Adaptation of focus to lexical tone and phrasing in Kammu

Anastasia M Karlsson, Jan-Olof Svantesson, David House

Dept. of Linguistics and Phonetics, Lund University; Dept. of Speech, Music and Hearing, KTH, Stockholm

anastasia.karlsson@ling.lu.se, jan-olof.svantesson@ling.lu.se, davidh@speech.kth.se

Abstract

Recently it has been proposed to base intonational typology on the way languages convey focus (pragmatic prominence). Generally, languages can enhance, add or delete phrase boundary tones (phrase languages) or add an extra pitch accent (intonation languages) to mark focus. Tone languages are somewhat problematic for this typology as it is difficult to make direct observations on their sentence intonation. Generally, tone languages are supposed to have a suppressed use of intonation on the sentence level. As a consequence, tone languages are usually described as a separate prosodic type.

We investigate Kammu, an Austroasiatic language spoken mainly in Laos. Kammu has two main dialects, one with lexical tones and one without tones; other phonological differences are minimal. In our previous studies we concentrated on elicited and controlled speech to explore the role of F0 in focusing and phrasing. We concluded that both dialects belong to the phrase language type. In the present study we use spontaneous material and investigate both F0 and duration for phrasing and focusing. We show that both F0 and duration are used for focusing in the two dialects. However, the prosodic realization of focus is restricted to the rightmost edge of the prosodic phrase, which coincides with the phrase boundary. The main conclusion is that the prevalence for edge marking influences the realization of focus, in accordance with prosodic typology. Further, the occurrence of lexical tones can still be combined with the phrasing pattern, and the tonal dialect shows the same strategies for focusing and phrasing as a phrase language.

Index Terms: lexical tone, intonation, phrasing, focusing, prosodic typology, Kammu

1. Introduction

Two main functions of intonation are phrasing and conveying pragmatic focus. Not all languages use intonation for focal accentuation, but rather convey information structure by syntactical and morphological means. When prosodic means are used, focusing can be achieved by different strategies. These differences are the grounds for separating languages into edge-prominence languages vs. head-prominence languages [1], corresponding to phrase languages vs. intonation languages [2]. In intonation languages, focus is realized by adding an extra pitch accent on the focused word. In phrase languages, information structure is most often conveyed by morpho-syntactic means, and focusing is achieved by changes in pitch level of phrasing tones, dephrasing or insertion of a new boundary tone. No new pitch accents are added to mark a focused word as the case is in intonation languages. According to this description, main Indian languages as Hindi, Bengali, Tamil and Malayalam [2], as well as Korean [1], West Greenlandic [3] and Mongolian [4] are typical phrase languages.

Languages with lexical tones have been treated as a separate prosodic category, as it is usually assumed that lexical tones restrict the use of sentence intonation. As described in [2], tone languages have ‘nearly no phrasal tones (correlating with rich lexical specification of tone)’, and phrase languages have ‘only phrasal tones (no pitch accent, no lexical tone)’. However, intonation in tone languages is difficult to observe due to the influence of lexical tones.

We explore how the occurrence of lexical tones influences intonation by investigating data from the Austroasiatic language Kammu, one of few languages with two dialects whose only major phonological difference is the presence or absence of lexical tones.

2. The Kammu language

Kammu (Kmhmù, Khmu’, etc.) belongs to the Khmuic branch of the Austroasiatic language family. It is the largest minority language in Laos, spoken by about 600,000 people mainly in the northern parts of Laos; there are also speakers in adjacent areas of Thailand, Vietnam and China.

There are two main dialects of Kammu. One dialect (Eastern Kammu) lacks lexical tones, while the other dialect (having two subdialects, Northern and Western Kammu) has a tone system consisting of one high and one low tone. An original contrast between voiceless and voiced stops and sonorants is preserved in the non-tonal dialect. In the tonal dialect, voiceless and voiced initial consonants merged, giving rise to high and low tone, respectively. Other differences between the dialects are marginal, and speakers of different dialects understand each other without difficulty [5], [6].

The phonetic distance between the tones is small, typically around 15–20 Hz (about 2–3 semitones) for male speakers [6], but perception experiments have shown that speakers of the tonal dialect are quite sensitive to F0 changes and use F0 as the main cue for discriminating minimal pairs such as klaay ‘eagle’ vs. klaay ‘stone’. Speakers of the non-tonal dialect, where these words are klaay and glaan, respectively, do not use F0 as a cue and do not hear the difference between such minimal pairs spoken by a tonal Kammu speaker. In the tonal dialect, stop consonants that developed from original voiceless and voiced stops are acoustically identical [7] while the difference in VOT is maintained in the non-tonal dialect [6]. Thus tonal Kammu uses F0 to distinguish words but non-tonal Kammu does not.

There are two kinds of syllables in Kammu, traditionally called major and minor syllables, and most monomorphemic words are either monosyllabic, consisting of one major syllable (e.g. tis ‘mushroom’), or sesquisyllabic, consisting of one minor and one major syllable (e.g. kinhmài ‘human being’, kmiài ‘silver’). Major syllables have a vowel nucleus and a tone. There are two kinds of minor syllables. One kind (e.g. kin in kinhmài) has a rhyme (consisting of a consonant coda) and an own tone. The second kind of minor syllable (exemplified by k in kmiài) has no rhyme and no tone. The tone is phonetically realized as the same tone on both
syllables, if the minor syllable is pronounced with a schwa vowel (e.g. k.múul [k.múul] ‘silver’); otherwise the tone is realized only on the major syllable, [kmúul], [8].

The contrastive load of tones is very high, and almost 900 minimal pairs for the tone contrast are registered in the Dictionary for Yuán, one of the tonal subdialects [9].

Two dialects that constitute a ‘minimal pair’ with and without lexical tones provide us with a rare experimental linguistic material for investigating the interaction between lexical tones and intonation.

3. Speakers and material

The investigation reported here is based on data collected in northern Laos in November 2007 and in northern Thailand in March 2008. A total of 24 speakers ranging in age from 14 to 72 years were recorded. 10 speakers (7 men and 3 women) represent the non-tonal dialect and 14 speakers (6 men and 8 women) the tonal dialect. The recordings comprise about 15 hours of speech. For various reasons, we could not get comparable data from each speaker for each kind of recording, so the number of participants differs in the different investigations.

The subjects were recorded with a portable Edirol R-09 digital recorder and a lapel microphone. The utterances were digitized at 48 kHz sampling rate and 16-bit amplitude resolution and stored in .wav file format. Most speakers were recorded in quiet hotel rooms. One speaker was recorded in his home and one in his native village.

The recorded material used in the investigations reported here is of two kinds:

Focus material. A non-scripted controlled material used for eliciting focus. To elicit focus we used pictures in which the oppositions eagle ~ lizard and black ~ red were systematically varied. For example, when showing a picture of a red eagle but asking Is this a black eagle?, we expected to get This is a RED eagle, with focus on RED, as the answer, but asking Is this a red lizard?, we expected to get This is a red EAGLE, with focus on EAGLE, as the answer. The questions were asked in (tonal) Kammu by Damrong Tayanin, a native speaker of tonal Kammu. The speakers answered in their own dialect. They were instructed to answer with a full sentence. Each answer was repeated three times by each speaker.

The combinations in the answers were:

- non-tonal: gii moh klaaj/daaj hajj/fim
- tonal: kii moh klaaj/taaj jajj/fim
- glosses: this is eagle/lizard black/red
- “This is a black/red eagle/lizard”

When the color asked for was wrong we should expect narrow final focus on the adjective (final word). When the animal asked for was wrong we expect narrow focus on the noun (non-final word). Controlling for lexical tones and the position of focus we got eight possible combinations.

Spontaneous rice monologues. The speakers were asked to explain how rice is grown and prepared, from the burning of the forest to make a rice field until the rice is eaten. All Kammu speakers are acquainted with this subject, but some of the younger speakers living in cities gave very brief accounts of it. Although spontaneous, this material is structured in the same way by all speakers: new information is given at the end of an utterance (most often a verb describing the next phase of the growing or preparing of rice) and then repeated in the immediately following utterance, for example: When the rice is ripe, we harvest it. After we have harvested the rice, we carry it home to the village. This allows us to make systematic comparisons of the same word in focused and non-focused position (as new and old information, respectively); see further [10].

The recorded material was checked, transcribed and translated into English by Damrong Tayanin. The Praat speech analysis program was used to display the waveform, spectrogram and fundamental frequency contour of the utterances. To be able to compare different speakers, male and female, all F0 values were converted to semitones.

4. Previous results

In previous work based on measurements of F0 in elicited speech ([11], [12]) and observations from spontaneous speech ([13], [10]) we found a characteristic high boundary tone at the right edge of prosodic phrases. This holds for both dialects. Thus, longer utterances in elicited material (sentences translated from Lao or Thai to Kammu by our speakers) are divided into phrases with a high (falling or truncated) boundary tone on their right edge. This boundary tone is also found utterance finally though it can be missing in elicited speech. In spontaneous speech the boundary tone is realized by all speakers. As this high peak is realized in utterance final position regardless of the placement of focal accent, it has other functions than prominence and is analyzed as a phrase boundary tone. Figure 1 illustrates the tonal course of the sentence traak [ʔɔʔ gɔɔ went] hmraaj [ʔɔʔ gɔɔ went] [buffalo I also buy, horse I also buy] “A buffalo I’ll buy and a horse I’ll buy” with focus on traak “buffalo” and hmraaj “horse”. The sentence is divided into four phrases with high boundaries on their right edges. The final phrase is also signaled by a high F0 peak. This pattern is found for all speakers. The most typical size of a prosodic phrase is two lexical words.

Phrasing patterns are difficult to observe in the tonal dialect. In [14] we could show, using the Focus material, that the utterance final phrase boundary occurs also in the tonal dialect though it is affected by the lexical tones. We found that the boundary tone occurs independently of the place of focus in both dialects. The F0 height of the boundary tone is affected by focus. The effect of focus in non-final position is rather
small, and we still find higher \( F_0 \) utterance finally, coinciding with the boundary tone.

The phrase pattern in the tonal dialect can also be observed in utterances with only high or only low tones. Such utterances are rare in our material, but one example is illustrated in Figure 2. We observe a semantically coherent speech unit chunked into six prosodic phrases: \([wàaj híc]\) \([kàaj kùum]\) \([wàaj kùum]\) \([kàaj rùŋ]\) \([wàaj rùŋ]\) \([kàaj pòʔ]\) “After pound, then winnow, after winnow, then steam, after steam, then eat”. All words except the second one have low lexical tone. The phrasing pattern with higher \( F_0 \) on the phrase final word can be observed. The last phrase gets the highest \( F_0 \) which is typical for topic shift in our Rice monologues [10].

![Figure 2: Phrasing pattern in spontaneous speech. The six prosodic phrases are \([wàaj híc]\) \([kàaj kùum]\) \([wàaj kùum]\) \([kàaj rùŋ]\) \([wàaj rùŋ]\) \([kàaj pòʔ]\) “After pound, then winnow, after winnow, then steam, after steam, then eat”. All words except the second one have low lexical tone. Arrows indicate boundaries of focused phrases. Male tonal speaker. Pitch in semitones (re 100 Hz).](image)

The main result of our previous investigations is that both tonal and non-tonal dialects have the same underlying intonation system. Lexical tones are prioritized and the right phrase edge is the default position for realization of focal intonation by enhancing the phrase boundary tone. Thus we established the following prosodic hierarchy in Kammu:

- 1. Lexical tone
- 2. Phrase-final boundary tone
- 3. Focal accent

5. Current investigation

5.1. Focus and \( F_0 \) in spontaneous speech

Based on our previous measurements of \( F_0 \) in controlled speech and observations from spontaneous speech, we concluded that the default position for focus is the phrase final one. To get a measure of the effect of focus on prosodic phrases we used the spontaneous Rice monologues. As described in Section 3, they are structured so that the same word often occurs first focused (as new information), and then repeated unfocused (as old information). In both cases the word is placed phrase finally, see illustration in Figure 2. The difference between the maximum \( F_0 \) in the focused and the unfocused word (in semitones) was taken as a measure of the focus effect in spontaneous speech. This study included 6 non-tonal speakers (all male) and 9 tonal speakers (5 male, 4 female). Altogether 124 word pairs were measured (56 for tonal and 68 for non-tonal speakers).

The mean focus effect was 4.0 semitones for the non-tonal speakers, and 1.6 semitones for the tonal speakers. Paired \( t \)-tests showed that the focus effect is significant \((p < 0.001\) in both dialects), and an unpaired \( t \)-test showed that the difference between the dialects is significant as well \((p < 0.001\).

The focus effect between two phrases in spontaneous speech is considerably larger than the elicited focus effect within a phrase. In elicited focus in the Focus material the mean focus effect was 1.0 semitone for the non-tonal speakers and 0.6 semitones (mean for all tone combinations) for the tonal speakers [14].

There is also a difference between the dialects, non-tonal speakers having a much larger focus effect than tonal speakers. This is consistent with [15], non-tonal speakers generally having a larger \( F_0 \) range than tonal speakers.

5.2. Focus and duration

5.2.1. Focus and duration in spontaneous speech

To investigate if focus is marked by duration, we used the spontaneous Rice monologues, where the same words occur in focused and non-focused position (as described in Section 3 and 5.1). The duration of the syllable rhyme (vowel nucleus and coda consonant) was compared in focused and non-focused position. Altogether 94 pairs, each consisting of the same word said by the same speaker in both focused and non-focused position were measured. The results are shown in Figure 3 and Table 1. The results show that words in focused position have much longer duration than non-focused words, on the average the rhyme is more than twice as long. For VN rhymes (with short vowel and a sonorant coda) the coda is lengthened much more than the vowel, perhaps because the vowel length contrast in Kammu prevents a short vowel from being lengthened too much. The variation is large, but paired \( t \)-tests show significant results for the lengthening of vowels and codas, except in two cases. The pattern is similar in both dialects.

![Figure 3: Average duration (ms) of the syllable rhyme in focused and non-focused position in spontaneous speech. 7 non-tonal speakers (all male) and 11 tonal speakers (5 male, 6 female). V = short vowel, VV = long vowel, N = sonorant coda, C = oral stop coda.](image)
Table 1: Average duration (ms) of the syllable rhyme in focused and non-focused position. The results of paired t-tests are shown, and the number of word pairs measured (n); for the t-tests, df = n–1. Same speakers as in Figure 3.

<table>
<thead>
<tr>
<th>Syllable type:</th>
<th>VN</th>
<th>VVN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>vowel</td>
<td>coda</td>
</tr>
<tr>
<td>non-tonal</td>
<td>focus</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>non-focus</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>5</td>
</tr>
<tr>
<td>tonal</td>
<td>focus</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>non-focus</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>0.140</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>4</td>
</tr>
</tbody>
</table>

5.2.2. Duration in utterances with elicited focus

We also measured the duration of the focused and non-focused words in the Focus material, presented in Section 3 above, where focus was elicited on either the final or pre-final word in the utterance. The durations of the whole words were measured (for words beginning with a stop, the occlusion phase was not included). The results are shown in Figure 4 and Table 2. In this case, there is lengthening in focused position in most cases, but it is much smaller than in the spontaneous material, less than 20%. The variation is large, and the difference is not always significant according to paired t-tests. The lengthening is somewhat smaller in the final than in the pre-final word, and in the non-tonal dialect there is no lengthening of the final word; the variation is large and the number of observations is small, however.

Table 2: Average duration (ms) of words in focused and non-focused position in elicited speech and the results of paired t-tests. The number of tokens (n) and the degrees of freedom are shown. Same speakers as in Figure 4.

<table>
<thead>
<tr>
<th>Word:</th>
<th>klaaŋ/klaaŋ</th>
<th>daaŋ/taaŋ</th>
<th>hiaŋ/jiaŋ</th>
<th>jim/jim</th>
</tr>
</thead>
<tbody>
<tr>
<td>focus</td>
<td>209</td>
<td>278</td>
<td>379</td>
<td>355</td>
</tr>
<tr>
<td>non-focus</td>
<td>178</td>
<td>241</td>
<td>405</td>
<td>354</td>
</tr>
<tr>
<td>p</td>
<td>0.078</td>
<td>0.033</td>
<td>0.589</td>
<td>0.862</td>
</tr>
<tr>
<td>n/df</td>
<td>209/178</td>
<td>176/187</td>
<td>187/198</td>
<td></td>
</tr>
<tr>
<td>tonal</td>
<td>293</td>
<td>278</td>
<td>394</td>
<td>370</td>
</tr>
<tr>
<td>focus</td>
<td>251</td>
<td>245</td>
<td>376</td>
<td>343</td>
</tr>
<tr>
<td>non-focus</td>
<td>251</td>
<td>245</td>
<td>376</td>
<td>343</td>
</tr>
<tr>
<td>p</td>
<td>0.100</td>
<td>&lt; 0.001</td>
<td>0.042</td>
<td>0.009</td>
</tr>
<tr>
<td>n/df</td>
<td>33/14</td>
<td>35/16</td>
<td>34/15</td>
<td>34/15</td>
</tr>
</tbody>
</table>

6. Discussion

Our results show that both the non-tonal and the tonal dialect mark prosodic phrases by a high boundary tone at the right edge. This boundary tone is also the tonal carrier of focus. When a phrase is focused, its boundary tone is prosodically enhanced both by lengthening and by higher F0. This is found for both dialects. The focus effect is smaller for tonal speakers as a result of narrower changes in F0. This is in accordance with our previous findings that tonal speakers have a narrower overall pitch range than non-tonal speakers [15]. It might be...
expected that tonal speakers compensate the restricted use of \( F_0 \) by using more lengthening to convey focus. However, this is not the case and both dialects show a similar degree of lengthening in focus.

Although focus can be marked by increased \( F_0 \) on constituents that are not phrase-final, such as the pre-final word in our *Focus material*, this increase is small both in absolute size and in comparison with the marking on phrase-final constituents [14]. A similar situation holds for focus marking by duration: focused words are lengthened considerably in phrase-final position, but focus elicited on a word that is not in phrase-final position, as the pre-final word in the *Focus material*, results in much less lengthening. In the *Focus material*, the lengthening is small in final position as well. We interpret this as showing that the final word is focused by default in this case, even if there is elicited focus on the pre-final word, so that eliciting focus on the final word has only a small effect on duration, or none at all (a similar situation holds for \( F_0 \)).

The main difference found between the two dialects is in pitch range. \( F_0 \) changes are considerably smaller for tonal speakers than for non-tonal speakers, as expressed by the focus effect (4.0 semitones for the non-tonal and 1.6 semitones for the tonal speakers). On the other hand, lengthening is similar in both dialects. This could be interpreted as the consequence of the occurrence of lexical tones. In the tonal dialect, \( F_0 \) has to express a lexical contrast in addition to phrasing and focusing. This restricts the freedom of using \( F_0 \) range. Duration is used for realizing the phonological length contrast of vowels in both dialects, but this has no direct connection to lexical tones, and we find similar lengthening patterns for tonal and non-tonal speakers.

The underlying interplay between signaling of phrase edges and focus is the same for both dialects though there is a strong influence of lexical tones on both focal accent and phrase boundary tone in the tonal dialect. The fact that Kammu marks prosodic phrases with a high boundary tone at the right edge can be related to the prosodic typologies proposed by Jun [1] and Féry [2], as described in the Introduction. Our results show that Kammu is an edge-marking language in Jun’s terminology, or a phrase language in Féry’s terms. In tonal languages, sentence intonation is usually described as global changes in tonal courses. We could show that a language with lexical tones can indeed use the same tactics for focusing and phrasing as non-tonal languages of the edge-marking type.

Another important aspect of the present study is the choice of material. Using spontaneous speech we could catch the natural structuring of information flow in the two dialects. The fact that speakers place the new focused element phrase finally by default could explain our previous findings of weak prosodic cues of non-final focus in elicited speech. Also in spontaneous speech we found a larger degree of focus effect than in elicited material both regarding \( F_0 \) and duration.

7. Conclusions

We have found that the intonational systems of the two Kammu dialects are basically identical, both being strongly edge-marking. Due to the prevalence of edge-marking, expression of focus is restricted to the phrase final position and is realized by prosodic enhancing of the accessible boundary tone. The main difference between the two dialects is the adaptation of intonation to lexical tones, which is expressed by narrower pitch range variation in the tonal dialect.

8. Acknowledgements

The work reported here was done within the project *Separating Intonation from Tone*, financed by the Bank of Sweden Tercentenary Foundation. We are grateful to Damrong Tayanin for his essential contributions to the project.

9. References