



Prosody helps L1 speakers but confuses L2 learners: Influence of L+H* pitch accent on referential ambiguity resolution

Chie Nakamura^{1,2,4}, Manabu Arai³, Yuki Hirose², Suzanne Flynn¹

¹ Massachusetts Institute of Technology, U.S.A.

² The University of Tokyo, Japan

³ Seijo University, Japan

⁴ Japan Society for the Promotion of Science, Japan

chienak@mit.edu, manabu-arai@seijo.ac.jp, hirose@boz.c.u-tokyo.ac.jp, sflynn@mit.edu

Abstract

Numerous studies have reported an effect of prosodic information on initial parsing decision. However, whether prosody functions in the same way in adult second language (L2) sentence processing is not known. In visual world eye-tracking experiments, we investigated the influence of contrastive intonation and visual context on processing locally ambiguous sentences with L1 speakers (native English speakers) and L2 learners (Japanese adult learners of English). Our results showed that referential visual context alone helped both L1 speakers and L2 learners to correctly analyze the sentence structure. Interestingly, however, the results also revealed that contrastive intonation accompanied by referential visual context facilitated the correct interpretation with L1 speakers but misled L2 learners “down a garden path”. L2 learners did not interpret the contrastive intonation as a cue that highlights a contrastive set in the visual scene. Instead, they interpreted the contrastive intonation as a simple emphasis and adopted the incorrect syntactic analysis.

Index Terms: Contrastive prosody, referential ambiguity resolution, garden-path, eye-tracking, second language acquisition

1. Introduction

In investigating the underlying mechanism of human language comprehension, a large number of studies have shown influences of various kinds of linguistic information on how language comprehenders construct a syntactic structure. Especially, studies that focused on comprehension in spoken language examined how listeners analyze the speech signal according to the spoken language-specific information, known as *prosody* [1,2], and provided evidence that listeners use prosodic cues to correctly analyze the sentence syntactic structure. For example, [2] examined sentences with a global ambiguity such as *Tap the frog with the flower*, and found that the location of the prosodic boundary directly affects listeners’ syntactic analyses between the two alternatively possible structures (see also [1, 3, 4, 5]). Also, [6] demonstrated that the presence of contrastive accent disambiguates the two alternatively possible interpretations for the same sentences such as *I asked the pretty girl who is cold* (ambiguous between an embedded question analysis and a relative clause analysis). These studies demonstrated that prosody plays an important role in correctly analyzing the syntactic structure of a sentence

and that language users make use of this information for an efficient conversation (see also [7]).

What is unknown, however, is whether contrastive intonation functions in the same way in adult L2 sentence processing especially in contexts where contrastive intonation can be used to resolve referential ambiguity. In order to address this issue, the current study examined the influence of contrastive L+H* pitch accent in combination with contextual information in processing locally ambiguous sentences in English with adult English learners of Japanese (henceforth Japanese EFL learners).

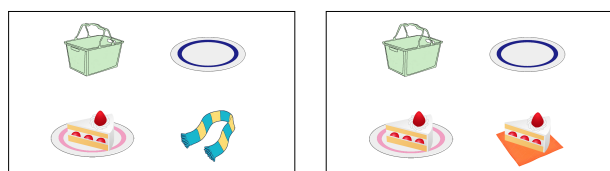
2. Experiment

The current study examined the influence of contrastive intonation in the processing sentences with a local ambiguity such as *put the apple on the towel in the box* [8]. Previous studies showed that contrastive pitch accents evoke a notion of contrast in a discourse context and facilitate the processes of identifying an upcoming referent in spontaneous dialog [9, 10]). Our study examined the impact of the contrastive L+H* pitch accent on noun within a prepositional phrase (PP) modifying NP1 in processing of locally ambiguous sentences such as (1).

(1) Put the cake on the PLATE_{L+H*} in the basket.

In addition to prosody, we also manipulated visual context using the visual world eye-tracking technique [12]. In the processing of sentences such as (1), the first PP is locally ambiguous as to whether it specifies the location of the object to be picked up (i.e., *put [the cake on the plate]...*, *NP-attached PP analysis*) or it denotes the location where the object is to be put (i.e., *[put the cake on the plate]*, *VP-attached PP analysis*). This temporary ambiguity between NP-attached and VP-attached PP for the first PP continues until listeners encounter the second PP *in the basket* (henceforth disambiguating PP). Using this type of garden-path structure [8] showed that when participants saw a contrastive object set in the visual scene (e.g., when there are two cakes in the visual scene, one on a plate and another one on a napkin as shown in Figure 1b), the context information facilitated the correct NP-attached PP analysis. Following their manipulation, the visual context in the current study was either with or without a contrastive object set in the scene. The visual context always depicted four entities, three of which corresponded to the referents in the

sentence; NP1 (a cake), the object in ambiguous PP (an empty plate), and the object in disambiguating PP (a basket). The fourth entity in one-referent condition was a distractor (a scarf; Figure 1a). In two-referent condition, it was an object that stood as a contrast to the NP-attached PP interpretation (i.e., another cake on a napkin as opposed to *the cake on the plate*; Figure 1b).



a. One-referent context b. Two-referent context

Figure 1: Example visual context.

If contrastive intonation helps to resolve referential ambiguity, it is predicted that L+H* pitch accent within the temporary ambiguous PP modifying NP1 (henceforth ambiguous PP) would be interpreted as a cue that highlights the contrastive set in a two-referent context, thus facilitates the correct NP-attached PP analysis. On the other hand, without a contrastive set in the visual scene, the same L+H* pitch accent within the ambiguous PP is expected to be interpreted as information that specifies the destination where the NP1 needs to be put. Hence, the L+H* accent in a one-referent context should encourage the incorrect VP-attached PP analysis, causing a strong garden-path effect in processing the disambiguating PP (*in the basket*).

2.1. Experiment 1: Native speakers of English

Participants

Twenty native speakers of English with normal visual acuity and hearing, recruited in Boston area, participated in the experiment for monetary compensation.

Stimuli

Twenty-four experimental items were created. Each item consisted of an auditory sentence and a corresponding visual scene. The auditory stimuli were recorded by a female native speaker of English. Figure 2 shows the F0 contours of the sentence (1) with and without L+H* pitch accent within the ambiguous PP. The sentence up to the onset of ambiguous PP in each item was identical between conditions with and without contrastive intonation.

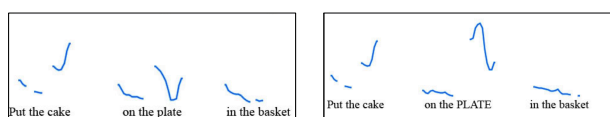


Figure 2: F0 contour of sentence with (right) and without (left) contrastive intonation.

The visual scenes were prepared using commercial clipart images. The position of objects was counter-balanced across the pictures. Four experimental lists were created following

Latin square design including 48 fillers. The 72 items in each list were presented in a pseudo-random order.

Procedure

Participants were first given brief instructions and underwent a calibration procedure. They were told to listen to auditory sentences carefully while paying attention to the picture on the computer monitor. After each sentence was finished, a mouse cursor appeared on the screen and participants were told to click on an object where the first-mentioned object was put (e.g., participants were expected to click on the picture of a basket in Figure 1). In each trial, an auditory sentence was presented 3000 ms after the picture onset on the computer screen. Participants' eye-movements on the screen were recorded with EyeLink 1000 (SR Research) at the sampling rate of 1000 Hz. The whole experimental session took approximately 30 minutes.

2.2. Data analysis and results

The fixation coordinates from the eye tracker were mapped onto four entities in the visual scene and were then converted to gazes. We manually marked the onset of each word in target sentences. For the analysis, we summed the gazes to each entity in the scene and calculated the logit of looks to the each entity out of looks to all the objects in a scene, including background. Statistical analyses for the duration of ambiguous PP (*on the plate*) and disambiguating PP (*in the basket*) were conducted using Linear Mixed Effects models [13]. We included Prosody (with or without L+H* accent within ambiguous PP) and Visual Context (one-referent or two-referent) as fixed effects as well as the interaction between the two factors. Participants and items were included as random factors and the best-fit model with an optimal random slope structure was selected using a backward selection approach.

We first report the analysis on the gazes made to the incorrect destination (e.g., an empty plate in Figure 1) for the duration of ambiguous PP (*on the plate*). Figure 3 shows the proportion of gazes to the incorrect destination entity from the ambiguous PP onset until 1500 ms. The dotted line marks the mean onset of disambiguating PP (1384 ms). We analyzed the gazes to the incorrect destination entity during the interval between the ambiguous PP onset to the minimum onset of the disambiguating PP (1057 ms). We excluded the first 200 ms from the analysis by taking into consideration the time required for reflecting a fixation in response to linguistic input.

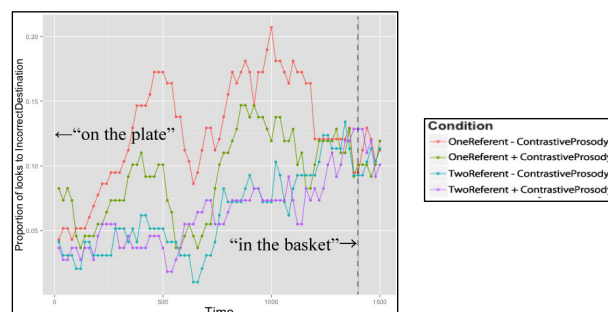


Figure 3: Proportion of gazes to the incorrect destination entity from the onset of ambiguous PP to 1500 ms.

Table 1 shows coefficients, standard error, t -values, and p -values from the model. The results showed a main effect of visual context; more looks to the incorrect destination entity were observed in one-referent condition than in two-referent condition. Consistent with the prediction, the results demonstrate that participants were more likely to adopt VP-attached PP analysis on hearing the ambiguous PP when there was no contrastive object in the visual scene, replicating the results of previous research [8].

Table 1. Analysis of looks to the incorrect destination entity for the duration of ambiguous PP.

	β	SE	t	p
Intercept	-5.45	0.21	-25.54	
Prosody	-0.44	0.34	-1.29	n.s.
Visual context	-1.21	0.43	-2.80	<0.001
Prosody*Visual context	1.05	0.77	1.37	n.s.

We next report the analysis on the gazes made to the correct destination (e.g., a basket in Figure 1) for the duration of disambiguating PP (*in the basket*). Figure 4 shows the proportion of gazes made to the correct destination entity from the disambiguating PP onset until 1500 ms.

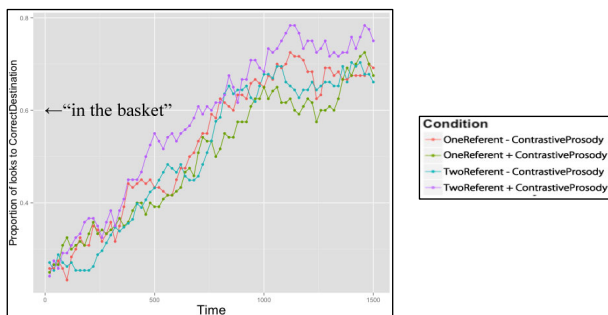


Figure 4: Proportion of gazes to the correct destination entity from the onset of ambiguous PP to 1500 ms.

Table 2 shows the results of analysis on gazes for the 1300 ms time interval of 200 to 1500 ms following the mean onset of disambiguating PP. We observed an interaction between prosody and visual context. Further analysis on a simple effect of prosody showed opposite pattern of results between one-referent and two-referent conditions. In one-referent condition, there were fewer looks to the correct destination entity when the sentence carried contrastive prosody compared to when it did not. In two-referent condition, on the other hand, more looks to the correct destination entity were observed when the sentence had contrastive prosody compared to when it did not. Consistent with the prediction, these results demonstrate that contrastive prosody on ambiguous PP facilitated the correct VP-attached PP analysis when there was a contrastive object set in the visual scene, whereas it was interpreted as a cue that emphasize the information that denotes the destination of the object without a contrastive set in the visual scene. As a result, in absence of a referential context, contrastive intonation within the ambiguous PP led native English speakers to incorrectly adopt NP-attached PP analysis.

Table 2. Analysis of looks to the correct destination entity for the duration of disambiguating PP.

	β	SE	t	p
Intercept	1.58	0.35	4.51	
Prosody	0.10	0.25	0.39	n.s.
Visual context	0.18	0.25	0.73	n.s.
Prosody*Visual context	0.97	0.49	1.98	0.048

2.3. Experiment 2: Japanese EFL learners

Participants

Twenty-nine adult Japanese EFL learners, age 18 to 22, participated in Experiment 2. They were all undergraduate students at the University of Tokyo, Japan. They had at least six years of English education before entering the university. We obtained English proficiency scores from all the participants. The mean score for our participants corresponded to proficiency level of intermediate to advanced.

Stimuli and Procedure

Stimuli and procedure were identical to those in Experiment 1 except that Tobii TX300 was used to record participants' eye-movements in Experiment 2.

2.4. Data analysis and results

Analyses on gazes to each entity were performed in the same manner for the same time windows as in Experiment 1. We first report the gazes made to the incorrect destination entity for the duration of ambiguous PP and next the gazes to the correct destination entity for the duration of disambiguating PP.

Figure 5 shows the proportion of gazes to the incorrect destination entity from the ambiguous PP onset until 1500 ms. We again analyzed gazes for 1057 ms time interval of 200 to 1257 ms, from the onset of ambiguous PP to the minimum onset of disambiguating PP. Table 3 summarizes the results.

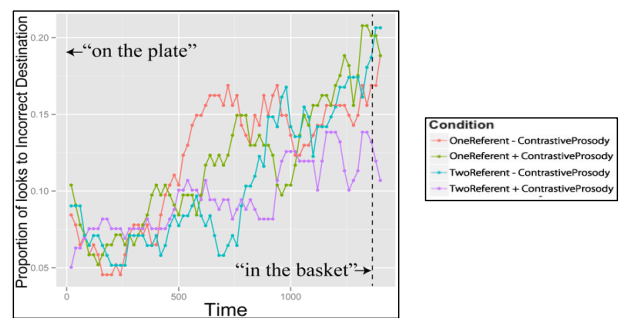


Figure 5: Proportion of gazes to the incorrect destination entity from the onset of ambiguous PP to 1500 ms.

The results showed a main effect of visual context; there were more looks to the incorrect destination entity in one-referent condition than in two-referent condition. Consistent with the results of native English speakers in Experiment 1, the results demonstrated that Japanese EFL learners were more likely to adopt the correct NP-attached PP analysis when the visual

context had a contrastive object set, providing evidence that referential visual context information helped Japanese EFL learners to correctly analyze the syntactic structure.

Table 3. *Analysis of looks to the incorrect destination entity for the duration of ambiguous PP.*

	β	SE	t	p
Intercept	-4.34	0.31	-13.86	
Prosody	-0.03	0.28	0.09	n.s.
Visual context	-0.99	0.28	-3.57	<0.001
Prosody*Visual context	-0.82	0.56	-1.47	n.s.

Figure 6 shows the proportion of gazes to the correct destination entity from the ambiguous PP onset until 1500 ms. We analyzed gazes made to the correct destination entity for the 1000 ms time interval of 200 to 1200 ms following the onset of disambiguating PP. The results are summarized in Table 4.

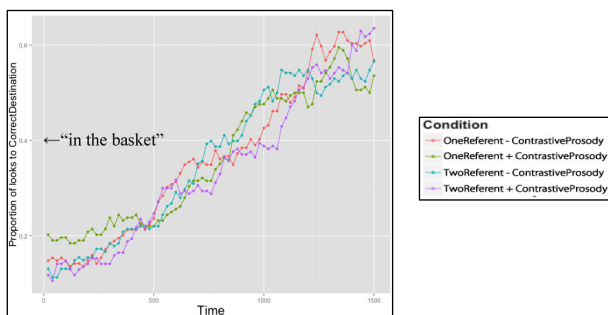


Figure 6: *Proportion of gazes to the correct destination entity from the onset of ambiguous PP to 1500 ms.*

The results showed an interaction between prosody and visual context. Further analysis on a simple effect of prosody revealed no effect of prosody in one-referent condition ($p > 0.1$). In two-referent condition, on the other hand, it was shown that there were fewer looks to the correct destination entity when the sentence had contrastive intonation compared to when it did not ($p = 0.07$). These results indicate that, contrary to the prediction and the pattern of results with native English speakers, contrastive prosody within the ambiguous PP, when accompanied by a contrastive set in the visual scene, did not help Japanese EFL learners to adopt the correct NP-attached PP analysis. Instead, it led them to incorrectly adopt the VP-attached PP analysis, making it difficult to recover from the incorrect syntactic interpretation on hearing the disambiguating information.

Table 4. *Analysis of looks to the correct destination entity for the duration of disambiguating PP.*

	β	SE	t	p
Intercept	-0.56	0.33	-1.69	
Prosody	-0.09	0.29	-0.30	n.s.
Visual context	0.07	0.42	0.18	n.s.
Prosody*Visual context	-1.19	0.57	-2.01	0.039

3. Discussion

The current study examined an influence of L+H* pitch accent on referential ambiguity resolution. The results from Experiment 1 and 2 showed that visual context information alone helped both native English speakers and Japanese EFL learners to correctly analyze the syntactic structure of locally ambiguous sentences, replicating the results of previous studies that showed an effect of referential visual context on syntactic structure analysis [8]. However, importantly, an effect of contrastive intonation within the ambiguous PP, when accompanied by a referential visual context, showed different pattern of results between native speakers and Japanese EFL learners. With native speakers, contrastive intonation on the temporary ambiguous PP was interpreted as a cue that highlights the contrastive set in the visual scene, thus facilitated the NP-attached PP analysis. On the other hand, Japanese EFL learners incorrectly adopted the VP-attached PP analysis when the ambiguous PP carried contrastive intonation and there were two potential referents in the visual context. As a result, they experienced stronger processing difficulty on hearing the disambiguating information.

To summarize, the current study showed that contrastive prosody provided in a referential context facilitated the syntactic analysis of locally ambiguous sentence structure with native speakers. However, such information did not have facilitatory effect with Japanese EFL learners and instead it led them to adopt the incorrect syntactic analysis. This suggests that prosodic information provided with a referential visual context put an extra load on the processing of temporary ambiguous sentences for L2 learners, and that they did not use this information to reach the correct syntactic analysis. This finding conflicts with the results in first language acquisition studies, which reported that young children use prosody to resolve referential ambiguity although an effect of prosody appeared to be delayed compared to L1 adults [14]. We argue that the contradictory results between the two different language learning populations come from the fact that L2 learning adult have much smaller input of spoken language utterances compared to L1 learning children. Although this remains still a speculation, it appears that adult L2 learners have less opportunity to learn how to encode prosodic information to syntactic structure due to the overall shortfall of spoken language input in the target language, and as a result have difficulty in integrating prosody into other information such as visual context in online sentence processing.

4. Conclusions

The results of this study indicated that referential visual context alone helped both native speakers and L2 learners to correctly analyze the sentence structure. However, importantly, L+H* pitch accent accompanied by two potential referents in visual context facilitated the correct interpretation with native speakers but misled L2 learners to the incorrect syntactic analysis. With L2 learners, the contrastive L+H* accent within the ambiguous PP was not interpreted as a cue that highlights the contrastive set in the visual context. Instead, it was interpreted as a simple emphases on information that specifies the destination, and as a result they experienced stronger processing difficulty in revising the initially adopted VP-attached PP structure for the correct NP-attached PP structure on hearing the disambiguating information.

5. References

- [1] Schafer, A., Speer, S., Warren, P., & White, S. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29, 169–182.
- [2] Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48, 103–130.
- [3] Kjelgaard, M., & Speer, S. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40, 153–194.
- [4] Snedeker, J., & Casserly, E. (2010). It is all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*, 25, 1234–1264.
- [5] Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25, 249–271.
- [6] Schafer, A., Carter, J., Clifton, C., Jr., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes*, 11, 135–163.
- [7] Marslen-Wilson, W. D., Tyler, L. K., Warren, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology*, 45, 73–87.
- [8] Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- [9] Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2012). Intonation facilitates contrast resolution: Evidence from Japanese adults and 6-year olds. *Journal of Memory and Language*, 66, 265–284.
- [10] Ito, K., & Speer, S. R. (2008). Anticipatory effect of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, 58, 541–573.
- [11] Weber, A., Grice, M., & Crocker, M. W. (2006b). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, 99, 63–72.
- [12] Cooper, R. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- [13] Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- [14] Snedeker, J. & Yuan, S. (2004). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, 58, 574–608.