In analyses of extraposition, the specificity constraint is understood to be a restriction on movement and extraction (barring extraction from specific NPs). If a similar restriction holds of NP-floating quantifier relations, this may add further support for a movement analysis of floating quantifiers in Thai. Additionally, and potentially relating to the issue of non-specificity, functionally the rightward extraposition of PPs and relative clauses in English is commonly used when an NP is introduced for the first time into the action described in a discourse situation, as for example in (43) and (44):

(43)  [A review ] appeared in the Times [of a new book about Roosevelt].

(44)  I met [a man ] yesterday [who had known your father in the 1960s]

As will later be discussed, floating quantifiers in Thai are also used frequently in presentational situations, increasing the parallels between extraposition and floating quantifier constructions.

Finally, the hypothesis that rightward syntactic movement is involved in the distribution of floating quantifiers in Thai is supported by the interaction of quantifier float and standard configurational restrictions on movement, i.e. ‘island phenomena’. As in Burmese, there are patterns indicating that the relation between an NP and a floating quantifier in Thai is regularly restricted by the occurrence of ‘island’ constituents, and it is not possible for an NP inside a constituent such as a relative clause, adjunct clause or other island type to relate to a floating quantifier located outside such a constituent in clause-final position, as illustrated in (45). This example is only acceptable with the continuation in (a) in which the floating quantifier is associated with the noun ‘man’ which is external to the relative clause island, and is not acceptable if relating to the noun ‘Rolls Royce’ inside the relative clause as in (b) (where the switch in classifier from the classifier for people khon to the classifier for vehicles khan makes the intended meaning clear, though unacceptable):

---

34 An NP is non-specific if its identity is unknown by both the hearer and the speaker prior to the action described in the sentence.
Floating Quantifiers in Burmese and Thai

If the sensitivity of a syntactic dependency to island phenomena is an indication that such a dependency is the result of movement, then the presence of island restrictions in floating quantifier constructions in Thai clearly suggests that movement of the floating quantifier occurs, separating it from the NP.

4.2 Consideration of a non-movement analysis: floating quantifiers as adverbs

While the distribution of floating quantifiers in Thai seems highly compatible with a rightward movement analysis similar to extraposition, this being supported by island phenomena and other parallels with extraposition, one might also consider an alternative potential analysis of floating quantifiers, that they are not moved to their surface position from any other underlying location but are inserted directly into their clause-final position as (or, rather, like) adverbs. Such a non-movement approach to floating quantifiers might be supported by the following range of observations.

First, VP-level adverbs occur in similar clause-final positions to those occupied by floating quantifiers in Thai:

(i)  kháw sŏng-khon  (Thai)   (ii)  thuṭō hna-yāuq  (Burmese)
they 2-CL   they 2-CL
‘the two of them’ ‘the two of them’

Parallel forms also occur in Mandarin Chinese, as shown in (iii) from Li (1999):

(iii)  tāmen liang-ge  (Mandarin)
they 2-CL
‘the two of them’

A reviewer of the paper notes that there are patterns in both Thai and Burmese which might suggest that even when a numeral-classifier pair are adjacent to a noun/NP, the numeral-classifier pair might not be part of the DP nominal constituent, hence might always be adverbal in nature. Specifically, it is observed that pronouns in Thai and Burmese can be followed by numeral-classifier pairs, as illustrated in (i) and (ii). If it were to be assumed that the pronoun replaces the full DP, one might have to conclude that numeral-classifier pairs are, at least in some instances of adjacency, not necessarily part of a DP:

(i)  kháw sóng-khon  (Thai)   (ii)  thuṭō hna-yāuq  (Burmese)
they 2-CL   they 2-CL
‘the two of them’ ‘the two of them’

Parallel forms also occur in Mandarin Chinese, as shown in (iii) from Li (1999):

(iii)  tāmen liang-ge  (Mandarin)
they 2-CL
‘the two of them’

Li (1999:83) argues convincingly that the pronoun in such sequences is in the D\(^0\) position (actually a fairly common syntactic assumption about pronominal elements), and that the numeral and classifier occur in the same post-D\(^0\) DP-internal positions that they occupy in other instances. There is consequently no need to conclude that the post-pronominal occurrence of numeral classifier pairs in examples such as (i-iii) indicates that such elements are DP-external and adverbal in nature. Furthermore, Simpson (2005) shows that elements which close off DPs, such as case-markers in Burmese, and demonstratives in Thai, can occur following post-nominal numeral-classifier pairs, confirming that such sequences are DP-internal elements, at least in certain occurrences, and hence would not be naturally analyzed as adverbial elements in all instances.
Second, there are modifiers constructed from classifier bases which do occur in adverb-like ways in clause-final position. In examples (47-48), the bolded, underlined elements containing classifiers are not floating quantifiers as they either do not contain a quantifier or do not relate to any overt NP in the sentence, and yet they are licensed to occur in similar positions to other floating quantifiers.

(47) a. khrāng-thiilɛ̄εw kháw maa [khon-diaw] time-last he come CL-single
   ‘Last time he came alone.’

   b. khɔ́ɔ pay dúay [khon] nā request go together CL PRT
   ‘Can I go too?’

(48) lɔ̏ɔn klȁay pen thiiprūksáa khōong kháw pay thūk-rṳ̄ọng lɛ̄εw
she change be advisor of he ASP every-CL ASP
   ‘She became his advisor in everything.’

Third, there may be some final floating quantifiers which can not occur adjacent to the NP they modify, hence which do not alternate with a non-floated form. For example, in (49) below, it is not possible for the floating quantifier sequence sáam-sȍp to occur adjacent to its NP associate phùak-nān as shown in (49b). The absence of such an adjacent NP quantifier sequence may suggest that the floated form is not derived from an NP quantifier unit via movement of the quantifier.36

(49) a. kháw yīng phùak-nān taay lɛ̄εw sáam-sȍp
   they shoot group-that die ASP 3-CL
   ‘They shot three of them dead.’

   b. ??*khāw yīng phùak-nān sáam-sȇp taay lɛ̄εw
   they shoot group-that 3-CL die ASP

It can also be noted that Thai floating quantifier patterns are different to those in Burmese (and Japanese, Korean; Kang 2002) in the patterning of sentences with prepositional phrases/PPs. In Thai it appears to be quite possible for a clause-final floating quantifier to relate to an NP located inside a PP. In other languages such as Korean, Japanese and Bengali (Simpson and Bhattacharya 2008) where floating quantifiers may not relate to an NP inside a PP this has been suggested to be because PPs in many languages may disallow extraction/movement. If this is the correct interpretation of PP-related data,

36 The complication in this example is that sȍp is the classifier for dead bodies and only seems to allow for use in counting corpses once it has been established that the relevant people are dead – hence in floated clause-final position after ‘died’ taay lɛ̄εw.
it might weaken the case for a movement analysis of floating quantifiers in Thai. Examples (50-52) show that quantifier units floated in clause-final position are free to associate with NPs contained in a range of PP types:

(50) kháw kêp ngan [ppcāak nāksāksāa] maa lēew kūap thūk-khon
    he collect money from student ASP ASP almost every-CL
    ‘He has collected money from almost all of the students.’

(51) chán khuy [ppkāp khēek] maa lēew sōong-sāam-khon
    I chat with guest ASP ASP 2-3-CL
    ‘I’ve talked with about 2 or 3 guests already.’

(52) kháw sōon-nāngsù [ppthii mahāawitayalaay] maa lēew sīi-kwāa hēeng
    he teach in university ASP ASP 4-over-CL
    ‘He has taught in more than 4 universities.’

4.3 The partitivity issue

Data relating to interpretations of partitivity provide further potential clues as to the derivation of floating quantifier structures. Sentences with floating quantifiers sometimes have different interpretations from those where numeral-classifier pairs are not floated. Consider examples (53) and (54), where quantifier float occurs in (54) (but not (53)):

(53) dēk sāam-khon taay lēew
    child 3-CL die ASP
    ‘The three children have died.’

(54) dēk taay (pay) sāam-khon lēew (or: dēk taay lēew sāam-khon)
    child die ASP 3-CL ASP child die ASP 3-CL
    ‘Three of the children have died (so far/already)’

(53) is commonly described (by speakers of Thai) as having the meaning that some group of three children known to the speaker and hearer (i.e. a definite group of three children) had died. (54), by way of contrast, is suggested to mean that three children from some group known to the speaker and hearer have died, and to imply that there are still other children from that group who may be in danger. This corresponds to a “partitive” interpretation in which the numeral+classifier quantifies over a definite set.

A similar difference in interpretation is found to occur when the NP is located in object position as in (55) and (56) (quantifier float takes place in (56)):

(55) kháw kin kāpkhàaw sōong-vāang sêt lēew
    he eat dishes 2-CL ASP ASP
    ‘He has finished eating the two dishes.’

(56) kháw kin kāpkhàaw sêt sōong-vāang lēew
    he eat dishes ASP 2-CL ASP
    ‘He has finished eating two of the dishes.’
(55) is characterized as meaning that there are only two dishes in total (and they are now consumed), whereas (56) implies that there is more food on the table and has the partitive-like interpretation ‘two of the dishes’.

The potential significance of this data is the following. If clause-final floating quantifiers are supposed to be derived by movement of a numeral+classifier from a position adjacent to the NP, it might not be expected that this movement would affect the meaning/interpretation of the numeral+classifier in such a clear way. In other words, why would a partitive interpretation be present with floating quantifiers but not non-floated numeral+classifier sequences if the former are simply derived from the latter? This difference in interpretation could be taken as an argument against analyzing floating quantifiers as the result of a movement transformation, as syntactic movement is regularly understood to preserve rather than alter meaning.

In trying to make sense of the alternations here, it is useful to reflect on how partitive interpretations generally may arise from syntactic structures. In English and many other languages, partitive interpretations occur when a numeral (or a universal quantifier such as ‘all’) is positioned external to the “DP” unit created by the addition of a determiner to an NP, as represented in (57):

\[
\text{QP two/all of [DP the [NP students]]}
\]

This contrasts with the interpretation which arises when a numeral occurs inside the DP, following ‘the’:

\[
[\text{DP the two students}]
\]

The sequence in (57) implies there are more students who are part of a group familiar to both speaker and hearer, while (58) refers to a group composed of just two students known to speaker and hearer, and there is no implication that other students belong to this particular group. This allows for the statement of a simple partitive generalization relating to the syntactic structuring of partitive phrases:

\[
\text{PARTITIVE GENERALIZATION}
\]

A numeral which is external to and quantifies over a definite DP/NP gives rise to a partitive interpretation.

A numeral which is internal to a definite DP/NP does not give rise to a partitive interpretation.

Now, if adjacent sequences of NP + numeral-classifier in Thai such as dek saam-khon (ex. 53) and kap-khaaw soong-yaang (ex. 55) do not give rise to partitive interpretations, this suggests that the numeral+classifier in such sequences has to be interpreted as being ‘internal’ to any definite DP and that Thai does not have a second possible position for numeral+classifiers equivalent to the position of the ‘outer’, DP-external numeral in English (57). Linearly adjacent NP + numeral-classifier sequences would therefore be assumed to always have the structure [DP NP quantifier.classifier] and not the structure [[DP NP ] quantifier classifier] (which would be expected to license partitive interpretations). Finally, as movement operations are assumed to conserve
fundamental aspects of meaning, if Thai only makes available a DP-internal non-partitive position for numeral+classifier pairs, it might be concluded that clause-final floating quantifiers associated with partitive interpretations cannot be moved from positions adjacent to the NPs they modify.

However, the fuller patterning of partitivity with nominal expressions is actually more complex still than the patterns seen above, and there is clear evidence that Thai in fact allows for two different positions of NP-adjacent numeral+classifier pairs, one of which appears to be an ‘outer’ position which can give rise to partitive readings. This is seen in examples (59) and (60) when possessor-phrases and relative clauses co-occur with numeral-classifier pairs in two alternating orders. In the first (a) order, the numeral and classifier are directly adjacent to the noun and there is no partitive interpretation, whereas in the second (b) order the numeral-classifier pair occurs further to the right, separated from the noun by the possessor phrase/relative clause, and partitive interpretations naturally arise:

(60) a. [bàan sáam-láng khoong phóm]  
   house 3-CL of I  
   ‘my three houses’

   b. [bàan khoong phóm sáam-láng]  
   house of I 3-CL  
   ‘three of my houses’

(61) a. [bàan sóong-láng[thì phóm sū nay Amerikaa]]  
   house 2-CL which I buy in America  
   ‘(the) two houses which I bought in America’

   b. [bàan [thì phóm sū nai Amerikaa] sóong-láng]  
   house which I buy in America 2-CL  
   ‘two (of the) houses which I bought in America’

If the (b) forms of show that there is a second ‘outer’ position for numeral+classifier pairs in nominal expressions and this position can give rise to partitive interpretations, it should clearly be possible for a clause-final floating quantifier to move from such an outer position, maintaining the partitive interpretation which is made available by its initial occurrence in the outer position. Considered further, therefore, the availability of partitive interpretations with floated numeral+classifier pairs ultimately does not provide a clear argument against the assumption that they are moved to their surface position from a position adjacent to the NP associate.

What does still require some explanation, however, is why partitive interpretations of linearly adjacent NP numeral+classifier sequences do not appear not to be possible, i.e. why (53) and (55) do not seem to easily allow a partitive reading. If there is indeed a second structural NP-adjacent position for numeral+classifier pairs which will allow for a partitive interpretation (given the patterns in (60b) and (61b)), why is it the case that the numeral+classifier pairs in (53) and (55) cannot be interpreted in such a position, giving rise to partitive meanings? A possible explanation for the lack of a (now) expected
partitive interpretation in (53) and (55) may be to attribute this to parsing preferences and
the cross-linguistic phenomenon of ‘Local Association’ (aka ‘Late Closure’ Frazier 1979).
Local Association is a preference in parsing to keep adjacent words analyzed as being
close together in the underlying syntactic structure constructed by hearers, and such a
preference principle may restrict the analyses and associated interpretations that hearers
find it easy to mentally construct. For example, both of the English sentences in (62) and
(63) below are ambiguous, but Local Association strongly leads hearers to make the
interpretation in (a), as this involves mentally analysing the adverb ‘yesterday’ and the
relative clause ‘who was on the balcony’ as modifying the nearest available unit:

(62) John said that Mary left yesterday.
   (a) strong parsing preference: ‘yesterday’ modifies ‘left’
   (b) less naturally available: ‘yesterday’ modifies ‘said’

(63) Someone shot the servant of the actress who was on the balcony.
   (a) strong parsing preference: ‘who was on the balcony’ modifies ‘the actress’
   (b) less naturally available: ‘who was on the balcony’ modifies ‘the servant’

Applied to the apparent lack of a natural partitive interpretation in examples such as (53)
and (55), it could be suggested that Local Association enforces the parsing attachment of
the numeral+classifier in the first mentally available position - the syntactically lower
internal position of numeral+classifier pairs - and that hearers are only able to make a
higher attachment (in the outer position) when there is material intervening between the
NP and the numeral classifier as in (60b) and (61b), which enforces high attachment of the
numeral+classifier in the outer position and the ensuing partitive interpretation.

4.4 Mid-way conclusions and partial summary
Although it has been seen that there are two potential analyses of the derivation of clause-
final floating quantifiers in Thai, it can be suggested that the balance of the evidence may
favor an analysis of movement, with repositioning of floating quantifiers from a position
adjacent to an associated NP to some clause/VP-final location. While an adverbial
analysis of floating quantifiers is theoretically possible, it is perhaps not strongly
supported, and is clearly challenged by the occurrence of island restrictions. Moving
forward with such assumptions, we can now partially summarize the differences which
seem to exist in floating quantifier patterns in Burmese, Thai and English, as established by
the answers to five questions probing significant parameters of variation in floating
quantifier phenomena.

QUESTION 1: Where does the NP associated with the floating quantifier occur?
In English, the NP occurs in subject position. In Thai, the NP associate occurs in regular
subject, object, indirect object, or object of preposition positions (hence the NP itself is not
moved to any special position). In Burmese, the NP commonly appears in a topic-like
position and often seems to have been displaced from a regular subject or object position.

QUESTION 2: Where does the floating quantifier/FQ occur?
In English, the FQ commonly occurs between auxiliary verbs and the main verb, before the
VP. In Burmese, FQs commonly occur in the regular position of the associated NP (e.g. in
regular subject or object position). In Thai, FQs occur in clause/VP-final positions which often do not correspond to the regular position of the associated NP, nor any position which the NP would have moved through.

**QUESTION 3:** *What is the definite/indefinite status of the “NP”?*

In English, the “NP” has to be definite (and is therefore actually a DP; e.g. ‘the students’, and cannot be a bare indefinite NP e.g. ‘students’). In Burmese, the NP is commonly a bare indefinite NP (e.g. *khetan* ‘pencil’). In Thai, the NP is often a bare indefinite NP (e.g. *phuuying* ‘woman’), but can also be more complex and definite (e.g. *pheuan khoong phom* ‘my friend(s)’).

**QUESTION 4:** *How does the floating quantifier structure seem to be syntactically created?*

In English, stranding appears to occur - the NP leaves behind the FQ when it moves to a higher position (the subject position). In Burmese, stranding also appears to occur - the NP leaves behind the FQ when it moves to a higher topic position. In Thai, it may appear that rightward movement takes place - the FQ appears to be moved away from the NP to a position to the right of the clause, similar to extraposition movement in English.

**QUESTION 5:** *What elements occur as floating quantifiers?*

In English, only quantifiers such as ‘all’ and ‘both’, and not numerals occur as FQs. In Burmese, numerals (combined with classifiers) and other quantifiers (not combined with classifiers, e.g. ‘aa-loun’ ‘all’) appear as FQs. In Thai, numerals (combined with classifiers) and other quantifiers combined with classifiers (e.g. *thuk* ‘every’, *laai* ‘several’, *baang* ‘some’) function as FQs.

Generally, then, floating quantifier patterns are not fully uniform across different languages and may vary according to a range of distinct properties. The hypotheses developed here concerning the syntactic mechanisms which result in separation of NP and FQ (i.e. question 4) are schematized below in (64). In section 5, we go on to consider what may functionally be responsible for this separation of NP and FQ.

(64) a. **ENGLISH**

```
   NP...........NP Q...........NP moves to subject position
         Q is stranded
```

b. **BURMESE**

```
   NP...........NP Q...........NP moves to topic-like position
         Q is stranded
```

c. **THAI**

```
   NP Q ...........Q...........Q moves to clause-final position
         NP is stranded
```
5. Why does separation of the NP and quantifier take place?
Having considered some of the structural properties and differences in floating quantifier constructions in Burmese and Thai (and English), we should now ask why this kind of splitting and separation of NP and quantifier ever occurs. The splitting of a constituent into two separate, distanced parts imposes considerable extra processing costs, as the reference value of a subject or object or other event participant can only be computed fully once both NP and quantifier are encountered in the processing of a sentence, and in instances where there is separation of an NP and an associated quantifier, hearers have to mentally ‘store’ the partial information provided by the NP until it can be combined with that of the associated quantifier and a full referential value for the subject/object etc can be arrived at. Given such extra processing costs, it is natural to ask what benefits may result from the use of floating quantifiers as compensation for the processing burden. In addition to such a functional question, one might also wonder whether there is any principled way to explain the syntactic differences observed above, or does it have to be concluded that the cross-linguistic variation found in floating quantifier constructions is simply random and fully unpredictable? In the remainder of the paper, it will be suggested that, in employing floating quantifier constructions, languages may frequently be trying to achieve the same basic effects in terms of information structure, but are constrained by differently configured local resources, and it is language-specific properties and restrictions which result in the range of differences noted above. FQ constructions will therefore be suggested to potentially contain elements of the universal interacting with elements of the language-specific, the universal here being the linear ordering of elements in information structure, and the language-specific being variation in basic word order type: V-O vs. O-V (i.e. head-initial vs. head-final). As a way to approach these issues, we will begin by considering how and when floating quantifier constructions are commonly made use of.

5.1 The functional use of FQs in Thai
In Thai, there are two particularly common contexts which give rise to the use of FQ constructions. The first of these is presentational sentences - floating quantifiers frequently occur when new referents are being introduced in a discourse, often with the existential verb mii or as the object of a verb, as illustrated in (65-66):

\begin{align*}
(65) & \text{mii } \text{dèk } \text{maa } \text{ngaanpaati } \text{raw } \text{sìisìp-kwàa-khon} \\
& \text{be } \text{child come } \text{party } \text{about } \text{40-above-CL} \\
& \text{‘More than forty children/young people came to the party.’}
\end{align*}

\begin{align*}
(66) & \text{phómphàng } \text{sùu } \text{nángsùu } \text{maa } \text{sòong-lém} \\
& \text{I just } \text{buy } \text{book } \text{ASP } \text{2-CL} \\
& \text{‘I just bought two books.’}
\end{align*}

The second context where FQs occur with significant regularity in Thai is in instances of ‘re-presentation’ and partitivity. Splitting and separation of NP and quantifier occurs in instances where the NP is not new information – the NP is definite in reference and already familiar to hearer and speaker – and splitting results in partitive interpretations with a frequent focus on what characterizes a certain number of the set represented by the NP, as seen in (67-68):

\begin{align*}
(67) & \text{mii } \text{dèk } \text{maa } \text{ngaanpaati } \text{raw } \text{sìisìp-kwàa-khon} \\
& \text{be } \text{child come } \text{party } \text{about } \text{40-above-CL} \\
& \text{‘More than forty children/young people came to the party.’}
\end{align*}

\begin{align*}
(68) & \text{phómphàng } \text{sùu } \text{nángsùu } \text{maa } \text{sòong-lém} \\
& \text{I just } \text{buy } \text{book } \text{ASP } \text{2-CL} \\
& \text{‘I just bought two books.’}
\end{align*}
Floating Quantifiers in Burmese and Thai

The main difference between the two common contexts for FQs cases is the referential familiarity of the NP - the NP is either old-familiar information being re-presented in a sentence, or is new information being presented for the first time. A generalization which unites the two contexts is that in both cases the quantifier itself represents important, new information.

GENERALIZATION ONE (Thai)
The quantifier in FQ constructions represents important, new information.

Because of (69), it is quite unnatural for demonstratives to occur floated in final position, as demonstratives regularly do not encode new information:

khāw sūu nāngsūu maa yōng-lēm/*?![lēm-nūi] lēw
he buy book ASP 2-CL/CL-this ASP

‘He has bought two books/this book.’

The unacceptability of ‘floating demonstratives’ here is similar to the unnaturalness of repositioning a demonstrative-marked NP to the right of a clause in English Heavy NP Shift constructions:

(71) I gave _ to Mary */[a book about elves]/*?[this book].

A second functional generalization which characterizes FQ patterns in Thai is that use of floating quantifiers is often felt to sound more natural when the quantifier is accompanied by some other qualifying/focus particle such as the following: khae ‘only’, tang ‘as many as’, thawnan ‘only’, keuap ‘almost’, raaw ‘approximately’. FQs are also judged to be natural-sounding when the numeral which occurs has a remarkable or high value:

mii khon maa tāng-hāa-sip-khon
be people come as-many-as-50-CL

‘As many as 50 people came.’

GENERALIZATION TWO (Thai)
FQ constructions often involve the occurrence of an additional focus or qualifying particle or remarkable/high-valued numerals.
These two observations in (69) and (73) support the view that floating quantifiers instantiate focused information. It can therefore be suggested that when there is natural pressure to stress the focal salience of new information represented by a quantifier, this may be achieved by positioning the quantifier away from its associated NP in clause-final position, where new information is most naturally positioned in a very wide range of languages.37

Continuing to examine such a functional view of floating quantifiers, in Thai in the frequent instances where there is presentation of a fully new quantified referent, it can be noted that there are actually two pieces of new, important information which occur in floating quantifier structures: (a) the noun/NP - i.e. the identity of the type of the referent: ‘student’, ‘book’ etc, and (b) the quantifier - the amount of the N: ‘two books’, ‘fifty students’ etc. In such instances it can be hypothesized that the splitting and separation of a constituent into two components (NP and FQ) may serve to highlight the two, separate parts of the constituent, and splitting may be a particularly useful solution, where syntactically available, to situations in which there are two pieces of adjacent information both of which are informationally new and in focus. If the two components are separated and certain linear space is created between them, this may potentially serve to increase the salience/prominence of both items. Elsewhere in studies of language there is much evidence that perceptual salience may be at a maximum at the beginnings and ends (the edges) of units. For example, it is known that the beginnings and ends of words and syllables are perceptually more salient than the internal parts of such units. In a similar way, it can be suggested that the splitting of an adjacent NP Q sequence (in which both NP and Q are new information) into a spatially separated NP.....Q order may serve to create a structure in which the perceptual salience of both parts is usefully heightened. In this regard, there may be similarities with other common splitting/separation constructions. As noted earlier, extraposition structures such as (74) are naturally used in presentational situations, where a new referent is introduced. Here the noun ‘man’ encodes information about the basic type of the new referent, and the extrapoosed relative clause adds further new information about this basic type. In such cases, splitting of the NP into two parts may functionally occur to enhance focal salience on both pieces of new information:

(74) A man entered the room who was wearing a black hat.

A key property of separation and splitting may therefore be to establish a certain distance between two new units of information, both of which are in need of emphasis. Whereas some languages have considerably flexibility in stress placement and the manipulation of stress for informational purposes, other languages (in particular tone

37 In many languages, the cross-linguistic tendency for new information to be introduced in sentence/clause-final location causes the occurrence of non-canonical word order patterns, sometimes with the re-ordering of major argument constituents. For example, the neutral SV(O) word order in languages such as Italian is regularly reordered as VS if the subject encodes new information, as in (i):

(i) É arrivato Gianni.

‘Gianni has arrived’ (a natural answer to the question: ‘Who has arrived?’)

The use of FQ constructions can therefore be seen as another manifestation of non-canonical word order to highlight new information in clause-final position.
languages, such as Thai and Burmese) have less flexibility, and may need to make use of special syntactic structures and movement/repositioning of elements to achieve similar functional ends. The occurrence of split, floating quantifier structures may consequently be the result of situations in which adjacent focal elements cannot both be naturally stressed and so constituents are split in two to allow for both parts to maximize their focal prominence.

5.2 The functional use of FQs in Burmese

When Burmese floating quantifier constructions are considered from a pragmatic, functional point of view, they are interestingly found to show similar focus properties to those observable in Thai, and floating quantifiers in Burmese occur very naturally with focus-type particles (e.g. -taun ‘as many as’). Indeed, various configurations involving floating quantifiers which speakers categorize as unacceptable/highly unnatural or even ungrammatical can be ‘rescued’ and made perfectly acceptable by the appropriate use of focus particles. This is an observation which has also been made about similar patterns in Japanese and Korean floating quantifier constructions (Kang 2002, Miyagawa and Arikawa 2007). In various earlier works on Japanese and Korean (e.g. Miyagawa 1989) the linear sequencing of a floating quantifier associated with a subject but following an object, as schematized in (75) and illustrated with Korean (76) was categorized as ungrammatical. However, if a focus particle is added to the floating quantifier, and/or a numeral quantifier is made into a large ‘remarkable’ number (hence inherently focused) as in (77), it has been noticed that the configuration in fact becomes perfectly acceptable (Kang 2002, Miyagawa and Arikawa 2007):

(75) NP\textsubscript{Subject-k} NPO\textsubscript{bject} Q\_k V

(76) *\textit{hakseyn-ji khempywuthe-lul twu-myeng} sassta
    student-NOM computer-ACC 2-CL bought
    ‘2 students bought a computer.’ (Kang 2002)

(77) \textit{hakseyn-ji khempywuthe-lul twu-myeng-ina/-man} sassta
    student-NOM computer-ACC 2-CL-as-many-as/only bought
    ‘As many as/only two students bought computers.’ (Kang 2002)

Alternatively, if it is ensured that the object NP in sequences such as (75) is not interpreted as new information (which might potentially distract attention away from the intended focus on the new information of the floated quantifier), such a strategy will also ‘save’ structures with the form in (75). Again, this confirms the required focal properties of floating quantifiers. While (78) is regularly judged as deviant in Japanese, if the object sake\-o ‘wine’ is pronominalized as sore\-o ‘that’ and so encodes old/given information as in (79), the sentence is accepted as well-formed and natural (Nakanishi 2008):

(78) ?*gakusei-ga sake-o san-nin non-da
    student-NOM wine-ACC 3-CL drank
    ‘3 students drank wine.’
Similar patterns occur in Burmese and reinforce the assumption that floating quantifiers are focused information in Burmese, as in Thai, and occur in final pre-verbal position in order to heighten their focal prominence, the pre-verbal position in Burmese being the position that other focused elements naturally occur in, as in many SOV type languages (e.g. Turkish, Hindi, Bangla).

Although the focused interpretation of floating quantifiers can thus be characterized as similar in Burmese and Thai (and Japanese and Korean), it can be noted that there is also a difference in the interpretation of the associate NP which frequently occurs in Burmese but not in Thai. Speakers of Burmese often note that there seems to be a natural sense of contrast implied in many cases of splitting and separation of the NP and a quantifier, and the implication of ‘lists’ in which items are compared and contrasted against each other. This is illustrated in (80). When presented with such sentences, speakers report that there is a natural implication that the subject also bought (different) quantities of other items too.

(80) dāqhkē-h ko candaw lēe-lūun we-de.
    battery-ACC I 4-CL buy-REAL
    ‘I bought batteries.’
    implication: I bought different quantities of other items too.

The interpretation of the NP in sentences such as (80) is therefore that of a contrastive topic. Contrastive topics are frequently both old and new in informational terms: their identity is generally known/familiar, but there is new information present in the fact that they are contrasted with other members of a particular set, as illustrated in (81) (Lee 1999):

(81) [[These]FOCUS examples]TOP I found [in Gundel].
    ‘these’=focal/contrastive
    ‘these examples’=old/known information

Burmese separation of NPs and associated quantifiers therefore involves both new information focus on the pre-verbal floating quantifier and frequent contrastive topic-like interpretation of the NP. In this patterning, floating quantifier separation constructions are similar to splitting constructions in languages such as German (also Polish, Russian). In German (Fanselow and Cavar 2001), the two parts of a single NP unit can be split apart as in (82),

(82) Autos besitzt er (nur)schnelle.
    cars owns he only fast
    ‘He owns only fast cars.’
    ‘As far as cars are concerned, he only has fast ones. As for motorcycles,...’
The same kind of contrastive interpretations that are often felt in Burmese floating quantifier constructions are common in such splitting, and may be a frequent property of many splitting constructions. However, they do not seem to be a common interpretation in Thai floating quantifier constructions, and this accordingly is an instance of some difference in the patterning of floating quantifiers in the two languages (to be returned to below).

Reflecting on the commonalities found with floating quantifiers in Thai and Burmese, and the functional question of why splitting of NPs and quantifiers occurs in the two languages, a general conclusion which it seems plausible to adopt is that splitting takes place in order to focus the quantifier and its new information in a prominent, final focus position as stated in (83).

(83) Functional generalization on FQ constructions in Thai and Burmese:
Splitting and distancing of NP and quantifiers coincides with and is appropriate for the encoding of (new information) focus on the quantifier.

We have now attempted to provide at least partial answers to the three questions in (84) we began this paper with, with special reference to Burmese and Thai, and comparisons made with patterns already reported in English, Japanese and Korean:

(84) i. How are the NP and quantifier related to each other in floating quantifier constructions?
ii. Why does separation of the NP and quantifier occur?
iii. Are floating quantifier constructions cross-linguistically uniform?

The answer to question (84iii) has been that there are in fact a number of differences in floating quantifier patterns across languages, even when one considers just the two languages Burmese and Thai. For example, it appears that in FQ constructions in Thai the quantifier is repositioned to the right, stranding its associated NP in situ, while in Burmese the NP is moved to the left, stranding the associated quantifier. An interesting question which we can now ask is whether such differences might in any way be predicted or accounted for by the answer to the other questions (i) and (ii)? The answer here may be ‘yes, quite possibly so’. Specifically, it will be suggested that differences such as those observed with Thai and Burmese may result from the interaction of ‘universal’ and language-specific properties. The ‘universal’ property relevant here is the observation drawn from general studies of information structure that there is a pervasive cross-linguistic tendency for elements representing new information to occur focused in clause-final positions (hence new referents are commonly introduced in object rather than subject positions). The language specific property we will consider here is the difference in basic, neutral word order in Burmese and Thai.

First, looking at patterns in Thai, let us consider the case of a subject which is represented by an NP and a quantifier whose content the speaker wishes to focus. Due to the basic SVO word order in Thai, if the quantifier remains adjacent to the NP, this will result in the linear sequencing in (85):
In this neutral word order, the quantifier which is to be focused is located far away from the clause-final position which new information focus most naturally occurs in. In order for the quantifier to occur in such a position, it is regularly moved/relocated to its right, stranding the NP, as in (86):

\[
(86) \quad [NP \, Q_{\text{focused}}]_{\text{Subject}} \, V \, NP_{\text{Object}} \, Q
\]

The stranded subject NP is not in a position which has a special focal status in the information structure of the sentence and simply receives the regular interpretation of an NP in subject position, hence there is no necessary/common interpretation of the NP as being a contrastive topic (unlike the NP frequently in Burmese FQ constructions). Displacement of the quantifier to the right occurs simply in order to position the quantifier in the clause-final, new information focus position.

Now turning to Burmese, let us consider the case of an object NP with a quantifier whose content a speaker wishes to focus. Given the SOV basic word order in Burmese, in cases where there is no NP-quantifier separation, and the NP and quantifier remain adjacent to each other, this will result in the linear sequencing depicted in (87):

\[
(87) \quad NP_{\text{Subject}} \, [NP \, Q_{\text{focused}}]_{\text{Object}} \, V
\]

In Burmese, as in many other SOV languages, the ‘clause-final’ position associated with new information focus is actually not fully clause-final, but the position immediately preceding the verb. Hence in the set of constituents which can be re-ordered (this not including the verb), a focused argument or adverbial is commonly placed in final position and preceded by other old and backgrounded information. In the neutral word order configuration in (87), the quantifier associated with the NP already naturally occurs in the pre-verbal focus position, so there is a natural convergence of position and information structure status for the quantifier. Where quantifier float patterns do occur, and result in a further heightened focus effect on the quantifier due to the splitting and separation effect, this results in a splitting away of the NP stranding the quantifier in focus position and placement of the NP further forward in the sentence, in topic position. As the NP which is regularly displaced leftwards to topic position here is a commonly a bare noun, such a bare, indefinite noun/NP is interpreted generically as representing the type of the noun/NP, and this in turn results in its interpretation as a contrastive topic, as non-contrastive topic interpretations are restricted to entities that are definite in reference (hence the oddness of sentences such as: ??A book, I bought yesterday.’).

Consequently it can be suggested that the different word orders of Thai and Burmese interacting with cross-linguistic pressures to place focused, new information in final positions conspire to cause the major observable differences between floating quantifier constructions in the languages, namely: (1) in Thai, the quantifier undergoes movement, while in Burmese it is the NP which is regularly moved away stranding the quantifier, and (2) in Burmese: there is frequently a contrastive topic interpretation of the NP, while in Thai: no special interpretation of the NP occurs in floating quantifier constructions.
A further common property of Q-float constructions which appears to be shared widely across languages is the patterning that when NP and quantifier are separated, it is common for the NP to linearly precede the quantifier (...NP.....Q....), and the opposite sequencing of quantifier preceding NP is quite uncommon (...Q.....NP....), though sometimes suggested to be grammatically possible in languages such as Japanese. 38 This common linear distribution of NP and quantifier can arguably also be attributed to aspects of information structure, and the strong cross-linguistic tendency for new information to be sequenced following old information. In instances where an NP and a quantifier are split apart in Q-float constructions/configurations, it is much more likely that the generic reference value of the noun/NP can be assumed to be familiar, contextually retrievable and more available as a topic-like center of interest than the value of the quantifier, which will frequently resist any topic-like licensing in the information structure of a sentence, as illustrated in the contrast in (88):

(88) a. As for apples, I want three.
   b. ??As for three, I want apples.

General principles of information structure interacting with parametrizable properties of languages therefore results in a range of quite predictable and understandable variation in the distribution of floating quantifiers and their associated NPs.

A final question we will consider here relating to the issue of word order, floating quantifiers and information status is the special occurrence of post-verbal elements in SOV languages and the syntactic status of ‘afterthought information’. Above it was mentioned that the basic word order of languages such as Burmese, Japanese and Korean is SOV, with the verb in clause-final position. In written forms of these languages, subjects, objects, obliques and other adjuncts may occur in a range of different orders preceding the verb, depending on their contextual information status as definite/indefinite, new/old, focused/topical material, but the verb is regularly final in its clause and does not participate

38 A striking illustration of the pressures to conform to a linear NP > Q ordering can be noted from Mandarin Chinese, where the quantifier *dou ‘all’ has grammaticalized in a fixed pre-verbal position, and is actually never combined with an associated NP in a single syntactic constituent. Whenever *dou quantifies over an object NP which would normally follow the verb in the basic SVO word order of Mandarin, the object NP is actually forced to undergo repositioning into some position to the left of *dou resulting in a linear NP>Q sequence. This may result in the NP object being immediately adjacent to *dou as in (c) below, or further to the left in pre-subject topic position (d). Examples (a) and (b) show that *dou cannot be combined with the NP object in post-verbal position, and that the object may not remain in situ following the verb if associated with *dou:

(a) *wǒ kàn-le dōu shū
   I read-ASP all book
(b) *wǒ dōu kàn-le shū *..Q…NP..
   I all read-ASP book
(c) wǒ shū dōu kàn-le ✓..NP…Q...
   I book all read-ASP
(d) shū wǒ dōu kàn-le ✓..NP…Q...
   book I all read-ASP
   ‘I read all the books.’
in any linear re-ordering with arguments and adjuncts (hence Burmese, Japanese and Korean are often referred to as ‘verb-final’ languages). In spoken forms of these languages, however, certain non-verbal elements such as subject and object NPs are found to optionally occur following the verb. This being the case, an interesting question is whether it is possible for floating quantifiers to be positioned following the verb resulting in a separation and splitting more similar to that in Thai, with floating quantifiers often occurring in sentence-final positions which do not correspond to regular locations of the associated NPs.

In Burmese and Japanese, it is in fact possible for floating quantifiers to occur following the verb in a clause, as illustrated in (89) and (90):

(89) manēcēka thuu zēe-hmaa sa-ōuq wē-te, thōun-ōuq
yesterday he market-in book buy-REAL 3-CL
‘Yesterday he bought books in the market, ..three to be precise’

(90) Taroo-wa Kinokuniya-de hon-o katta, san-satsu
Taroo-TOP Kinokuniya-in book-ACC bought 3-CL
‘Taroo bought books in Kinokuniya, ..three it was.’

However, such post-verbal elements in Burmese and Japanese are commonly interpreted in a particular way as ‘afterthoughts’ - information which a speaker adds to a sentence often in the way of further clarification – and it is often assumed that, as afterthoughts, post-verbal elements in verb-final languages such as Japanese are not syntactically integrated into the preceding clause (Kuno 1978, Sells 1999, Soshi and Hagiwara 2004). The existence of such postposed ‘afterthought’ quantifiers in Burmese and Japanese raises a question about Thai where floating quantifiers regularly occur in sentence-final positions. Specifically, we may ask whether the patterns found in Thai are really different from those in (88) and (89) and whether clause/sentence-final floating quantifiers in Thai might be ‘afterthought’ additions to the sentence like post-verbal quantifiers in Burmese and Japanese? The answer to the latter question is ‘no’. Thai floating quantifiers are indeed clearly integrated into the syntactic structure of the sentence and are not just added on as ‘afterthoughts’ in sentence-final position. First, Thai floating quantifiers can in fact (optionally) precede sentence-final particles such as laew (see ex.39), showing that they occur within the main syntactic structure of the clause. Second, the prosodic break/intonational pause between verb and floating quantifier which characterizes the occurrence of post-verbal quantifiers in Burmese and Japanese, setting the quantifier off from the rest of preceding clause is not present with clause-final floating quantifiers in Thai, and these are intonationally integrated into the sentence without any separating pause. Third, although Burmese and Japanese post-verbal quantifiers may represent new information in some sense, they are not interpretable as obviously focal new information, unlike the situation in Thai. Because of this, there are clearly different restrictions on what kinds of quantifiers can occur post-verbally in Burmese/Japanese and what may occur
clause-finally in Thai. Significantly, in Burmese and Japanese, focused and interrogative quantifiers are unacceptable in post-verbal position:39

(91) *thuu zēe-hmaa saōuq wε-th-lē, behna-ōuq?
   he market-in book buy-REAL-Q how-many-CL
   Intended interpretation: ‘How many books did he buy in the market?’

(92) a. *kare-wa hon-o katta no, nan-satsu?
   he-TOP book-ACC bought Q how-many-CL
   Intended: ‘How many books did he buy?’
   b. *kare-wa hon-o kawanakatta san-satsu-shika
      he-TOP book-ACC bought-NEG 3-CL-only
      Intended: ‘He only bought three books.’

This contrasts with patterns in Thai, where interrogative and focused quantifiers in final position are natural and frequent in their occurrence:

(93) jon-bāad-nīi mii nāksʉ̏ksáa maa lɛ̄εw/thʉ́ng kīi-khon?
   up-until-now be student come ASP/arrive how-many-CL
   ‘How many students have arrived so far?’

(94) kháw hày ngon Daeng khɛ̀ε sóong-bȁat (thàwnān)
   he give money Daeng only 2-JBaht (only)
   ‘He gave Daeng only two Baht.’

Thai is therefore clearly distinctive in this patterning, not only from SOV Burmese and Japanese, but also from SVO Chinese which permits ‘afterthought’-type VP-final floating quantifiers, as illustrated in (95), but not focused or interrogative floating quantifiers, as seen in (96):

(95) yŏu rén lái zhăo nĭ, sān-ge
   be person come find you 3-CL
   ‘There were some people who came looking for you, three people.’

(96) *Zhāngsan xĭang-zhīdào yŏu rén lái zhăo tā ji-ge
   Zhangsan want-to-know be person come find him how-many-CL
   Intended: ‘Zhangsan wanted to know how many people came looking for him.’

The conclusion from such contrasts is that Thai clause-final floating quantifiers are not simple afterthought elements and are integrated into the syntactic structure of the sentence in a way quite different to afterthoughts, which may simply be linear concatenations not syntactically connected to what precedes them as parts of a single sentence (Sells 1999, Soshi and Hagiwara 2004).

39 Note that if the focused and interrogative quantifier-classifier pairs in (91) and (92) are positioned before the verb, these sentences are grammatical. There is just a special restriction on these elements when they occur in the post-verbal ‘afterthought’ position.
Having thus considered how aspects of the functional use of floating quantifier structures may bear on their distribution within the clause, and how certain interpretations are associated with quantifiers floated in various positions, we now close the paper with a brief summary of this exploration of Q-float phenomena in Thai and Burmese.

6. Summary
This paper set out to document and analyze patterns involving displaced, ‘floating’ quantifiers (and classifiers) in two neighboring languages of southeast Asia, Thai and Burmese, as a way to provide further potential insight into three general questions associated with floating quantifier constructions. First, what syntactic mechanisms result in the separation and linear distancing of a noun/NP and its associated quantifier? Second, what pragmatic/functional motivations might underlie the optional use of Q-float forms, licensing their occurrence? Third, from a comparative, typological viewpoint, do the mechanics and triggers of Q-float show signs of patterning in a uniform way across languages, or are there differences, and how might these be accounted for? Thai and Burmese were selected as the two principal languages of this micro-study from southeast Asia as both languages clearly exhibit the phenomenon of Q-float in appropriate contexts and allow for the separation of quantifier-classifier pairs from similar adjacent linear sequences of [NP quantifier classifier], hence seem to display parallel base resources in this regard. With regard to word order patterns at the clausal level, however, Thai and Burmese show significant differences, Thai being S-Aux-V-O and Burmese S-O-V-Aux in neutral sentences. One intended focus of the study was to look at how this difference in basic word order might potentially influence the way Q-float is manifested in a language. As the investigation of Thai and Burmese proceeded, it was concluded that both languages separate nouns/NPs from modifying quantifiers by mechanisms of movement (hence subject to syntactic restrictions commonly associated with movement, such as island constraints), and that Q-float in both languages regularly appears to be linked to the expression of focused new information. However, Thai and Burmese were seen to differ with regard to the element of the NP/quantifier pair that undergoes movement/displacement in Q-float constructions, in Thai the quantifier shifting rightwards to a clause-final position in a way similar to extraposition or Heavy NP Shift in English, while in Burmese it is the NP which undergoes a leftwards displacement in sentences exhibiting Q-float. This major difference between Thai and Burmese was attributed to a tension between principals determining linear word order - in this instance, information structure - and those regulating hierarchical syntactic structure – here, the syntactic organisation of clauses in a head-initial or head-final way, resulting in SVO and SOV type languages. Both languages attempt to achieve an optimal ordering of separated NP and quantifier for the purposes of information structure and focus, and bring this about through the displacement of different elements in opposite directions in the clause, (new) focused constituents cross-linguistically favoring a rightward, clause-final position. The comparison of Q-float patterns in Thai and Burmese therefore illustrates how linear strings with an important shared property (..NP…Q…) may be produced in distinct ways by languages with different syntactic properties, underlining the fundamental importance of linear sequencing for (certain) aspects of interpretation despite the dominance of hierarchical structure in other areas of syntax and construal. The paper also provides the first substantial description of floating quantifier patterns inThai and Burmese and so extends the available empirical coverage of this area of syntax in Asian languages, adding...
it as a resource to previous insightful studies of Japanese and Korean and the issues discussed in these works (Kang 2002, Miyagawa 1989, Miyagawa and Arikawa 2007, Nakanishi 2008, Ko 2007 among others).

References


THE EVOLUTION OF POLYFUNCTIONALITY OF DÂY CONSTRUCTION IN THAI: SPLIT PATTERNS OF POSSIBILITY-RELATED MODAL CONCEPTS*

Kiyoko Takahashi
Kanda University of International Studies
<kiyoko@kanda.kuis.ac.jp>

0 Abstract
This study aims to investigate historical changes in Thai modal concepts marked by the irrealis marker câp (≡ câk), especially possibility-related concepts denoted by periphrastic constructions containing the morpheme dây. I have examined irrealis expressions in Thai inscription corpora from the end of the 13th century through the 20th century, and found that Thai modal concepts in the present have gradually emerged. In semantic extensions of Thai modals, an original modal concept does not disappear even after a newly derived modal concept has established itself. Rather, it is as if one single line split into two lines both of which would continue extending. In particular, I have identified two directions of semantic changes involved in split patterns of possibility-related modal concepts in Thai: (1) less subjective > more subjective; (2) non-volitive (participant-external) > volitive (participant-internal). The latter direction is opposed to the direction that has been postulated in the literature of historical semantic change. I have also found split patterns with little change in subjectivity, to which most of previous studies have not paid due attention.

Keywords: modals, historical semantics, epigraphy

1 Introduction
This paper examines historical changes in the types of modal concepts marked by the irrealis marker câp (≡ câk) in Thai. Specifically, this paper analyses split patterns of the modal concepts by using discourse corpora of Thai inscriptions from the end of the 13th century through the 20th century (see the list of the inscription corpora at the end of this

* Earlier versions of this paper were presented at LSJ134 (the 134th General Meeting of the Linguistic Society of Japan), Chiba, June 16-17, 2007 and at ICHL18 (the 18th International Conference on Historical Linguistics), Montréal, August 6-11, 2007. I would like to thank Rumiko Shinzato, Atsuhiko Kato, Heiko Narrog, and Bruce Horton for their helpful comments and suggestions on the earlier versions of this paper. I am also grateful to an anonymous reviewer of JSEALS (the Journal of the Southeast Asian Linguistics Society) for constructive and helpful comments on a draft of this paper. Errors and shortcomings are all my own.

40 This kind of historical semantic change is often called ‘layering’ (Hopper 1991: 22) in the study of grammaticalization, i.e., the process whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalized, continue to develop new grammatical functions (Hopper & Traugott 1993: xv).
I use the term ‘split patterns’ to mean patterns of semantic extension in which a more specific modal concept emerges out of a less specific modal concept, and the original concept does not readily disappear and typically remains long after the new concept becomes established, as if one single line split into two lines both of which keep extending (shown in Diagram 1 in Section 5).

In this paper, special consideration is given to split patterns of the possibility-related types of modal notions such as ‘ability (participant-internal, volitive possibility)’ and ‘circumstantial possibility (participant-external, non-volitive possibility)’. Those possibility-related concepts are expressed by specific constructions ending up with the morpheme 

I have found that in Thai, ‘ability (participant-internal, volitive possibility)’ evolved out of ‘circumstantial possibility (participant-external, non-volitive possibility)’, and in turn ‘circumstantial possibility’ derived from ‘circumstantial impossibility’. The direction of the former split pattern (circumstantial possibility > ability) is opposed to the direction found in most of the existing studies on the evolution of modal concepts, namely ‘ability > circumstantial possibility (root possibility)’ or ‘participant-internal modality > participant-external modality’ (Bybee 1988, Bybee & Pagliuca 1985, Bybee et al. 1994, Heine & Kuteva 2002, Traugott 2006, Traugott & Dasher 2002, van der Auwera & Plungian 1998, inter alia).43 Furthermore, the latter split pattern (circumstantial impossibility >

41 It is generally believed that the oldest Thai inscription (King Ram Khamhaeng Inscription) was made in 1292. There is an opposing view, however, that it was made in 1354-1376 (Prasithrathsint 2006: 129). Since I do not have knowledge of the inscription-dating, in this study I simply follow the former general view that the oldest one was made in 1292 (the end of the 13th century).

Note that the number of Thai inscriptions is limited to a few hundreds and each inscription is not long. Moreover, most of their contents is about the Buddhist religion or royalty-related matters such as politics and laws. Since the inscription data thus have the limitations in terms of quantity and genre-variety, it might be possible that some missing types of modal expressions with cák / cãp (and dây) in earlier periods were actually in use but unfortunately are not attested due to the limitations of the data. This study, however, uses the inscription data because the data nonetheless have a very good point. That is, the production years of most of the inscriptions have been estimated ranging from the end of the 13th century throughout the 20th century, which enables us to at least roughly see which types of Thai expressions were getting more and more (or less and less) common in a relatively long time span.

42 Matisoff (1991: 419-420) and Bisang (1996: 569-570) among others postulate that the original meaning of the Thai morpheme dây is ‘get’. Enfiled (2001: 279-280, 2004: 276) posits that the original meaning of the Lao morpheme dâj and the corresponding morpheme in other Tai languages is ‘come to have, acquire’.

43 However, van der Auwera et al. (2009) examine semantic extension of so-called ‘acquisitive’ modals (e.g. ‘I get to watch TV tonight’) and acknowledge its plausible bidirectionality between ‘participant-external’ and ‘participant-internal’ possibility. It also deserves mentioning that Li (2004) and Shinzato (2008) report the details of historical semantic change from ‘participant-external’ to ‘participant-internal’ possibility observed in Chinese and Japanese, respectively. Li argues for the following semantic change of a Chinese modal for possibility déde: (a) ‘participant-external non-deontic’ > (b) ‘deontic/epistemic’ > (c) ‘participant-internal possibility’. In a similar vein, Shinzato discusses the following stages of semantic change of two
The Evolution of dāy Construction in Thai

circumstantial possibility) has scarcely been discussed in the relevant literature. I believe that the findings of this study shed some light on the theory of the directionality of the evolution of modal concepts.

This paper is organized as follows. Sections 2 and 3, respectively, explain the nature of the irrealis marker in Thai and the semantic map of modal concepts to be used for visualizing a seven-century course of gradual change in Thai modal concepts. Section 4 presents quantitative data regarding historical changes in Thai modal concepts. Section 5 accounts for split patterns of possibility-related modal concepts in Thai, some of which have hitherto been rarely documented. In Section 6, I will give concluding remarks.

2 Thai irrealis marker

Thai, an isolating language, has a morpheme indicating the irrealis status of the situation represented by the following verb phrase, namely cā?). Thai lacks the obligatory coding of grammatical categories, and therefore cā?) is not an obligatory marker of irrealis. However, in present-day Thai cā?) is necessarily used to express counterfactuality (Srioutai 2004) and to form an irrealis complement with the complementizer or relativizer thîi (Diller (2001) calls cā?) in this usage ‘irrealis-complement formative’). Accordingly, I regard cā?) in modern Thai as the irrealis marker proper.

The irrealis marker cā?) derives from the verb cāk meaning ‘to intend, consider’ (Diller 1988). The word form cāk began to change into cā?) in the middle of the 14th century, and cā?) came to be commonly used in succeeding ages (Diller 1988, Takahashi 2007). Diller (2001) further hypothesizes that the two morphemes, cāk and cā?), have undergone the process of grammaticalization proceeding along the path indicated in (1), which is based on the attested grammaticalization path of the English auxiliary ‘will’ (Bybee et al. 1991: 26-29) as shown in (2).

(1) Grammaticalization path of cāk / cā?) in Thai (Diller 2001):
    desire > intention > future > irrealis

(2) Grammaticalization path of ‘will’ in English (Bybee et al. 1991):
    desire > intention > future > probability, imperative

However, Diller (2001) did not present Thai diachronic data to verify this hypothesis. This study, on the other hand, uses a data-driven approach. Previously I have gathered modal expressions marked by the irrealis marker from the aforementioned inscription corpora (Takahashi 2007). With this diachronic corpus data, in this paper I will examine historical changes in the types of the modal concepts. I consider a new modal concept encoded by a certain conventionalized construction to be fully established when it becomes amenable to marking by irrealis. From the corpus data, I see that, over time, the number of specific types of modal concepts marked by the irrealis marker has gradually increased. Put

Japanese modals for possibility naru and dekiru: (a) ‘disabling conditions exist external to non-specific individuals’ > (b) ‘disabling conditions exist external to specific individuals temporarily’ > (c) ‘disabling conditions exist internal to specific individuals temporarily’ > (d) ‘disabling conditions exist internal to specific individuals permanently’ > (e) ‘enabling conditions exist internal to specific individuals permanently’.

44 The semantic shift from negative to positive possibility of Japanese and Thai modals for possibility is reported in Takahashi & Shinzato (2005), Shinzato (2008) and Takahashi (2008b).
differently, specific modal concepts with clear semantic boundaries in the present have gradually emerged from inclusive and ambiguous modal concepts with fuzzy boundaries in the past. I have also found that the two morphemes, câk and câʔ, have been compatible with quite a wide range of modal situations since as early as the 14th century (shown in Figure 3 in Section 4). This fact makes it clear that they had become an irrealis marker proper before that period. Therefore, if Diller’s (2001) assumption about the grammaticalization path of the two morphemes (i.e. desire > intention > future > irrealis) holds, this process of grammaticalization must have been accomplished prior to the 14th century. In this paper, however, I am not able to validate his assumption because of the lack of relevant diachronic data in the corpora. What I can do instead is to reconstruct the most reasonable split patterns of Thai modal concepts marked by the full-fledged irrealis marker from the 14th century onward (see Section 5).

3 Semantic map for modal concepts
I follow Narrog’s (2005: 683-690) idea that modal concepts can be classified in terms of the dimensions of (a) ‘speaker-orientation’ or ‘subjectivity’ and (b) ‘volitivity’. Figure 1 illustrates a semantic map for modal concepts, where six representative modal concepts hold respective positions in terms of the two dimensions (cf. Figure 3 in Narrog 2005: 694).

<table>
<thead>
<tr>
<th>(Hearer-oriented, Intersubjective)</th>
<th>Imperative</th>
<th>Conditional concessive</th>
</tr>
</thead>
<tbody>
<tr>
<td>↑ Speaker-oriented, Subjective</td>
<td>Obligation</td>
<td>Prediction</td>
</tr>
<tr>
<td>↓ Event-oriented, Objective</td>
<td>Volition</td>
<td>Apparent imminence</td>
</tr>
</tbody>
</table>

Figure 1. An illustrated semantic map for modal concepts

‘Speaker-oriented’ or ‘subjective’ situations are directly linked to the speaker’s own modal judgment at the time of speech in the given speech situation. In contrast, ‘event-oriented’ or ‘objective’ situations are concerned with conditions on a participant of the described event, independent of the speaker and the present speech act. However, the distinction between subjective and objective situations is a gradient. For instance, ‘volition’ and ‘apparent imminence’ are less subjective than ‘obligation’ and ‘prediction’. On the other hand, the distinction between ‘volitive’ and ‘non-volitive’ situations is a discrete one since they constitute a dichotomy as to whether or not the ‘element of will’ (Jespersen 1924: 320-321) is involved, or to put it another way, whether or not the person concerned has an ‘interest’ (Heine 1995: 29) or ‘preference’ (Givón 1990: 528-530) in an event occurring or not occurring. In this regard, it is evident that deontic (e.g. obligation) and boulomaic (e.g. volition) modal situations are volitive, while epistemic (e.g. prediction) and evidential (e.g. apparent imminence) modal situations are non-volitive.

It is known that languages with the irrealis-marking system may differ in the range of modal contexts in which the irrealis marker appears, that is, the types of modal concepts may cross-linguistically differ (Mithun 1995, Elliott 2000). Figure 2 below indicates Thai modal concepts that are attested to be compatible with the irrealis marker câk / câʔ.
From Figure 2 we can see how close or distant each modal concept is to the others. Take the concepts of ‘ability’ (2.4) and ‘circumstantial possibility’ (1.6) for example. They show the same degree of subjectivity, but they are contrastive in their volitivity values; that is, ‘ability’ is volitive, while ‘circumstantial possibility’ is non-volitive.

It should be noted that although I accept the subjectivity and volitivity dimensions posited by Narrog (2005), the position of ‘ability’ in my semantic map of modals (Figure 2) is not the same as that in his semantic map of modals (cf. Figures 4 and 7 in Narrog 2005: 695, 702). He considers ‘ability’, like ‘apparent imminence’, as non-volitive and less subjective than ‘circumstantial possibility’, whereas I consider it to be volitive and no less subjective than ‘circumstantial possibility’. In my opinion, the concept of a human being’s physical or mental ‘ability’ entails his volition or desire to become able to do something, which is evidently distinct from the purely non-volitive concept of ‘circumstantial possibility’. Moreover, I assume that the concept of ‘possibility’ in general, be it volitive or non-volitive, always involves the speaker’s evaluation of the possibility. That is to say, when mentioning some possibility (or impossibility), the speaker should have evaluated in what sense it is possible (or impossible). This is the reason why I regard ‘ability’, ‘circumstantial impossibility’ and ‘circumstantial possibility’ to be more subjective than ‘volition’, ‘desire’, ‘conditionals’ and ‘apparent imminence’ which are more objective and closer to mere propositional concepts.

Thai expressions for all the modal concepts listed in Figure 2 above, which are from the inscription corpus data, are given in examples (3) to (17) below.\(^{46}\)

---

\(^{45}\) Narrog (personal communication) comments that although especially learnt or acquired ability might be subject to the person’s volition, there is good cross-linguistic evidence that ability also is non-volitive and many languages express it through a non-volitive ‘out-of-control’ (‘spontaneous’) marker.

\(^{46}\) Sample expressions with câk / câ? cited in this paper are transcribed into phonetic equivalents in present-day Thai. Abbreviations for functional morphemes in the English glosses are: IRR(ealis), NEG(ative), POSSI(bility), PART(icle), TOP(ic marker), REL(ativizer), CONJ(unction), CAUS(ative), PERF(ective), COP(ula), and NOM(inalizer).
1. Non-volitive
1.1 Conditional concessive ‘Even if ...’

(3) suuu  cάk  nάp  kσο  ле  mί?  thуān

Even if you would count them, you could not finish counting them. [5] (1361) 47

1.2 Prediction ‘... will ...’

(4) cά?  pay  sùu  ?abaayathǔk  sί  plàaw

They will go to hell in vain. [91] (1734)

1.3 Inevitability ‘... will ...; It is inevitable to ...’

(5) sάŋкèet  wеeаa  prάʔsόŋ  cά?  лoŋ  sūat  mоn  nау  bооt

They notice the time when monks will come to recite a sutra in temple. [190] (1782-1925)

1.4 Negative conviction (Conclusion) ‘It is never possible to ...’

(6) chάу  cά?  nίyом  yинdιи  lұtém  sάy  lάttθι?  sάatsάnά

It is impossible that we would willingly believe in other religions than the Buddhism. [187] (1782-1925)

1.4 Ironical interrogative (Conclusion) ‘Wherever can one find ...?’

(7) cά?  hάа  mίt  mуńаn  cάw  thи  nάу  dάy

Wherever can I find a good friend like him? [278] (1925-1978)

1.5 Circumstantial impossibility ‘... is not achieved; it is not possible to ...’

(8) cά?  phanntnаа  bοо  mί?  dάy  lгęy

It is impossible to describe it. [86] (1528)

1.6 Circumstantial possibility ‘... is achieved; it is possible to ...’

(9) mάу  mίi  pratuу  thи  cά?  ?όk  pay  phаay  нόk  dάy

There is no door through which we can go out. [146] (1782-1925)

---

47 The number in the brackets attached to the end of each English translation (e.g. [5]) is the registration number for the identification of inscriptions that have been discovered in and around Thailand, and the number in the parentheses (e.g. (1361)) is the estimated production year of each source inscription.
1.7 Conditionals ‘If ...’

(10) phìɁ cà cà cà càkk kk náp dûay duan dài yìp mṳuun sìi phan hòk sip 
duan 
    month 
    If we count by the month, it will be 24,060 months. [3] (1357)

1.8 Apparent imminence ‘... is about to ...’

(11) mɯaŋ sùkhǒothay níi mii dài cà cà cà càkk kk tɛ̀ɛk 
    city Sukhothai TOP noisy as if 
    The city of Sukhothai is so noisy as to be nearly broken. [1] (1292)

2. Volitive

2.1 Appropriateness ‘... should ...; It is appropriate to ...’

(12) khuan cà cà cà càɁɁ ɁɁ pen thîi chɯ̂ɯn chom yindii 
    should IRR COP NOM love be glad 
    It should be loved. [178] (1782-1925)

2.2 Permission ‘It is possible for one to ...’

(13) câ? phùuk nay bâan kɔ̂ɔ dài 
    IRR perform a ceremony in home CONJ POSSI 
    One can perform a ceremony (of establishing the Sangha communion area) in the home. [193] (1925-1978)

2.3 Obligation ‘One must ...’

(14) phûu thîi tham chûa câ cà cà càɁɁ ɁɁ tɔ̂ŋ ráp thúk 
    person REL do bad IRR must receive sorrow 
    Those who commit a sin must suffer. [256] (1925-1978)

2.4 Ability ‘One is able to ...’

(15) Ɂàat sǎamâat câ? câ? thathãan dûay khaathãa bôt day bôt núŋ 
    be able IRR pray with Pali verse a certain paragraph 
    One is able to pray by saying a certain full paragraph of a Pali verse. [257] (1925-1978)

2.5 Volutive ‘One will ...’

(16) phûta câ? háy Ɂaanaaprachaarãatsaɗɔɔn thàŋ puâŋ 
    in order to IRR CAUS the people in general 
    ... to let the general public look at it respectfully. [187] (1782-1925)
2.6 Desire ‘One wants to...’

(17) mii sàthaa càʔ khrây sâaŋ ?aaraam
    have faith iRR desire build temple

    *With faith in the Buddhism, they want to build a temple. [86] (1528)*

4 Quantitative data

This section provides quantitative data on occurrences of càk and càʔ in Thai inscriptions, which give an entire picture of gradual increase in the number of specific types of Thai modal concepts marked by càk / càʔ.

Table 1 shows the number of occurrences of the two morphemes in different semantic contexts of the inscription discourses. For convenience’s sake, I divide the documented history of the Thai language into the following four periods according to different dynasties:

- Period I: the Sukhothai period (1292-1438)
- Period II: the Ayutthaya-Thonburi period (1438-1782)
- Period III: the first half of the Rattanakosin period (1782-1925)
- Period IV: the latter half of the Rattanakosin period (1925-1978)

A total of 635 tokens of càk and càʔ were found in the inscriptions, though 15 of them were not decodable due to the unreadability of some inscriptions in Period I. The ratio comparing the two morphemes’ occurrences in each period is also indicated at the bottom of Table 1.

Table 1 reveals that before Period III the ratio of occurrences of the two morphemes in the volitive contexts exhibits higher values than that in the non-volitive contexts, but after Period III the two ratios do not differ very much. Note that càk was used as a verb in Period I. Since the number of occurrences of its verbal usage is too small (only four), I cannot tell exactly what verbal meaning it had. At any rate, the important point is that since Period I the two morphemes have been capable of functioning as an irrealis marker occurring not only in the contexts of ‘desire’, ‘volition (intention)’ and ‘prediction (future)’, which are named in (1) above, but also in the contexts of ‘conditional concessive’, ‘inevitability’, ‘conditionals’ and ‘apparent imminence’.
Table 1. Occurrences of càk and càɁ in different semantic contexts of Thai inscription discourses

<table>
<thead>
<tr>
<th></th>
<th>Period I Sukhothai</th>
<th>Period II Ayutthaya &amp; Thonburi</th>
<th>Period III Rattanakosin King Rama 1-6</th>
<th>Period IV Rattanakosin King Rama 7-9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1292-1438</td>
<td>1438-1782</td>
<td>1782-1925</td>
<td>1925-1978</td>
</tr>
<tr>
<td>1. Non-volitive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1 Conditional concessive</td>
<td>50</td>
<td>21</td>
<td>73</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>38%</td>
<td>28%</td>
<td>53%</td>
<td>48%</td>
</tr>
<tr>
<td>1.2 Prediction, 1.3 Inevitability, 1.4 Conclusion (including Negative conviction, Ironical interrogative)</td>
<td>32</td>
<td>14</td>
<td>56</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>24%</td>
<td>18%</td>
<td>41%</td>
<td>39%</td>
</tr>
<tr>
<td>1.5 Circumstantial impossibility, 1.6 Circumstantial possibility</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3%</td>
<td>2%</td>
<td>4%</td>
</tr>
<tr>
<td>1.7 Conditionals, 1.8 Apparent imminence</td>
<td>12</td>
<td>4</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>9%</td>
<td>5%</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>2. Volitive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1 Appropriateness, 2.2 Permission, 2.3 Obligation</td>
<td>0</td>
<td>4</td>
<td>10</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>5%</td>
<td>7%</td>
<td>16%</td>
</tr>
<tr>
<td>2.4 Ability</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8%</td>
</tr>
<tr>
<td>2.5 Volition, 2.6 Desire</td>
<td>77</td>
<td>49</td>
<td>54</td>
<td>91%</td>
</tr>
<tr>
<td></td>
<td>59%</td>
<td>66%</td>
<td>39%</td>
<td>33%</td>
</tr>
<tr>
<td>VERB càk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Total</td>
<td>131 +15</td>
<td>100%</td>
<td>74 100%</td>
<td>137 100%</td>
</tr>
<tr>
<td>càk : càɁ</td>
<td>119 : 27</td>
<td>44 : 30</td>
<td>1 : 136</td>
<td>4 : 274</td>
</tr>
</tbody>
</table>

Figures 3 through 6 below are semantic maps designating a variety of Thai modal concepts marked by the irrealis marker càk / càɁ in each of the four periods. From the first two Figures (Figures 3 and 4) we can see that the irrealis marker began to be used in the contexts of ‘appropriateness’ and ‘circumstantial impossibility’ in Period II (1438-1782).

---

48 I have rounded off the decimal fractions of the percentages indicated in Table 1 and Figures 3 to 6, and so the total of the percentages in each part may be slightly under or over 100%.
Volition 38%, Desire 21%  
Conditional concessive 5%  
Prediction 11%, Inevitability 14%  
Volitive  

Volition 54%, Desire 12%  
Conditional concessive 1%  
Prediction 9%, Inevitability 9%  
Circumstantial impossibility 3%  
Appropriateness 5%  
Conditionals 5%  
Non-volitive

Figure 3. Semantic map for Thai modal concepts marked by the irrealis marker in Period I

Figure 4. Semantic map for Thai modal concepts marked by the irrealis marker in Period II

Figure 4 above and Figure 5 below show that the irrealis marker began to be used in the contexts of ‘permission’, ‘conclusion’ and ‘circumstantial possibility’ in Period III (1782-1925).

---

Although the concept ‘apparent imminence’ disappears in the semantic maps of Period II-IV (Figures 4 to 6), I do not think that the irrealis marker was incompatible with the meaning of ‘apparent imminence’ during that time. The occurrence frequency of ‘apparent imminence’ expressions in the inscriptions is low, presumably because inscription discourses normally have nothing to do with situations characterized as ‘apparent imminence’. In fact, the irrealis marker in present-day Thai is still compatible with the meaning of ‘apparent imminence’. There are some entrenched forms with the irrealis marker for the concept (e.g. kamlaŋ càʔ ‘be about to’, klây càʔ ‘nearly’) which are used in oral discourses more frequently than in written discourses (Takahashi 2002).
The Evolution of ตาย Construction in Thai

<table>
<thead>
<tr>
<th>Appropriateness 7%, Permission 1%</th>
<th>Conditional concessive 8%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volition 32%, Desire 7%</td>
<td>Prediction 13%, Inevitability 24%, Conclusion 2%</td>
</tr>
<tr>
<td></td>
<td>Circumstantial impossi. 1%, Circumstantial possi. 1%</td>
</tr>
<tr>
<td>Volitive</td>
<td>Conditionals 2%</td>
</tr>
</tbody>
</table>

### Figure 5. Semantic map for Thai modal concepts marked by the irrealis marker in Period III

Figure 5 above and Figure 6 below show that the irrealis marker began to be used in the contexts of ‘obligation’ and ‘ability’ in Period IV (1925-1978).

<table>
<thead>
<tr>
<th>Appropriateness 14%, Perm. 1%, Obligation 1%</th>
<th>Conditional concessive 2%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability 3%</td>
<td>Prediction 14%, Inevitability 24%, Conclusion 1%</td>
</tr>
<tr>
<td>Volition 24%, Desire 8%</td>
<td>Circumstantial impossi. 2%, Circumstantial possi. 2%</td>
</tr>
<tr>
<td>Volitive</td>
<td>Conditionals 3%</td>
</tr>
</tbody>
</table>

### Figure 6. Semantic map for Thai modal concepts marked by the irrealis marker in Period IV

Thus, specific modal concepts in Thai have gradually emerged, at least, since the 15th century. In addition, the irrealis marker came to be less and less used in the contexts of ‘volition’ (38% → 54% → 32% → 24%) and ‘desire’ (21% → 12% → 7% → 8%), while it came to be more and more used in the contexts of ‘inevitability’ (14% → 9% → 24% → 24%), ‘prediction’ (11% → 9% → 13% → 14%) and ‘appropriateness’ (0% → 5% → 7% → 14%). It follows that formerly the irrealis marker was more likely to mark less subjective types of irrealis situations, but recently it has been used to mark more subjective types.

### 5 Split patterns of Thai modal concepts related to possibility

In this section, I will closely examine historical split patterns of the possibility-related types of modal concepts (such as ‘circumstantial possibility’ and ‘ability’) expressed in Thai. Here I limit the discussion to the issue of split patterns of possibility-related concepts because I could not find a sufficient number of tokens of expressions for modal concepts other than those related to possibility in the inscription data.

Examples (18) through (23) provide samples of Thai modal expressions of the possibility-related types which I have collected from the inscription corpora. To identify possible split patterns involving these possibility-related concepts, I have analysed similarities in their syntactic forms and semantics.
1.5 Circumstantial impossibility ‘… is not achieved; it is not possible to …’

(18)=(8)  
\[ \text{CA? phannanaa b?o m?} \]  
\[ \text{d?y} \]  
\[ \text{IRR describe NEG POSSI PART} \]  

It is impossible to describe it. [86] (1528)

1.4 Negative conviction (Conclusion) ‘It is never possible to …’

(19)=(6)  
\[ \text{cab? niyom yindii luam s?y l?thi? s?atsan?a} \]  
\[ \text{NEG IRR favor be glad believe in ideology faith religion} \]  
\[ \text{?uuun n?ok caak phr?ph?uththas?atsan?a n?n h?a m?} \]  
\[ \text{d?y} \]  
\[ \text{other besides Buddhism TOP seek NEG POSSI} \]  

It is impossible that we would willingly believe in other religions than the Buddhism. [187] (1782-1925)

1.6 Circumstantial possibility ‘… is achieved; it is possible to …’

(20)=(9)  
\[ \text{mi? mi pratuu th?i ca? ??ok pay phaay n?ok d?y} \]  
\[ \text{NEG there is door REL IRR exit go outside POSSI} \]  

There is no door through which we can go out. [146] (1782-1925)

2.2 Permission ‘It is possible for one to …’

(21)=(13)  
\[ \text{cab? phuuk nay baan k?o d?y} \]  
\[ \text{IRR perform a ceremony in home CONJ POSSI} \]  

One can perform a ceremony (of establishing the Sangha communion area) in the home. [193] (1925-1978)

2.4 Ability ‘One is able to …’

(22)=(15)  
\[ \text{aat saam?aat cab? ath?thaan d?ay khaath?a b?t day b?t n?n} \]  
\[ \text{be able IRR pray with Pali verse a certain paragraph} \]  
\[ \text{h?y khr?p d?y} \]  
\[ \text{CAUS fully POSSI} \]  

One is able to pray by saying a certain full paragraph of a Pali verse. [257] (1925-1978)

1.4 Ironical interrogative (Conclusion) ‘Wherever can one find …? ’

(23)=(7)  
\[ \text{cab? h?a m?t m?uan cab? thi? n?y} \]  
\[ \text{IRR seek friend be like he where POSSI} \]  

Wherever can I find a good friend like him? [278] (1925-1978)

It is noteworthy that the syntactic forms for these possibility-related concepts in Thai all include the functional morpheme \textit{d?y} meaning ‘possibility’. Having analysed diachronic corpus data of \textit{d?y} expressions which I have complied from Thai inscriptions (Takahashi 2005, 2006), I hypothesize \textit{d?y}’s grammaticalization from a verb into a modal, as follows (Takahashi & Shinzato 2005, Takahashi 2008b).\textsuperscript{50} The verb \textit{d?y} originally expressed the meaning of ‘emergence’ in affirmative assertions and ‘non-emergence’ in

\textsuperscript{50} Enfield (2001, 2003, 2004) has proposed different grammaticalization paths for so-called ‘aquire’ words which, he assumes, subsume the Thai verb \textit{d?y}. 
negative assertions. When the meaning of ‘non-emergence’ was extended from the nominal domain (non-emergence of an object) into the verbal domain (non-achievement of an event) and the construction expressing the latter more abstract sense (VP \textit{NEG} \textit{dây}) became entrenched, \textit{dây} gained the function of signalling ‘circumstantial impossibility’ (viz. it is not possible to do/be something due to certain circumstances). In the case of example (18) above, the entity that the writer wants to describe has very special properties in quality or quantity, and so he cannot perfectly describe its specialty (such as its wonderfulness or numerosness).

Plausible derivations of the other possibility-related constructions exemplified in (18) to (23) above are as the following.

(a) circumstantial impossibility (e.g. (18)) \textgreater{} negative conviction (e.g. (19)):

The verb \textit{hǎa} ‘seek’ came to be included in the construction for circumstantial impossibility (\textit{hǎa \textit{NEG} \textit{dây} ‘seeking is not achieved; it is not possible to find out’), and another clause beginning with \textit{chây} (\textit{chây} (NP) VP ‘it is not …’) came to precede the construction, which resulted in the construction for negative conviction (\textit{chây} (NP) VP, \textit{hǎa \textit{NEG} \textit{dây} ‘It is never possible to …’}).

(b) circumstantial impossibility (e.g. (18)) \textgreater{} circumstantial possibility (e.g. (20)):

The negative in the construction for circumstantial impossibility came to be omitted, which gave rise to the construction for circumstantial possibility (VP \textit{dây} ‘… is achieved; it is possible to …’).

(c) circumstantial possibility (e.g. (20)) \textgreater{} permission (e.g. (21)):

The verb phrase in the construction for circumstantial possibility came to refer to an action to be engaged in by someone, which led to the construction for permission ((\textit{agent}-NP) VP \textit{dây} ‘It is possible for one to …’).

(d) circumstantial possibility (e.g. (20)) \textgreater{} ability (e.g. (22)):

The two loanwords \textit{Ɂàat} ‘brave’ (from Khmer) and \textit{sǎamâat} ‘hope, desire’ (from Pali) came to be added to the construction for circumstantial possibility. At the same time, the verb phrase in the construction came to refer to someone’s action. This resulted in the construction for ability ((\textit{agent}-NP) \textit{Ɂàat sǎamâat} VP \textit{dây} ‘One is able to …’).

(e) circumstantial possibility (e.g. (20)) \textgreater{} ironical interrogative (e.g. (23)):

The interrogative \textit{thîi nãy} ‘where’ came to be included in the construction for circumstantial possibility, which yielded the construction for ironical interrogative (VP \textit{thîi nãy \textit{dây} ‘Wherever can one …?’}).

The derivational relationships among the five split patterns (a) to (e) is graphically shown in Diagram 1 below. In Period II, the concept of ‘circumstantial impossibility’ (e.g. (18)) was established. In Period III, ‘negative conviction (conclusion)’ (e.g. (19)) and

\footnote{Although in present-day Thai the morpheme \textit{chây} means ‘yes, that is correct’, originally it was a negative marker preceding a noun phrase (Takahashi 2008a). It came to be used as a constituent of some formulaic expressions like (19) before changing into the non-negative meaning in the present.}
‘circumstantial possibility’ (e.g. (20)) derived from ‘circumstantial impossibility’, and also ‘permission’ (e.g. (21)) derived from ‘circumstantial possibility’. And in Period IV, ‘ability’ (e.g. (22)) and ‘ironical interrogative (conclusion)’ (e.g. (23)) derived from ‘circumstantial possibility’.

<table>
<thead>
<tr>
<th>Period II 1438-1782</th>
<th>Period III 1782-1925</th>
<th>Period IV 1925-1978</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Circumstantial impossibility</td>
<td>←Negative conviction</td>
<td>←Circumstantial possibility</td>
</tr>
<tr>
<td>a.</td>
<td>b.</td>
<td>c.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>←Permission</td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td>←Ability</td>
</tr>
<tr>
<td>e.</td>
<td></td>
<td>←Ironical interrogative</td>
</tr>
</tbody>
</table>

Diagram 1. Split patterns of possibility-related modal concepts in Thai

(24) and (25) summarize the directionality of semantic changes involved in these split patterns occurring during Periods II-IV (1438-1978).

(24) Directionality of changes in subjectivity of possibility-related concepts in Thai:
less subjective > more subjective

(25) Directionality of changes in volitivity of possibility-related concepts in Thai:
non-volitive > volitive\(^\text{52}\)

Figures 7 to 10 below graphically depict different types of change in the values of subjectivity and volitivity associated with the split patterns (a) to (e) indicated in Diagram 1 above.

\(^\text{52}\) In her study on developments of Thai verbs into auxiliaries, Meesat (1997: 178) states that the verb \(\text{təŋ} \) ‘hit, fit’ began to be used as an auxiliary expressing ‘obligation’ (i.e. volitive, deontic modality) in the reign of King Rama I (1782-1809), and then gained another auxiliary function to express ‘conclusion’ (i.e. non-volitive, epistemic modality) in the reign of King Rama 4 (1851-1868). Similarly, Diller (1988: 291) mentions that the Khmer-derived verb \(\text{Ɂàat} \) ‘brave’ first developed into a modal for ‘ability’ (i.e. volitive, dynamic modality) and then came to function as a modal for ‘probability (possible conclusion)’ (i.e. non-volitive, epistemic modality). Unfortunately, however, my corpus data of irrealis expressions in the inscriptions dated in 1292-1978 do not include evident samples of the ‘conclusion’ usage of \(\text{təŋ} \) \(\text{càɁ} \text{təŋ} \text{VP} \) ‘it is concluded by inference that …’) and the ‘probability’ usage of \(\text{Ɂàat} \) \(\text{Ɂàat} \text{càɁ} \text{VP} \) ‘it might appear that …’), and therefore in this study, which deals with rather limited language data, I cannot attest the direction of these late semantic changes of Thai modals, viz., ‘volitive > non-volitive’. 
The three split patterns in Figure 7, (a) ‘circumstantial impossibility > negative conviction (conclusion)’, (c) ‘circumstantial possibility > permission’ and (e) ‘circumstantial possibility > ironical interrogative (conclusion)’ exhibit changes in subjectivity, namely ‘less subjective > more subjective’. This direction is consistent with the hypothesis called ‘unidirectionality of semantic change’ or ‘subjectification’ (Traugott 1982, 1989).

In contrast, the two split patterns in Figure 8, (b) ‘circumstantial impossibility > circumstantial possibility’ and (d) ‘circumstantial possibility > ability’ exhibit little change in subjectivity. It appears that judging to be possible is no more subjective than judging to be impossible; judging to be physically or mentally possible is no more subjective than judging to be circumstantially possible; and so on. So far the literature on semantic change has offered little in-depth analysis of these split patterns.53

The two split patterns in Figure 9, (c) ‘circumstantial possibility > permission’ and (d) ‘circumstantial possibility > ability’ exhibit shift in volitivity, namely ‘non-volitive >

---

53 Shibuya (1993) and Shinzato (2008), which investigate the evolution of Japanese modals for possibility, are a couple of exceptions.
volitive’. This shift is opposed to the widespread view that non-volitive modal concepts arise from volitive (agent-oriented) ones (Bybee et al. 1994). However, my investigation into the Thai diachronic corpus data reveals that this view is not applicable to Thai. Aside from the recent semantic extensions of ต่าย (obligation > conclusion) and ผ่าต (ability > probability) (see Footnote 52), semantic changes of Thai modals in the past are largely ‘non-volitive > volitive’. This supports Narrog’s (2005) claim that actually this direction of semantic change is pervasively found in many languages.

<table>
<thead>
<tr>
<th>Subj.</th>
<th>Ironical interrogative</th>
<th>Negative conviction</th>
</tr>
</thead>
<tbody>
<tr>
<td>↑</td>
<td>Circumstantial possibility ← Circumstantial impossibility</td>
<td>(e)</td>
</tr>
<tr>
<td>↓</td>
<td>(b)</td>
<td>(a)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obj.</th>
<th>Volitive</th>
<th>Non-volitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>↑</td>
<td>Ironical interrogative</td>
<td>Negative conviction</td>
</tr>
<tr>
<td>↓</td>
<td>Circumstantial possibility ← Circumstantial impossibility</td>
<td>(e)</td>
</tr>
<tr>
<td>(a)</td>
<td>(b)</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 10.** Semantic map for split patterns with no change in volitivity

The three split patterns in Figure 10, (a) ‘circumstantial impossibility > negative conviction (conclusion)’, (b) ‘circumstantial impossibility > circumstantial possibility’ and (e) ‘circumstantial possibility > ironical interrogative (conclusion)’ exhibit no change in volitivity. Curiously, as far as Thai possibility-related concepts are concerned, non-volitive concepts more frequently split and became diversified than volitive concepts did.

6 Concluding remarks

In this study, I have examined actual discourses in Thai diachronic corpus data and found a number of plausible split patterns of Thai modal concepts related to possibility.

In order to explain the facts presented here, I suggest that the two opposing directions of semantic change between the two types of possibility-related concepts, i.e. ‘ability > circumstantial possibility (root possibility)’ and the other way around, may correspond to the typological dichotomy of the characteristic way of describing a situation, namely ‘person- vs. situation-focus’ (Hinds 1986; cf. also Teramura 1976 and Ikegami 1991). Person-focus languages (like English) tend to focus on the agent (i.e. conscious, willful and responsible person) in describing a situation, whereas situation-focus languages (like Japanese and Thai) tend to describe the whole event without placing a special focus upon the agent. It is reasonable to assume that in person-focus languages, the domain of volitive modal concepts, rather than the domain of non-volitive modal concepts, has rich split patterns, since lexical items expressing an agent’s volitive action, ability, desire and the like are apt to evolve into modal markers, the process of which Langacker (1999: 297-315) calls ‘attenuation’. In situation-focus languages, on the other hand, the domain of non-volitive modal concepts, rather than the domain of volitive modal concepts, has rich split patterns, since the lexical items for volition-related concepts are less likely to be employed for the expression of a modal concept. Whether this is true or not remains a matter of future research.

In conclusion, following Narrog (2005), I suggest that the directions of semantic change, except for the seemingly universal direction ‘less subjective > more subjective’, may vary across languages. I also suggest that variation in the directions may reflect
variation not only in the linguistic structures (morphosyntactic principles and lexical systems) but also in the speakers’ preferred style of the description of a state of affairs.

References


The Evolution of dây Construction in Thai


**Corpus data**

The Prime Minister’s Secretariat, Thailand. 1924/1978. *Corpus of Thai Inscriptions, Installment 1*. Bangkok, The Prime Minister’s Secretariat.

The Prime Minister’s Secretariat, Thailand. 1965. *Corpus of Thai Inscriptions, Installment 3*. Bangkok, The Prime Minister’s Secretariat.


The Prime Minister’s Secretariat, Thailand. 1978. *Corpus of Thai Inscriptions, Installment 6.2*. Bangkok, The Prime Minister’s Secretariat.

JSEALS style guide for authors

Papers should be submitted electronically as .DOC or .RTF files. Initial drafts should be prepared with minimal formatting, and where possible with just one Unicode compliant font. After papers have been reviewed, revised and approved, authors are required to format their papers to style.

Download the template from the homepage at www.jseals.org, and type or paste 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 – 12 point Times New Roman. If your version of Times New Roman does not accommodate the characters you require, you may substitute with DoulosIPA (freely available from sil.org). For non-roman fonts please consult with the Managing Editor first.

Line spacing: 14 pt exactly in normal text paragraphs

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.

Do not link references, headings, tables or figure numbers in WORD.

Use footnotes, not endnotes.

All papers must have an abstract of approximately 100 words, plus 3 keywords (preferably consistent with the keywords used at http://sealang.net/sala/).

If you have any queries, please consult the website at www.jseals.org or contact Paul Sidwell via <paulsidwell@yahoo.com> or <paul.sidwell@anu.edu.au>.