Prosodic Cues to Noun and Verb Categories in Infant-Directed Mandarin Speech

Aijun Li [1], *Rushen Shi* [2], *Wu Hua* [1]

[1] Institute of Linguistics, Chinese Academy of Social Sciences, China
[2] Université du Québec à Montréal, Canada

liaj@cass.org.cn, shi.rushen@uqam.ca

Abstract

Mandarin Chinese, a Sino-Tibetan language, has distinct syntactic and morphological structures in comparison to Indo-European languages. This study concerns Chinese infants' initial derivation of grammatical categories. We examined the prosodic properties of nouns and verbs of the maternal input speech. Non-word disyllabic noun-verb homophones were created and embedded in frequent carrier phrases. Mandarin-speaking mothers read these stimuli to their babies during a play session. Prosodic properties of noun and verb productions by these mothers were analyzed. The results show that isolated homophone nouns and verbs were identical prosodically. However, when these items were embedded in noun and verb carrier phrases, they exhibited some prosodic distinctions. Specifically, mean F_0 of the second syllable of the non-words was significantly different in verb versus noun productions, and the duration ratios of the two syllables of the noun productions also differed for that of the verb productions. These results suggest that maternal speech contains some prosodic cues to nouns versus verbs, which might support infants' acquisition of grammatical categories.

Index Terms: infant, lexical category, prosodic cues, verb and noun

1. Introduction

Infants start building their mental lexicon shortly before one year of age. For each word, they must learn its grammatical category, for example, noun, verb, preposition, etc. This knowledge is essential for both lexical and syntactic acquisition. Several theories have been proposed in the literatures regarding the mechanisms underlying infants' initial learning of grammatical categories, e.g., semantic based theories [1,4,10], distributional based theories (e.g., [7,8]), and prosodic-phonological based theories (e.g.[2,13,15]).

There is evidence in the literatures that input speech contains prosodic and phonological cues to grammatical categories. In English nouns are more likely to be trochaic whereas verbs tend to be iambic, e.g. [5]. Nouns and verbs differ prosodically in infant-directed speech in English, and infants have been shown to be sensitive to such differences [3]. Infant-directed speech in French has been shown to contain prosodic differences in noun versus verb productions [13]. Connectionist studies that model category learning in English, French, Dutch and Japanese showed that phonological differences between nouns and verbs play an essential role in the categorization of these two classes of words [9]. In neural network simulations using parental speech, prosodic and phonological cues were found to contribute importantly to the derivation of lexical versus functional categories in Mandarin, Turkish and English [13, 17], and even newborn infants use such cues to categorize these two fundamental syntactic categories [14]. But it is not yet clear if prosodic cues to nouns versus verbs exist in input speech in Mandarin.

Most studies on the acquisition of syntactic categories in Chinese focus on older children's production [6]. There is an emerging interest in the categorization ability of preverbal and early verbal Chinese-learning infants. A recent perceptual experimental study [18] examined the possible mechanisms underlying the acquisition of syntactic categories. In particular, the study [18] tested if Chinese infants can use distribution regularities to categorize novel nouns and verbs. While it is crucial to also test infants' use of prosodic cues in grammatical categorization, it is necessary to understand whether prosodic distinctions exist in input in Chinese. In the present paper we report the analysis results of prosodic features of noun and verb productions in parental speech to infants, using a well-controlled procedure. Existence of any prosodic cues would imply that it is plausible for infants to use such cues to derive their initial noun and verb categories.

2. Method

2.1. Materials

Disyllabic non-word homophones were created in the present experiment. The use of non-words can eliminate prosodic variations caused by word frequency, word phonological and grammatical structures. Table 1 displays the four homophone noun and verb pairs adopted in the study. The four non-words cover various tones in Mandarin Chinese, e.g., Tone 1 (high level, or HH phonologically), Tone 2 (rise, or LH phonologically, Tone 4 (fall, or HL phonologically). Because monosyllables in Chinese are often themselves individual morphemes, we constructed each non-word noun with two monosyllabic noun-morphemes (e.g., WN2 in Table 1 below consists of "dun4", "jie1", meaning "blockage", "street" respectively) and each non-word verb with two monosyllabic verb-morphemes (e.g., WV2 consists of "dun4", "jie1", meaning "simmer", "open up" respectively). These differences were reflected in the character writing, although the pinyin phonetic notation was identical.

Table 1. *Homophone nouns and verbs*

code	nouns	code	verbs	Pinyin
WN1	茶宋	WV1	查送	cha2 song4
WN2	盾街	WV2	炖揭	dun4 jie1
WN3	师灯	WV3	失登	shi1 deng1
WN4	价陆	WV4	驾录	jia4 lu4

Table 2. *Selected noun and verb carrier phrases*
(N for NW1~NW4, V for VW1~VW4)

No.	Noun phrases	Verb phrases
1	wo de N (My N.)	wo ye V (I also V.)
2	zhe ge N (This N.)	ni bie V (You shouldn't V.)
3	yi ge N (A/One N.)	bu neng V (Can't V.)

The verb and noun carrier phrases were frequently used utterances in parental speech to infant. As shown in table 2, three noun and three verb carrier phrases were selected. Target non-words were placed at the final position of each utterance. Each utterance thus contained two carrier syllables and the disyllabic target non-words. The target non-word in

table 1 each occurred with all three carrier phrases, giving 12 noun utterances and 12 verb utterances respectively, e.g., 'wo de chasong, wo ye chasong, etc.'. In total, we obtained 32 stimuli prompts including 8 homophone nouns and verbs produced in citation forms and 24 in noun and verb utterances.

2.2. Participants and recording

Participants were 20 mother-infant dyads, whose information is listed in table 3. All the mothers were speakers of Standard Chinese and had lived in Beijing ever since they were born.

The recording was conducted in a sound proof room with many soft toys and infant story books, the room was large and comfortable enough (about $15m^2$) for the mother and the infant to play together. There was a screen in the recording room; the prompts were shown as large-print slides on this screen, which was controlled by the recording staff in the monitoring room. The recording staff in the adjacent room listened to the mother's production through headphones and clicked a computer key to present new prompts when the mother completed the reading of the materials on each slide.

The mother wore a wireless microphone (AKG WMS 40 PRO) on her collar. Mothers were required to read the prompts shown on the screen in the way as she would usually address their babies. They were allowed to repeat the reading if they wished. The order of utterance presentation was counterbalanced across mothers, such that half of the mothers read the noun uses before the verb uses, and the other half of the mothers read the verb uses first. Each utterance was displayed twice. The sampling frequency was 44.1KHz.

The stimuli recorded for prosodic analysis include 160 isolated homophone nouns and verbs (4*2*20) and 960 noun and verb carrier phrases (24*2*20).

Table 3. *Information of 20 mother-infant dyad (month :m; the mothers' codes represent the testing dates)*

Mother code	Infant age	Mother code	Infant age
090224	12m	090509	14m
090227	16m	090613	11m
090301	15m	090617	12m
090312	13m	090622	14m
090317	12m	090627	15m
090321	19m	090629	12m
090326	24m	090629	13m
090404	11m	090704	15m
090408	16m	090704	12m
090409	13m	090913	15m

3. Prosodic analysis

3.1. Prosodic cues

All the target non-words, both isolated and within carrier phrases, were annotated and analyzed in Praat (http://*www.fon*.hum.uva.nl/praat/). After discarding those invalid sounds (e.g., creaky sounds, sounds containing interfering signals such as infant's overlapping voice and noise produced by the infant), we obtained 947 target words. Before the extraction of the data, F_0 data were manually checked to ensure their accuracy. All the duration of the first and second syllables of the non-words, and their F_0 were extracted. F_0 is transferred into semitone scale by the selection of 75Hz as the reference frequency.

Examples of the prosodic measures of a homophone target non-word are shown in figure 1, the F_0 and duration of the same word "cha2song4" in verb carrier phrase "wo ye chasong"(I also Cha4song4.) uttered two different times by the same speaker. Obvious variations can be observed between

these two productions, which are caused by speaking variability (i.e., speaking rate, high or low voice) or other paralinguistic factors, e.g., mood or emotion. Therefore, when comparing the prosodic properties between the nouns and verbs, some normalized or relative cues should be employed. The prosodic measures in this study included the following:

The duration ratio of the first syllable to the whole word: *dur1ratio*= dur1/durtotal; the duration ratio of the second syllable to the whole word: *dur2ratio*= dur1/durtotal; the duration ratio of the first syllable to the second syllable:*Dur1Rdur2*= dur1/dur2;

Where *dur1, dur2, durtotal* are the durations of the first syllable, the second syllable and the whole word.

Each mother's F_0 was normalized by using the z-score method.

ZF_0 *z-score*=$(F_0-\mu_{F0}) / \sigma_{F0}$;

Where μ_{F0} is the mean value of a mother's all F_0 and σ_{F0} is the standard deviation.

ZF_0max, ZF_0min are the maximum and minimum ZF_0 of each word, ZF_0rang is the varying range got by ZF_0max-ZF_0min. $ZF_0meanS1$, $ZF_0meanS2$ are the mean of the first and second syllable.
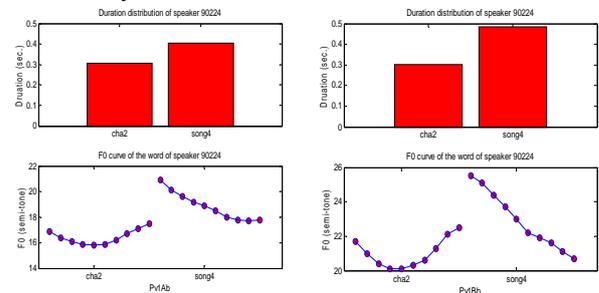

Figure 1: *Examples of prosodic measures of verb 'cha2 song4' in "wo ye chasong" (I also cha2song4) uttered in two separate times by one speaker*

3.2. Prosodic analysis for isolated words

3.2.1. F_0 and Duration of isolated homophone nouns and verbs
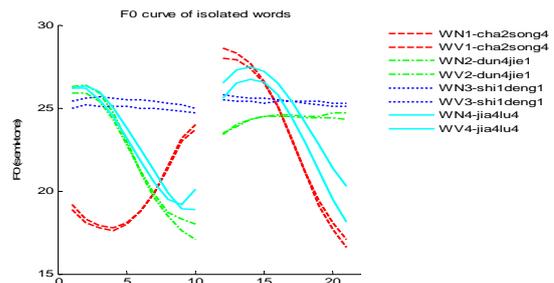

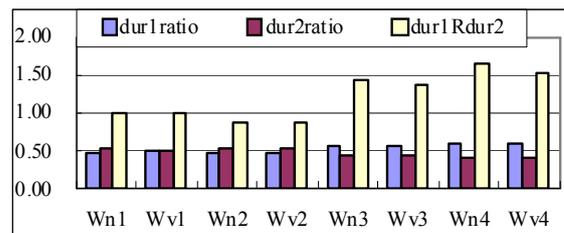Figure 2:F_0 contours of isolated homophone nouns and verbs

Figure 3: *Duration of isolated homophone nouns and verbs*

Figure 2 is the mean F_0 contours of 8 isolated homophone nouns and verbs: Wn1-Wn4 and Wv1-Wv4, the noun and verb pairs show similar contours. F_0 range is about 12st for 'W1:cha2song4' and 'W4:jia4lu4', and 9st for 'W2:dun4 jie1' (the first falling tone is realized as a half falling tone).

Figure 3 is the duration ratio of the isolated homophone nouns and verb, showing that the homophone nouns and verbs have identical duration relations. First syllables of W3 and W4 are longer than the second syllables.

3.2.2. Statistic analysis on prosodic cues of isolated homophone nouns and verbs

Prosodic cues, including F_0 and duration of isolated nouns and verbs, were statistically analyzed. Paired Samples Test was conducted for the prosodic cues of all nouns and verbs of each speaker. Since W3 has a H+H tone combination, the ZF_0min values were not compared. Results show that all prosodic measures were non-significant between noun and verb productions. That is, ZF_0max (p=.54), ZF_0range (p=.89), ZF_0mean1 (p=.39), ZF_0mean2 (p=.32), dur1R (p=.33), dur2R (p=.33) and dur1Rdur2 (p=.37) were not significantly different between homophone nouns and verbs. Therefore, even though the Chinese characters clearly revealed the noun versus verb status of each syllabic morpheme, the mothers did not produce any prosodic differences for the targets in citation forms.

3.3. Prosodic analysis for homophone nouns and verbs in phrases

3.3.1. F_0 and duration of homophone nouns and verbs in carrier phrases

Figures 4 and 5 show Dur1R and Dur1Rdur2 for noun and verb productions. Mean Dur1R appears greater for nouns than for verbs. Dur1Rdur2 appears greater for nouns than for verbs.

Figure 6 shows the F_0 contours of homophone nouns and verbs in the phrasal condition. The contours were plotted based on the average values of all speakers. Based on the visualization of the contours of the four homophone word pairs, F_0 of the second syllable of phrasal noun appears higher than that of the phrasal verb.
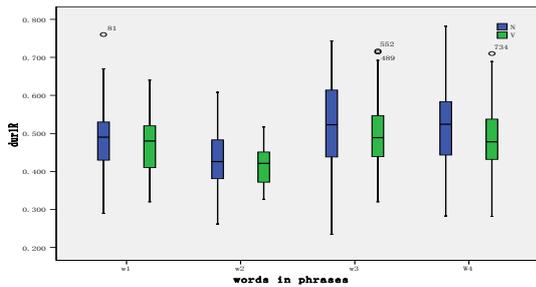

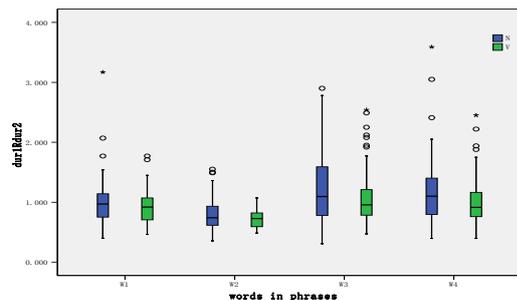
Figure 4: *Dur1R of phrasal homophone nouns and verbs*

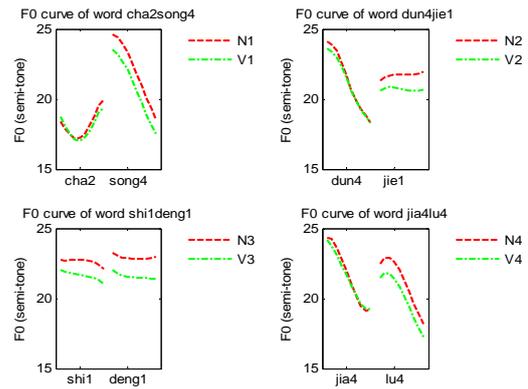Figure 5: *Dur1Rdur2 of phrasal homophone nouns and verbs*

Figure 6: *F_0 contours of phrasal homophone nouns and verbs (N1-N4 and V1-V4).The F_0 values on the contours are the means of all speakers.*

3.3.2. Statistical analysis for prosodic cues of phrasal nouns and verbs

As mentioned earlier, every mother read 24 phrasal nouns and verbs, but some of syllables were creaky voiced. These creaky sounds were excluded from the analysis. Only 13 mothers' prosodic data were used in the paired sample T-Test, the results are shown in table 4. ZF_0mean2, the mean F_0 of the second syllable, displays significant difference between the phrasal nouns and verbs (p <0.05), and Dur1Rdur2, the duration ratio of two syllables, tended to be greater for phrasal nouns than for phrasal verbs (p=0.073).

Table 4. *Paired sample-Test of prosodic properties of target words in noun carrier phrases versus verb carrier phrases*

paired Prosodic cues	t	df	Sig. (2-tailed)
ZF_0max	1.477	13	.164
ZF_0range	.509	13	.619
ZF_0mean1	.740	13	.472
ZF_0mean2	2.173	13	.049**
dur1R	1.675	13	.118
dur2R	-1.675	13	.118
dur1Rdur2	1.950	13	.073*

ONEWAY ANOVA was then conducted on prosodic cues of each phrasal homophone noun and verb pair for the two measures that showed significant and near significant differences, i.e., $ZF_0meanS2$, Dur1Rdur2. The results are shown in table 5.

Table 5. *ANOVA for individual pairs of phrasal homophone nouns and verbs for the mean F_0 of the second syllable ($ZF_0meanS2$) and the duration ratio (Dur1Rdur2)*

	Prosodic cues	F	Sig.
WN1 & WV1	$ZF_0meanS2$	5.874	0.016**
	Dur1Rdur2	1.714	0.192
WN2 & WV2	$ZF_0meanS2$	4.104	0.044**
	Dur1Rdur2	6.98	0.009**
WN3 & WV3	$ZF_0meanS2$	7.77	0.006**
	Dur1Rdur2	5.11	0.025**
WN4 & WV4	$ZF_0meanS2$	7.038	0.009**
	Dur1Rdur2	5.46	0.021**

In sum, the prosodic cues of the mean F_0 of the second syllable ($ZF_0meanS2$) and the duration ratio (Dur1Rdur2) were the main distinctive cues found between phrase homophone nouns and verbs. The analysis of individual non-word pairs shows that the $ZF_0meanS2$ was significantly

distinctive for all noun and verb pairs. For the measure of Dur1Rdur2, significant differences of noun versus verb productions were found in 3 out of 4 non-word pairs.

4. Discussion and conclusion

The morpho-syntactic structures of Chinese are distinct from those of European languages. Grammatical categories in Chinese are not defined morphologically. Shen proposed recently that Chinese grammatical categories are constructed distinctively from those of Indo-European, in that Chinese displays a kind of "constitution" relation between categories, while English reveals a 'realization' relation [11, 12]. This suggestion can explain why Chinese lacks morphological forms from the underlying cognition system. Shen illustrates this idea with the following examples. When the concept of 'tiger' is expressed linguistically, Chinese realizes it as one common noun word form '老虎' (laohu), while the concept is realized as multiple linguistic expressions in English such as 'a tiger', 'the tiger', 'tigers'. The following three English sentences can correspond to one Chinese sentence '老虎生活在丛林中' (Laohu shenghuo zai conglin zhong; tiger-live-at-jungle-in). The English verb 'live' here has two realizations, 'live', 'lives'. In contrast, Chinese employs only one verb form '生活' (live).

The tiger lives in the jungle.
A tiger lives in the jungle.
Tigers live in the jungle.

Thus, grammatical categories in English respect morphological relations across syntactic constituents. The grammatical categories themselves (such as nouns and verbs) can be realized in both morphological marking and in syntactic relations. Grammatical categories in Chinese, on the other hand, are realized in terms of syntactic constructions, rather than morphological marking.

These distinct characteristics have direct implications to the acquisition of grammatical categories. Unlike English-learning infants, Chinese-learning infants cannot rely on morphological markings due to the lack of inflectional morphology in Chinese. Other types of evidence must be available in the input for Chinese-learning infants to construct grammatical categories. For example, prosodic and/or phonological cues to grammatical categories, or the combination of these cues with distributional cues may be important for grammatical category formation [12, 13].

The present study examined the prosodic properties of the production of nouns versus verbs in infant-directed speech in Mandarin Chinese, using well controlled stimuli. Acoustic analyses revealed that while isolated homophone nouns and verbs do not have distinct prosodic cues, nouns and verbs in multi-word phrasal structures are marked by distinctive F_0 and duration cues in parental speech input. It is relevant to note that parental speech to infants contains primarily multi-word utterances [13]. In our experiment we found that the mean F_0 of the second syllable were different for phrasal homophone nouns and verbs, and the duration ratio of the two syllables also tended to differ. These prosodic cues can potentially play an essential role in the initial categorization of nouns and verbs in Chinese-learning infants. We are presently conducting perceptual experiments with Chinese infants to determine the types of information (distributional, prosodic or phonological) that infants rely on to categorize nouns and verbs.

5. References

[1] Braine, M., Children's first word combinations. In Monographs of the Society for Research in Child Development, 41, No. 1. 1976.

[2] Christophe, A., Guasti, T., Nespor, M., Dupoux, E., & Ooyen, B. V. Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. Language and Cognitive Processes, 12, 585–612. 1997.

[3] Conwell, E., & Morgan, J.L. (submitted). When parents verb nouns: Resolving the ambicategoricality problem.

[4] Gentner, D. (1982). Why nouns are learned before verbs : Linguistic relativity versus natural partitioning. In S. Kuczaj II (Ed.), Language development, Volume 2: Language, thought and culture. Hillsdale, NJ: Lawrence Erlbaum, 1982.

[5] Kelly, M. H. Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. Psychological Review, 99. 349–364. 1992.

[6] Lee, T., The acquisition of syntactic categories in Chinese: issues of bootstrapping and productivity. International Workshop on Grammar & Evidence, Taiwan. 2007.

[7] Maratsos, M. P., & Chalkley, M. A. The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (ed.), Children's Language (vol. 2). New York: Gardner Press, 127–214, 1980.

[8] Mintz, T. H. Frequent frames as a cue for grammatical categories in child directed speech. Cognition, 90, 91–117. 2003.

[9] Monaghan, P., Christiansen, M.H., Chater, N. The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. Cognitive Psychology, 55, 259–305.2007.

[10] Pinker, S. Language Learnability and Language Development. Cambridge, MA: Harvard University Press. 1984

[11] Shen, J., Chinese none and verb, HaZangYu Xuebao, first issue, Shanwu Yinshuguan, 2007.

[12] Shen, J., My View of Word Classes in Chinese，IACA-16，2008.

[13] Shi, R., Morgan, J., & Allopenna, P., Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. Journal of Child Language, 25, 169-201, 1998.

[14] Shi, R., Werker, J., & Morgan, J. Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. Cognition, 72, B11-21. 1999.

[15] Shi, R. Early syntactic categories in infants' language. In Cohen, Henri & Claire Lefebvre, (eds.), Handbook of categorisation in cognitive science (pp. 481-495). Elsevier. 2005.

[16] Shi, R., and Moisan, A., Prosodic Cues to Noun and Verb Categories in Infant-Directed Speech, 2008.

[17] Shi, R. 1996. Perceptual correlates of content words and function words in early language input (Ph.D. dissertation, Brown University, 1995). *Dissertation Abstracts International, 56, 3108A.*

[18] Zhang, Z., Shi, R., Li, A., Zheng, L., Rudimentary Form Class Categorization of Nouns and Verbs in Chinese-Learning Infants, NCMMSC 2009.