



Influence of lexical tones on intonation in Kammu

Anastasia Karlsson¹, David House², Jan-Olof Svantesson¹, Damrong Tayanin¹

¹Dept of Linguistics and Phonetics, Centre for Languages and Literature, Lund University, Sweden

²Dept of Speech, Music and Hearing, School of Computer Science and Communication, KTH, Stockholm, Sweden

Anastasia.Karlsson@ling.lu.se, davidh@speech.kth.se, Jan-Olof.Svantesson@ling.lu.se, Damrong.Tayanin@ling.lu.se

Abstract

The aim of this study is to investigate how the presence of lexical tones influences the realization of focal accent and sentence intonation. The language studied is Kammu, a language particularly well suited for the study as it has both tonal and non-tonal dialects. The main finding is that lexical tone exerts an influence on both sentence and focal accent in the tonal dialect to such a strong degree that we can postulate a hierarchy where lexical tone is strongest followed by sentence accent, with focal accent exerting the weakest influence on the F0 contour.

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

1. Introduction

The aim of this study is to investigate how the presence of lexical tones influences the realization of focal accent and sentence intonation. For this purpose we investigate realization of focal accent in the context of final and non-final focus in short utterances in tonal and non-tonal dialects of Kammu.

A problem which is often encountered in works on intonation in tone languages is the difficulty in separating tone from intonation, which may lead to uncertainty about what in the intonation contour is the contribution of lexical tones and what is intonation. In many previous studies, this problem has necessitated designing speech material where constraints are placed on the tonal composition such that all tones are present in all positions and the effects of intonation can be factored out by comparing the different tones.

The present study is a part of a larger investigation using a novel approach to this problem by analysing a language that has developed lexical tones rather recently, from the point of view of language history. 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 of this kind. One of the main dialects of this language is a tone language with a tone (high or low) on each syllable, while the other main dialect lacks lexical tones. The dialects differ only marginally in other respects. Tones arose in connection with loss of the contrast between voiced and voiceless initial consonants in the tonal dialect [1].

Generally, tonal phenomena in tone languages are described by analyzing lexical tones and sentence intonation as two separate levels: lexical tones are analysed as local events, superimposed on the intonational contour. The dominating aspect of work on intonation in Chinese is still how lexical tones are changed by phrase intonation, and this is also the case for work on other tone languages, such as Thai [2] and Vietnamese [3]. Fujisaki's model for speech generation, originally developed for Japanese, has recently

been adjusted for application in tone languages, such as Chinese, Thai and Vietnamese [4,5,6].

Xu [7] and Liu and Xu [8] question whether the use of the superpositional approach is satisfactory for tone languages and suggest the PENTA (Parallel encoding and target approximation) model in which tonal contours are decomposed into articulatory and functional components of F0, and mechanisms of integration of lexical and postlexical tonal features are described.

In previous work on Kammu, using material from read [9] as well as spontaneous speech [10], we found that both dialects exhibit a pronounced high sentence final accent which encompasses several functions, such as signaling of utterance boundary, speaker engagement and focusing. In this paper, we use controlled material to examine how focus is realized in final and non-final position in both dialects, and how lexical tones affect its realization.

2. Method

2.1. Speech material

To elicit focus in non-scripted material, four pictures were used. The pictures show a black eagle, a red eagle, a black lizard and a red lizard. Speakers were asked to respond to eight questions about the pictures in which the oppositions eagle-lizard and black-red were systematically varied. The answer was one of the following four sentences, which contain a final noun phrase (noun + adjective), denoting the object on the picture, in which all four combinations of lexical tones occur (the transcription in the left column is for the tonal dialect and in the right column for the non-tonal dialect):

[kii məh kláaŋ jíaŋ] this is eagle black	[gii məh klaaŋ hiaŋ]
---	----------------------

[kii məh kláaŋ jím] this is eagle red	[gii məh klaaŋ jím]
--	---------------------

[kii məh taaŋ jíaŋ] this is lizard black	[gii məh daaŋ hiaŋ]
---	---------------------

[kii məh taaŋ jím] this is lizard red	[gii məh daaŋ jím]
--	--------------------

In each sentence, the place of focus (on the noun or on the adjective) was varied by asking different questions. For example, when showing a picture of a red eagle but asking if this is a black eagle, we expected to get "RED eagle," with focus on "RED," in the answer, but asking if this is a red lizard, we expected to get "red EAGLE," with focus on

“EAGLE,” in the answer. The material thus consists of eight sentences controlled for lexical tones and position of focus.

Each answer was repeated three times by each speaker. 5 non-tonal speakers (all men) and 10 tonal speakers (6 women and 4 men) are included in this study. The speakers ranged in ages from 17 to 57.

2.2. Recording and analysis

The subjects were recorded in Laos and Thailand using 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.

The resulting utterances were checked and transcribed by one of the authors, Damrong Tayanin, who is a native speaker of Kammu.

For acoustic measurements, the *Praat* program was used to display the waveform, spectrogram and fundamental frequency contour of each utterance. Maximum F0 values were annotated manually for successive syllables for each utterance. The measurements were restricted to F0 maxima since we found in a previous investigation [9] that the focal accent is signaled by a high tonal peak.

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 [11]. A fixed semitone scale is used where the unit St is defined by

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

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

Plots 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. The three points are the average F0 maxima in the last three words (the copula *məh* “is” and the two words of the final noun phrase). For the non-tonal speakers, utterances with focus on the final adjective, and those with focus on the pre-final noun are shown as two groups. For the tonal speakers, lexical tones are taken into consideration and all eight possible combinations of lexical tones and focus position are shown separately.

The means of the maximal F0 values (in normalized semitones) for the last two words, the size of the final rise (the difference between the F0 maximum on the final adjective and on the preceding noun) and the “focus effect”, the average difference between the final rise when the final adjective is focused and the rise when the pre-final noun is focused) are given in Table 1 for non-tonal speakers and in Table 2 for tonal speakers.

3.1. The non-tonal dialect

Speakers with a non-tonal dialect have a pronounced high gesture on the final word in all sentences, both with final and non-final focus. The final rise is, however, larger when focus is on the final adjective, 4.5 St, compared to 3.5 St when the pre-final noun is focused. The focus effect, i.e. the difference between the final rises, is thus on the average 1.0 St for the non-tonal speakers.

Table 1. Average maximal tonal values, tonal rise and focus effect for non-tonal speakers (semitones).

focus	maximum on the noun	maximum on the adj.	final rise	focus effect
on adj.	0.2	4.7	4.5	1.0
on noun	0.5	4.0	3.5	

3.2. The tonal dialect

For the tonal speakers, there is a final rise for all tone combinations except for the combination high-low, where there is, on the average, a fall (i.e. a negative rise) when the pre-final noun is focused, and the F0 levels are equal when the final adjective is focused. The focus effect is, however, always positive, i.e. for all tone combinations, the final rise is higher when the final adjective is focused than when the pre-final noun is focused.

It can be observed that for each tone combination, the maximum F0 value is higher on a focused word than on the same word in non-focused position. This holds both for the noun and the adjective, except for the tone combination high-high, where the non-focused final adjective has a higher F0 (by 0.2 St) than the same word in focused position.

It can also be observed from Table 2 and Figure 2 that the identities of the lexical tones are preserved in all positions. Both for the pre-final nouns and for the final adjectives, the average maximal F0 values (in normalized semitones) are always higher for the word with high tone than for that with low tone, independent of the place of focus. For the noun, the gap between the lowest high tone and the highest low tone is 1.2 St and for the adjective it is 0.2 St. When the pre-final noun has low lexical tone, its F0 is lower than on the preceding copula *məh* “is”, but F0 is higher than on the copula when the lexical tone of the noun is high (Figure 2). This can be compared to the non-tonal dialect, where the noun has a slightly higher F0 than the copula (Figure 1).

Table 2. Average maximal tonal values, tonal rise and focus effect for tonal speakers (semitones). l=low lexical tone, h=high tone, L/H=focus.

context	maximum on the noun	maximum on the adj.	final rise	focus effect
l_l	-1.3	1.5	2.8	0.7
L_l	-1.0	1.1	2.1	
h_l	0.6	0.6	0.0	0.9
H_l	1.1	0.2	-0.9	
l_H	-1.8	2.5	4.3	0.5
L_h	-1.4	2.4	3.8	
h_H	0.2	1.7	1.5	0.2
H_h	0.6	1.9	1.3	

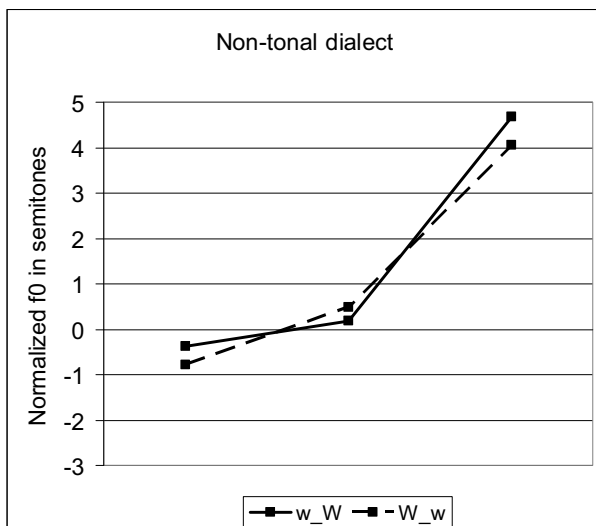


Figure 1: Normalized F0 measurement points for the non-tonal dialect. Measurement points are F0 maxima in the three last words (copula + [noun+adjective]). *w_W* indicates focus on the last word (adjective), and *W_w* indicates non-final focus on the pre-final word (noun).

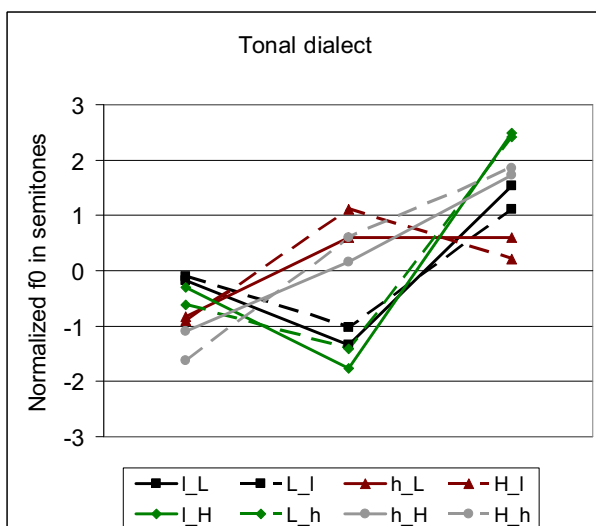


Figure 2: Normalized F0 measurement points for the tonal dialect. Measurement points are F0 maxima in the three last words (copula + [noun+adjective]). Lexical tone and focal status are indicated for the last two words. *l*=low lexical tone, *h*=high tone, *L/H*=focus.

4. Discussion

In the non-tonal dialect, we found a characteristic sentence accent realized as a final rising gesture. This is consistent with the results of our previous study [9]. The rising gesture occurs independently of the place of focus, which does, however, affect the scope of the rise, but only to a limited degree.

This sentence accent is also pervasive in the tonal dialect, but there is a strong influence of lexical tone on its realization. The two lexical tones are clearly separated both in the pre-final noun and in the final adjective, and the intonation

contour is strongly influenced by the tone combination on the two final words in the sentence.

The low-high tone combination is completely compatible with the rising gesture, and in this case sentence intonation is realized as a final rise of a size similar to the rise in the non-tonal dialect.

On the contrary, the high-low tonal configuration is in direct conflict with the rising gesture, which is neutralized and even becomes a fall when the pre-final noun is focussed.

In the case of the high-high combination, the pre-final noun has high lexical tone, and thus gets relatively high F0, limiting the range of realization of the high tone on the final adjective. This may explain the relatively low focus effect of 0.2 (see Table 1).

The realization of the sentence accent for the low-low tone combination is less limited since the low lexical tone on the pre-final noun leads to a relatively low F0 there.

In a previous investigation [9], speakers of the non-tonal dialect showed consistently greater F0 range than the speakers of the tonal dialect. We have speculated that this is due to restrictions imposed by the tonal system comprising two level tones which differ only slightly in absolute F0 [1]. This restricted range is also apparent in the present investigation, and it is noteworthy that these speakers are able to systematically maintain several simultaneous contrasts within this restricted range.

5. Conclusions

The main results of this investigation can be seen to define a prosodic feature hierarchy in both non-tonal and tonal dialects of Kammu in which focal accent is subordinate to a rising sentence final accent. The effects of focal accent are apparent in our material as a raising of the maximum F0 level. This raising is not, however, large enough to alter the basic sentence final rising configuration as clearly seen in the results for the non-tonal dialect, and the non-conflicting tones of the tonal dialect. On the other hand, the marking of lexical tones by F0 in the tonal dialect is the strongest prosodic feature and does indeed alter the basic final rising sentence accent when the tones are in conflict with the basic accentual gesture. Thus we find that lexical tones in Kammu are superordinate to sentence intonation. Instead of variants of sentence intonation affecting the realization of lexical tones as postulated in many analyses of the interaction between tone and intonation, in Kammu we find the reverse: lexical tones affecting sentence intonation. This can possibly be partly explained by the fact that sentence intonation is quite predictable in Kammu even to the point that the sentence final rise overrides focal accent.

By comparing the non-tonal and tonal dialects we have shown the strong influence of lexical tones on both focus and sentence accent. However all three features are still preserved by the speakers of the tonal dialect. The prosodic feature hierarchy (1. lexical tone, 2. sentence accent, 3. focal accent) demonstrates the importance of the lexical tones in production and realization of the pitch contour in the tonal dialect, but at the same time shows that the underlying interplay between sentence accent and focus is the same for both dialects.

These findings can have implications for methods of analysis of the interaction between tone and intonation in other tone languages.

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] Svantesson, J-O. and House, D., “Tone production, tone perception and Kammu tonogenesis”, *Phonology*, 23:309-333, 2006.
- [2] Luksaneeyanawin, S. “Intonation in Thai”, in Hirst, D. and Di Cristo, A [Eds], *Intonation systems*, 376-394, Cambridge: Cambridge University Press, 1998.
- [3] Dô, T.D., Trán, T.H. and Boulakia, G., “Intonation in Vietnamese”, in Hirst, D. and Di Cristo, A [Eds], *Intonation systems*, 395-416, Cambridge: Cambridge University Press, 1998.
- [4] Fujisaki, H., “Information, Prosody, and Modelling – with emphasis on Tonal Features of Speech”, in *Proceedings of international Conference on Speech Prosody*, March 23-26, Nara, Japan, 1-10, 2004.
- [5] Mixdorff, H., “Quantitative Tone and Intonation Modelling across Languages”, in *Proceedings of International Symposium on Tonal Aspects of Languages*, Beijing, 137-142, 2004.
- [6] Seresangtakul, P. and Takara, T., “Study on Pitch Contour of Thai Polysyllabic Tone Sequences Using a Generative Model”, in *Proceedings of international Conference on Speech Prosody*, March 23-26, Nara, Japan, 483-486, 2004.
- [7] Xu, Y., “Transmitting Tone and Intonation Simultaneously: The Parallel Encoding and Target Approximation (PENTA) Model”, in *Proceedings of International Symposium on Tonal Aspects of Languages*, Beijing, 215-220, 2004.
- [8] Liu, F. and Xu, Y., “Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation”, *Phonetica* 62:70-87, 2005.
- [9] House, D., Karlsson, A., Svantesson, J-O. and Tayanin, D., “The Phrase-Final Accent in Kammu: Effects of Tone, Focus and Engagement”, in *Proceedings of Interspeech*, Brighton, UK, 2439-2442, 2009.
- [10] Karlsson, A. M., House, D., Svantesson, J-O. and Tayanin, D., “Prosodic Phrasing in Tonal and Non-tonal Dialects of Kammu”, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1309-1312, 2007.
- [11] Fant, G. and Kruckenberg, A., “Analysis and synthesis of Swedish prosody with outlooks on production and perception”, in Fant, G., Fujisaki, H., Cao, J. and Xu, J., [Eds], *From traditional phonology to modern speech processing*, 73-95, Foreign Language Teaching and Research Press, Beijing, 2004.