



Intonational Realization of Declarative Questions in Bai

Zenghui Liu^{1,2}, Hans Van de Velde³, Aoju Chen¹,

¹Utrecht University, Netherlands

²Yunnan University, China

³Fryske University, Netherlands

Zenghui.liu@hotmail.com, aoju.chen@uu.nl, HVandeVelde@fryske-akademy.nl

Abstract

This study investigates intonational realization of declarative questions in Southern Bai, a Sino-Tibetan tone language spoken in the southwest of China, by using a semi-spontaneous experimental approach. Our data shows that Bai speakers use prosody by lengthening the duration, expanding the pitch span, and raising the pitch maximum and minimum of the sentence-middle constituents to distinguish declarative questions from statements, regardless of focus condition. However, they vary the use of pitch span and pitch minimum with different lexical tones. These results thus suggest that sentence-medial prosody is different between statements and declarative questions in Bai. Furthermore, the modification of prosodic cues for encoding interrogativity appears to be sensitive to lexical tones in Bai.

Index Terms: intonation, declarative question, Bai

1. Introduction

Despite much research, there is still no consensus on whether the prosodic difference between statements and questions is limited in sentence-final position [1, 2, 3, 4, 5, and 6]. For example, it has been suggested that intonational realization of questions is characterized by an increase in pitch register (i.e., mean pitch over the whole sentence) [1, 6]. However, many studies have provided evidence that the intonational difference between questions and statements is limited in sentence-final position [2, 4, 5, and 7]. These conflict findings, especially in studies that concern the same language, might be due to the fact that the information structure (e.g., focus) and different types of questions (e.g., declarative question, wh-question, and yes/no question) were not controlled for in the previous examinations.

Previous studies have shown intonational realization of questions can differ in different focus conditions and this can in turn differ in different languages [1, 8, 9, and 10]. For example, [8] reported an interaction between focus and question intonation in the form of a boost to the pitch rising by the question starting from the focused word in Mandarin. However, [9] presented evidence that the pitch difference between statements and declarative questions becomes salient starting from the stressed syllable of the first content word, whether or not it is focused in English. Further, question intonation can vary the local pitch target. For example, studies of non-tone languages, such as English [9], showed that the underlying pitch targets of the stressed syllables change from high or falling in statements to rising in question. For tone

languages, question intonation can vary with lexical tones [3, 11].

The questions that arise for Bai are thus (1) whether and how declarative questions are prosodically distinguished from statements sentence-medially; (2) whether and how the declarative question intonation is affected by focus; (3) whether and how the declarative question intonation is affected by tone. We addressed these questions by examining the production of six native speakers of the southern variety of Bai, a well-studied variety of Bai for lexicon and lexical tones [12, 13]. The southern variety of Bai is hereafter referred to as Bai.

2. Methodology

2.1. Picture-marking game

We developed a picture-marking game to elicit semi-spontaneously produced declarative questions in different focus conditions. In the game, the participant needed to seek information of various scopes from the experimenter by asking a declarative question in each trial. In the game, two piles of pictures and a marker were used: the first pile of pictures ordered in a certain sequence was put in front of the experimenter and participant; the second pile of pictures was stored in a box and put next to the experimenter; the marker was held by the participant. In the first pile of pictures, there was always information missing, like a subject, an action (verb), an object, or all three. In the top right corner of each picture of the first pile, a small picture was covered by a sticker. This small picture provided the missing part which could be correct or incorrect. The second pile of pictures (stored in the box) all contained a complete and correct event. The participant's task was to mark the correctness of the small picture on the sticker. Here is a detailed example of a trial eliciting a target declarative question in the NF-i condition (narrow-focus on the subject NP in sentence-initial position): First, the experimenter and the participant looked at a picture of the first pile (e.g., a tree and a price label) together. The experimenter 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 looks like someone sells the tree." This was done to make sure that the entity in the picture was referentially given to the participant. Second, the experimenter directed the participant to take a look at a small picture covered by a sticker (e.g. "Could you open the sticker and take a look at the small picture? Then you can ask me a question, I will help you to check in my box."). Third, the participant took a look at the small picture covered by the

sticker and asked a question to seek confirmation (e.g. “[THE CAT] sells the tree?”). Fourth, the experimenter checked the pictures stored in the box (second pile) and gave a response (e.g. “Yes. The small picture is correct. Now you can make a mark.”). Finally, the participant could make a mark on the sticker to indicate the correctness of the small picture.

2.2. Experimental materials

Declarative questions in four focus conditions were elicited via the picture-marking game: 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), and broad focus (BF). The focus condition was set up by varied context, as illustrated in examples (1) to (5), where the focal constituent appears in square brackets. Each focus condition was realized in 24 SVO sentences.

(1) *Experimenter: Look! The tree. There is also a price label. It looks like someone sells the tree. Could you open the sticker and take a look at the small picture? Then you can ask me a question, I will help you to check in my box.*

Participant: [THE CAT] sells the tree?

a⁵⁵ mi⁵⁵ tw²¹ kw²¹ tsu^{33?}
Cat one (quantifier) sell tree?
 (NF-i: Narrow focus on the subject)

(2) *Experimenter: Look! The rabbit and the tree. It looks like the rabbit does something to the tree. Could you open the sticker and take a look at the small picture? Then you can ask me a question, I will help you to check in my box.*

Participant: The rabbit [SELLS] the tree?

th⁵⁵ lo⁵⁵ tw²¹ kw²¹ tsu^{33?}
Rabbit one (quantifier) sell tree?
 (NF-m: Narrow focus on the verb)

(3) *Experimenter: Look! The bear, it stands behind a shelf. It looks like the bear sells something. Could you open the sticker and take a look at the small picture? Then you can ask me a question, I will help you to check in my box.*

Participant: The bear sells [THE TREE]?

co⁴² tu²¹ kw²¹ tsu^{33?}
Bear one (quantifier) sell tree?
 (NF-f: Narrow focus on the object)

(4) *Experimenter: Look! This picture is very blurry. I can't see anything clearly. Could you open the sticker and take a look at the small picture? Then you can ask me a question, I will help you to check in my box.*

Participant: [THE CAT SELLS THE TREE]?

a⁵⁵ mi⁵⁵ tw²¹ kw²¹ tsu^{33?}
Cat one (quantifier) sell tree?
 (BF: Broad focus)

In order to keep the experiment within a feasible length, we included one lexical tone for each of the three tonal types: level tones (Tones 55, 44, and 33), falling tones (Tones 42, 31, 32, and 21) and the rising tone (Tone 35). The included lexical tones were well spread over the tonal space of Bai's tone

system: Tone 55 represented the level tones; Tone 21 represented the falling tones, and the rising tone: Tone 35.

The SVO 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, a noun needs to be followed by a quantifier to form a noun phrase as a subject of a sentence, but the quantifier of the noun phrase can be omitted when the noun phrase is an object of a sentence [14]. Four subject-nouns were selected, which were followed by the same low falling-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 resulted in 96 VPs, and then four subject nouns were evenly distributed over these 96 VPs to form 96 SVO target sentences. These 96 target sentences were split into two lists, and each list contained all the four focus conditions realized in different sentences, and all the six representations of the tones, but only half of the V+O combinations. This resulted in 48 sentences per list and participant.

2.3. Participants and procedure

Six (near-) monolingual speakers of Southern Bai (age range: 23 to 24 years; mean age: 23.67 years old; SD = 1.03; five male and one female) participated in our study. They were recruited from villages in Xizhou County, Dali Bai Autonomous Prefecture, China. Although all the participants learned Standard Mandarin and/or Dali Mandarin as their second language, they predominantly used Bai in their daily lives. The participants all met the following criteria: (1) acquire Bai as the first language from birth; (2) using Bai on a daily basis with self-estimated daily use exceeding 60%; (3) not having lived outside the Bai-speaking community over the last ten years at the time of testing; (4) not having actively used Mandarin or other languages for a long period on a daily basis; and (5) having no self-reported speech and hearing impairments. They were all paid a small fee for their participation. Participants were randomly assigned to one of the two lists and tested individually by a female experimenter, who was a native speaker of Southern Bai (age = 26). The game lasted 20 to 25 minutes per participant. All the test sessions were conducted in Bai, in a quiet room at a villager's private home in Jinguisi Village, Xizhou County. 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-recorded for future training purpose.

2.4. Acoustic annotation

The auditory recordings from the participants were first orthographically annotated. An elicited question was discarded for analysis if: (1) it contained self-corrections; (2) it contained hesitations (defined as a long “em” sound produced by the participants before asking questions); (3) it deviated from the target question in word choice or word order; (4) it was severely overlapped with background noise coming from the farm. In total 87% of the obtained responses (n = 251) were included for further analysis.

The verbs were acoustically annotated by examining the waveform, wide-band spectrum, and pitch track in Praat [15] in combination with auditory impressions [16]. Two pitch-related landmarks and two segmental landmarks were labelled

in each verb: pitch maximum, pitch minimum, word onset, and word offset.

The pitch values (in Hz) at the pitch-related landmarks and time values (in seconds) at the segmental landmarks were automatically extracted by means of Praat scripts. Four measures were obtained from each verb: pitch maximum, pitch minimum, pitch span (i.e., the difference between the maximum pitch and the minimum pitch), and word duration. In 45 of the usable responses (18%), an accurate measurement of pitch values was not possible. These responses were thus excluded from the analysis of pitch-related measurements. The verbs were later compared to the verbs in lexically identical statements produced by same speakers from [17] ($n = 227$).

3. Statistical analyses and results

3.1. Statistical analyses

Statistical analyses were conducted using mixed-effects modelling in R [18]. In all models, the sentence type, tone, and focus condition were included as fixed factors, while the speaker (i.e., the participants) and sentence (i.e., the elicited declarative questions) were included as random factors. Sentence type had two levels (i.e., “statement” and “question”), focus condition had four levels (i.e., “NF-i”, “NF-f”, “NF-m” and “BF”), and tone referred to the lexical tones of the target verbs, which had three levels (i.e., “Level”, “Rise”, and “Fall”). Outcome variables were the duration, pitch span, pitch maximum, and pitch minimum of the verbs. Following [19], our models were constructed and evaluated in a stepwise fashion. When building the models, only the factors and interactions that significantly improved the fit of the model were retained until the best fit model was determined.

3.2. Pitch span

For the analysis of pitch span, we found that the best-fit model was the model contained the main effects of sentence type, $\chi^2(1) = 23.902$, $p < .001$, and tone, $\chi^2(2) = 65.629$, $p < .001$, and a two-way interaction between sentence type and tone, $\chi^2(2) = 16.984$, $p < .001$. By further exploring the two-way interaction between sentence type and tone, we found that the pitch span of the verbs was wider in rising question than their counterparts in statements, but not in level tone ($p = .247$), regardless of focus condition, as shown in Figure 1.

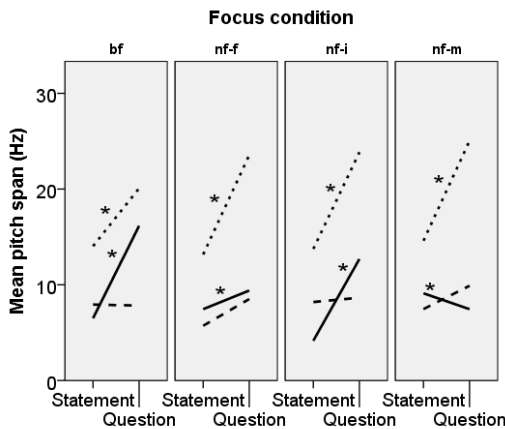


Figure 1: Mean pitch span of verbs in statement vs. question in three lexical tones and four focus conditions, $n = 6$, $N = 396$. Significant differences are marked with an asterisk.

3.3. Pitch maximum

For the analysis of pitch maximum, we found that the best-fit model was the model contained the main effects of sentence type, $\chi^2(1) = 83.46$, $p < .001$, and tone, $\chi^2(2) = 102.94$, $p < .001$. The main effect of sentence type was such that the pitch maximum of the verbs was significantly higher in declarative questions than in statements (164.2 Hz vs. 141.1 Hz, $p < .001$), regardless of tone and focus condition, as shown in Figure 2.

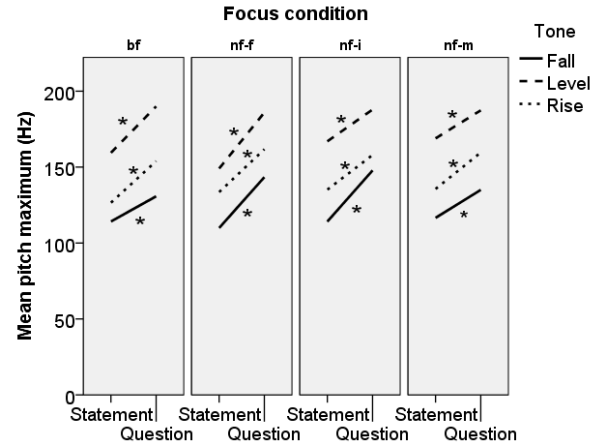


Figure 2: Mean pitch maximum of verbs in statement vs. question in three lexical tones and four focus conditions, $n = 6$, $N = 396$. Significant differences are marked with an asterisk.

3.4. Pitch minimum

For the analysis of pitch minimum, we found that the best-fit model was the model contained the main effects of sentence type, $\chi^2(1) = 44.044$, $p < .001$, and tone, $\chi^2(2) = 123.22$, $p < .001$, and a two-way interaction between sentence type and tone, $\chi^2(2) = 14.348$, $p < .001$. By further exploring the two-way interaction between sentence type and tone, we found that the pitch minimum of the verbs in declarative question was higher in level tone ($p < .001$) and rising tone ($p < .05$) than in statements, but not in falling tone ($p = .227$), regardless of focus condition, as shown in Figure 3.

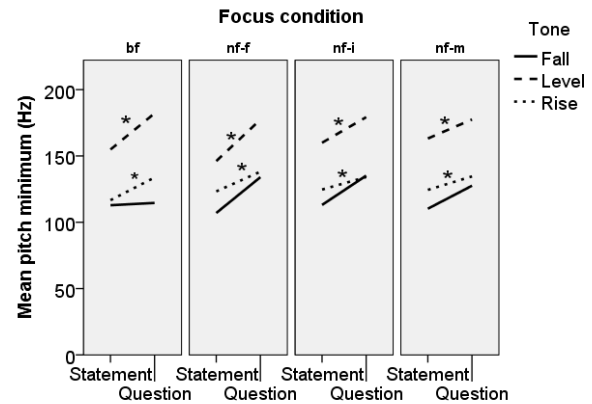


Figure 3: Mean pitch minimum of verbs in statement vs. question in three lexical tones and four focus conditions, $n = 6$, $N = 396$. Significant differences are marked with an asterisk.

3.5. Duration

For the analysis of duration, we found that the best-fit model was the model contained the main effects of sentence type, $\chi^2(1) = 11.589$, $p < .001$, and tone, $\chi^2(2) = 84.104$, $p < .001$. The main effect of sentence type was such that the duration of the verbs was significantly longer in declarative questions than in statements (167.0 ms vs. 151.5 ms, $p < .001$), regardless of tone and focus condition, as shown in Figure 4.

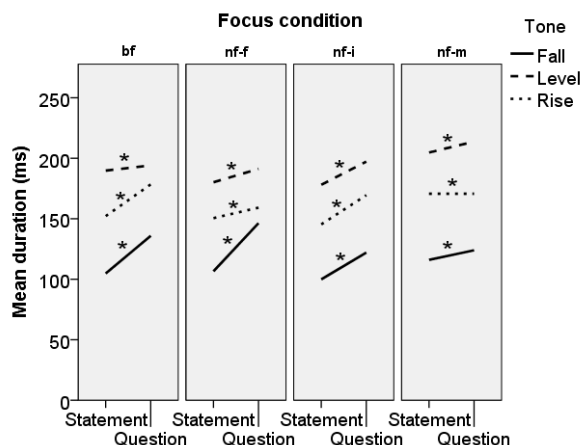


Figure 4: Mean duration of verbs in statement vs. question in three lexical tones and four focus conditions, $n = 6$, $N = 478$. Significant differences are marked with an asterisk.

4. Discussion and conclusion

Our results showed that Bai speakers lengthened the duration of the sentence-medial constituent in declarative questions relative to its counterpart in statements, regardless of focus condition and tone. Further, they raised the pitch maximum of the sentence-medial constituent in declarative questions relative to its counterpart in statements, regardless of focus condition and tone. However, they only expanded the pitch span of the sentence-medial constituent in rising and falling tones and raised the pitch minimum of the sentence-medial constituent in level and rising tones for distinguishing declarative questions from statements.

The present study shows that the prosodic difference between declarative questions and statements is present in sentence-medial position, similar to findings from English [9], Mandarin [6], Dutch [1] and Danish [7]. However, we did not find evidence that focus interacts with question intonation in Bai, different from studies of Mandarin [8]. Related to this is the fact that Bai only exploits duration to distinguish focal from post-focal information [17]. The present study thus provides evidence that the use of prosodic cues for encoding different linguistic components, such as encoding information structure and conveying interrogative meaning, can be independent from each other. Further, the present study suggests that the modification of prosodic cues for encoding declarative questions varies with lexical tones in Bai. Specifically, for level tone, Bai speakers raise the pitch maximum and minimum for realizing declarative question.

However, for falling tone, Bai speakers expand the pitch span by raising the pitch maximum without lowering the pitch minimum for realizing declarative question. This might be explained by that expanding the pitch span can lead to the change of level tone's tonal identity. Similarly, lowering the pitch minimum of Tone 21 (the falling tone we examined in the present study) might run the risk of changing the tonal identity of the falling tone, as there are four falling tones (i.e., Tones 42, 32, 31, and 21) in Bai.

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 Yanzhen Zhao for their feedback on Bai.

6. References

- [1] J. Haan, "Speaking of questions: A exploration of Dutch question intonation," Doctoral dissertation, Radboud University, Nijmegen, 2002.
- [2] D. R. Ladd, *Intonational phonology*. Cambridge: Cambridge University Press, 1996.
- [3] W.-S. Lee, "The effect of intonation on the citation tones in Cantonese," in *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, 2004.
- [4] M. Lin, "On production and perception of boundary tone in Chinese intonation," in *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, 2004.
- [5] J. B. Pierrehumbert, "The phonology and phonetics of English intonation," Doctoral dissertation, Massachusetts Institute of Technology, 1980.
- [6] J. Yuan, C. Shih, and G. P. Kochanski, "Comparison of declarative and interrogative intonation in Chinese," in *Speech Prosody 2002*, Aix-en-Provence, France, 2002.
- [7] N. G. Thorsen, "A study of the perception of sentence intonation—Evidence from Danish," *The Journal of the Acoustical Society of America*, vol. 67, no. 3, pp. 1014–1030, 1980.
- [8] F. Liu and Y. Xu, "Parallel encoding of focus and interrogative meaning in Mandarin intonation," *Phonetica*, vol. 62, no. 2–4, pp. 70–87, 2005.
- [9] F. Liu and Y. Xu, "Question intonation as affected by word stress and focus in English," in *Proceedings of the 16th international congress of phonetic sciences*, Saarbrücken, Germany, 2007, pp. 1189–1192.
- [10] S. J. Eady and W. E. Cooper, "Speech intonation and focus location in matched statements and questions," *The Journal of the Acoustical Society of America*, vol. 80, no. 2, pp. 402–415, 1986.
- [11] N. T. Chang, "Tones and intonation in the Chengtu dialect (Szechuan, China)," *Phonetica*, vol. 2, no. 1–2, pp. 59–85, 1958.
- [12] B. Allen, W. (苏玮雅) Sū, and M. (尹曼芬) Yīn, "Báiyǔ xīzhōuzhèn shēngdiào de cèshì fēnxī [The experimental study on Xizhou Bai lexical tones, 白语喜洲镇话声调的测试分析]," *Dàlǐ xuéyuàn xuébào [Journal of Dali University: general edition, 大理学院学报: 综合版]*, vol. 2, pp. 54–58, 1997.
- [13] Y. (赵衍荪) Zhào and L. (徐琳) Xú, *Bái-hàn cídiǎn [Bai-Chinese dictionary, 《白汉词典》]*. Chengdu: Sichuan minzú chūbǎn shè [Sichuan Minzu Press, 四川民族出版社], 1996.
- [14] Y. Zhào, "Zhàozhuān Báiyǔ cānkǎo yǔfǎ [Zhaozhuan Bai reference grammar, 《赵庄白语参考语法》]," Doctoral dissertation, Minzu University of China, 2009.
- [15] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer*. 2006.
- [16] A. Turk, S. Nakai, and M. Sugahara, "Acoustic segment durations in prosodic research: A practical guide," in *Methods in empirical prosody research*, S. Sudhoff, Ed. Berlin: Mouton de Gruyter, 2006, pp. 1–28.

- [17] L. Z. H. Liu, A. Chen, and H. Van de Velde, "Prosodic focus marking in Bai," in *Speech prosody*, 2014, vol. 7, pp. 628–631.
- [18] R Core Team, "R: A language and environment for statistical computing," *R Foundation for Statistical Computing, Vienna, Austria*, 2014.
- [19] A. Field, J. Miles, and Z. Field, *Discovering statistics using R*. Los Angeles: SAGE Publication Ltd, 2012.