Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Prosodic Focus Marking in Bai-Mandarin Sequential Bilinguals' Mandarin

Zenghui Liu[1], *Aoju Chen*[1], *Hans Van de Velde* [2]

[1] Utrecht University
[2] Fryske Akademy
l.z.h.liu@uu.nl, aoju.chen@uu.nl, HVandeVelde@fryske-akademy.nl

Abstract

This study investigates the prosodic marking of focus in sequential bilinguals speaking Bai as their first language and Mandarin Chinese as their second language. Mandarin SVO sentences with varying information structure were elicited through a picture-matching task. The participants were primary school teachers in a Bai-speaking community. Our data shows that in their Mandarin the Bai-Mandarin sequential bilinguals lengthen the duration of the focal constituents in comparison to non-focal constituents. In addition, they expand the pitch range to distinguish focal constituents from post-focal constituents. However, focus types differing in size and contrastivity are not distinguished in duration or pitch. The present study thus provides evidence that sequential bilinguals mark focus prosodically in their L2 Mandarin, although their use of pitch as a prosodic cue for marking focus is less systematic in comparison to monolingual speakers of Mandarin.

Index Terms: focus, prosody, sequential bilingualism

1. Introduction

An important function of prosody is to highlight new information in a sentence (i.e. focus). In languages that use prosody for marking focus, the prosodic cues involved differ from one language to another. For instance, some languages (such as English [1], Dutch [2, 3] and Mandarin [4, 5]) not only expand the pitch range and lengthen the duration of the focal constituent, but also compress the pitch range and intensity of the post-focal constituent. However, other languages only exploit duration for the same purpose (such as Cantonese [6], Tsat [7] and Bai [8]).

Due to these observed cross-linguistic differences in the manner of marking focus prosodically, recently the acquisition of prosodic focus marking in L2 learners who acquire another language after early childhood (i.e. late bilinguals) and bilinguals who acquire two languages in early childhood (sequential bilinguals) has received considerable attention ([9, 10, 11, 12]). It has been found that the similarity between the learners' L1 and L2 did not necessarily lead to the attainment of native-like prosodic focus marking in late bilinguals in L2. For instance, [10] found that Mandarin learners of Dutch did not acquire phonetic implementation of focus accent in Dutch, although Mandarin and Dutch mark focus in a similar manner in terms of expanding the pitch range and the duration of the focal constituent and compression the pitch range and duration of the post-focal constituents. Similarly, [9] found that native English speakers learning Mandarin as L2 did not show native-like patterns of in-focus changes in intensity and pitch in all the tones.

Prosodic focus marking not only seems difficult to acquire for late bilingluals, but also for sequential bilinguals. For example, [7] examined the Mandarin production of native speakers of Tsat, a tone language of the Austronesian language family spoken in Sanya, Hainan, China by the Utsuls. They found that Tsat-Mandarin sequential bilinguals did not mark focus prosodically in the same way as Mandarin monolinguals did. Specifically, Tsat-Mandarin sequential bilinguals did not expand the pitch range of the focal constituents, neither in their native language Tsat, nor in their second native language Mandarin. However, they consistently lengthened the duration of the focal constituents in both languages. [12] examined the prosodic focus marking produced by Min-Mandarin sequential bilinguals from different age groups in their Min and Mandarin. These speakers spoke Quanzhou Southern Min as their L1 and acquired Mandarin as L2 in mainland China. It was found that younger bilinguals could produce Beijing-Mandarin-like prosodic focus marking, but older bilinguals could not. [12] suggested that more Beijing-Mandarin-like input and intensive training of Putonghua (Standard Mandarin) could explain the monolingual-like production by younger bilinguals.

Sequential bilinguals can differ considerably in their proficiency in L2. It is still not clear whether sequential bilinguals can attain native-like prosodic ability in L2. We address this question by examining teachers of Mandarin in schools who are native Bai speakers and acquired Mandarin as L2 in early childhood. Bai, a Tibetan-Burman tone language spoken by the Bai minority group in China, only exploits duration to mark focus prosodically [8]. It is worth mentioning that monolingual speakers of Bai lengthened the duration of the focal constituent for marking focus, while pitch range is barely varied ([8]). These school teachers either graduated from teachers' colleges, or held certificates for teaching and for advanced Mandarin proficiency. They are seen as representatives of the Mandarin speaking population in the Bai speaking communities. Therefore, the present investigation of prosodic focus marking in Mandarin by Bai school teachers can inform us about the ultimate attainment of Mandarin by sequential Bai-Mandarin bilinguals.

More specifically, we examined (1) whether sequential bilinguals mark focus prosodically in their Mandarin, i.e. the effect of focus; (2) whether they use prosodic cues to distinguish focus types that differ in the size of the focal constituent, i.e. the effect of size; and (3) whether they distinguish contrastive focus from non-contrastive focus using prosodic cues, i.e. the effect of contrastivity.

2. Methodology

2.1. Participants

Six Bai-Mandarin sequential bilinguals (4 women and 2 men,

 doi: 10.21437/SpeechProsody.2016-195

between 29 and 51 years old) participated in our experiment. All these participants were native speakers of Southern Bai and acquired Mandarin as their second language at about age 6. These participants were 6 out of the 7 teachers who taught *Chinese* in Jinhe Primary School, Xizhou County, Dali Bai Autonomous Prefecture, China.

2.2. Experimental materials

The data collection was led by a female (27-year-old) native speaker of Mandarin Chinese. It aimed to elicit SVO Mandarin sentences in five focus conditions: (1) narrow focus on the subject NP in sentence-initial position: NF-i; (2) narrow focus on the verb in sentence-medial position: NF-m; (3) narrow focus on the object NP in sentence-final position: NF-f; (4) broad focus: BF; and (5) contrastive focus on the verb in sentence-medial position: CF-m. The focus conditions were 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.

(1) Experimenter: Look! The pen. There is also some soil. It seems someone buries the pen. Who buries the pen?

Participant: [THE CAT] buries the pen. (NF-i)

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

Participant: The cat [BURIES] the pen. (NF-m)

(3) Experimenter: Look! The cat. There is also some soil. It seems like that the cat buries something. What does the cat bury?

Participant: The cat buries [THE PEN]. (NF-f)

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

Participant: [THE CAT BURIES THE PEN]. (BF)

(5) Experimenter: Look! The cat and the pen. It seems like that the cat does something with the pen. I guess the cat cuts the pen.

Participant: The cat [BURIES] the pen. (CF-m)

80 sentences were elicited from each participant. Among these, each focus condition (n=5) was realized in 16 SVO sentences. High level tone (T1), rising tone (T2), dipping tone (T3) and falling tone (T4) were systematically balanced in subject noun phrase, the verb and the object noun. The realizations of these sentence-medial verbs were analyzed and compared over the five focus conditions.

2.3. Data elicitation

The experiment was designed as a picture-matching game. In this picture-matching game, three piles of pictures were used: the experimenter and the participant each held a pile of pictures ordered in a sequence; the third pile of pictures were scattered 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 in finding out which pictures from her own pile matched with pictures from the third pile. This method was also used in a study on Mandarin [5] enabling a better comparison between the adult production of Bai-Mandarin sequential bilinguals and Mandarin monolinguals. Our participants were informed that the picture matching game was designed for children, thus the task could be quite simple for adults.

2.4. Recording procedure and equipment

The participants did the experiment individually in a quiet room in Jinhe Primary School. The game was split into two separate sessions, and each session took approximately 20 minutes. 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.

2.5. Data selection and acoustic analysis

The target SVO sentences were selected and evaluated using Praat [13], and the following criteria were used for including in the data set: (1) it was a response to the target question; (2) contained no self-corrections or hesitations; and (3) did not deviate from the target sentences in lexical choice and word order. In total, 440 out of 480 SVO target sentences (92%) were obtained from six speakers. The verbs were the items for acoustic and statistical analysis, as they had played multiple roles in the five focus conditions. They were focal constituent in the BF, NF-m and CF-m conditions, but they can also be a pre-focal constituent in the NF-f condition and a post-focal constituent in the NF-i condition.

Each of the selected sentences was segmented into words, and landmarks were inserted demarcating verb onset and offset, in addition to the locations of pitch-maximum and pitch-minimum within the verb. Since the target verb was a monosyllabic word, the syllable duration equaled to the word duration. In addition, the pitch range of the target verb was obtained by subtracting the pitch-minimum from the pitch maximum.

3. Analysis and Results

In order to investigate the effect of focus and focus types, we compared the word duration and pitch range of the target verb in various focus conditions. Specifically, to investigate the effect of focus, we compared the measurements of the verbs between unfocused conditions and focused condition, i.e. NF-m (focus) vs. NF-i (post-focus); and NF-m (focus) vs. NF-f (pre-focus). To investigate the effect of focus type that differs in size of the focal constituent, we compared narrow focus (NF-m) with broad focus condition (BF). To investigate the effect of contrastivity, we compared contrastive focus (CF-m) with non-contrastive focus (NF-m).

Statistical analyses were conducted using mixed effect modelling in R [14, 15]. We were interested in how well the focus conditions (*Focus Condition*) and the lexical tone of the verbs (*Tone Verb*) could explain variation in word duration and pitch range on the target verbs, i.e. the phonetic measurements of the target verbs.

In all models, *Focus Condition* and *Tone Verb* were included as fixed factors, while *Speaker* and *Sentence* were included as random factors. *Focus Condition* always contained two levels in every comparison listed above to answer specific research questions; and *Tone Verb* had four levels which referred to four lexical tones of the target verbs. The experimental design contained one representation for each lexical tone, i.e. only one word containing one specific lexical tone was included. Dependent variables were word duration and pitch range of the target verbs.

When building the models, only factors that significantly improved the previous model were included in subsequent models. The improvement of the model fit was assessed by the difference in -2LL (log likelihood), i.e. a statistically significant difference between these two models was an indication of a significant effect of the added fixed factor. Then we excluded the models that did not lead to a significant improvement over the previous model to get the best fit model. Using this procedure we could assess the effect of the factors listed, as well as their interactions.

3.1. Effect of focus: narrow focus vs. non-focus

3.1.1. Duration

The duration data obtained from the verbs in the NF-m and NF-i conditions showed that the verbs were on average 22.2 ms longer when focused (NF-m) than when not focused and following a focused constituent (NF-i). Mixed-effect modeling was used to assess the effect of *Focus Condition* on the duration of the verbs, as described above. It revealed a main effect of *Focus Condition* ($\chi^2(1) = 8.006$, $p<0.01$). The best fit model was the one containing main effects of *Focus Condition* and *Tone Verb.* However, there was no two-way interaction between *Focus Condition* and *Tone Verb* ($p= 0.76$). This suggested that the duration of the focal constituent was lengthened in all the lexical tones in comparison to the post-focal constituent. The use of duration for distinguishing the focal constituent from the post-focal constituent is shown in Figure 1.

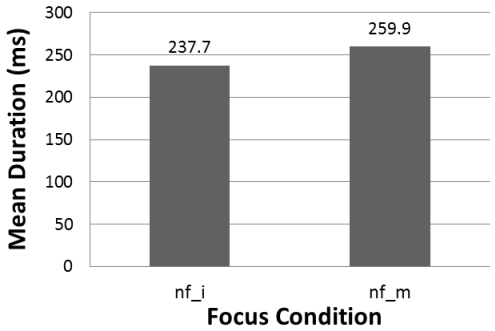

Figure 1: *Mean duration (in ms) of post-focal constituent vs. focal constituent (nf_i = verb following a focused constituent, nf_m = verb in sentence medial focused position)*

The duration data obtained from the verbs in the NF-m and NF-f conditions showed that the verbs were on average 21.8 ms longer when focused (NF-m) than when not focused and preceding a focused constituent (NF-f). Mixed-effect modeling revealed a main effect of *Focus Condition* ($\chi^2(1) = 6.4$, $p<0.05$). The best fit model contained main effects of *Focus Condition* and *Tone Verb.* This suggested that the duration of the focal constituent was lengthened in all the lexical tones in comparison to the pre-focal constituent. The use of duration for distinguishing the focal constituent from the pre-focal constituent is shown in Figure 2.

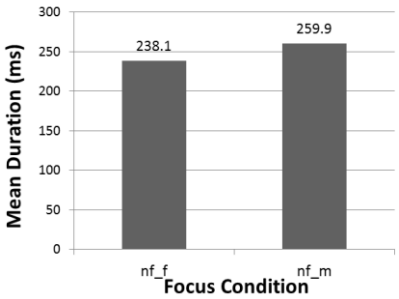

Figure 2: *Mean duration (in ms) of post-focal constituent vs. focal constituent (nf_f = verb preceding a focused constituent, nf_m = verb in sentence medial focused position)*

To summarize, these results suggested that in our sequential bilinguals' production of Mandarin, duration was used to differentiate focus from non-focus, regardless of the lexical tones of the verbs. Specifically, the use of duration was not only found in the comparison between the focal constituent and the pre-focal constituent, but also in the comparison between the focal constituent and the post-focal constituent. Our sequential bilinguals lengthened the duration of the focal constituents in comparison to non-focal constituents.

3.1.2. Pitch Range

Our data showed that the pitch range of the verbs were on average 7.5 Hz larger when focused (NF-m) than when not focused and following a focused constituent (NF-i). Mixed-effect modeling revealed a main effect of *Focus Condition* ($\chi^2(1) = 4.338$, $p<0.05$). The best fit model contained main effects of *Focus Condition* and *Tone Verb.* This suggested that the pitch range of the focal constituent was expanded in all the lexical tones in comparison to the post-focal constituent. The use of pitch range for distinguishing the focal constituent from the post-focal constituent is shown in Figure 3.

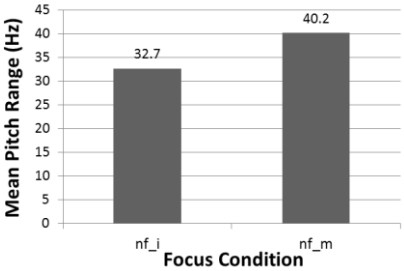

Figure 3: *Mean pitch range (in Hz) of post-focal constituent vs. focal constituent (nf_i = verb following a focused constituent, nf_m = verb in sentence medial focused position)*

Regarding the comparison between focus (NF-m) and pre-focus (NF-f). The best fit model only contained a main effect of *Tone Verb*.

The above results provided evidence that our Bai-Mandarin sequential bilinguals only used pitch to differentiate focal constituents from post-focal constituents, but not to differentiate focal constituents from pre-focal constituents.

3.2. Effect of size: narrow focus vs. broad focus

Regarding the comparison between narrow focus and broad focus, the duration data obtained from the verbs in the BF and NF-m conditions showed that the verbs were on average 11.9 ms longer when in the narrow focus condition (NF-m) than

when in the broad focus condition (BF). In addition, the pitch data obtained showed that pitch range of the verbs were on average 4.3 Hz larger when the verbs were in narrow focus condition (NF-m) in comparison to the broad focus condition (BF). For both dependent variables, the best fit models we obtained were the models only containing the main effect of *Tone Verb*. There was thus no evidence that our sequential bilinguals varied the duration and pitch range in their Mandarin to distinguish focus types that differ in the size of the focal constituent.

3.3. Effect of contrasitivity

Regarding the comparison between non-contrastive focus (NF-m) and contrastive focus (CF-m), the duration data obtained from the verbs in the CF-m and NF-m conditions showed that the verbs were on average 7.1 ms longer when in the contrastive focus condition (CF-m) than when in the non-contrastive focus condition (NF-m). Additionally, the pitch data obtained showed that pitch range of the verbs were on average 3.3 Hz larger when the verbs were in contrastive focus condition in comparison to the non-contrastive condition. The best fit model only contained a main effect of *Tone Verb*. There was thus no evidence that our sequential bilinguals used duration or pitch range to distinguish focus types that differ in the contrastiveness of the focal constituent.

4. Discussion

The present study examined prosodic focus marking produced by sequential bilinguals whose L1 is Bai and acquired Mandarin Chinese at about the age of six. Our results show that this type of Bai-Mandarin sequential bilinguals marks focus prosodically in their Mandarin, but are not completely native-like in their Mandarin.

With regard to our first research question, our results show that Bai-Mandarin sequential bilinguals increase the duration of the focal constituent in comparison to pre-focal and post-focal constituents in their Mandarin production. In addition, they also expand the pitch range of the focal constituent in comparison to post-focal constituent. But this effect is absent in the comparison between focal and pre-focal constituent. The consistent use of duration for marking focus might be attributed to positive L1 transfer, as their native language Bai also exploits duration as a prosodic cue for the same purpose. However, the sequential bilinguals in our study also expand pitch range for distinguishing focal constituents from post-focal constituents in their Mandarin, as monolingual speakers of Mandarin do ([5]). As Bai does not exploit pitch as a prosodic cue for marking focus, the use of pitch cannot be attributed to positive L1 transfer. Thus, the mastery of pitch range by Bai-Mandarin sequential bilinguals can only be explained by successful acquisition of prosodic focus marking in Mandarin.

With regard to our second research question, our results show that neither duration nor pitch is used for differentiating focus types that differ in size by sequential bilinguals in their Mandarin. It means that the duration and pitch range of the focal constituents do not differ when the scope of the focus is on a specific constituent of the utterance (narrow focus) from when the scope of the focus is on the whole utterance (broad focus). In contrast, [5] showed that monolingual Mandarin speaker use duration for distinguishing narrow focus from broad focus.

With regard to our third research question, our results show that neither duration nor pitch is used to make a distinction between contrastive focus and non-contrastive focus by sequential bilinguals in their Mandarin. Similarly, [5] showed that native Mandarin adult speakers neither use duration nor pitch range for differentiating the contrastive focal constituent from the non-contrastive focal constituent in spontaneous speech.

5. Conclusions

Our study has revealed both similar and different results compared to previous studies of bilinguals' prosodic focus marking in L2. Firstly, Bai-Mandarin sequential bilinguals do mark focus prosodically in their Mandarin. In comparison with previous studies of American English learners of Mandarin [9] and Mandarin learners of Dutch [10], our Bai-Mandarin speakers successfully mastered the use of duration and pitch for distinguishing the focal constituent from the non-focal constituent. Although American English and Dutch are acoustically similar to Mandarin in terms of exploiting prosodic cues for encoding focus, L2 learners seem to have difficulties in the acquisition of prosodic focus marking [9, 10]. The difference between the sequential bilinguals in the current study and those late bilinguals in previous studies is that the former group was exposed to Mandarin from an early age. The monolingual-like production of prosodic focus marking by the speakers in the present study might be attributed to their early input of Mandarin as Bai-Mandarin sequential bilinguals.

Secondly, our results are consistent with the previous results from Min-Mandarin bilinguals [12] in that sequential bilinguals successfully acquire the use of prosodic cues for encoding focus. The native-like production of prosodic focus marking in the younger group of Min-Mandarin bilinguals Mandarin was suggested as results of more Beijing-like inputs and intensive training of Standard Mandarin. In our opinion, the intensive training of Standard Mandarin received by our Bai-Mandarin sequential bilinguals as teachers plays an important role in their successful acquisition. However, our results also show that the use of prosodic cues, such as pitch and duration by Bai-Mandarin sequential bilinguals is still not completely monolingual-Mandarin-like. Duration is used more consistently in comparison to pitch range. It might relate to the fact that Bai-Mandarin sequential bilinguals have less access to Beijing Mandarin in daily communication, except for the input from the media.

To conclude, bilinguals' prosodic focus marking in L2 appears to depend on various factors, such as age of exposure to L2, age of the bilinguals, the daily contact of L2, but to a lesser degree of similarities and differences in prosodic focus marking between L1 and L2.

6. Acknowledgements

We are grateful to all the bilingual participants from Jinhe Primary School and to the children's parents for their cooperation. We also thank Shanpeng Duan and teachers of Jinhe Primary School for their enormous support for the field work, and Anqi Yang and Anna Sara H. Romøren for their feedback. This study is supported by a scholarship from the Chinese Scholarship Council to the first author and a VIDI grant (276-89-001) from the Netherlands Organization for Scientific Research to the second author.

7. References

[1] W. E. Cooper, S. J. Eady and P. R. Mueller. Acoustical aspects of contrastive stress in question–answer contexts. *J. Acoust. Soc. Am.* 77(6), pp. 2142-2156. 1985.

[2] H. Rump and R. Collier. Focus conditions and the prominence of pitch-accented syllables. *Lang. Speech* 39(1), pp. 1-17. 1996.

[3] A. Chen. The phonetics of sentence-initial topic and focus in adult and child Dutch. *Phonetics and Phonology: Interactions and Interrelations* pp. 91-106. 2009.

[4] Y. Xu. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27(1), pp. 55-105. 1999.

[5] A. Yang and A. Chen. Prosodic focus marking in child and adult Mandarin Chinese. Paper presented at *The 4th International Symposium on Tonal Aspects of Language*, pp.54-58. 2014.

[6] W. L. Wu and Y. Xu. Prosodic focus in Hong Kong Cantonese without post-focus compression. *Proceedings of the Speech Prosody* 2010.

[7] B. Wang, C. Li, Q. Wu, X. Zhang, B. Wang and Y. Xu. Production and perception of focus in PFC and non-PFC languages: Comparing Beijing Mandarin and Hainan Tsat. Paper presented at the *Interspeech*. 2012.

[8] Z. Liu, A. Chen and H. van de Velde. Prosodic focus marking in Bai. Proceedings of the *Speech Prosody*. 2014

[9] Y. Chen and S. Guion-Anderson. Prosodic realization of focus in Mandarin by advanced American learners of Chinese. *J. Acoust. Soc. Am.* 131(4), pp. 3234-3234. 2012.

[10] X. He, J. Hanssen, V. J. van Heuven and C. Gussenhoven. Phonetic implementation must be learnt: Native versus Chinese realization of focus accent in Dutch. Presented at *Proceedings of the XVIIth International Congress of Phonetic Sciences*. 2011, .

[11] E. Nava and M. L. Zubizarreta. Prosodic transfer in L2 speech: Evidence from phrasal prominence and rhythm. *Proceedings of Speech Prosody* pp. 335-338. 2008.

[12] Y. Chen, Y. Xu and S. Guion-Anderson. Prosodic realization of focus in bilingual production of Southern Min and Mandarin. *Phonetica* 71(4), pp. 249-270. 2014.

[13] P. Boersma. Praat, a system for doing phonetics by computer. Glot International 5:9/10, 341-345. 2001

[14] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. 2014

[15] D. Bates, M. Maechler, BM. Bolker and S.Walker S. "lme4: Linear mixed-effects models using Eigen and S4." ArXiv e-print; submitted to _Journal of Statistical Software_, <URL: http://arxiv.org/abs/1406.5823>. 2014