



# Acquisition of prosodic focus marking by three- to six-year-old children learning Mandarin Chinese

Qianyutong Zhang, Kexin Lyu, Zening Chen, Ping Tang

Nanjing University of Science and Technology, Nanjing, Jiangsu, China

zhangqianyutong@126.com, lyukexinkkk@163.com, chenzeningxxx@163.com,  
ping.tang@njjust.edu.cn

## Abstract

Prosodic focus plays an important role during speech communication, delivering speakers' pragmatical intention to emphasize key information, especially in contrastive scenarios. Previous studies exploring children's acquisition of prosodic focus have generally focused on Germanic and Romance languages, while it was unclear when children learning Mandarin Chinese were able to correctly interpret the pragmatic meaning of prosodic focus and integrate it into speech comprehension. The current study explored Mandarin-learning 3-6-year-olds' online interpretation of prosodic focus to identify contrastive referents. Twenty 3-4-year-olds, 23 5-6-year-olds, and 22 adult controls were tested. The visual-world paradigm was adopted, where participants were instructed to search for target pictures while listening to contrastive objects in discourse sequences, e.g., *Find the red cat. Now, find the PURPLE/purple cat*, where the second adjective was produced with or without prosodic focus. Participants' fixation patterns were recorded via eye-trackers. The results showed that while adults and 5-6 years showed faster fixation toward target pictures in the presence of prosodic focus, this was not the case for 3-4 years. These results indicated that Mandarin-learning children at 5-6 years have acquired the pragmatic meaning of prosodic focus and utilize it to guide their identification of contrastive referents.

**Index Terms:** prosodic focus, contrastive referents, eye-tracking, Mandarin-learning children

## 1. Introduction

In daily speech communication, speakers tend to highlight the key information of an utterance by increasing its prosodic prominence, especially in contrastive scenarios. For example, "A: Do you want a black puppy? B: No, I want a WHITE puppy!". In this case the speaker B emphasizes the adjective "white" in contrast to the already mentioned adjective "black", using a series of prosodic approaches, e.g., producing it with higher overall pitch, longer duration and/or greater intensity [1-4]. To become successful communicators, children have to perceive these prosodically focused expressions and correctly interpret the pragmatic implications, e.g., the emphasized color. Studies exploring children's comprehension of prosodic focus have primarily focused on Germanic and Romance languages, finding that it might take 3 to 6 years for children to fully acquire the pragmatic meaning of prosodic focus and integrate it into speech comprehension such as identifying contrastive referents [5-7]. However, despite the fact that more than 60% of languages in the world are tonal [8], only a few studies have explored children's acquisition of prosodic focus

in tonal languages such as Mandarin Chinese, with various research focuses [10-12]. Therefore, the current study explored when children learning Mandarin Chinese can acquire the pragmatic meaning of prosodic focus and utilize it to guide their speech comprehension.

The comprehension of prosodic focus has been examined on children learning English, French, and German, while the detailed age of acquisition is contingent on the language in question and the method being used. For example, based on a sentence correction task, Szendrői et al. [5] found that, as early as 3 years old, children learning English, French or German have already shown sensitivity to prosodic prominence as a means of determining the identity of a focused subject phrase [5]. Moreover, English-learning children relied more on prosodic information in focus comprehension as compared to children learning the other two languages. Using an eye-tracking-based visual-world paradigm [9], Arnold further demonstrated that English-learning children at 4-5 years old have already shown adult-like sensitivity to accented information during their online interpretation of referential expressions [6]. However, Ito et al. found that although English-learning children at 6 years were able to use the prominence accent to facilitate their detection of contrastive referent in discourse sequences (e.g., red cat – PURPLE cat), their speed was much slower than adults even at 11 years [7]. A potential reason for their slower processing might be related to the complexity of the target picture used in [7], where 18 objects were presented simultaneously on a picture, which might be especially challenging for children to find the target object rapidly.

In tonal languages such as Mandarin Chinese, the pitch information not only marks prosodic focus via expanding intonation contour, but also deliver lexical meaning at word level by changing its tonal identity. However, only a few studies have explored the comprehension of prosodic focus by children learning Mandarin Chinese, with various research focuses. For example, two studies compared the perceptual weight of prosody and word order in the perception of focus by Mandarin-learning 3-5 years [10, 11]. While [10] found that children mostly used word order to identify focus as compared to prosodic cues, [11] found the reversed pattern. However, none of them directly tested the role prosody played in the process of interpreting the pragmatic implicature of a focus. Zhou examined Mandarin-learning 4-5-year-old's comprehension of prosodic information in semantic ambiguity scenarios, e.g., "Zhiyou XIAOMING de naozhong shi huangsede vs. Zhiyou Xiaoming de NAOZHONG shi huangsede" *Only XIAOMING's clock is yellow vs. Only Xiaoming's CLOCK is yellow* [12]. The results showed that these children were able to use the prosodic information to guide their interpretation of these semantical ambiguity

sentences. However, the target sentences always started with a semantic operator “zhiyou” *only*, which might facilitate children’s interpretation of the pragmatic implicature of prosodic focus. Therefore, so far it was still unclear if Mandarin learning children were able to use prosody alone to interpret the pragmatic meaning of the focus.

The present study, therefore, aimed to explore whether children learning Mandarin Chinese were able to correctly interpret the pragmatic implicature of prosodic focus during speech comprehension. Specifically, we asked if 3-6-year-olds can use prosodic focus to facilitate their identification of contrastive referents in discourse sequences (e.g., red cat – PURPLE cat; a similar method to Nakamura et al. [13]). Three to six years were tested since it has been shown that the youngest age at which children learning Germanic and Romance languages can interpret prosodic focus with its pragmatic function falls generally in this range [5-7]. We predicted that, as compared to younger children (e.g., 3 years), older children (e.g., 6 years) might better utilize the prosodic information to guide their identification of contrastive referents.

## 2. Method

### 2.1. Participants

Forty-three 3-6-year-old Mandarin-learning children were recruited from public kindergartens of Nanjing, Jiangsu of China (number of children: 3-year-olds = 2, 4-year-olds = 18, 5-year-olds = 13, and 6-year-olds = 10). In the analysis, they were further divided into two groups to maximize the statistical power, i.e., 3-4-year-olds (20 children) and 5-6-year-olds (23 children). They do not have any hearing, language, or cognitive disorder according to reports of kindergartens. Besides, 22 adults were also included as controls. They were all undergraduate students without any history of speech, hearing, or mental disorder according to self-report.

### 2.2. Stimuli

#### 2.2.1. Audio stimuli

A series of sentence pairs were adopted as stimuli which consisted of two sentences carrying two “color + animal” phrases, e.g., “Zhaodao hongse de mao. Xianzai, zhaodao zise de mao.” *Find the red-color cat. Now, find the purple-color cat.* The two sentences differed in the color of the animal. Thus, the first sentence set up an informational context where a type of color and an animal have been mentioned (hereafter “contextual sentence”), whereby the color/adjective in the second sentence was contrasted with that in the contextual sentence and supposed to carry the contrastive focus (hereafter “target sentence”).

All sentences were recorded by a female undergraduate student majoring in broadcasting. In each sentence pair, the contextual sentence did not carry a narrow focus on any word, while the adjective in the target sentence was produced in two different focus conditions, either with a contrastive accent (accent condition) or with a neutral accent/broad focus (unaccented condition):

(1) Accented condition: “Zhaodao hongse de mao. Xianzai, zhaodao ZISEDE mao” *Find the red-color cat. Now, find the PURPLE-color cat.*

(2) Unaccented condition: “Zhaodao hongse de mao. Xianzai, zhaodao zise de mao” *Find the red-color cat. Now, find the purple-color cat.*

An acoustic analysis performing on adjectives in the target sentences indicated that they exhibited higher pitch and longer duration in accented condition as compared to unaccented condition. Thus, adjectives in the target sentences carried the expected prosodic focus in the accented condition.

Table 1: *Pitch and duration of adjectives in target sentences in accented and unaccented conditions.*

Conditions	Pitch (Hz)	Duration (s)
Accented	223.26	0.52
Unaccented	202.62	0.32

#### 2.2.2. Visual stimuli

Each sentence pair was yoked with a picture illustrating six animals (see Figure 1), corresponding to target animals mentioned in contextual and target sentences (e.g., red cat and purple cat), two competitors (with the same color as target animals, but differed from them in animal type, e.g., red horse and purple deer), and two distractors that differed from target animals in both color and type (e.g., pink horse and green cow).

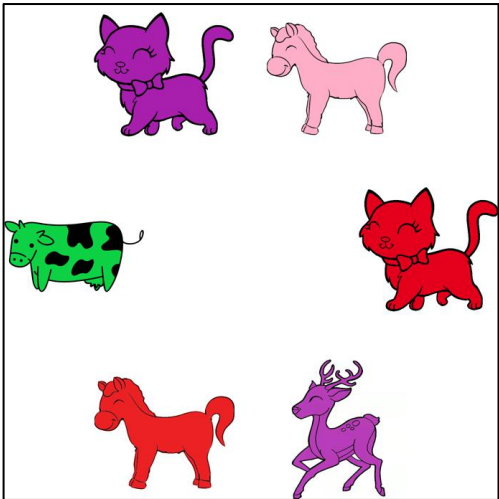


Figure 1: The picture yoked with the sentence pair “Zhaodao hongse de mao. Xianzai, zhaodao zise de mao” *Find the red-color cat. Now, find the purple-color cat.*

### 2.3. Procedure

All participants were tested individually in quiet rooms (for children) or sound-proof rooms (for adults). Prior experiment, a picture-naming test was conducted to ensure that all participants were able to correctly recognize the color and animal in target sentences.

Each perceptual experiment lasted for around five minutes, consisted of 18 trials, including seven accented trials (accented condition), seven unaccented trials (unaccented condition), and four filler trials. In the filler trials, target animals in the target sentence were all distractors, produced either with or without prosodic focus. The experiment always started with two practice trials (an accented trial and an unaccented trial),

followed by 16 test trials in pseudo-random order. Only accented and unaccented trials underwent further analysis.

In each trial, participants were required to listen to audio stimuli while looking at the target picture presented on a computer screen. Each trial lasted for 15 seconds. In each trial, participants were firstly presented with the picture for one second, followed by the audio for the contextual sentence/sentence 1 (two seconds). After two seconds, an animation (rotation, lasted for one second) was played on the target animal, followed by two seconds of silence. Then, the audio of the target sentence/sentence 2 was played (three seconds), followed by an animation (lasted for one second) on the target animal after two seconds.

Participants' eye movements during the test phases were recorded using a Tobii Pro X3-120 portable eye-tracker (for children) or an EyeLink 1000 Plus desktop eye-tracker (for adults). Nine-point calibrations were conducted prior testing for all participants.

#### 2.4. Statistical analysis

Participants' fixation curve toward target pictures (second-mention animals) within 800ms from the noun onset of the target sentence was analyzed. Second-order polynomials were fitted on each curve to compare it between accented and unaccented conditions for each group. Generalized linear mixed-regression models were then performed to compare three curve parameters between conditions, i.e., the height (reflected by the intercept), slope (reflected by the 1<sup>st</sup> order polynomial coefficient), and curvature (reflected by the 2<sup>nd</sup> order polynomial coefficient). Statistical models were performed using the *lme4* package in R [14].

### 3. Results

Figure 2 shows participants' fixation curves toward target pictures in the target sentence across groups and conditions. A second-order generalized linear-mixed regression model was performed on the fixation curves for each group, with a fixed factor "Condition" (two levels: unaccented and accented; reference level: unaccented) and two random factors "Participant" and "Trial".

The results are presented in Table 2, which showed that adults and 5-6-year-olds demonstrated higher and more rising fixation curves in the accented condition as compared to the unaccented condition, while there was no evidence showing any curve difference between accented and unaccented conditions for 3-year-olds. These results indicated that 5-6-year-olds identified the target picture/referent faster in the accented condition as compared to the unaccented condition, just like adults, while 3-4-year-olds did not show any fixation difference between the two conditions.

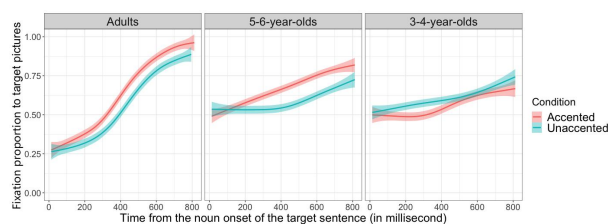


Figure 2: Fixation proportion to target pictures within 800ms from the noun onset of the target sentence.

Table 2: Results of generalized linear mixed regression models for each group.

Group	Parameter	$\beta$	SE	z	p
Adults	Intercept	1.08	0.34	3.15	0.002**
	Slope	11.83	0.38	30.78	<0.001***
	Curvature	1.84	0.34	5.42	<0.001***
	Accented	0.66	0.32	2.07	0.039*
	Slope $\times$ Accented	3.03	0.47	6.38	<0.001***
	Curvature $\times$ Accented	0.45	0.45	1.01	0.311
5-6-year-olds	Intercept	0.94	0.23	4.07	<0.001***
	Slope	3.93	0.26	14.91	<0.001***
	Curvature	-0.08	0.26	-0.29	0.769
	Accented	0.54	0.25	2.17	0.030*
	Slope $\times$ Accented	2.09	0.32	6.62	<0.001***
	Curvature $\times$ Accented	-0.33	0.31	-1.06	0.290
3-4-year-olds	Intercept	-0.12	0.22	-0.53	0.595
	Slope	2.22	0.26	8.68	<0.001***
	Curvature	0.45	0.25	1.78	0.075
	Accented	-0.19	0.21	-0.95	0.343
	Slope $\times$ Accented	0.13	0.32	0.41	0.683
	Curvature $\times$ Accented	0.14	0.32	0.44	0.657

### 4. Discussion

The current study examined whether 3-6-year-olds learning Mandarin Chinese were able to utilize the prosodic information to guide their identification of contrastive referents. The results showed that while the presence of prosodic focus facilitated 5-6-year-olds to identify contrastive referents more quickly, just like adults, it did not show any facilitative effect for 3-4-year-olds.

Our results extended the previous English study which showed that English-learning children at 4-5 years were able to correctly perceive prosodic focus and use it to guide their online reference comprehension [6], suggesting that Mandarin children also showed similar ability around age 5. The current study also extended the previous study of Mandarin Chinese [12], showing that children's ability to interpret the pragmatic meaning of pragmatic is not contingent on the presence of the semantic operator "Zhiyou" only.

However, our results are inconsistent with the study of [5] which found that children learning English, French or German at 3 years have already shown sensitivity to prosodic prominence as a mean of determining the identity of a focused subject phrase. A potential reason leading to this inconsistency

might be related to the nature of focus processing being tested, i.e., online processing in this study vs. offline processing in [5]. This suggested that even if children might correctly interpret the pragmatic meanings of prosodic focus eventually, their processing differed from adults and older children. The result can only be revealed through fine-grained online measurements such as the eye-tracking task used in this study. Another reason might be related to the language difference. Unlike Germanic and Romance languages, pitch in Mandarin Chinese deliver both intonational and tonal information. Children learning Mandarin Chinese might thus have more difficulties in understanding prosodic focus, because they need the combination of abilities to comprehend both the lexical meanings and the pragmatic meanings of pitch information, which is the main approach to prosodic focus.

The reason why 3-4-year-olds did not use prosodic focus to guide their online identification might be two-fold. First, it might be the case that children at this age could not correctly perceive prosodic focus. But this might not be plausible since there have been extensive evidence showing that children at age 3 are already able to perceive prosodic prominence. Second, it is also possible that 3-4-year-olds were not able to map prosodic focus to its pragmatic implicature. During the process of identifying the target objects, children need the ability of general executive control [15] to maintain and compare the information across utterances. Since executive control develops with age, older children may perform better in switching attention to new objects than their younger peers [16]. This suggests that it might take children at least 5 years to fully exploit the mapping between prosody and pragmatics.

There are several limitations of this study that could be addressed in future studies. First, the current study only placed prosodic focus on new information (which is supposed to be marked with prosodic focus). Future studies could test children's perception of prosodic focus marked on already-mentioned information to see if this could exert a garden-path effect as found in [7, 16]. This would enable us to better understand children's predictive use of prosodic information in speech comprehension. Second, due to the limited number of participants, the current study only grouped children into two groups, i.e., 3-4-year-olds and 5-6-year-olds, thus future studies could test more children and run more fine-grained analysis to explore the more detailed time point at which children can fully master the adult-like comprehension of prosodic focus.

## 5. Conclusions

This study found that children learning Mandarin Chinese at 5-6 years have acquired the pragmatic meaning of prosodic focus and can utilize prosodic focus to guide their online interpretation of contrastive referents. This suggests that children learning Mandarin Chinese at 5-6 years have acquired the mapping between prosody and pragmatics, enriching our understanding of the acquisition of prosodic focus more broadly.

## 6. Acknowledgments

This research was funded by The National Social Science Fund of China (20CYY012). This paper is also funded by the National Undergraduate Training Program for Innovation and Entrepreneurship. Item number: 202010288135Z

## 7. References

- [1] Y. Chen, and B. Braun, "Prosodic realization of information structure categories in standard Chinese," *Paper presented at the Speech Prosody*, 2006.
- [2] Y. Chen, and C. Gussenhoven, "Emphasis and tonal implementation in Standard Chinese," *Journal of Phonetics*, vol. 36, no 4, pp. 724–746, 2008.
- [3] I. C., Ouyang, and E. Kaiser, "Prosody and information structure in a tone language: An investigation of Mandarin Chinese," *Language Cognition and Neuroscience*, vol. 30, no 1–2, pp. 57–72, 2014.
- [4] Y. Xu, "Effects of tone and focus on the formation and alignment of F0 contours," *Journal of Phonetics*, vol. 27, no.1, pp. 55–105, 1999.
- [5] K., Szendrői, et al., "Acquisition of prosodic focus marking by English, French, and German three-, four-, five- and six-year-olds," *Journal of Child Language*, vol. 45, no. 1, pp. 219–241, 2018.
- [6] J. E. Arnold, "THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases," *Cognition*, vol.108, no.1, pp. 69-99, 2008.
- [7] K. Ito, et al., "Interpretation of contrastive pitch accent in six- to eleven-year-old English-speaking children (and adults)," *Journal of Child Language*, vol. 41, no. 1, pp. 84–110, 2014.
- [8] M. Yip, *Tone*. Cambridge: Cambridge University Press, 2002.
- [9] M. K. Tanenhaus, et al., "Integration of visual and linguistic information during spoken language comprehension," *Science*, vol.268, pp.1632–1634, 1995.
- [10] H. Chen, K. Szendrői, S., Crain, and B. Hoehle, "Understanding prosodic focus marking in mandarin chinese: Data from children and adults," *Journal of Psycholinguistic Research*, vol. 48, no.1, pp. 19-32, 2019.
- [11] S.-H. E. Chen, "Surface cues and the development of given/new interpretation," *Applied Psycholinguistics*, vol. 19, no. 4, pp. 553–582, 1998.
- [12] P. Zhou, et al., "Children's use of phonological information in ambiguity resolution: a view from Mandarin Chinese," *Journal of Child Language*, vol. 39, no. 4, pp. 687–730, 2012.
- [13] C. Nakamura, et al., "LISTENERS' BELIEFS INFLUENCE PROSODIC ADAPTATION: ANTICIPATORY USE OF CONTRASTIVE ACCENT DURING VISUAL SEARCH," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, 2019, pp. 447-451.
- [14] D. Bates, et al., *The lme4 package*. R package version, 2(1), vol. 74, 2007.
- [15] M. Davidson, et al., "Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching," *Neuropsychologia*, vol. 44, pp. 2037 – 2078, 2006
- [16] K. Ito, et al., "Intonation facilitates contrast resolution: Evidence from Japanese adults and 6-year olds," *Journal of Memory and Language*, vol. 66, no. 1, pp. 265–284, 2012.