

# Prosodic focus marking in Bai

Zenghui Liu<sup>1</sup>, Aoju Chen<sup>1,2</sup> & Hans Van de Velde<sup>1</sup>

Utrecht University<sup>1</sup>, Max Planck Institute for Psycholinguistics<sup>2</sup>

l.z.h.liu@uu.nl, aoju.chen@uu.nl, h.vandevelde@uu.nl

## Abstract

This study investigates prosodic marking of focus in Bai, a Sino-Tibetan language spoken in the Southwest of China, by adopting a semi-spontaneous experimental approach. Our data show that Bai speakers increase the duration of the focused constituent and reduce the duration of the post-focus constituent to encode focus. However, duration is not used in Bai to distinguish focus types differing in size and contrastivity. Further, pitch plays no role in signaling focus and differentiating focus types. The results thus suggest that Bai uses prosody to mark focus, but to a lesser extent, compared to Mandarin Chinese, with which Bai has been in close contact for centuries, and Cantonese, to which Bai is similar in the tonal system, although Bai is similar to Cantonese in its reliance on duration in prosodic focus marking.

**Index Terms:** focus, prosody, duration, pitch, Bai

## 1. Introduction

This is a study of prosodic focus marking in Bai. The Bai language is spoken in the Southwest of China by more than one million people of the Bai ethnic group. It has eight lexical tones from three tonal categories: level (55, 44, 33), rise (35), and fall (42, 21, 32, 31) [1, 2, 3]. The term ‘focus’ refers to the part of a sentence that conveys new information on a topic, following [4] and [5]. Focus can differ in the size of the focused constituent and contrastivity. In terms of size, focus can be on a whole sentence (broad focus), or on a lexical word (narrow focus). If narrow focus also conveys an explicit contrast to alternatives in the context, it is termed as contrastive focus [6]. There is some debate on the position of Bai within the Sino-Tibetan group [2], but this discussion is not crucial for the topic of our paper. Pertinent to the current study is that Bai has been in close contact with Mandarin Chinese for centuries [7].

Pitch is used to encode focus in many non-tone languages [6, 8, 9]. Previous studies have shown that pitch also plays an important role in signaling focus in some tone languages. For example, in Stockholm Swedish, a lexical pitch accent language with two contrasting lexical accents, a separate high tone is added to the lexical accent to mark focus, making pitch relevant for both lexical and post-lexical distinctions [10]. In Mandarin Chinese, a tone language with four lexical tones, pitch is used as a major prosodic cue to realize focus in addition to duration [11]. According to [11], the pitch range at the focus is substantially expanded; the pitch range after the focus is lowered as well as compressed; and the pitch range before the focus does not really deviate much from the neutral focus condition. In addition to that, in Mandarin Chinese, syllable duration increases significantly under focus, regardless of the position of the syllable in the utterance. The data presented in [11] reveals that the pre-focus constituents undergo little change in pitch and duration compared to the same constituents in focus in Mandarin. The same is true for Vietnamese, a tone language with six lexical tones [12, 13]. However, other tone languages do not use pitch to mark focus. For example, in Cantonese, a language with six lexical tones, pitch variation is not systematically modified to mark focus. According to [13], duration and intensity are the main acoustic correlates of focus in Cantonese; both are increased significantly in the on-focus words in any word location for all

lexical tones. Besides, no decrease in mean pitch range is found in the post-focus words. In Yucatec Maya, a language with two lexical tones, pitch is only used at the lexical level and focus is not prosodically encoded [15, 16, 17].

Such differences in prosodic focus marking among tone languages suggest that which cues are used to what extent can vary from language to language and is not related to the total number of lexical tones in a language. Against this background, we investigate how pitch and duration may be used to mark focus in the southern variety of Bai by adopting a semi-spontaneous experimental approach. The southern dialect of Bai is chosen because it is well studied at the segmental and lexical level compared to other varieties of Bai [1, 2, 3, 18, 19]. The southern variety of Bai is hereafter referred to as Bai.

## 2. Methodology

### 2.1. Experimental materials

The production experiment aimed to elicit SVO sentences in five focus conditions: narrow-focus on the subject NP in sentence-initial position (NF-i), narrow-focus on the verb in sentence-medial position (NF-m), narrow-focus on the object NP in sentence-final position (NF-f), broad focus (BF) and contrastive-focus on the verb in sentence-medial position (CF-m). The focus condition was set up by a WH-question or a statement from the experimenter, as illustrated in examples (1) to (5), where focused constituents appear in square brackets.

*Target sentence:* ɕo<sup>42</sup> tu<sup>21</sup> ku<sup>21</sup> tsu<sup>33</sup>.

*Bear one (quantifier) sell tree.*

*The bear sells the tree.*

(1) *Experimenter:* Look! The tree. There is also a price label. It seems someone sells the tree. Who sells the tree?

*Participant:* [THE BEAR] sells the tree. (NF-i)

(2) *Experimenter:* Look! The bear and the tree. It seems like that the bear does something with the tree. What does the bear do with the tree?

*Participant:* The bear [SELLS] the tree. (NF-m)

(3) *Experimenter:* Look! The bear, it stands behind a shelf. It seems like that the bear sells something. What does the bear sell?

*Participant:* The bear sells [THE TREE]. (NF-f)

(4) *Experimenter:* Look! This picture is very blurring. I can't see anything clearly. What has been depicted in the picture?

*Participant:* [THE BEAR SELLS THE TREE]. (BF)

(5) *Experimenter:* Look! The bear and the tree. It seems like that the bear does something with the tree. I guess the bear wipes the tree.

*Participant:* The bear [SELLS] the tree. (CF-m)

Each focus condition was realized in 30 SVO sentences. The lexical tones of verbs were strictly controlled. The verbs were the items for acoustic and statistical analysis, as they can have multiple roles played in the five focus conditions. For example, the verb could be a focused constituent in the BF, NF-m and CF-m conditions; it can also be a pre-focus constituent in the NF-f condition and a post-focus constituent in the NF-i condition. The property of the verbs in the present setting thus provided us with an opportunity to investigate effects of both focus (focused vs. unfocused) and focus type (narrow focus vs. broad focus vs. contrastive focus) on pitch and duration in Bai.

In order to keep the experiment within a feasible length, three lexical tones were included, representing the three tonal categories existing in Bai: level, falling, and rising tones. Tone in Bai can be considered as a complex combination of pitch, phonation type, and degree of tenseness [2]. The present study has selected lexical tones that were well spread over the tonal space of Bai's tone system. Specifically, 55 was selected as a representation of level tones, 21 as a representation of falling tones, and 35 as the rising tone.

The target sentences were constructed in such a way that each was a unique combination of a subject-noun and a VP (verb + object-noun). Six verbs were included, two in each tonal category. In Bai, the noun needs to be followed by a quantifier to form an NP as a subject in a sentence, but the quantifier of the NP can be omitted when the NP is an object in a sentence [20]. Four subject-nouns were selected, which followed by a same low fall-tone quantifier in all the target sentences. Four level-tone object-nouns were selected. The six verbs and four object nouns formed 24 VPs, each of which appeared in each focus condition. This gave us 120 VPs. The subject nouns were evenly distributed over the 120 VPs to form 120 target sentences. To make sure that the duration of the experiment was manageable and reasonable for the participants, the 120 target sentences were split into two lists. Each list contained all the five focus conditions realized on different sentences, and all the six representations of the tones, but only half of the V+O combinations. This results in 60 items per list and participant.

## 2.2. Data elicitation

In the picture-matching game, three piles of pictures were used: the experimenter and the participant each held a pile of pictures ordered in a certain sequence; the third pile of pictures were scattered around on a table. In the experimenter's pictures (the first pile), there was always something missing, like a subject, an action (verb) or an object. The participant's pictures (the second pile) all contained a complete event. The participant's task was to help the experimenter with sorting out pictures from her own pile and the third pile that went together. Here is a detailed example of a trial eliciting a target sentence in the NF-i condition: First, the experimenter took a picture (e.g. a tree) from her own pile, drew the participant's attention to the picture and established what the picture was by saying, e.g. *"Look! The tree! There is also a price label. It seems like that someone sells the tree."* This was done to make sure that the entity in the picture was referentially given to the participant before the utterance of the question. Second, the experimenter asked a question about the picture (e.g. *"Who sells the tree?"*). Third, the participant took a complete picture from his or her pile and looked at it. The experimenter then repeated the question, followed by an answer from the participant (e.g. *"[THE BEAR] sells the tree."*). Fourth, the experimenter found the picture containing the missing information in the third pile and pairs it up with her own picture. The participants were explicitly instructed (1) to respond in full sentences and (2) not to show their own pictures to the experimenter. Prior to the picture-matching game proper, the experimenter conducted six practice trials with the participant to familiarize him or her with the game.

In order to ensure the consistency in the participants' word choice, the picture-matching game was preceded by a picture-naming task, which was designed to familiarize the

participants with the target words and the entities in the pictures used in the game.

## 2.3. Participants and procedure

Five native speakers of Bai (four male and one female, aged between 23 and 25) took part in the experiment. The participants all met the following criteria: (1) using Bai on a daily basis with self-estimated daily use exceeding 60%; (2) not having lived outside the Bai speaking community for the past 10 years; (3) not having used Chinese or other languages for a long period on a daily basis; (4) having no self-reported speech and hearing impairments.

Every participant was randomly assigned to one of the two lists. The game lasted 20 to 25 minutes per participant. The participants were tested individually by a female experimenter, who was a native speaker of Bai, in a quiet room in a villager's private home. The experiments were recorded using a portable ZOOM H1 digital recorder at a 44.1 kHz sampling rate and 16 bit accuracy. Each session was also video-taped.

## 3. Analysis and Results

### 3.1. Analysis

The auditory recordings from each participant were first orthographically annotated so that the participant's responses could be selected. A strict selection criterion of the usable data was applied, i.e. a sentence was considered usable only if it contained no self-correction and hesitation and was uttered as a response to the target question. In total 80.3% of the obtained responses (N=241) were included in further analysis. The usable sentences were subsequently acoustically annotated in Praat [21]. A textgrid with four interval tiers (word, tone, sentence, comment), and two point tiers (pitch, duration) was created for each target sentence. Every sentence was segmented into words in the 'word' tier, then landmarks demarcating verb onset and offset, and the locations of pitch-maximum and pitch-minimum within the verb were added to the 'duration' and 'pitch' tiers. The landmarks for the onset and offset of verbs were determined according to the information in the waveform and spectrogram.

The pitch values of the pitch landmarks and the time values of the word boundaries were subsequently extracted via Praat scripts. Two measures from these values were calculated: word duration (i.e. offset time minus onset time) and pitch range (i.e. the difference between the maximum pitch and the minimum pitch). In 55 of the usable responses, the pitch values could not be reliably measured. These responses were thus excluded from the analysis on pitch range.

In order to investigate how focus is prosodically realized in Bai, several analyses were done to find out the effect of focus (focused vs. unfocused), focus type, and the interaction between these variables and the tone of the verb on the duration and pitch range of the verbs. To find out the effect of focus and its interaction with tone, we compared the duration and pitch range of the focused verbs with these measures of the verbs in the unfocused conditions: NF-m (focus) vs. NF-i (post-focus); NF-m (focus) vs. NF-f (pre-focus). To find out the effect of focus type and its interaction with tone, we compared the verbs in the NF-m condition with the same verbs in the BF and CF-m conditions.

## 3.2. Results

### 3.2.1. Duration

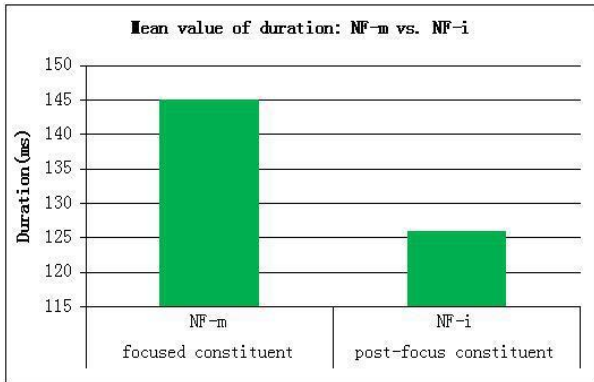


Figure 1: Mean duration (in ms) of verbs in focus vs. non-focus position. (NF-m=verb in sentence medial focused position, NF-i=verb preceding a focused constituent)

The duration data obtained from the verbs in the NF-m and NF-i conditions showed that the verbs were on average 19.1 ms longer when focused (NF-m) than when not focused and following a focused constituent (NF-i). To assess the effect of focus (focused vs. unfocused) on the duration of the verbs, we built a mix-effect model with ‘focus’ as the fixed factor (independent variable), ‘speaker’ and ‘verb’ as the random factors; and another mix-effect model with only the random factors. The variation between speakers was corrected in the models per focus condition. A statistically significant difference between the two models was taken as the evidence for a main effect of the fixed factor at issue. In the comparison between NF-m and NF-i, the model including the fixed factor differed significantly from the model with only the random factors ( $p < 0.01$ ). This indicated that the speakers used duration to distinguish focus from non-focus when the unfocused verb was in post-focus position. Furthermore, we also used mixed-effect-modeling to assess the effect of the interaction between ‘focus’ and ‘lexical tones of verbs’. The model involving the interaction did not differ significantly ( $p = 0.31$ ) from the model without the interaction. Thus, the speakers used duration to distinguish focus from non-focus regardless of the tonal category of the verbs.

The duration data obtained from the verbs in the NF-m and NF-f conditions showed that the verbs were on average 12.5 ms longer when focused (NF-m) than when not focused and preceding a focused constituent (NF-f). Mixed-effect modeling was used to assess the effect of ‘focus’ on the duration of the verbs, as described above. It did not reveal a main effect of ‘focus’ ( $p = 0.13$ ). Furthermore, the model involving the interaction did not differ significantly ( $p = 0.31$ ) from the model without the interaction. This indicated that there was no main effect of ‘focus’, i.e. the speakers did not use duration to distinguish focus from non-focus (i.e. pre-focus in NF-f). This suggested that the duration in the pre-focus constituents hardly changed relative to the constituents in focus.

With regard to the effect of focus type, mixed-effect modeling revealed no main effect ( $p = 0.16$ ) of ‘focus type’ (referring to the three types of focus). Further, there was no interaction between ‘focus type’ and ‘lexical tones of verbs’ ( $p = 0.49$ ). Thus, duration was not used to differentiate the three focus types regardless of the tones of the verbs. To find out

whether duration was used to distinguish NF-m and CF-m, two focus types with a smaller focus-constituent size, from BF, we grouped NF-m and CF-m and built new models. The models showed that NF-m and CF-m did not differ from BF in the duration of the verbs regardless of the tone of the verbs ( $p = 0.33$ ). Finally, models were built to see whether NF-m and CF-m could differ in duration. Again we found no significant difference in duration.

### 3.2.2. Pitch range

The mean pitch ranges hardly differed across conditions, i.e. 8.78Hz in the BF condition, 7.99Hz in the CF-m condition, 7.89Hz in the NF-f condition, 10.78Hz in the NF-i condition, and 9.12Hz in the NF-m condition. Mix-effect-modeling confirmed that pitch range was not used in any way in focus marking in Bai.

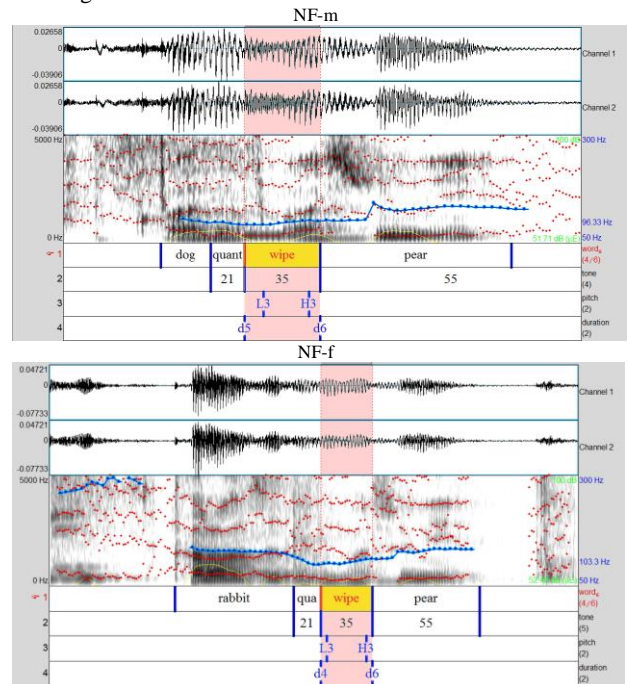


Figure 2: Pitch contour (in Hz) of verbs in focus vs. non-focus position. (NF-m=verb in sentence medial focused position, NF-f=verb following a focused constituent)

## 4. Discussion and Conclusions

The present study shows that speakers of the southern variety of Bai increase the duration of the focused constituent and reduce the duration of the post-focus constituent to encode focus, similar to speakers of Mandarin Chinese [11] and Cantonese [14]. Further, they do not vary the duration of a pre-focus constituent compared to the same constituent in a focused position, again similar to speakers of Mandarin Chinese and possibly Cantonese. However, they do not use pitch variation in any way in focus marking, different from speakers of Mandarin Chinese but similar to speakers of Cantonese. In addition, speakers of this variety of Bai do not use duration to distinguish focus types differing in size and contrastivity, different from speakers of Mandarin Chinese [11] and Cantonese [14], who use duration to distinguish focus types differing in the size of the focused constituent. These results suggest that Bai uses duration in prosodic focus-marking to a lesser extent than Mandarin Chinese and

Cantonese. Related to this is the fact that Bai also exploits word order and morphological topic marker to distinguish focal information from non-focal or topical information. Specifically, the canonical word order in Bai is SVO, the word order OSV can be used to highlight the topic status of the object. Further, the topical status of a subject can be optionally marked by topic markers, such as ‘nu<sup>55</sup>’ and ‘lu<sup>44</sup>’ [19]. The use of these non-prosodic cues may explain the modest use of prosody in focus marking in Bai. Furthermore, the results add to the existing findings on prosodic focus-marking in tone languages and show that there is no relationship between the cues used to mark focus prosodically and the number of lexical tones in a language.

Our study suggest two topics for future research. Bai has been in close contact with Mandarin Chinese for centuries [7], which has led to a large number of Chinese-loan words in Bai, and deep influence from Chinese syntactic structure on the syntax of Bai [2, 19]. However, in spite of the lexical and syntactic influence from Mandarin Chinese, Bai is by and large more similar to Cantonese in prosodic focus-marking in that both Bai and Cantonese, duration is the major prosodic cue rather than pitch. This puts forward an interesting hypothesis for future research. That is, prosodic focus-marking may not easily undergo changes as a result of language contact. Further, considering the dialectal differences in Bai language, it is not clear whether prosodic focus marking is similar across Bai dialects. Future research on the northern variety of Bai can shed light on this question.

## 5. Acknowledgements

We are grateful to Wenju He for collecting data in Xizhou County, and to the participants in the village. We also thank Liquan Yang and Yanzen Zhao for their feedback on Bai, Paula Cox, Anqi Yang, Anna Sara Romøren for their help. This study is supported by a scholarship grant from the Chinese Scholarship Council to the first author and a VIDI grant (276-89-001) from the Netherlands Organisation for Scientific Research to the second author.

## 6. References

- [1] 赵衍荪, 徐琳, 白语, 中国社会科学院, & 民族研究所. (1996). 白汉词典 四川民族出版社.  
(Zhao, Xu, Bai-Chinese dictionary, 1996)
- [2] Allen, B. (2007). Bai dialect survey. SIL International,
- [3] 艾磊, 苏玮雅, & 尹曼芬. (1997). 白语喜洲镇话声调的测试分析. 大理学院学报 (社会科学版), 2, 011.  
(Allen, Su and Yin, The experimental study on Xizhou Bai lexical tones, 1997)
- [4] Vallduví, E., & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics*, 34(3), 459-520.
- [5] Lambrecht, K. (1996). Information structure and sentence form: Topic, focus, and the mental representations of discourse referents Cambridge University Press.
- [6] Gussenhoven, C. (2004). The phonology of tone and intonation Cambridge University Press.
- [7] Hefright, B. E. (2011). Language Contact as Bilingual Contrast among Bǐi Language Users in Jiànchūn County, China,
- [8] Ladd, D. R. (2008). Intonational phonology Cambridge University Press.
- [9] 王蓓, 吐尔逊, 卡得, & 许毅. (2013). 维吾尔语焦点的韵律实现及感知. 声学学报, 38(1), 92-98.  
(Wang, Tursun and Xu. Prosodic encoding and perception of focus in Uyghur, 2013)
- [10] Bruce, G. (1982). Textual aspects of prosody in Swedish. *Phonetica*, 39(4-5), 274-287.
- [11] Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, 27(1), 55-105.
- [12] Jannedy, S. (2007). Prosodic focus in vietnamese. *Interdisciplinary Studies on Information Structure*, 8, 209-230.
- [13] Jannedy, S. (2008). The effect of focus on lexical tones in vietnamese. *Experimental Linguistics ExLing* 2008, , 113.
- [14] Wu, W. L., & Xu, Y. (2010). Prosodic focus in hong kong cantonese without post-focus compression. *Speech Prosody* 2010,
- [15] Kügler, F., & Skopeteas, S. (2007). On the universality of prosodic reflexes of contrast: The case of yucatec maya.
- [16] Gussenhoven, C. (2006). Yucatec maya tone in sentence perspective. Poster Presented at LabPhon10, Paris,
- [17] Gussenhoven, C., & Teeuw, R. (2008). A moraic and a syllabic H-tone in yucatec maya. *Fonología Instrumental: Patronos Fónicos y Variación*, , 49-71.
- [18] 邓瑶, & 何稳菊. (2012). 云南大理喜洲白族居民语言生活调查. 民族翻译, 3, 017.  
(Den and He. The language attitude survey on Bai in Xizhou, Dali. 2012)
- [19] 徐琳. (2008). 大理丛书·白语篇.  
(Xu, The Dali series. Bai language. 2008)
- [20] 赵燕珍. (2009). 赵庄白语参考语法. Doctoral dissertation. (Zhao. Zhaozhuan Bai reference grammar. 2009)
- [21] Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341-345.

