

The Phrase-Final Accent in Kammu: Effects of Tone, Focus and Engagement

David House¹, Anastasia Karlsson², Jan-Olof Svantesson², Damrong Tayanin²

¹ Dept of Speech, Music and Hearing, School of Computer Science and Communication, KTH, Stockholm, Sweden

² Dept of Linguistics and Phonetics, Centre for Languages and Literature, Lund University, Sweden
davidh@speech.kth.se, Anastasia.Karlsson@ling.lu.se, Jan-Olof.Svantesson@ling.lu.se, Damrong.Tayanin@ling.lu.se

Abstract

The phrase-final accent can typically contain a multitude of simultaneous prosodic signals. In this study, aimed at separating the effects of lexical tone from phrase-final intonation, phrase-final accents of two dialects of Kammu were analyzed. Kammu, a Mon-Khmer language spoken primarily in northern Laos, has dialects with lexical tones and dialects with no lexical tones. Both dialects seem to engage the phrase-final accent to simultaneously convey focus, phrase finality, utterance finality, and speaker engagement. Both dialects also show clear evidence of truncation phenomena. These results have implications for our understanding of the interaction between tone, intonation and phrase-finality.

Index Terms: Kammu, intonation, tone, accent, focus

1. Introduction

The realization of pitch contours at the ends of phrases has received considerable attention in intonation research for a variety of reasons. The marking of a boundary is a fundamental function of prosody, and the final boundary tone is an important component. Bruce [1] for example demonstrated that the Swedish focal accent in phrase-final position contains the focal rise and the terminal fall, but that the focal accent rise becomes separated from the fall when the accent occurs earlier in the phrase. The phrase-final accent seems to be privileged in the sense that it can contain a multitude of simultaneous prosodic signals. In spontaneous speech, for example, a phrase-final focal rise can function as a prosodic feature conveying a friendly question [2], or function as a request address as when asking a favor of someone [3]. The combinations of prosodic functions found at the ends of phrases have also given rise to the analysis of the phenomena of truncation or compression of F0 that occur when the final syllable ends in non-sonorant segments (e.g. [4][5]). The aim of this study is to look more closely at some of the multiple functions of the phrase-final accent with an interest in the potential interactions between intonation, tone, focus and speaker engagement. To this purpose, we are specifically studying the phrase-final accent in Kammu.

Kammu, a Mon-Khmer language spoken primarily in northern Laos by approximately 600,000 speakers, but also in Thailand, Vietnam and China, is a language that has developed lexical tones rather recently, from the point of view of language history. Tones arose in connection with loss of the contrast between voiced and voiceless initial consonants in a number of dialects [6]. One of the main dialects of this language is a tone language with high or low tone on each syllable, while the other main dialect lacks lexical tones. The dialects differ only marginally in other respects. This makes

the different Kammu dialects well-suited for studying the influence of lexical tones on the intonation system.

In previous work using material gathered from spontaneous storytelling in Kammu, the utterance-final accent stands out as being especially rich in information showing two types of focal realizations depending on the expressive load of the accent and the speaker's own engagement [7]. In another study of the non-tonal Kammu dialect, it was generally found that in scripted speech, the highest F0 values were located on the utterance-final word [8]. In this paper, we examine the influence of tone, focus and to a certain extent speaker engagement on the phrase-final accent by using the same scripted speech material recorded by speakers of a non-tonal dialect and by speakers of a tonal dialect of Kammu.

2. Method

2.1. Data collection and speech material

In November, 2007, data collection was carried out during a two-week excursion in Laos. A total of 13 speakers (5 women and 8 men ranging in age from 14 to 72) were recorded. 10 speakers represented non-tonal dialects and 3 speakers tonal dialects. The recordings comprise approximately ten hours of both elicited and spontaneous speech material. In February and March, 2008, data collection was carried out during a two-week excursion to northern Thailand and northwest Vietnam. In Thailand, a total of 11 speakers of tonal dialects were recorded (6 women and 5 men ranging in age from 16 to 42). The recordings represent similar material to that recorded in Laos and comprise about five hours of speech. In Vietnam, a small number of spontaneous dialogues were recorded involving speakers of a non-tonal dialect.

The speech material was comprised of 47 read sentences. The sentences were composed in order to control for lexical tone, to elicit focus in different positions and to elicit phrasing and phrase boundaries. Kammu speakers are bilingual with Lao or Thai being their second language. Since Kammu lacks a written script, informants were asked to translate the material from Lao or Thai to Kammu. This resulted in some instances of slightly different but still compatible versions of the target sentences. The resulting utterances were checked and transcribed by one of the authors, Damrong Tayanin, who is a native speaker of Kammu. The speakers were requested to read each target sentence three times.

Non-scripted material to elicit focus was recorded by asking speakers to respond to eight questions about pictures in which the oppositions black-red and eagle-lizard were systematically varied, e.g. "Is this a red eagle?" asked about a picture of a black eagle for which the expected answer was, "No it is a BLACK eagle." The speakers were additionally requested to provide a spontaneous account of growing and

preparing rice elicited by photographs and also to give a spontaneous account of three series of drawings depicting a cooking fire which spreads to and burns down a house, hunting a deer and fishing. The purpose of the final task was to elicit informational structuring in a simple story.

For the present investigation six of the 47 read sentences were chosen for analysis. The sentences are transcribed below using the transcription convention for the tonal dialect. Sentences 1 and 2 contain only words with a low lexical tone while sentences 3 and 4 contain only words with a high lexical tone. Sentences 2 and 4 differ from 1 and 3 only in that they end with an additional color adjective following the noun (red and black respectively). Sentences 2 and 4 were designed to elicit focal accent on the final syllable. Sentences 5 and 6 convey a listing of three nouns (animals). The nouns all have high lexical tone in sentence 5 and low lexical tone in sentence 6.

- 1) nàa wèet hmràŋ
(*she bought a horse*)
- 2) nàa wèet hmràŋ yǐm
(*she bought a red horse*)
- 3) téek pháan tráak
(*Teeek killed a buffalo*)
- 4) téek pháan tráak yáŋ
(*Teeek killed a black buffalo*)
- 5) Ò àh tráak, àh sǎŋ, àh hyár
(*I have a buffalo, a pig and a chicken*)
- 6) Ò àh hmràŋ, àh mǝew, àh prùul
(*I have a horse, a cat and a badger*)

2.2. Speakers

Of all the speakers recorded, one non-tonal speaker and four tonal speakers were excluded from this study as they had problems reading and translating from the Lao/Thai script. All the other speakers were able to fluently translate the Lao/Thai script into Kammu and were included in the analysis. Thus there were 9 non-tonal speakers (2 women and 7 men) and 10 tonal speakers (6 women and 4 men) included in this study. The speakers ranged in ages from 14 to 72.

2.3. Recordings and analysis

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 of the speakers were recorded in quiet hotel rooms. One speaker was recorded in his home and one in his native village.

Using the WaveSurfer speech analysis program [9] the waveform, spectrogram and fundamental frequency contour of each utterance was displayed. Maximum and minimum F0 values were then annotated manually for successive syllables for each utterance. For sentences 1-4, there was generally very little F0 movement in the syllables leading up to the penult (pre-final syllable). Therefore, measurements were restricted to maximum F0 values on the pre-penultimate syllables, while both maximum and minimum values were measured on the

penultimate and ultimate syllables. For sentences 5 and 6, the initial syllable and the three nouns often showed considerable F0 movement and were therefore provided with both maximum and minimum F0 values. For the other syllables which showed little F0 movement, measurements were again restricted to F0 maximum values.

To be able to compare the different speakers, the measured F0 values in Hz were converted to semitones (St) and then normalized in accordance with [10]. A fixed semitone scale is used where the unit St is defined by

$$St = 12[\ln(\text{Hz}/100)/\ln 2] \quad (1)$$

which results in a reference value of St=0 semitones at 100 Hz, St=12 at 200 Hz and St=-12 at 50 Hz. Normalization is performed by subtracting each subject's average F0 in St (measured across the three utterances of each target sentence) from the individual St values.

3. Results

3.1. Focus and sentences 1-4

Plots for sentences 1-4 showing the F0 measurement points in normalized semitones are presented in Figure 1 for the non-tonal dialect and in Figure 2 for the tonal dialect. Alignment is from the end of the sentences. Both dialects exhibit a pronounced rise-fall excursion on the phrase-final syllable. There is, however, a clear difference in F0 range between the two dialects. The non-tonal dialect exhibits a much wider range of the final excursion (6-7 St) than does the tonal dialect (3-4 St).

If we are to find evidence of the influence of focus on the F0 excursions, we would expect differences between sentence pairs 1 and 2 on the one hand and 3 and 4 on the other hand where the final words in 2 and 4 receive focal accent. In the non-tonal dialect (Figure 1) we find clear evidence of this difference. The final F0 maximum of sentence 1 is 1.73 St while for sentence 2 it is 3.13 St. A single factor ANOVA shows this difference to be significant ($p < 0.05$). The final F0 maximum of sentence 3 for the non-tonal dialect is 1.49 St while for sentence 4 it is 2.84 St. Here again a single factor ANOVA shows this difference to be significant ($p < 0.05$). In the tonal dialect there is some evidence of the effect of focal accent but the difference is considerably reduced and not statistically significant. The final F0 maximum of sentence 1 is 0.56 St and for sentence 2 it is 0.76 St. A single factor ANOVA shows this difference to be non-significant ($p = 0.48$). The final F0 maximum of sentence 3 is also 0.76 St. and for sentence 4 it is 1.54 St. This difference is also not significant ($p = 0.05$).

If we test for the effects of tone in the non-tonal dialect, we find (as we should expect) no significant differences between the sentence pairs 1 and 3 ($p = 0.56$) nor between the sentence pairs 2 and 4 ($p = 0.66$). Testing for effects of tone in the tonal dialect does not result in a significant difference between sentences 1 and 3 ($p = 0.49$) but does show a significant difference between sentences 2 and 4 ($p < 0.05$). This may be the result of the combination of focus and a high tone on the final word of sentence 4.

3.2. Sentences 5-6

Plots for sentences 5 and 6 showing the F0 measurement points in normalized semitones are presented in Figure 3 for the non-tonal dialect and in Figure 4 for the tonal dialect.

Alignment is from the end of the sentences. Both dialects show a similar intonation pattern exhibiting rise-fall excursions on each of the three nouns comprising the listing of the three animals in each sentence. As is the case for sentences 1-4, there is also here a clear difference in F0 range between the two dialects. The non-tonal dialect exhibits a much wider overall range (6-9 St) than does the tonal dialect (3-4 St).

The nouns in sentence 5 have high tone in the tonal dialect, while those in sentence 6 have low tone. A comparison of the F0 maximum of the nouns in the three positions for the non-tonal dialect (Figure 3) shows that the noun in the first and third position has a higher F0 in sentence 6 than in sentence 5, i.e. the opposite to what we would expect

if this dialect were tonal. The noun in the second position has a higher F0 in sentence 5 than in sentence 6, i.e. in line with the tonal distinction. A single factor ANOVA shows these differences to be significant for the first position (F0 for sentence 5 is 1.60 St and for sentence 6 is 2.63 St, $p < 0.05$) and for the third position (F0 for sentence 5 is 2.02 St and for sentence 6 is 4.00 St, $p < 0.05$), but not for the second position (F0 for sentence 5 is 2.47 St and for sentence 6 is 1.86 St, $p = 0.17$).

A comparison of the F0 maximum of the nouns in the three positions for the tonal dialect (Figure 4) shows that the F0 maximum for the high tone (sentence 5) is higher than the low tone (sentence 6) in the first and second position but not

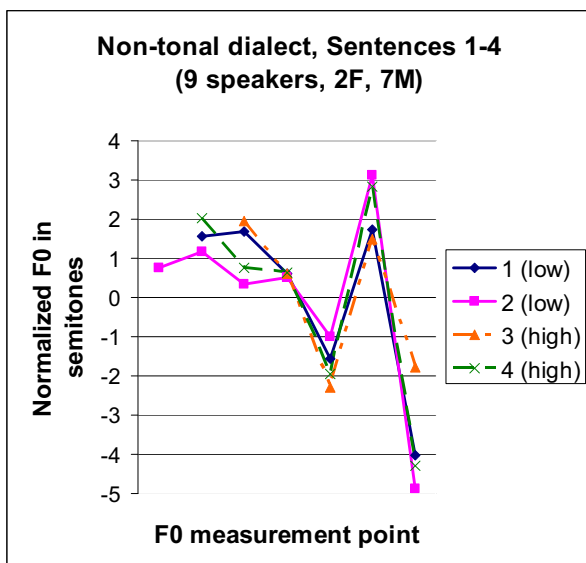


Figure 1: Normalized F0 measurement points for sentences 1-4 from nine speakers of the non-tonal dialect. Lexical tone in parenthesis refers to the tonal dialect.

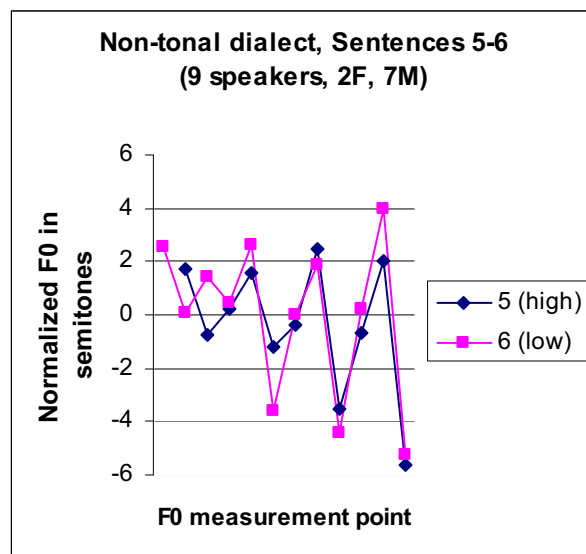


Figure 3: Normalized F0 measurement points for sentences 5-6 from nine speakers of the non-tonal dialect. Lexical tone in parenthesis refers to the tonal dialect.

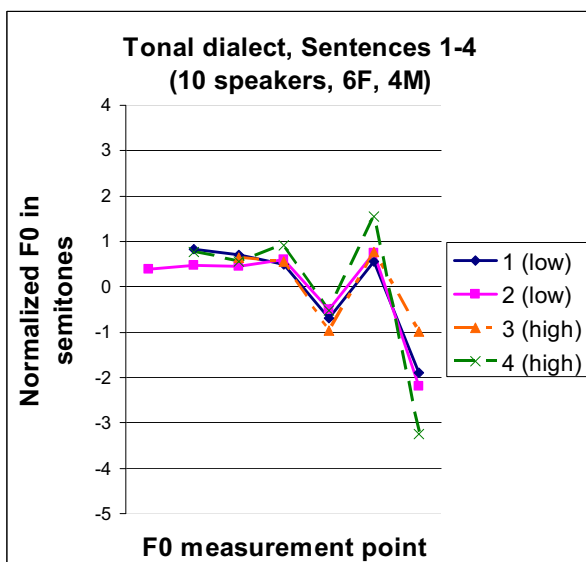


Figure 2: Normalized F0 measurement points for sentences 1-4 from ten speakers of the tonal dialect. Lexical tone is indicated in parenthesis.

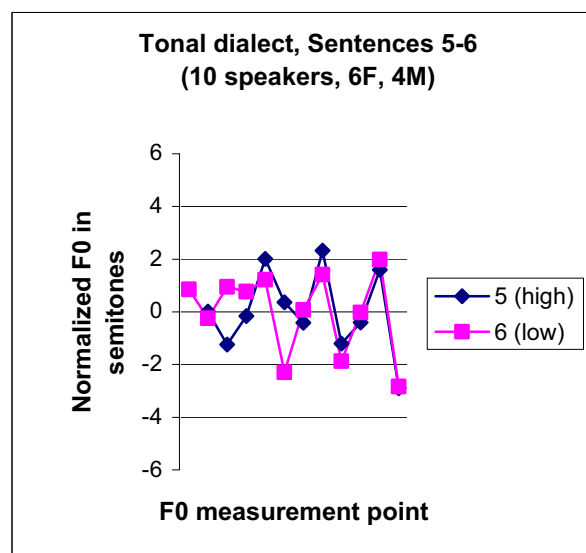


Figure 4: Normalized F0 measurement points for sentences 5-6 from ten speakers of the tonal dialect. Lexical tone is indicated in parenthesis.

in the third position. A single factor ANOVA shows these differences to be significant for the first position (F0 for sentence 5 is 2.01 St and for sentence 6 is 1.22 St, $p < 0.05$) and for the second position (F0 for sentence 5 is 2.33 St and for sentence 6 is 1.42 St, $p < 0.05$), but not for the third position (F0 for sentence 5 is 1.59 St and for sentence 6 is 1.98 St, $p = 0.30$).

Finally, there is evidence of a truncation phenomenon in both dialects. In Figures 1 and 2, the final F0 value for sentence 3 is considerably higher than for the other sentences. The final word in sentence 3 ends in a voiceless stop, while the other sentences end in sonorants.

4. Discussion

In general terms, the results presented here show considerable similarities in the basic intonation patterns of the phrase-final accent in the two dialects. There is a pronounced final rise-fall excursion which marks the end of the utterance. The presence of lexical tone in the tonal dialect does, however, restrict the intonation in certain specific ways. The most apparent overall difference is found in the restricted range of the tonal dialect. This is an interesting, but perhaps not such an unexpected difference. As the tonal differences have lexical meaning for the speakers of the tonal dialect, it may be important for speakers to maintain control over the absolute pitch of the syllable which can result in the general reduction of pitch range. In a previous study of Kammu, it was found that listeners of the tonal dialect were sensitive to extremely small pitch variation [6].

The lexical tones and the reduction of pitch range seem to have implications for the realization of focal accent. In the non-tonal dialect, the final color adjectives of sentences 2 and 4 showed a much higher F0 maximum than did the final nouns of sentences 1 and 3. Here the speakers are free to use rather dramatic F0 excursions to mark focus. The tonal speakers, on the other hand, seem to be restricted from doing this. It is only the final color adjective of sentence 4 which is given a markedly higher F0 maximum than the counterpart noun in sentence 3. Since the adjective of sentence 4 has high lexical tone, this fact seems to allow the speakers to additionally raise the maximum F0. As the adjective of sentence 2 has low lexical tone, the speakers are not free to raise this F0 maximum. Here we see evidence of interplay between lexical tone and intonation.

In the listing of animals in sentences 5 and 6, there is a large difference in the F0 maximum of the final word in the non-tonal dialect. The word “badger” in sentence 6 is spoken with a much higher F0 maximum than the word “chicken” in sentence 5. This can be explained by the fact that the word “badger” is semantically marked compared to the other common farm animals in the list. It is quite natural in Kammu farming culture to have a buffalo, a pig, a chicken, a horse and a cat, but not a badger! Some of the speakers even asked to confirm what the word was, and therefore it is not surprising if the word often elicited additional speaker engagement. This extra engagement also shows up in the tonal speakers’ versions of “badger” raising the low lexical tone to a higher F0 maximum than the word “chicken” in sentence 5 which has high lexical tone. This explains why significant tonal differences were only found in the first two positions in Figure 4, and not in the last. Here, speaker engagement is seen to override the tonal restriction, although the overall pitch range is still restricted compared to the non-tonal dialect. Thus we see an interaction between speaker engagement, tone and intonation.

Finally, it is interesting to note that the phenomenon of truncation seems to apply to both dialects, truncating the F0 movement when the syllable ends in a voiceless stop in both sets of sentences. This could indicate that the F0 maxima are more important than the minima.

5. Conclusions

In this study of the phrase-final accent in tonal and non-tonal dialects of Kammu, we see that the general patterns of intonation for these sentences are similar in the two dialects. However, there is clear evidence of the lexical tones of the tonal dialect restricting the pitch range and the realization of focus, especially when the lexical tone is low. Speaker engagement can have a strong effect on the phrase-final accent, and can even neutralize pitch differences of high and low lexical tone in certain cases. The results of this study are also consistent with the role of the phrase-final accent as able to signal many simultaneous linguistic functions.

6. Acknowledgements

The work reported in this paper has been carried out within the research project, “Separating intonation from tone” (SIFT), supported by the Bank of Sweden Tercentenary Foundation with additional funding from Crafoordska stiftelsen.

7. References

- [1] Bruce, G. “How floating is focal accent?” In K. Gregersen and H. Basbøll (Eds.), *Nordic Prosody IV*, 41-49, Odense University Press, 1987.
- [2] House, D. “Phrase-final rises as a prosodic feature in wh-questions in Swedish human-machine dialogue.” *Speech Communication*, 46, 268-283, 2005.
- [3] Ambrazaitis, G. “On final rises and fall-rises in German and Swedish.” In *Proceedings, FONETIK 2008*, 81-84, Department of Linguistics, University of Gothenburg, 2008.
- [4] Grabe, E. “Pitch accent realization in English and German.” *Journal of Phonetics* 26, 129-143, 1998.
- [5] Hanssen, J., Peters, J. and Gussenhoven, C. “Phrase-final pitch accommodation effects in Dutch.” *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1077-1080, 2007.
- [6] Svantesson, J-O. and House, D. “Tone production, tone perception and Kammu tonogenesis.” *Phonology*, 23, 309-333, 2006.
- [7] Karlsson, A. M., House, D., Svantesson, J-O., and Tayanin, D. “Prosodic Phrasing in Tonal and Non-tonal Dialects of Kammu.” *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1309-1312, 2007.
- [8] Karlsson, A., House, D., and Tayanin, D. “Recognizing phrase and utterance as prosodic units in non-tonal dialects of Kammu.” In *Proceedings, FONETIK 2008*, 89-92, Department of Linguistics, University of Gothenburg, 2008.
- [9] Sjölander, K., and Beskow, J. “WaveSurfer - an open source speech tool.” In *Proceedings of ICSLP 2000*, 6th Intl Conf on Spoken Language Processing, 464-467, Beijing, 2000.
- [10] Fant, G. and Kruckenberg, A. “Analysis and synthesis of Swedish prosody with outlooks on production and perception.” In G. Fant, H. Fujisaki, J. Cao, and Y. Xu (Eds.), *From traditional phonology to modern speech processing*, 73-95, Foreign Language Teaching and Research Press, Beijing, 2004.