Asymmetry in L1 and L2 listeners’ use of prosody for PP-attachment disambiguation

Yuanyuan Zhang\textsuperscript{1,2}, Hongwei Ding\textsuperscript{3*}

\textsuperscript{1}School of Foreign Languages, East China University of Science and Technology, Shanghai, China
\textsuperscript{2}MARCS Institute, University of Western Sydney, Penrith, NSW, Australia
\textsuperscript{3}Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China
evelyn_yuanyuan@163.com, hwding@sjtu.edu.cn

Abstract

The role of prosody in sentence processing has received little attention in L2 spoken-language comprehension research, and studies of prosodic disambiguation of PP-attachment ambiguity by L2 learners are rare. We investigated the effects of prosody (accompanied by visual context from the pictures) on PP-attachment ambiguity resolution by native Australian English speakers and Chinese-speaking learners of English, comparing their attachment preferences to examine L1-L2 sentence processing differences. Our results show that prosody can guide both native speakers and L2 learners towards the resolution of sentence-level ambiguity. However, the observed effects varied. For native speakers, prosody overrode bias for the NP-attachment induced by the visual context; L2 learners in contrast showed a preference for the VP-attachment independent of prosodic cues, indicating that they did not detect the ambiguity within the structure from the visual context and that it is difficult for them to effectively integrate different domains of information in ambiguity resolution. These L1-L2 differences might reflect L2 learners’ limited processing capacity for detecting ambiguity and integrating multiple domains of information during L2 sentence processing.

Index Terms: PP-attachment ambiguity, prosody, syntax, L2 learners, sentence processing

1. Introduction

Prosodic cues (e.g., prosodic boundary) have been shown to guide listeners to resolve the structural ambiguity at the initial stage of sentence processing and override their parsing bias [1, 2]. Moreover, native speakers can effectively integrate prosodic information and other domains of information to resolve the ambiguity [3, 4, 5, 6, 7, 8]. In comparison to the multitude of research on native language (L1), not much work has investigated the effects of prosody on the second language (L2) learners’ spoken language comprehension, especially in the presence of visual context, and no consensus has been reached with regard to how L2 learners make use of prosodic information in sentence processing.

Previous studies have reported that although L2 learners were able to access multiple types of information in sentence processing, such as lexical [9], semantic [10], prosodic [11, 12, 13], and visual context information [4, 6], they were slower and less accurate [14, 15, 16, 17, 13]. L2 learners have been reported to use prosodic information to resolve globally ambiguous sentences [18], but they might rely more heavily on prosodic rather than syntactic information, particularly when interpreting ambiguous sentences with conflicting prosodic and syntactic cues [11]. There is also evidence showing that L2 learners experienced difficulty in integrating prosodic information and other domains of information, and they were less able to integrate multiple domains of information during on-line sentence processing compared to native speakers [14, 19, 4, 6, 17]. In addition, L2 learners’ sentence processing was affected by the ambiguity structure and task demands as well. Those who could perceive prosodic cues might not be able to effectively integrate prosodic information with other types of information in complex structures [20], and the degree to which L2 processing converges to the native depends on the complexity of the task [12].

To account for the existing differences between native speakers and L2 learners, the Shallow Structure Hypothesis (SSH) proposes that L2 learners’ parsing is ‘shallower’ than that of native speakers in that it is more likely to be affected by lexical, semantic, and discourse context information, whereas that of the native speakers’ is more syntactically driven [21, 22, 23, 24, 25]. The Interface Hypothesis suggests that L2 learners have difficulty when sentence processing involves an interface between syntax and other domains of information [26]. L2 learners’ processing has been shown to be less efficient when syntactic information needs to be integrated and coordinated with other domains of information, such as contextual discourse-pragmatic information. The present study aims to investigate whether Chinese adult L2 learners can utilize prosodic information to resolve PP-attachment ambiguity accompanied with visual context, and to delve into the differences in prosodic disambiguation by L2 learners and native speakers.

2. Method

The present study involved both reading (Task 1) and listening tasks (Task 2). As Chinese-speaking learners of English tend to learn American English since primary school, the stimuli for the L2 learners were recorded by a female native English speaker from America, and for the L1 by a female native English speaker from Australia. Both of the native speakers were phonetically trained to disambiguate these ambiguous sentences with pause.

2.1. Participants

The participants were divided across the experimental group (L2) and the control group (L1). Thirty adult Chinese-
speaking learners of English and thirty native speakers of Australian English were recruited. The L2 learners were second-year university students, and they learned English as a foreign language in Shanghai, China (mean age = 19.367, range 19-22). All of them had passed CET-4 (College English Test Band 4) when they participated in the experiments, but none of them had studied English in any English-speaking country, or majored in English. The native speakers of Australian English were students at universities in Sydney, Australia (mean age = 23.667, range 19-36). No participants reported any speaking or hearing disorders.

2.2. Materials and design

Sixteen pairs of sentences with double prepositional phrases (PP-attachment ambiguity) were created based on previous studies [27, 28]. For instance, in Put the dog in the basket on the mat, the first prepositional phrase (PP1) in the basket can be attached to the verb phrase put (VP) (high VP-attachment) or the noun phrase the dog (NP) (low NP-attachment) to respectively mean Put the dog in the basket that is on the mat or Put the dog that is in the basket on the mat. In both tasks, each sentence was accompanied with two pictures so that participants could give their answers for their interpretation. Sixty-four additional filler items with a variety of structures but no ambiguity were also created based on the picture of the target sentences.

The same set of pictures was employed in both tasks for the target sentences. Each picture was composed of a set of objects, e.g. in the above sample sentence, object to be moved (dog), object as a modifier or goal (basket and mat), and an arrow indicating which object to move and where to move [28], in that the ambiguity within the picture supports both the VP- and NP-attachment of the sentence. This allowed us to present each pair of ambiguous sentences with the same picture display, and to exclude the effects of other cues in the listening task.

In task 1, the sentences were presented to participants in written text (see Figure 1 for the sample picture). All the target sentences had the same structure as the sample. In task 2, the sentences were presented to participants through headphones along with pictures on the screen. The pictures presented in Task 2 were the same as those in Task 1 except that no text was provided above the pictures and only two options (A and B) were available. Participants were asked to listen carefully to the utterances and to choose a picture that best interpreted its meaning.

The recording of the stimuli was performed in the professional recording studio. In total, thirty-two target utterances and sixty-four fillers were produced, and digitized at 44 kHz with 16-bit resolution into a computer. Because the speakers were trained to disambiguate these sentences with pause and duration, and duration pattern has been shown to play a greater role in the interpretation of structurally ambiguous strings and its effects on language users’ interpretation of the ambiguity have been identified in the absence of pitch variation [29, 30]; thus, durations of pause and pre-boundary word were extracted and analysed. The data extraction was performed with a Praat script ProsodyPro [31]. Since each sentence has different numbers of words and might be produced with different speech rates, the duration ratio instead of the absolute value was thus obtained by dividing the duration of the word from that of the entire utterance. All the sentences were regarded as a single structure for the VP- and NP-attachment respectively. For each interpretation, the duration ratio of each word was then averaged across the same component in all the target utterances.

As indicated in Figure 2, both American (AM) and Australian (AU) native English speakers employed the same prosodic cues to mark prosodic phrasing. They inserted a pause (P1) immediately after N1 and lengthened N1 in the VP-attachment. Meanwhile, a lengthening was also observed on N2 in the NP-attachment. Statistical analyses by means of Linear Mixed model fit by REML were implemented using ‘lmer’ function in the lme4 package in R [32]. Results showed that N1 in Australian English speaker’s speech was significantly longer in VP-attachment than its correspondent in NP-attachment (ß = -.074, t = -11.99, p < .001), and N2 in NP-attachment was significantly longer than its correspondent in VP-attachment (ß = .059, t = 12.19, p < .001). American English speaker employed the same prosodic features by significantly lengthening N1 in VP-attachment (ß = .059, t = 9.208, p < .001), and N2 in NP-attachment (ß = .045, t = 6.115, p < .001).

2.3. Procedure

Task 1 and Task 2 were administered on a laptop via E-prime (2.0), lasting about thirty minutes in total. The stimuli were presented to participants via slides on the screen in Task 1, and via a set of headphones in Task 2. Task 1 and Task 2 were conducted in two sessions, and in two blocks for Task 2 to avoid the participants’ fatigue. In each block of Task 2, only one member of each pair appeared with a mixed assignment of half VP-attachment and half NP-attachment. The target and
filler items appeared in a pseudo-random order, with at least one filler item being inserted between two target items, in order to discourage response strategies.

Participants were tested individually in a quiet room. They were seated comfortably in front of the computer monitor. At the beginning of the experiment, the experimenter showed participants the objects that would appear in the pictures and their English names on the slides, familiarizing them with the materials. Each session began with a short practice block containing six sentences to get participants familiarized with the experimental paradigm. In the initial familiarization phase, the experimenter explained the procedure to the participants orally but provided no cues to the ambiguity.

2.3.1. Task 1

In Task 1, participants were asked to read the sentences on the screen and to select the picture or pictures that matched the candidate meaning of the sentences. Before each trial, they would see a red fixation (i.e., “+”) in the middle of the screen which would last 300 ms. Then they would see a slide as shown in Figure 1. Participants were asked to pay careful attention to the meaning of each sentence and the differences between the two pictures and to give their answers after they understood the sentence by pressing the corresponding button on the keyboard. The participants’ responses were recorded. This task also familiarized participants with the testing materials in the listening task.

2.3.2. Task 2

In Task 2, the participants’ task was to choose the picture that best interpreted the intended meaning of the utterances that they heard through headphones, but only A and B were provided, indicating that only one picture could match the meaning of the utterance because the disambiguating prosody was provided. None of them received any explicit instructions on how to disambiguate the sentences. The slides and the sound stimuli appeared at the same time. The sounds were played only once, while the picture would remain on the screen until a response was made. There was no time limit for participants to respond. The participants’ responses were recorded.

3. Data analysis

Participants’ responses to the practice trials were not included in data analysis. In total, 480 responses (30 participants * 16 sentences) in Task 1 and 960 responses (30 participants * 2 conditions * 16 sentences) in Task 2 were collected from each group. Analyses and comparisons of the mean accuracy score were made between the VP- and NP-attachment and between the L1 and L2 groups. Statistical analyses were conducted using R with logistic mixed-effects models of regression probability [32], treating the choice of interpretation as a binomial variable because it is discontinuous. An alpha level of .05 was adopted throughout. For each measure, participants’ accuracy score was set as a fixed effect, with the correct answer coding as 1, and the incorrect answer coding as 0. Condition and group were set as fixed predictors respectively, and participants and items were simultaneously set as random effects because both of them were selected from large populations.

4. Results

Participants’ decisions for each choice in Task 1 are summarized in Figure 3. It can be observed that for native speakers, 51.7% of the target sentences were ambiguous, 25.2% were VP-attachment in which PP1 was interpreted as the goal of the verb, and 23.1% were NP-attachment where PP1 was the modifier of NP. For L2 learners, 54.8% of the target sentences were VP-attachment, 18.8% were NP-attachment, and only 25.2% were ambiguous. This suggests that in the reading task, most of the native speakers treated the sentences as ambiguous, and their preference for one interpretation over the other was not robust. In contrast, most of the L2 learners were not aware of the ambiguity and preferred the VP-attachment.

Figure 3: Mean score (%) of L1 and L2 readers’ preference for each interpretation of the ambiguous sentences in Task 1.

Two observations can be made in Task 2, as shown in Figure 4. First, native speakers had a higher accuracy rate for the NP-attachment (80.2%) than the VP-attachment (74.9%), whereas L2 learners showed a higher accuracy rate for the VP-attachment (78.1%) than the NP-attachment (60.8%). Second, native speakers performed better (77.53%) than L2 learners (69.45%) averagely. The results indicated that both native speakers and L2 listeners employed prosodic cues to resolve the ambiguous sentences, yet the effects of prosody varied.

Figure 4: Mean accuracy score (%) for each interpretation by L1 and L2 listeners in Task 2.

To further examine the differences in accuracy score in Task 2 between each interpretation, statistical analyses were conducted by means of the Generalized Linear Mixed model implemented using ‘glmer’ (binomial family: correct vs. incorrect) function in the lme4 package in R (Version 1.1.463) [32]. The dependent variable was response: correct (coded as 1) or incorrect (coded as 0), fixed effects were group (categorical predictor: L1 coded as 1 and L2 coded as -1), condition (categorical predictor, VP-attachment coded as 1 and NP-attachment as -1), and the interaction of group and condition. Participants and items were random intercepts. With the deviation coding of two-level categorical factors (i.e.,
coded as 1 and -1), the reference for categorical data was set to the overall mean, the coefficient of each individual category was calculated compared to the overall effect of the categorical variable, and the grand mean was mapped onto the intercept, allowing us to test the main effects of these predictors. The results are given in Table 1.

Table 1: Generalized Linear Mixed model for accuracy score.

| Task | Fixed effects | Estimate | SE | z-value | Pr (>|z|) |
|------|---------------|----------|----|---------|----------|
|      | Intercept     | 1.175    | .117 | 10.016  | .000     |
|      | Group         | .229     | .104 | 2.198   | .028     |
|      | Condition     | .142     | .075 | 1.889   | .065     |
|      | Group*Condition | -3.10   | .055 | 5.595   | .000     |

Significant main effects of group, the interaction between group and condition, and a marginally significant main effect of condition were revealed. The significant main effect of the group indicates a significantly higher accuracy rate by the L1 listeners than that by the L2 listeners. The interaction between group and condition suggests that L1 listeners showed lower accuracy rate in the VP-attachment, but higher in the NP-attachment, and the difference between the VP- and NP-attachment by L2 listeners was larger than that by L1 listeners.

Given the significant interaction effects in the previous model, Tukey post-hoc comparisons using lsmeans function in R were conducted to compare the accuracy score in each condition within and across each group [33]. Results were given on the log odds ratio scale. Table 2 summarizes the results: L2 listeners showed a significantly higher accuracy score in the VP-attachment than in the NP-attachment. However, L1 listeners did not show a significantly different accuracy score between the NP- and VP-attachment.

Table 2: Summary of post-hoc pairwise comparisons.

<table>
<thead>
<tr>
<th>Task</th>
<th>Contrast</th>
<th>Estimate</th>
<th>SE</th>
<th>z-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L1VP-L1NP</td>
<td>-.334</td>
<td>.192</td>
<td>-1.745</td>
<td>.080</td>
</tr>
<tr>
<td></td>
<td>L1VP-L2VP</td>
<td>-.161</td>
<td>.237</td>
<td>-.681</td>
<td>.494</td>
</tr>
<tr>
<td></td>
<td>L2VP-L2NP</td>
<td>.904</td>
<td>.182</td>
<td>4.961</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>L1NP-L2NP</td>
<td>1.077</td>
<td>.235</td>
<td>4.588</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

In summary, different processing patterns between the L1 and L2 groups are observed. L1 listeners showed a lower accuracy rate in the VP- than in the NP-attachment, indicating their difficulty in analysing the VP-attachment. In contrast, a significantly higher accuracy rate in the VP- than in the NP-attachment was observed among the L2 listeners, showing their parsing preference for the VP-attachment. The results indicated that both L1 and L2 listeners could attend to prosodic cues in resolving sentence-level ambiguity, but the effects of prosody varied between these two groups.

5. Discussion & Conclusion

This study investigated the effects of prosody (accompanied with the visual context) on PP-attachment ambiguity resolution by Chinese L2 English learners and compared their attachment preference with those of native speakers. The results revealed that, in the absence of prosodic information, some of the L1 group detected the ambiguity with implications from the visual context on the screen, and did not show a preference for the VP- or NP-attachment. Whereas, the L2 group preferred the VP- over the NP-attachment, indicating their failure to detect the ambiguity within the structure from the visual context. When disambiguating prosody was provided, both groups could employ prosodic cues to guide their sentence parsing. Native speakers effectively distinguished the alternative meaning of the ambiguous sentence. However, L2 learners still showed a preference for the VP-attachment as in the listening task. The result stands in contrast with previous studies which predicted that native English speakers preferred the VP-attachment in reading tasks [34, 23]. This difference appeared to be evoked by the visual context information from the pictures. On the one hand, the pictures provided alternative interpretations of the ambiguous sentences. On the other, the object display in the picture might introduce a parsing bias towards the NP-attachment [27]. L2 learners’ preference for the VP-attachment, however, suggests that they did not detect the ambiguity from the pictures, and their interpretation seems to be constrained by the Minimal Attachment principle [35]. The L2 learners’ non-native like sentence parsing strategy may result from the structural asymmetry of PP-attachment ambiguity in English and Chinese, which constrains their awareness of the ambiguity within the structure. Therefore, they tended to interpret the sentences in a ‘good-enough’ approach (i.e. structure-based strategy) to avoid a deeper reanalysis of the sentences which requires more cognitive resources [36]. Another account to explain L2 learners’ performance is that they had difficulty in integrating different domains of information to revise their initial misanalysis, as has been predicted by the Interface Hypothesis [26].

This finding is inconsistent with the previous finding that both contextual and prosodic cues were robust in constraining the L2 learners’ parsing strategy, and the contextual effects appeared more robust [23, 24, 13]. The discrepancy might come from the processing which requires an integration of prosodic, syntactic, and visual context information, or the type of contextual information in our study that was provided in the pictures instead of the discourse. It is thus suggested that unlike native speakers, L2 learners in our experiment could not effectively integrate prosodic, syntactic, and visual context information in resolving sentence-level ambiguity, as has previously been suggested by [20, 4, 6]. Therefore, L2 learners appeared to rely more on the structure-based parsing principles.

The results in the present study suggest that the observed differences in sentence processing between native speakers and L2 learners tend to result from the L2 learners’ lack of ambiguity awareness and their limited ability to integrate multiple domains of information in L2 sentence processing.

6. Acknowledgements

We thank Professor Anne Cutler for her insightful comments and valuable suggestions to our work, Martin Ho Kwan Ip, Mary Youkanna, Laurence Bruggeman, and Juqiang Chen for their technique help. Our thanks also go to the participants. This research was funded by the major Program of National Social Science Foundation of China (No. 18ZDA293), and the China Scholarship Council [201806230185].
7. References


