Do cognitive constraints drive second-language listeners' attention to prosodic information in speech?

Yuanyuan Zhang[1, 2], Hongwei Ding[1]*, Hui Zhang[1]

[1]Institute of Cross-Linguistic Processing and Cognition, School of Foreign Languages
Shanghai Jiao Tong University, China
[2]The MARCS Institute and ARC Centre of Excellence for the Dynamics of Language, Western Sydney University, Penrith South, NSW 2751, Australia
yyuan.zhang@sjtu.edu.cn, hwding@sjtu.edu.cn

ABSTRACT

Can Chinese learners of English use prosodic cues, as native listeners do, to resolve syntactic ambiguity in spoken-language comprehension? This question was addressed with two speech perception tasks, one in which listeners were aware of potential ambiguity, and one in which they were unaware of the ambiguity.

Accuracy scores showed that learners' use of prosodic information in disambiguation depended on awareness of ambiguity. Without awareness, the listeners' preference for one meaning over the other could not be reversed by the presence of prosodic cues; after receiving explicit information that there was ambiguity, however, listeners succeeded in identifying the intended utterance meaning by exploiting the available prosody. This result suggests that L2 learners' often-reported failure to use prosodic information in listening should not be ascribed to inability to use such information, but may instead result from a deliberate choice to restrict levels of processing, in order to simplify the listening task.

Keywords: syntactic ambiguity, prosodic disambiguation, L2 language comprehension.

1. INTRODUCTION

Prosody can convey multiple types of information in speech communication. To get the intended meaning of the speaker, the listener needs to recognize the lexical words, to assign the syntactic structure of the utterance, and to determine the semantic relation between words [1, 2]. Syntactic structure determines the relation of individual words in the sentence. The structural representation built up from sentence components also determines the sentence meaning [3]. The mapping between prosodic structure and syntactic structure can group words or phrases into utterances, thus deciding the relation between these components in spoken language. Therefore, an effective comprehension of spoken language requires to integrate multiple types of information. This integration exploits cognitive resources, but a limited attentional capacity might constrain these cognitive operations [4, 5].

Syntactic ambiguity, especially the type in which words can be grouped in alternate ways, allows multiple interpretations. Ambiguity resolution has been central to investigate the mechanisms of language comprehension. For example, it is argued that a full reanalysis of garden-path sentences requires more cognitive resources, so that readers settle for "Good Enough" interpretations to avoid costly reprocessing [3, 6, 7].

Prosodic phrasing can resolve structural ambiguity in native (hereafter L1); both speakers and listeners can identify such disambiguating prosodic cues, with correct identification rates ranging from 70% to 84% for listeners in different tasks [8–14]. Prosodic information has been shown to affect the early stages of ambiguity interpretation [15–19], and even when a sentence is semantically biased to one interpretation, disambiguating prosodic cues can reverse listeners' preference [20]. Studies of event-related potentials (ERPs) found that prosodic information guides syntactic analysis at the initial stage, and suffices to override syntactic parsing preferences [21–23]. It has been shown that in resolving prepositional phrase-attachment (PP-attachment) ambiguity, second language (L2 hereafter) learners tend to ignore prosodic cues because of interpretative bias or when other cues are available, except when they listen explicitly for prosodic violations [24]. Other studies found that learners can refer to prosodic cues in disambiguation, though their specific use of these correlates is different from that of the native speakers [25-28].

Studies showed that Chinese L2 learners of English showed different ERP effects from the native speakers [29], and asymmetric accuracy rate and reaction time in prosodic disambiguation have been reported in Chinese L2 English and native English [30]. Multiple theories and hypotheses have been proposed to explain L2 language processing. It is suggested that target language processing is less automatic and more effortful than L1 processing [31]. This study will investigate whether Chinese

*Corresponding author

learners of English can use prosodic cues in L2 spoken language comprehension before and after they were informed of the syntactic ambiguity.

2. METHOD

The present study is composed of three tasks, using stimuli randomly selected from the previous production study [32]. Task 1 is a speech perception task in which learners were asked to select the right picture to match the meaning of the utterances they heard through headphones. Task 2 is the same as Task 1, except that the sentences were presented in text on the screen instead. Task 3 is the same as Task 1 except that subjects received explicit information about the ambiguity.

2.1. Participants

Thirty Chinese learners of English who were undergraduates from Shanghai participated in our experiment, and were paid for their participation. All were native speakers of Mandarin Chinese aged between 18 and 21, and had been learning English for approximately 10 years. None reported any speech or hearing disorders.

2.2. Materials

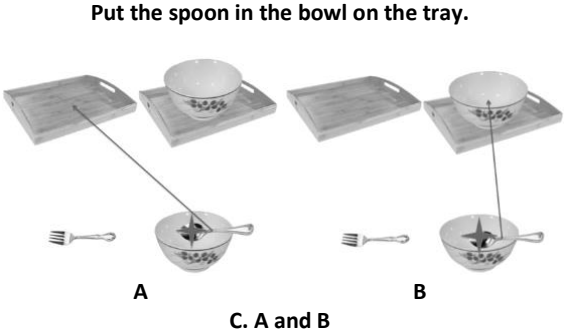
The target stimuli in this study were 10 pairs of utterances with PP-attachment ambiguity, and each of the pair was randomly selected from each native English speaker [32]. The ambiguity is embodied in the structure "*Put NP1 in NP2 on NP3*", which can be interpreted as "*Put NP1 / in NP2 on NP3*" (early juncture marked with condition A hereafter) or "*Put NP1 in NP2 / on NP3*" (late juncture marked with condition B hereafter) with different prosodic phrasings. All the utterances were appropriately disambiguated by inserting a pause at the boundary and lengthening the pre-boundary component, and have been correctly recognized by the phonetically trained confederate [32]. In addition, 15 filler items, which had similar sentence structure as the target ones but no syntactic ambiguity, were also constructed. One female native speaker who participated in the production task uttered the filler items. Totally, 35 utterances were used in this study.

The stimuli were grouped into two blocks in each task to avoid the fatigue of subjects. In each block, only one member of each pair appeared with mixed assignment of half condition-A and half condition-B utterances. Target items and filler items appeared in a pseudo-random order, with at least one filler item being inserted between two target items. The order of picture selections and answers were counterbalanced in each block.

In Task 1 and Task 3, learners were presented with two pictures indicating the alternative interpretations of the utterance that they would hear through headphones. In Task 2, sentences were presented in text on the screen. Examples of the target display in Task 1 and Task 3 are shown in Figure 1, and that of Task 2 is shown in Figure 2. To familiarize them with the experimental paradigm, the subjects should first complete 6 practice items with no syntactic ambiguity. All sounds and pictures were programmed into E-prime (2.0).

Figure 1: Schematic sample set of the target display for sentence "Put the spoon in the bowl on the tray" in Task 1 and Task 3.

Figure 2: Schematic sample set of the target display for sentence "Put the spoon in the bowl on the tray" in Task 2.



2.4. Procedure

Three forced-choice identification tasks were run on a laptop via E-prime (2.0). Subjects were tested individually in a quiet room. Sound stimuli were presented over headphones (AKG, K240MkII), and text and picture were displayed on the screen.

2.4.1. Task 1

In the initial familiarization phase, the experimenter firstly explained the procedure to the subjects orally. In the first block, the subjects would see the experiment instruction on the screen informing them of the procedure and their task. Before each trial, subjects would see a red fixation signal (i.e. "+") in the middle of the screen, and meanwhile hear a beep via the headphones (AKG, K240MkII) which would

last for 500 ms. Then they would hear the utterance, and see pictures meanwhile. Their task was to select the right picture matching the utterance meaning with prosodic cues by pressing "q" for answer "A", "p" for answer "B" as soon as they understood the whole utterance. The audio sound was played once, while the pictures would remain on the screen until response was made. Subjects were told to listen to the utterances and to look at the pictures carefully in each display.

2.4.1. Task 2 and Task 3

Task 2 and Task 3 were conducted one week after Task 1 was finished. Subjects in Task 1 were not informed about doing Task 2 and Task 3 a week later, so two subjects were not brought back when being informed of the new tasks. In Task 2, subjects were told to look at the pictures on the screen and to read the sentence above in each display as shown in Figure 2. Their task was to judge which picture corresponded to the meaning of the sentence according to their understanding by pressing "q" for answer "A", "p" for answer "B", and "c" for both answers A and B.

After Task 2, the experimenter asked questions concerned with the ambiguity, finding that only five of the subjects knew how to interpret the stimuli alternatively. The experimenter then gave them explicit information about the ambiguity, but no prosodic cues were provided. After the subjects understood the structural ambiguity and had a short rest, Task 3 was conducted in the same procedure as in Task 1.

2.2. Data analyses

Subjects' responses to practice trials and filler trials were not included in data analyses. As two of the subjects who participated in Task 1 did not take part in Task 2 or Task 3, their responses were discarded in data analyses. Thus, a total of 560 responses were obtained (28 participants * 2 conditions * 10 sentences) for Task 1 and Task 3 respectively, and 280 (28 participants * 10 sentences) responses were obtained for Task 2. Analyses and comparisons were initially made among the means of accuracy score. Then statistical analyses were carried out using R with logistic mixed-effects model to analyse the accuracy score in Task 1 and Task 3. The model included subjects' response accuracy score (1 for congruent vs. 0 for incongruent) as dependent variable; task, condition, and the interaction of task and condition as fixed predictors. Random effects were also included for participants and sentences. In Task 1 and Task 3, A or B reflected subjects' accurate understanding of the utterance when prosody was imposed, and the total is an average of their accurate selection. While in Task 2, the total indicated subjects' understanding of ambiguity in these tokens, and A or B showed their preference for one interpretation over the other when no prosodic cues were provided. Thus, we did not compare the result in Task 2 with those in Task 1 and Task 3 in statistical analyses.

3. RESULTS

The accuracy rate for each task is summarized in Table 1. In Task 1 and Task 3, A and B indicate subjects' accurate understanding of the utterance. In Task 2, the total reflects subjects' understanding of ambiguity, A or B indicates their interpretative bias.

Table 1: Accuracy rate for Task 1, Task 2, and Task 3.

Task	A (%)	B (%)	Total (%)
1	65.7	42.9	54.3
2	42.5	34.3	23.2
3	74.6	89.3	82.0

It can be observed that in Task 1, when heard target utterances, subjects showed a preference for A (65.7%). The total accuracy rate 54.3% indicates their failure to use prosodic cues in ambiguity resolution. Total accuracy rate in Task 2 (23.2%) shows that subjects even did not realize the ambiguity with the target sentences. Rather, accuracy rates for A (42.5%) and B (34.3%) indicate their interpretative bias for A. Accuracy rates for A (74.6%), B (89.3%), and the total (82%) in Task 3 suggest that they could identify the intended meaning by exploiting the available prosody after being informed of the ambiguity.

Logistic mixed-effects models were conducted to analyse the accuracy score in Task 1 and Task 3. Results show significant main effects of task (β = -.727, SE = .075, $p < 0.001$) and the interaction of task and condition (β = .507, SE = .075, $p < 0.001$), though no significant effect is found in condition (β = -.023, SE = .074, p = .757). To explore the interaction of task and condition, post-hoc pairwise comparisons were computed, and the results are presented in Table 2. It can be observed that in Task 1 the accuracy score for A is significantly higher than that for B ($p < .0001$), while in Task 3 it is significantly higher for B than that for A ($p < .001$). The subjects' preference for A in Task 1 tends to show a preference over the early juncture interpretation, while the significant higher accuracy score for B in Task 3 indicates that prosodic cues can reverse their interpretation preference. Thus, it is suggested that listeners could exploit the available

prosodic information to distinguish the alternative interpretation of the ambiguous structure.

Table 2: Results of post-hoc pairwise comparisons.

Contrast	*β*	*SE*	*z*. ratio	*p*
1A - 1B	0.968	0.178	5.447	<0.0001
1A - 3A	-0.441	0.189	-2.335	0.090
3A - 3B	-1.060	0.239	-4.436	<0.001
1B - 3B	-2.468	0.232	10.652	<0.0001

In text-reading task, when Chinese learners of English were not informed of the ambiguity, most of them did not know there might be two alternative interpretations for one sentence. In listening comprehension task when they did not receive instructions about ambiguity, most of them preferred an early juncture interpretation for the ambiguity, suggesting that they did not attend to prosodic cues in sentence interpretation. After being informed of the ambiguity, however, most of them could employ the available prosodic cues to identify the intended meaning of the ambiguous sentences.

4. DISCUSSION AND CONCLUSION

In this paper, the role of prosody in L2 spoken language comprehension was investigated with two speech perception tasks and one text-reading task. Accuracy scores were analysed to show whether learners could attend to prosodic cues in speech processing. Task 1 (speech perception) showed that learners failed to resolve the ambiguity with prosodic cues, and they tended to interpret the ambiguous PP-attachment with an early juncture, while ignoring prosodic cues. Results from the text-reading task in Task 2 and a brief interview with the subjects suggested that learners even had no awareness of the ambiguity, and they did not realize that there might be two alternative interpretations for one sentence. Nevertheless, in Task 3 after receiving explicit information about ambiguity with the sentence structure, the learners could attend to prosodic cues, and reversed their interpretative bias to resolve the syntactic ambiguity. Statistical analyses also revealed significant differences in the accuracy scores among these tasks. This result is consistent with that of Berkovits [24] in which learners were found to ignore prosodic information when other types of information were available unless they listened explicitly for it.

Effective comprehension of spoken language requires integration of all available information, resulting in a greater computational burden within the limits of cognitive resources [33]. One possible reason why the learners failed to attend to prosodic cues in Task 1 is that they did not notice the ambiguity at all. They tended to choose an interpretation to avoid reprocessing that required greater cognitive resources. It is also assumed in the Good Enough approach that listeners tend to build up structure and to interpret sentences in a quick and efficient manner [7]. Task 2 provides evidence for this possibility. In Task 3, the learners could attend to the available prosodic cues, indicating that limited attentional capacity might constrain their attention to prosodic information in Task 1 and Task 2.

Our result suggests that learners have problems in information integration. This might due to the fact that they even did not notice the ambiguity, and therefore tended to interpret the sentence as quickly and efficiently as possible, to avoid using more cognitive resources and incurring a greater computational burden. Different types of ambiguity and experiment paradigms might cause the differences between this result and other studies. In response to the original question, our data indicates that the learners' failure to use prosodic cues in disambiguation might due to their cognitive constraints, but not an inability to process the available prosodic information.

5. ACKNOWLEDGEMENT

Our sincere thanks go to the participants and professor Anne Cutler for her insightful comments. This study is sponsored by Shanghai Social Science project (2018BYY003), the interdisciplinary program of Shanghai Jiao Tong University (14JCZ03), and the Chinese Scholarship Council.

6. REFERENCES

[1] Cutler A., Dahan D., and van Donselaar W., "Prosody in the comprehension of spoken language: a literature review.," *Lang. Speech*, vol. 40, no. 2, pp. 141–201, 1997.

[2] Caplan D., Alpert N., and Waters G., "PET studies of syntactic processing with auditory sentence presentation," *Neuroimage*, vol. 9, no. 3, pp. 343–351, 1999.

[3] Slattery T. J., Sturt P., Christianson K., Yoshida M., and Ferreira F., "Lingering misinterpretations of garden path sentences arise from competing syntactic representations," *J. Mem. Lang.*, vol. 69, no. 2, pp. 104–120, 2013.

[4] Wingfield A. and Grossman M., "Language and the aging brain: patterns of neural compensation revealed by functional brain imaging," *J. Neurophysiol.*, vol. 96, no. 6, pp. 2830–2839, 2006.

[5] Wingfield A. and Tun P. A., "Cognitive supports and cognitive constraints on comprehension of

spoken language," *J. Am. Acad. Audiol.*, vol. 18, no. 7, pp. 548–558, 2007.

[6] Christianson K., Williams C. C., Zacks R. T., and Ferreira F., "Misinterpretations of garden-path sentences by older and younger adults," *Discourse Process.*, vol. 42, no. 2, pp. 205–238, 2006.

[7] Ferreira F., Bailey K. G., and Ferraro V., "Good-enough representations in language comprehension," *Curr. Dir. Psychol. Sci.*, vol. 11, no. 1, pp. 11–15, 2002.

[8] Tree J. E. F. and Meijer P. J., "Untrained speakers' use of prosody in syntactic disambiguation and listeners' interpretations," *Psychol. Res.*, vol. 63, no. 1, pp. 1–13, 2000.

[9] Kraljic T. and Brennan S. E., "Prosodic disambiguation of syntactic structure: For the speaker or for the addressee?," *Cogn. Psychol.*, vol. 50, no. 2, pp. 194–231, 2005.

[10] Lehiste I., "Phonetic Disambiguation of Syntactic Ambiguity," *J. Acoust. Soc. Am.*, vol. 53, no. 1, pp. 380–380, 1973.

[11] Price P. J., Ostendorf M., and Park M., "The use of prosody in syntactic disambiguation," *J. Acoust. Soc. Am.*, vol. 90, no. December 1991, pp. 2296–2970, 1991.

[12] Schafer A., "Prosodic parsing: The role of prosody in sentence comprehension," University of Massachusetts at Amherst, 1997.

[13] Snedeker J. and Trueswell J., "Using prosody to avoid ambiguity: Effects of speaker awareness and referential context," *J. Mem. Lang.*, vol. 48, no. 1, pp. 103–130, 2003.

[14] Warren P., Schafer A. J., Speer S. R., and White S. D., "Prosodic resolution of prepositional phrase ambiguity in ambiguous and unambiguous situations," 2000.

[15] Beach C. M., "The interpretation of prosodic patterns at points of syntactic structure ambiguity Evidence for cue trading relations," *J. Mem. Lang.*, vol. 30, pp. 644–663, 1991.

[16] Kjelgaard M. M. and Speer S. R., "Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity," *J. Mem. Lang.*, vol. 40, no. 2, pp. 153–194, 1999.

[17] Marslen-Wilson W. D., Tyler L. K., Warren P., Grenier P., and Lee C. S., "Prosodic effects in minimal attachment," *Q. J. Exp. Psychol.*, vol. 45, no. 1, pp. 73–87, 1992.

[18] Nagel H. N., Shapiro L. P., Tuller B., and Nawy R., "Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing.," *J. Psycholinguist. Res.*, vol. 25, no. 2, pp. 319–344, 1996.

[19] Speer S. R., Kjelgaard M. M., and Dobroth K. M., "The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities.," *J. Psycholinguist. Res.*, vol. 25, no. 2, pp. 249–71, 1996.

[20] Misono Y., Mazuka R., Kondo T., and Kiritani S., "Effects and limitations of prosodic and semantic biases on syntactic disambiguation," *J. Psycholinguist. Res.*, vol. 26, no. 2, pp. 229–245, 1997.

[21] Pannekamp A., Toepel U., Alter K., Hahne A., and Friederici A. D., "Prosody-driven sentence processing: An event-related brain potential study," *J. Cogn. Neurosci.*, vol. 17, no. 3, pp. 407–421, 2005.

[22] Pauker E., Itzhak I., Baum S. R., and Steinhauer K., "Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis," *J. Cogn. Neurosci.*, vol. 23, no. 10, pp. 2731–2751, 2011.

[23] Steinhauer K., Alter K., and Friederici A. D., "Brain potentials indicate immediate use of prosodic cues in natural speech processing," *Nat. Neurosci.*, vol. 2, no. 2, pp. 191–196, 1999.

[24] Berkovits R., "Perception of intonation in native and non-native speakers of English," *Lang. Speech*, vol. 23, no. 3, pp. 271–280, 1980.

[25] Hwang H. and Schafer A. J., "Prosodic effects in parsing early vs . late closure sentences by second language learners and native speakers," in *Speech Prosody*, 2006, no. September, pp. 6–9.

[26] Nickels S. and Steinhauer K., "Prosody–syntax integration in a second language: Contrasting event-related potentials from German and Chinese learners of English using linear mixed effect models," *Second Lang. Res.*, vol. 34, no. 1, pp. 9–37, 2018.

[27] Nickels S., Opitz B., and Steinhauer K., "ERPs show that classroom-instructed late second language learners rely on the same prosodic cues in syntactic parsing as native speakers," *Neurosci. Lett.*, vol. 557, no. PB, pp. 107–111, 2013.

[28] Holzgrefe-Lang J. *et al.*, "How pitch change and final lengthening cue boundary perception in German: converging evidence from ERPs and prosodic judgements," *Lang. Cogn. Neurosci.*, vol. 31, no. 7, pp. 904–920, 2016.

[29] Li X., Yang Y., and Lu Y., "How and when prosodic boundaries influence syntactic parsing under different discourse contexts: An ERP study," *Biol. Psychol.*, vol. 83, no. 3, pp. 250–259, 2010.

[30] Ip M. H. K. and Cutler A., "Asymmetric Efficiency of Juncture Perception in L1 and L2," in *Speech Prosody*, 2018, no. June, pp. 289–293.

[31] Reichle R. V., Tremblay A., and Coughlin C., "Working memory capacity in L2 processing," *Probus*, vol. 28, no. 1, pp. 29–55, 2016.

[32] Zhang Y. et al., "Prosodic disambiguation by Chinese EFL learners in a cooperative game task," in *Speech Prosody*, 2018, no. June, pp. 979-983.

[33] Fultz A. L., "The Use of Prosody for Disambiguation in English-French Interlanguage," in *In 9th Generative Approaches to Second Language Acquisition Conference*, 2008, no. Gasla 2007, pp. 130–139.