Syntactic reanalysis and lingering misinterpretations in L2 sentence processing*

Jonghyeon Lee · Jeong-Ah Shin**

(Seoul National University · Dongguk University)

Lee, Jonghyeon and Jeong-Ah Shin. 2016. Syntactic reanalysis and lingering misinterpretations in L2 sentence processing. Linguistic Research 33(Special Edition), 53-79. Readers often misinterpret sentences and reanalyze them, but their initial misinterpretations sometimes linger even after they finish reading the sentences. This lingering effect has been argued as evidence that comprehenders create interpretations to some extent that it seems “good enough” for them to understand the meaning (Christianson, Hollingworth, Halliwell, and Ferreira 2001). This study examined L2 learners’ processing of garden path sentences, investigating whether they show a lingering effect as L1 speakers do by employing the gender mismatch paradigm (Sturt 2003). In the self-paced reading experiment, participants read garden path sentences such as While Mary washed the boxer who was tall coughed and warmed herself with a blanket. If participants correctly rule out an initial misinterpretation from the temporary ambiguity, the reflexive herself would create a gender mismatch effect (i.e., reading times are slowed in the gender mismatch condition). After reading the sentences, participants were asked to answer follow-up comprehension questions such as Did Mary wash the boxer? The results showed that L2 learners were able to notice the temporary local ambiguity in a sentence and reanalyze the structure as L1 speakers do. However, the rate of accurate responses for comprehension questions was significantly lower in the ambiguous condition than in the unambiguous condition, indicating that the L2 learners’ initial misinterpretations lingered. L2 learners do not seem to lack detailed syntactic representation, and their lingering misinterpretation could be explained by the good enough approach as for L1 readers. Participants eventually recovered from the garden path, showing the gender mismatch effect, but their global comprehension was significantly influenced by the initial misinterpretation, since they gave more incorrect answers when they read garden path sentences. (Seoul National University · Dongguk University)

Keywords L2 sentence processing, garden path, syntactic reanalysis, lingering effect, good enough processing, gender mismatch effects

* We would like to thank two anonymous reviewers of Linguistic Research for their invaluable comments and suggestions.

** Corresponding author
1. Introduction

Garden path sentences containing syntactic ambiguity have been extensively investigated in psycholinguistic research to examine the human language processing system. Most researchers hold that readers reanalyze garden path sentences when they encounter disambiguating regions (e.g., Frazier 1987; Frazier and Rayner 1982). When people read the canonical example sentence for temporary ambiguities, *the horse raced past the barn fell*, and see the word *fell* (i.e., the disambiguating point), they tend to slow down, discard their initial interpretation, and reanalyze the structure of the sentence. Although considerable controversy has existed concerning the sources of information that help resolve the ambiguity (Ferreira and Clifton 1986; Ferreira and Henderson 1990; MacDonald, Pearlmutter, and Seidenberg 1994; Trueswell, Tanenhaus, and Garnsey 1994), most theories of language processing assume that readers process language as fully as possible in real time.

Recently, however, a series of studies have questioned this assumption (Christianson, Hollingworth, Halliwell, and Ferreira 2001; Christianson, Williams, and Ferreira 2006; Ferreira, Christianson, and Hollingworth 2001; Malyutina and Ouden 2015; Patson, Darowski, Moon, and Ferreira 2009; Staub 2007; Slattery, Sturt, Christianson, Yoshida, and Ferreira 2013; Sturt 2007; van Gompel, Pickering, Pearson, and Jacob 2006). They found that people’s initial misinterpretations often persist even after they finish reading disambiguating regions. In other words, the initial misinterpretations linger and influence global comprehension of the sentence, which is termed lingering effects. In particular, Christianson, Hollingworth, Halliwell and Ferreira (2001) found that syntactic manipulations of the garden path influence accuracy rates on follow-up comprehension questions as follows:

1. While Anna dressed the baby that was cute and cuddly played in the crib.

2. Did Anna dress the baby?

3. Did Anna dress herself?

When a garden path sentence in (1) is given, if readers recover from the first
misinterpretation—the baby as a direct object of dressed—and correctly construct the sentence structure—as a subject of played, the answer to a follow-up question like (2) should be “No,”; however, many readers incorrectly answered “Yes.” When they were given the question in (3) after reading the sentence (1), they also correctly responded, “Yes.” The results indicates that readers did not fail to reanalyze the garden path sentence appropriately. Rather, incorrect responses to a question like (2) seemed to result from the early misinterpretation’s persistence.

In van Gompel et al. (2006), the lingering effects were investigated with a structural priming paradigm. In the experiment, participants read temporarily ambiguous sentences such as (1) as a prime, or unambiguous sentences that were disambiguated by a comma between the verb, dressed, and the noun, the baby. Then they were asked to read aloud sentence fragments such as When the doctor was visiti... and to complete them. As a result, they produced more transitive sentences after encountering ambiguous sentences like (1) than unambiguous sentences. They argued that it was because the initial transitive analysis remained within a memory trace and was activated when participants completed sentence fragments. Patson et al. (2009) also provided evidence for lingering misinterpretations using a paraphrasing methodology. When participants were asked to paraphrase a garden path sentence such as (1), their final interpretations were likely to be incorrect. In addition, Sturt (2007), using an eye-tracking experiment, showed the lingering effects were not restricted to the specific structure such as (1), but can occur for cases in which reanalysis is relatively straightforward. When participants read a sentence such as (4a), where the first misinterpretation (“the explorers found the South Pole”) was semantically inconsistent with the final interpretation (the explorers did not find the South Pole), they took more reading time than in (4b), where that disambiguated the ambiguity.

(4) a. The explorers found the South Pole was actually impossible to reach.
   b. The explorers found that the South Pole was actually impossible to reach.

While it seems uncontroversial that lingering effects exist, it is not yet known why the misinterpretation lingers and how it affects language processing. Christianson et al. (2001) suggested that a failure to fully reanalyze a sentence accounts for the lingering effect, arguing for the good enough processing approach, which posits that comprehenders often create a structure that is “good enough” to
understand the meaning of sentences rather than building an idealized structure with every detail included. In this regard, because readers, when they meet a temporarily ambiguous sentence, reanalyze the sentence only to the extent that they understand it, the previous syntactic structure is not completely disregarded, but rather it lingers. On the other hand, Kaschak and Glenberg (2004) claimed that the persistence of the first misinterpretation originates from memory traces of the initial parsing. Meanwhile, Sturt (2007) suggested that the effects are due to semantic persistence, which is constructed at the early parsing.

Many psycholinguists have recently explored the good enough approach on the basis of first language processing. However, attempts to apply this approach to second language processing are relatively rare. In one such study, Lim and Christianson (2013a) tried to explain L2 sentence processing with the good enough approach, using plausible and implausible subject relative clauses and object relative clauses (also see Lim and Christianson 2013b, 2015). From a self-paced reading task and a translation task, the study found that L2 comprehenders also utilize the good enough strategy for L2 sentence processing. Their interpretation from the syntactic route was often overridden by the plausible interpretation from the semantic route (The dog that the man bit is often misinterpreted as ‘The dog bit the man’) and differences in reading goals affected their parsing route; they paid more attention to syntactic structure in a translation task. These results were predicted by the good enough approach.

In spite of the results, it is still not convincing that L2 learners use the same good enough strategy. From another perspective, this incomplete sentence parsing was not a result of good enough processing but rather a result of L2 readers’ deficient language comprehension skills. Clahsen and Felser (2006a, 2006b) claimed that the syntactic representations of L2 learners are shallower and less detailed than those of native speakers (the shallow structure hypothesis, SSH). Even highly proficient learners seem to fail to acquire native-like, detailed syntactic representations. That is, L2 learners tend to miss or depend less on detailed syntactic information, such as structural relations, subject-verb agreement, or binding restrictions, during sentence processing. Rather, they seem to rely primarily on world knowledge and semantic cues when interpreting sentences. In fact, Lim and Christianson (2013a) also found that L2 speakers were more dependent on the semantic route for parsing compared to native speakers.

Nevertheless, Clahsen and Felser (2006a, 2006b) did not suggest that the idea of
shallow processing is something entirely distinct from the “good enough” representation of native speakers. Rather, it suggests that the good enough representation is a native speakers’ version of shallow processing. However, from the perspective of the good enough approach, it differs from SSH, in that while the good enough strategy is one option for computing a sentence structure for native speakers, L2 learners are largely limited to this option in L2 processing, or are more dependent on good-enough representations (Jacob and Felser 2016).

However, it remains to be examined whether L2 learners lack detailed syntactic representation and whether they are restricted to shallow processing. Since most evidence for L2 processing has been derived from online reading time data, the previous studies mainly argued for shallow processing of L2 based upon the results that L2 learners showed different patterns in reading times than native speakers, particularly in terms of a larger plausibility effect on L2 comprehenders (Felser, Roberts, Gross, and Marinis 2003; Papadopoulou and Clahsen 2003; Roberts and Felser 2011). However, there is also some evidence for L2 speakers’ syntactic processing being similar to that of native speakers, especially when the learners are highly proficient (Jackson 2008; Kim 2015; Williams 2006). Moreover, plausibility also affects L1 processing to a greater or less degree (Christianson et al. 2010; Ferreira 2003; Slattery et al. 2013; Pickering and Traxler 1998).

In this regard, this study explored the nature of L2 sentence processing in terms of the “good enough” approach and the shallow structure hypothesis, addressing the following research questions:

Research Question 1: Do the initial misinterpretations of garden path sentences linger and influence L2 learners’ comprehension as they do in L1 speakers’ sentence processing?
Research Question 2: If so, is this due to a lack of detailed syntactic representations of L2 learners, or a “good enough” nature of reanalyzing a sentence structure similar to that of L1 speakers?
Research Questions 3: Does learners’ L2 proficiency modulate syntactic reanalysis and/or lingering effects?

To examine the research questions, a self-paced reading experiment was conducted. Participants read sentences that were designed to utilize the garden path
sentence and the gender mismatch effect. The gender mismatch effect refers to a reading slow-down when the gender of reflexive mismatches the stereotypical gender of an antecedent (Kreiner, Sturt, and Garrod 2008; Sturt 2003). This effect is known to be observed only if the link between the reflexive and its antecedent is grammatical. Thus, the gender mismatch effect would indicate that the comprehenders construct a detailed syntactical representation (Slattery et al. 2013).

After reading sentences, participants were asked to answer follow-up yes or no questions to check their comprehension. The accuracy rate for responses was calculated as an index of the influence of the initial misinterpretations on the readers’ comprehension. Although this forced-choice response paradigm has been questioned its reliability (van Gompel et al. 2006), the results were consistent with other studies using different paradigms such as picture selecting, structural priming, and paraphrasing task (Malyutina and Ouden 2015; Patson et al. 2009; Sturt 2007; van Gompel et al. 2006). Furthermore, since the accuracy rate is relatively clearer than the results of other paradigms, this paradigm seems to be more appropriate in L2 sentence processing research, where only a few studies have recently begun to examine L2 good enough sentence processing.

2. Experiment

The experiment used a 2 (ambiguous vs. unambiguous) x 2 (gender match vs. mismatch) factorial design as in (5). Participants received one of the four counter-balanced lists. They first read sentences such as those in (5) while their reading times were recorded by the computer, and then they were asked to answer a follow-up comprehension question such as (6). The sentences consisted of one subordinate clause and one main clause with the same words. The only differences among conditions were the gender of a reflexive pronoun in the main clause and the presence of a comma after the subordinate clause, which were designed to manipulate the gender match condition and ambiguity condition.

(5) a. Garden Path / Match
   While Mary washed the boxer who was tall coughed and warmed himself with a blanket.
b. Garden Path / Mismatch
While Mary washed the boxer who was tall coughed and warmed herself with a blanket.

c. Non-Garden Path / Match
While Mary washed, the boxer who was tall coughed and warmed himself with a blanket.

d. Non-Garden Path / Mismatch
While Mary washed, the boxer who was tall coughed and warmed herself with a blanket.

(6) Did Mary wash the boxer?

In the garden path conditions, a subordinate clause causes ambiguity, in that at the point of the verb washed, whether the verb is transitive or intransitive is ambiguous. In the non-garden path conditions, the structure is unambiguous because the presence of the comma enables the embedded verb only to be interpreted as intransitive. In the gender match condition, the stereotypical gender of the matrix subject, the boxer, and its reflexive, himself, match each other, but the first subject, Mary, as always having the opposite gender of the matrix subject, does not match the reflexive. In the gender mismatch condition, the reflexive herself does not match the second subject but the first subject in terms of gender.

With this design, we will test our predictions as follows. First, if L2 learners ever notice the garden path ambiguity, slower reading times at the disambiguation verb (coughed) and its adjunct words (and warmed) will be observed in the garden path condition. If this garden path effect differs from that of the native control group, an interaction will be also found at these words. Second, if L2 learners succeed in analyzing the second noun phrase (the boxer who was tall) as the subject of the main clause, the gender mismatch condition leads to greater reading times on the reflexive and its adjunct words. Most importantly, if they construct a detailed syntactic structure after reanalysis, the gender mismatch effect should be shown in both ambiguity conditions, resulting in a main effect of gender match with no interaction with ambiguity. In addition, if this pattern is similar to that of the native control, no interaction with L1 participants will be found.

However, if they fail to fully reanalyze a detailed syntactic structure, there are two possibilities. First, if the second NP is not parsed as the matrix subject until this
point, it would lead to greater reading times on the reflexive in the garden path conditions compared to the non-garden path conditions. Also, since the reflexive does not have an accessible antecedent at this point, the gender mismatch effect would be absent in the ambiguous conditions, or at least weaker than in the unambiguous conditions. Thus, this predicts a main effect of ambiguity at the reflexive region, with an interaction between gender and ambiguity (Slattery et al. 2013). Second, if the first subject (Mary) continues to play the role of the antecedent, reverse gender mismatch effects will be observed, leading the participants to read more slowly in the matching conditions.

As for the comprehension questions, if the first misinterpretation lingers, the response accuracy will be lower in the garden path conditions than in the non-garden path conditions, consistent with the native control. How the gender match conditions will affect the comprehension accuracy has two possible consequences. First, the accuracy for the gender match conditions will be even lower. Due to the gender mismatch, participants may consider their reinterpretation to be incorrect, weighting the first misinterpretation. Alternatively, the rate of correct responses can be higher, since participants might believe themselves to misremember the second NP. When they encounter the follow-up question (Did Mary wash the boxer?), they would correctly answer “no,” just because the boxer is what they have misremembered.

In short, if participants reanalyze a sentence structure, there would be a main effect of the garden path at the disambiguation and its adjunct region. Second, if the parser construct a detailed syntactic structure, a gender mismatch effect with no interaction with ambiguity will be observed at the reflexive and its spill-over region, whereas if the full reanalysis fails, there would be a main effect of ambiguity with an interaction between ambiguity and gender match. For comprehension accuracy, if the initial parsing persists after reanalysis, the accuracy rate will be lower in the ambiguous conditions than in the unambiguous conditions.

2.1 Method

2.1.1 Participants

For the L2 learners group, twenty eight undergraduate and graduate students
from two universities in Seoul participated in the experiment (age range: 18-29; mean age: 23.46; 16 females and 12 males). They all were Korean native speakers and were paid 5,000 won for their participation. No participant except one had experience living in an English-speaking country more than a year. Only one participant lived for about seven years in the United States (this participant was included in the analysis, since the analysis including him did not change the overall results). They all learned English in an academic setting for more than 10 years and did not report any problems in understanding the English sentences during the experiment. Their proficiency varied from intermediate to advanced levels. They were divided into two different proficiency groups by the standardized English test score (TEPS or TOEIC). For the advanced group, the participants whose TEPS score (or its equivalent) ranged from 800 to 990 (near-native level) were included (12 participants; mean score: 857.41; mean age: 23.33), while those who scored from 501 to 700 (mid and high intermediate levels) were classified as the intermediate group (16 participants; mean score: 633.94; mean age: 23.56).

For the native control group, twenty-seven participants were recruited via two crowdsourcing web platforms, Amazon’s Mechanical Turk and Click Worker (age range: 18-65; mean age: 30.70; 14 females and 13 males). They were self-reported native speakers of English. Only participants with US locations were allowed to participate. Participants were paid $0.70.

2.1.2 Materials

Sixteen items were created and adapted from Slattery et al. (2013) and Christianson et al. (2001). They were modified to be more comprehensible for English learners, by using more frequent words and shortened relative clauses for noun phrases (the girl who was tall, instead of the child that was blond and pudgy). Sixteen experimental items were selected among 32 items based on the results of norming tests, which were performed by L2 participants who did not participate in the main experiment (as rated more plausible—above 3 points in a scale of 1 to 4). All the sentences have the same structure across conditions. We also included 32 filler items. The comprehension accuracy for the fillers was used to identify the participants who did not seem to fully understand the experimental sentences. We only used fillers because the global comprehension of the target sentences was
expected to be influenced by the lingering effect from the experimental design.

All subordinate clause verbs were reflexive absolute transitive (RAT) verbs, such as bathe, dress, wake up, and wash (Christianson et al. 2001). Verbs in this class have some advantages, in that they are obligatorily semantically transitive, even in the absence of a direct object. For example, Mary washed means that she washed herself, although there is no overt object mentioned. Thus, RAT verbs can partly rule out the possibility that participants would answer the comprehension question based on pragmatic inference, which can be drawn with optionally transitive verbs such as hunt, telephone, and write. That is, when the participants were given a sentence such as while the man hunted the deer ran into the woods, they can infer the man hunted the deer, even after they correctly parsed the deer as the subject of the main clause, because the man was hunting in general and the deer might be hunted. However, since RAT verbs obligatorily take a reflexive in the absence of a direct object, such an inference would not be drawn if the participants completely reanalyze the syntactic structure. Among RAT verbs, ten verbs were selected with the consideration of word familiarity to Korean speakers.

Comprehension questions simply asked if the subordinate NP did the described action to the object, which was actually the subject NP of the main clause (e.g., Did Mary washed the boxer? for (5)). Thus, the answer to the experimental items always required a “no” response, while twenty fillers required a “yes” response and twelve fillers required a “no” response. To reduce the possible effect of participants’ test-taking strategies, four pseudo-experimental sentences were also included among 32 filler sentences, where the embedded clause had a direct object and then the answer to the questions was “yes” (e.g., When Josh bathed the baby, he happily giggled and threw himself into the bath tub).

To maximize the gender mismatch effect, the subject nouns in the main clauses were chosen among the words whose mean rate of gender stereotype was highest in Kennison and Trofe (2003). They were tested again in the norming test by native Korean speakers who did not participate in the main experiment but were likely to share the cultural gender stereotypes with the participants. They included both stereotypical nouns (e.g., soldier, cheerleader) and definitional nouns (e.g., nun, boy), since stereotypical and definitional gender nouns seem to lead to a similar mismatch-cost when the gender role noun is presented earlier than the reflexive (Kreiner, Sturt, and Garrod 2008). Meanwhile, the subjects in embedded clauses
were always proper names with the other gender of the main clause subject (e.g., *John* for the subordinate clause and *the nurse* for the main clause).

### 2.1.3 Procedure

The experiment consisted of a lab-based experiment and a web-based experiment. In the lab-based experiment, twenty Korean participants were individually tested in a quiet room. The self-paced reading experiment was implemented on a laptop computer using Linger software (Rohde 2003). In the web-based experiment, eight Korean participants and English native participants were tested, using the Ibex web interface for psycholinguistic experiments (Drummond 2012). There is much evidence that even online measures such as a self-paced reading experiment administered over the web replicate results obtained in lab-based experiments (Demberg 2013; Enochson and Culbertson 2015; Keller, Gunasekharan, Mayo, and Corley 2009). Some self-paced reading experiments have been recently conducted based on the web platform using the Ibex web interface (Fine and Jaeger 2016; Myslin and Levy 2016).

The procedure for the experiment was consistent across both formats. Before the experiment began, they were asked to fill out a consent form and were given instructions on how to perform the task. A practice session was then carried out to help participants to be accustomed to the task. The sentence was presented in a self-paced, noncumulative, word-by-word fashion, from the left to right of the screen. It was first hidden with a series of underscores, which were replaced by words one at a time when participants pressed a space bar. A comprehension question was given after each sentence was read and disappeared. Participants answered the question by pressing one of two designated keys for "yes" and "no" on the keyboard. Reading times and answering accuracy were recorded. When the question was answered, the next sentence was presented. The experiment lasted for about 15 minutes.

### 2.2 Results

Prior to analysis, nine L2 learners and nine native English speakers were excluded from the data analysis due to their low accuracy for filler sentences (using 70 percent accuracy rate as a criterion, adapted in most studies (e.g., Jiang 2013)) or abnormally long reading times (about 16.5 seconds per one word). In the end, 37
participants were included in data analysis (11 advanced learners, 8 intermediate learners, 17 native speakers). Then data points for reading times that were 3 standard deviation above the region mean were removed, which resulted in an 1.81% data loss. For the purposes of data analysis, sentences were divided into several critical regions (Table 1). For analyzing reading times of each region and comprehension question accuracy, a generalized linear mixed model was conducted, using the lme4 library (Bates, Maechler, Bolker and Walker 2014) in the R environment (R Core Team 2014). The model first assumed fixed effects for ambiguity, gender match, and language to ascertain differences between native speakers and non-native speakers. Then proficiency was included as a fixed effect instead of language for proficiency analysis. Participants and items were treated as random effects in both analyses.

<table>
<thead>
<tr>
<th>Table 1, Critical regions of test sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1</td>
</tr>
<tr>
<td>Pre-critical</td>
</tr>
<tr>
<td>(…) the boxer coughed and warmed himself with a blanket.</td>
</tr>
</tbody>
</table>

2.2.1 Comprehension question

Every item was followed by comprehension questions. Figures (1a) to (1d) show the average accuracy rates with the relative 95% confidence intervals for the answers. As shown in the figures, the percentage of correct responses (“no” responses) was significantly higher in the unambiguous condition than in the ambiguous condition (estimate=1.067, SE=0.112, z=9.484, p<0.001). There was also a main effect of language, where L1 speakers responded more correctly than L2 learners (estimate=-0.628, SE=0.168, z=-3.738, p<0.001). The language difference can be accounted for with a significant interaction between ambiguity and language (estimate=0.356, SE=0.108, z=3.311, p<0.001). There was no significant difference between languages in the unambiguous condition (estimate=-0.2753, SE=0.2085, z=-1.321, n.s.), but there was a significant difference in the ambiguous condition (estimate=-122.853, SE=55.438, z=-4.746, p<0.001). On the other hand, the response times for the comprehension questions did not vary across conditions. No main effect of language, ambiguity or match was found. There was only a significant interaction between language, ambiguity, and gender match (estimate=-122.853, SE=55.438, t=-2.216, p<0.05).
As for the data of the second language speakers, there was a main effect of ambiguity (estimate=1.513, \( SE=0.180, z=8.410, p<0.001 \)), where ambiguous conditions caused much higher error rate than unambiguous conditions. There was also a significant interaction between gender match and proficiency (estimate=0.448, \( SE=0.180, z=-2.488, p<0.05 \)). While the advanced level group was more likely to choose a wrong answer in gender mismatch conditions, the intermediate group was more likely to show errors in gender match conditions. In particular, the tendency seemed more salient within ambiguous conditions, which was evidenced by a three-way marginal interaction between ambiguity, gender match and proficiency (estimate=-0.274, \( SE=0.158, z=-1.739, p=0.082 \)). For response times, there was no main effect of proficiency, ambiguity, or match.

Figure 1. The answering accuracy for comprehension questions
2.2.2 Reading times

2.2.2.1 Disambiguation region and spill-over 1 region

At the disambiguation region, there was a significant main effect of language (estimate=244.218, SE=69.637, \( t=3.507, p<0.001 \)), indicating that native speakers tended to read faster than non-native speakers, and a main effect of ambiguity (estimate=-91.068, SE=24.065, \( t=-3.784, p<0.001 \)), where the garden path condition yielded longer reading times compared to the non-garden path condition. There was also an interaction between language and ambiguity (estimate=-63.349, SE=24.068, \( t=-2.632, p<0.01 \)). It seemed that the garden path effect was slightly greater for L2 speakers, compared to L1 speakers. Similar patterns were found at its spill-over region (spill-over 1), in that there was a significant main effect of language (estimate=252.377, SE=49.558, \( t=5.093, p<0.001 \)) and ambiguity (estimate=-85.727, SE=21.661, \( t=-3.958, p<0.001 \)). There was also a marginal interaction between language and ambiguity (estimate=-38.072, SE=21.732, \( t=1.752, p<0.1 \)). Again, the garden path effect was numerically greater in the RTs of the L2 group.

For the non-native speakers group, the results showed a significant main effect of ambiguity at both regions (disambiguation: estimate=-153.69, SE=45.50, \( t=-3.378, p<0.001 \); spill-over 1: estimate=-125.766, SE=38.324, \( t=-3.282, p<0.001 \)). L2 participants read the words more slowly in the ambiguous conditions than in the unambiguous conditions. There was also an interaction between ambiguity and proficiency (disambiguation: estimate=-95.08, SE=45.42, \( t=-2.093, p<0.05 \); spill-over 1(marginal): estimate=70.885, SE=38.315, \( t=1.850, p<0.1 \)). The interaction seemed to result mainly from the advanced group’s faster reading in the unambiguous conditions. At the disambiguation region, the advanced group read numerically 200ms faster than the intermediate group in the unambiguous conditions (\( t=-1.476, df=98.429, p=0.143, n.s. \)) but there was little difference in the ambiguous conditions (\( t = -0.044, df = 125.2, p<1, n.s. \)).

2.2.2.2 Reflexive region and spill-over 2 region

The gender mismatch effect was not found at the reflexive region, since there was no main effect of gender (estimate=-10.577, SE=9.168, \( t<1, n.s. \)). Instead, there
was a main effect of ambiguity (estimate=-26.745, SE=9.169, t=-2.917, p<0.01) and a significant effect of language (estimate=68.574, SE=20.471, t=3.350, p<0.01). This showed that the garden path effect was likely to be prolonged to the reflexive region. In particular, there was an interaction between language and ambiguity (estimate=-18.292, SE=9.176, t=-1.993, p<0.05). It seemed that the garden path effect influenced only L2 learners, or at least the effect was larger for L2 learners. There was, in fact, no difference in the ambiguity when only the L1 group was included in the regression model (estimate=-7.998, SE=8.551, t<1, n.s.).

In contrast to the prior region, there was no main effect of ambiguity (estimate=11.467, SE=11.291, t<1, n.s.) at its spill-over region (spill-over 2). The garden path effect would not cause a global slow-down for participants. At some point, participants recovered from the garden path and may have completed their syntactic reanalysis. Instead, a significant main effect of gender was found at this region (estimate=63.604, SE=11.295, t=5.631, p<0.001). Gender mismatch conditions yielded significantly longer reading times than gender match conditions. Furthermore, there was no interaction between gender and ambiguity (estimate=6.604, SE=11.296, t<1, n.s.). In both garden path conditions, participants took longer times in the mismatch condition than in the match condition. In addition, there was no interaction between gender and language (estimate=-10.052, SE=11.296, t<1, n.s.), indicating that the gender match effect did not vary between the two language groups. There was only a marginal effect of language (estimate=47.388, SE=28.983, t=1.635, p<0.1).

For the non-native group, the results showed a significant main effect of ambiguity at the reflexive region (estimate=-44.648, SE=16.022, t=-2.787, p<0.01). The garden path effect appeared to be delayed up to this region. At the spill-over 2 region, however, there was a significant main effect of gender match (estimate=74.918, SE=16.705, t=4.485, p<0.001) with no interaction between gender and ambiguity (estimate=5.051, SE=16.786, t<1, n.s.). This meant that in both ambiguity conditions, reading times were greater for the gender mismatch condition than for the match condition. A significant interaction between gender and proficiency (estimate=48.368, SE=17.722, t=2.729, p<0.01) was also found at the region. The gender mismatch effect seemed to be greater for the intermediate group.
### Table 2. Reading times in ms (SE).

<table>
<thead>
<tr>
<th>Region</th>
<th>GP Match</th>
<th>GP Mismatch</th>
<th>NGP Match</th>
<th>NGP Mismatch</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Region 1 (disambiguation): coughed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>458.67 (30.35)</td>
<td>528.57 (35.52)</td>
<td>440.82 (24.15)</td>
<td>433.45 (26.38)</td>
</tr>
<tr>
<td>L2</td>
<td>999.11 (109.91)</td>
<td>1147.50 (141.86)</td>
<td>836.24 (95.36)</td>
<td>776.09 (78.00)</td>
</tr>
<tr>
<td>L2 advanced</td>
<td>954.36 (121.70)</td>
<td>1187.38 (194.43)</td>
<td>654.30 (63.18)</td>
<td>794.19 (108.69)</td>
</tr>
<tr>
<td>L2 intermediate</td>
<td>1061.77 (203.57)</td>
<td>1095.16 (209.31)</td>
<td>1086.41 (202.87)</td>
<td>751.78 (111.79)</td>
</tr>
<tr>
<td><strong>Region 2 (spill-over 1): and warmed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>872.48 (47.26)</td>
<td>961.01 (53.52)</td>
<td>827.94 (43.21)</td>
<td>822.69 (32.84)</td>
</tr>
<tr>
<td>L2</td>
<td>1429.56 (91.08)</td>
<td>1555.55 (102.31)</td>
<td>1242.49 (72.48)</td>
<td>1274.86 (70.77)</td>
</tr>
<tr>
<td>L2 advanced</td>
<td>1535.29 (134.28)</td>
<td>1600.86 (153.42)</td>
<td>1171.30 (88.35)</td>
<td>1265.02 (98.76)</td>
</tr>
<tr>
<td>L2 intermediate</td>
<td>1286.32 (110.94)</td>
<td>1494.16 (123.94)</td>
<td>1340.38 (121.51)</td>
<td>1288.38 (100.82)</td>
</tr>
<tr>
<td><strong>Region 3 (reflexive): himself</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>430.68 (22.91)</td>
<td>421.46 (20.12)</td>
<td>418.90 (21.45)</td>
<td>402.34 (21.92)</td>
</tr>
<tr>
<td>L2</td>
<td>609.39 (38.98)</td>
<td>592.39 (40.04)</td>
<td>533.73 (32.75)</td>
<td>490.90 (21.65)</td>
</tr>
<tr>
<td>L2 advanced</td>
<td>653.26 (58.59)</td>
<td>633.36 (67.27)</td>
<td>551.65 (51.82)</td>
<td>483.86 (31.29)</td>
</tr>
<tr>
<td>L2 intermediate</td>
<td>548.55 (44.13)</td>
<td>538.62 (26.69)</td>
<td>509.66 (32.84)</td>
<td>501.71 (27.12)</td>
</tr>
<tr>
<td><strong>Region 4 (spill-over 2): with a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>785.85 (28.37)</td>
<td>868.36 (43.88)</td>
<td>778.42 (35.35)</td>
<td>903.51 (43.75)</td>
</tr>
<tr>
<td>L2</td>
<td>841.35 (25.77)</td>
<td>987.54 (43.07)</td>
<td>870.68 (30.80)</td>
<td>1020.82 (42.24)</td>
</tr>
<tr>
<td>L2 advanced</td>
<td>873.53 (37.30)</td>
<td>950.86 (47.83)</td>
<td>915.27 (50.01)</td>
<td>974.21 (50.95)</td>
</tr>
<tr>
<td>L2 intermediate</td>
<td>798.09 (32.83)</td>
<td>1045.18 (81.35)</td>
<td>809.38 (21.71)</td>
<td>1083.97 (71.04)</td>
</tr>
</tbody>
</table>

Note. GP: Garden Path, NGP: Non Garden Path

3. Discussion and conclusion

This study examined L2 learners’ processing of garden path sentences, focusing on whether L2 learners’ initial misinterpretations of garden path sentences linger and influence comprehension. Moreover, this study explored the nature of L2 sentence processing compared to L1 processing by employing the gender mismatch paradigm, and it also investigated whether learners’ L2 proficiency modulate syntactic reanalysis and/or lingering effects. Overall, the RT results showed a main effect of ambiguity at the region of disambiguation and its spill-over. This indicates that L2 learners also notice the temporary local ambiguity in a sentence and reanalyze the
structure as L1 readers do. This result is in line with several previous L2 studies (Dussias and Scaltz 2008; Felser, Roberts, Gross, and Marinis 2003; Robert and Felser 2011). What differed from the L1 processing was that the garden path effect was greater and lasted longer. It seemed that residual difficulty from the garden path continued to spill over at the reflexive region, where a main effect of ambiguity was also found. The second NP did not seem to be yet successfully reanalyzed as the matrix subject at this point. This shows that L2 learners have more difficulty in recovering from temporary ambiguity than L1 speakers, which is also consistent with earlier findings (Jacob and Felser 2016). Nevertheless, this does not mean that there was a complete failure of reanalysis because the residual difficulty eventually disappeared after the reflexive region.

More importantly, regarding the first research question, even though L2 learners’ reanalysis was completed as seen above, the results of accuracy data for comprehension questions showed that their initial misinterpretations lingered and influenced their comprehension of the sentences. The rate of accurate responses for comprehension questions was significantly lower in the ambiguous condition than in the unambiguous condition, which is consistent with that of native speakers. This might be because the first misinterpretation lingers after a full reanalysis, not because L2 learners did not know that RAT verbs obligatorily took a reflexive in the absence of a direct object. Note that their answering accuracy in non-garden path conditions was as high as for the native readers. However, compared to the native group, their accuracies in the ambiguous condition were even lower than for the native speakers. It seems that the lingering effects are greater for L2 learners, probably because the garden path effect was also greater for L2 comprehenders. The longer they hold the initial misinterpretation, the stronger the lingering effects could be. In Christianson et al. (2001), native participants were more likely to give a wrong answer when they were presented with a longer garden path sentence, in which the matrix subject would be mistakenly interpreted as a direct object for longer times before disambiguation. Alternatively, their reanalysis might be incomplete so that they might not reject the first misinterpretation. This account, however, is ruled out by the evidence of the gender mismatch effect, which will be explained in detail below.

As for the second research question addressing the nature of L2 processing, L2 learners seemed to construct a detailed syntactic structure despite the relatively late completion of the reanalysis. Even though there was no main effect of gender match
at the reflexive region, a significant main effect of gender match was found at its spill-over region. Since the gender mismatch effect is present only if the link between the reflexive and its antecedent is grammatically permitted, participants were able to construct a syntactic relation between the second NP and the reflexive. Moreover, their patterns did not differ from those of the native speakers, in that there was no interaction between language and gender match at this region. Of course, there are other possibilities for how the interpretation might vary according to what is defined as a detailed representation. In other studies, L2 learners showed different relative-clause attachment preferences (Papadopoulou and Clahsen 2003) or were less sensitive to morphological cues than native speakers (Mueller and Jiang 2013). Thus, depending on types of structures, the results might differ, and many other different structures remain to be analyzed in future research. Also, related to the structure used in the current experiment, one concern can be raised that all of the comprehension questions focused only on the point of the temporary ambiguity which appeared in the relatively early section of each sentence. This might cause the participants to form a strategy of focusing more on the words in the early section and less on the rest of the sentence. However, even though it was possible that they paid attention to the reflexive section relatively less than garden path parts, the results showed a gender mismatch effect still occurred at the spill-over region, and this might be evidence for their considerable syntactic analysis during processing.

As for the third research question on the role of L2 proficiency in L2 processing, there was no overall difference between proficiency levels. Less proficient learners tended to read more slowly, but a main effect of proficiency group was not found at any region. However, there was an interaction between gender and proficiency at the spill-over 2 region. The gender mismatch effect seemed to be larger for the less proficient group in both ambiguous and unambiguous conditions, who were expected to have less detailed syntactic representations. Thus, less proficient learners were also able to construct a syntactic relation, but they took more time to read sentences due to their insufficient reading skills. Note that almost half the participants were excluded from the data analysis for their low comprehension accuracies, and in particular, most of them were intermediate participants, whose responses to filler sentences could not meet the 70% accuracy criterion. Likewise, among the intermediate participants included in the data analysis, their overall insufficient reading skills still might lead to greater reading times.
More importantly, the gender mismatch effect also seemed to influence global comprehension of the sentence. The gender mismatch effect could be a factor affecting the readers’ comprehension, since gender mismatch information seems to degrade the reliability of the reanalysis. It would either increase or decrease the comprehension accuracy depending on how readers use this mismatch information for sentence parsing. Assuming that the first misinterpretation lingers and it competes with the reanalyzed interpretation for global comprehension, the gender mismatch information would be considered as evidence for rejecting the second interpretation. Thus, comprehenders would respond even less accurately in gender mismatch conditions. That is what we found in the results of the advanced learners. Their accuracy rate for mismatch and garden path condition was only 9%, which was even lower than the 32% in the match and garden path conditions. However, the same tendency was not observed in the native speakers. The native speakers did not show any difference between the gender match and mismatch conditions. As they were able to recover from the temporary ambiguity earlier, reinterpreting the matrix subject against the biased sex (e.g. the boxer as a woman) with ease, they had no necessity to reject the second interpretation.

Interestingly, the intermediate group showed a lower accuracy rate in the match condition than in the mismatch condition, especially in the ambiguous conditions. This is inconsistent with previous findings. According to Robert and Felser (2011), it is more likely that the intermediate group responded less accurately in the mismatch condition, because less proficient learners are less dependent on syntactic cues in parsing the structure, at least not more than proficient learners. However, we obtained the opposite pattern, and we have two possible explanations. First, it is possible that while the intermediate students read the reflexive and its spill-over region slowly (the data showed their longer reading time data), their first misinterpretations decayed and did not linger any more. This can be supported by the claim that the lingering effect originates from memory traces which remain in the first parsing (Kaschak and Glenberg 2004; Slattery et al. 2013; Sturt 2007). From this perspective, the intermediate group took longer to read the spill-over region, which caused their initial misinterpretation to stop lingering and helped them reject it. In fact, in the match condition where their accuracy rate was lower, their reading times were numerically shorter than for the advanced group. This might suggest the possibility that the
first misinterpretation decayed over time while reading. Second, given that the gender mismatch effect was greater for the intermediate learners, it is possible that they paid more attention to reading the sentences, and this might lead to more accurate response in the mismatch condition. As shown in the previous studies, differences in reading goals (translation compared to normal reading) or reading strategies (sub-vocal simulation compared to normal reading) could facilitate readers’ attention to syntactic structure or improve their comprehension (Lim and Christianson 2013a; Zhou and Christianson 2016)

In conclusion, the accuracy data for comprehension questions provided evidence that the initial misinterpretation lingers and affects L2 learners’ comprehension, which is consistent with previous L1 processing studies. This is not because L2 learners failed to reanalyze the correct structure or because they lack a detailed syntactic representation. Instead, based upon the gender mismatch effect observed in both garden path and non-garden path conditions, they indeed successfully reanalyzed the detailed syntactic structure. Therefore, we can argue that L2 learners do not lack a detailed syntactic representation but routinely compute the syntactic relations to some extent, for which it seems “good enough” for them to comprehend.

References


Clahsen, Harald and Claudia Felser. 2006b. Continuity and shallow structures in language
Syntactic reanalysis and lingering misinterpretations in L2 sentence processing 73


Drummond, Alex. 2012. Ibex: A web interface for psycholinguistic experiments. [https://github.com/addrummond/ibex]


Appendix

Below are listed the target sentences which are used in the experiment.

1. While Mary washed, the boxer who was big coughed and warmed himself with a cover.
2. When Tiffany hid, the boy who was smart ran and covered himself with the blanket.
3. As Andy dressed, the bride who was pretty smiled and prepared herself for the wedding.
4. While John woke up, the nurse who was tall sighed and dragged herself out of bed.
5. When Mary bathed, the boy who was cute giggled and enjoyed himself with a rubber duck.
6. As Monica groomed, the boy who was little smiled and entertained himself with a book.
7. While David woke up, the nun who was old yawned and prepared herself for the service.
8. When John dressed, the girl who was pretty laughed and prepared herself for the school.
9. As James dressed, the girl who was blond sighed and served herself the whole-grain cereals.
10. While James calmed down, the babysitter who was short smiled and prepared herself to go home.
11. When Sarah groomed, the boy who was sleepy cried and covered himself with the blanket.
12. As Jessica washed, the firefighter who was tall smiled and served himself a strong drink.
13. While John dried off, the cheerleader who was skinny frowned and lowered herself below the desk.
14. When Jessica undressed, the boy who was small shouted and threw himself into the couch.
15. As Monica hid, the detective who was smart laughed and declared himself an
expert on finding.

16. While Sarah shaved, the soldier who was short frowned and reminded himself of the mission.

Below are listed the filler sentences. For 21-32, one of four sentences was used for the experiment.

1. When James scratched the ballerina she frowned and told herself a short story.
2. While Joan calmed down the secretary he cried again and told himself something.
3. As Kyle settled down her, the woman who was fat frowned and lowered herself.
4. When Josh bathed the baby, he happily giggled and threw himself into the bath tub.
5. The coach hugged by Brian and Nicole had an amazing career.
6. After Brad was soaked by the clown he never went to another circus again in his life.
7. Colin bothered by the salesman decided to shop at another store.
8. The rats contaminated by the rabid dog had to be destroyed.
9. Joan registered by the attendant left to find her classroom.
10. After Tony was tested by the doctor, he was allowed to finish the game.
11. After Alexander was crushed by the car door he had to have surgery on his hand.
12. After David was taught by his mother, he finally got straight A in the semester.
13. Before Sally was cleaned by her mother she was covered in spaghetti made for dinner.
14. Before Kate was hired by the owner, she used to work seven nights per week.
15. Soon after Cindy was kissed by the celebrity she got so excited that she literally fainted.
16. When Jenny was kidnapped by the stalker, she managed to escape.
17. The cook criticized by Charlie and his wife used too much salt.
18. Susan embraced by the grandparent felt safe and warm.
19. The parent worried by Timmy decided to talk to him.
20. Jack assessed by the supervisor was not right for the job.
21. James anticipated his opponent’s move would be just as expected.
   James anticipated his opponent’s move would be quite unexpected.
   James anticipated that his opponent’s move would be just as expected.
   James anticipated that his opponent’s move would be quite unexpected.
22. James Bond read the agent’s name was actually Christopher John Whitehouse.
   James Bond read the agent’s name was actually never going to be disclosed.
   James Bond read that the agent’s name was actually Christopher John
Whitehouse.

James Bond read that the agent’s name was actually never going to be disclosed.

23. The audience appreciated the new opera was actually a true modern masterpiece. The audience appreciated the new opera was actually not going to be performed. The audience appreciated that the new opera was actually a true modern masterpiece. The audience appreciated that the new opera was actually not going to be performed.

24. Morse revealed the murder weapon had been a Russian-made revolver. Morse revealed the murder weapon had been impossible to determine. Morse revealed that the murder weapon had been a Russian-made revolver. Morse revealed that the murder weapon had been impossible to determine.

25. Old women recalled the events of the war had been quite upsetting to them. Old women recalled the events of the war had been too long ago to remember. Old women recalled that the events of the war had been quite upsetting to them. Old women recalled that the events of the war had been too long ago to remember.

26. Mary perceived the other people were definitely in the room too. Mary perceived the other people were definitely not in the room. Mary perceived that the other people were definitely in the room too. Mary perceived that the other people were definitely not in the room.

27. The explorers found the South Pole was actually right at their feet. The explorers found the South Pole was actually impossible to reach. The explorers found that the South Pole was actually right at their feet. The explorers found that the South Pole was actually impossible to reach.

28. The astronomer saw the full moon was definitely clearly visible that night. The astronomer saw the full moon was definitely not visible that night. The astronomer saw that the full moon was definitely clearly visible that night. The astronomer saw that the full moon was definitely not visible that night.

29. The mayor disclosed the budget was going to involve increases in spending. The mayor disclosed the budget was going to be announced later that week. The mayor disclosed that the budget was going to involve increases in spending. The mayor disclosed that the budget was going to be announced later that week.

30. The scientist observed the insects were actually flying very fast.
The scientist observed the insects were actually too small to see.
The scientist observed that the insects were actually flying very fast.
The scientist observed that the insects were actually too small to see.
31. The little boy overheard his parents were now trying to get a divorce.
The little boy overheard his parents were now missing, presumed dead.
The little boy overheard that his parents were now trying to get a divorce.
The little boy overheard that his parents were now missing, presumed dead.
32. The director acknowledged his assistant had been a great help to him.
The director acknowledged his assistant had been no help at all to him.
The director acknowledged that his assistant had been a great help to him.
The director acknowledged that his assistant had been no help at all to him.

Jonghyeon Lee
Department of English Language and Literature
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
E-mail: museeq@snu.ac.kr

Jeong-Ah Shin
Department of English Language and Literature
Dongguk University
30 Pildong-ro 1 gil, Jung-gu, Seoul 04620, Korea
E-mail: jashin@dongguk.edu

Received: 2016. 04. 04.
Revised: 2016. 08. 02.
Accepted: 2016. 08. 02.