Revision of Predictive Inferences and Japanese EFL Learners’ Text Comprehension Processes: A Study of Eye Movements During Reading

Shingo NAHATAME
Graduate School, University of Tsukuba / The Japan Society for the Promotion of Science

Abstract

Readers often make inferences about the likely outcomes of events described in the text (i.e., predictive inferences). However, these inferences are sometimes inconsistent with the following text. In this case, readers need to revise their inferences for successful comprehension. This study investigated Japanese university EFL students’ text comprehension when encountering the context disconfirming their predictive inferences. In the experiment, participants read several short narratives where a predictive inference was initially induced but later disconfirmed. Participants’ eye movement data were collected to examine real-time comprehension processes during reading. Participants’ text memory was also assessed by a sentence recognition task after reading. The results of the recognition task showed that readers had difficulty eliminating disconfirmed predictive inferences from text memory. More importantly, eye movement analysis suggested that both higher and lower proficiency readers immediately detected inconsistencies between drawn inferences and the context disconfirming the inferences during reading; however, lower proficiency readers experienced more difficulty in integrating the disconfirming context into current comprehension. These results suggest the difficulty specific to less proficient EFL readers.

1. Background

When reading a narrative text, readers often make inferences about the likely consequence of the event described therein. For example, when reading the sentence As the pitcher released the ball, the batter raised his bat and the ball went directly towards him, most readers would infer that the batter will hit the ball. This kind of inference is called a predictive inference. Predictive inferences are drawn not only in first language (L1) reading, but also in second language (L2) reading (e.g., Horiba, 1996; Yoshida, 2003).

However, the inferences readers make are not always confirmed by the following text. For instance, in the above example the text may be followed by a sentence disconfirming the drawn inference, such as Suddenly, the ball dropped in front of the bat and fell in the catcher’s mitt. In
this case, the readers need to revise their initial inferences to achieve appropriate text comprehension. Ushiro (2010) asserted that flexibly revising text comprehension during reading is required for L2 learners to become successful readers. To provide insights into such flexible text comprehension processes among Japanese learners of English as a foreign language (EFL), the present study investigated text comprehension when predictive inferences are initially drawn but later disconfirmed in reading. Specifically, it aimed to confirm and extend the findings of the previous study (Nahatame, 2014), by collecting learners’ eye movement data during reading. The following sections provide a review of related studies.

1.1 Predictive Inferences in Reading

Predictive inference is beneficial for readers in ways such as easing the processing of the subsequent context, promoting construction of situation models, and encouraging active engagement with the text (Allbritton, 2004; Linderholm, 2002). Previous studies found that predictive inferences are automatically drawn during L1 reading under some conditions, such as when the context strongly constrains a single, specific inference (Klin, Guzmán, & Levine, 1999) or when the inference relates to motivation for a narrative character’s action (Klin, Murray, Levine, & Guzmán, 1999). Most previous L1 studies investigated this type of inference generation by analyzing readers’ response times to inference-related words (e.g., lexical decision times to hit) that were collected immediately after reading the passage.

Predictive inferences are also drawn during L2 reading. Horiba (1996) and Yoshida (2003) analyzed L2 readers’ think-aloud protocols and suggested that they make several types of inferences, including predictive inferences, during reading. These studies also revealed that readers with higher L2 proficiency are more likely to make inferences. Moreover, Nahatame (2013) revealed that predictive inference generation during L2 reading is facilitated by text characteristics (i.e., strong contextual constraint and relatedness to characters’ motivation) based on analyses of recognition times to inference-related words and recall protocols.

In contrast, only a limited number of studies have investigated how readers deal with predictive inferences that are disconfirmed by the subsequent context. For example, in a study of L1 Japanese readers, Iseki (2006) analyzed response times to inference-related sentences, and demonstrated that the activation of predictive inferences was decreased (i.e., suppressed) after reading the context disconfirming the inferences. Nahatame (2014) conducted a similar experiment with Japanese university EFL students. In this experiment, participants engaged in a memory test (i.e., sentence recognition) in addition to the response task. The analysis of response times suggested that readers failed to suppress the activation of disconfirmed inferences during reading, regardless of L2 proficiency. Furthermore, the results of sentence recognition indicated that readers had some difficulty eliminating the disconfirmed inferences from text memory: Although they correctly recognized the sentence describing disconfirmed inferences (e.g., The boy
hit the ball) as not written in the text, they were less confident of their judgments. These results suggest the difficulty of revising predictive inferences for L2 readers.

However, some methodological limitations remain in Nahatame’s (2014) experiment. First, participants read experimental passages sentence-by-sentence on the computer screen before the response task. This prevented participants from looking back to prior context during reading, which might increase the cognitive load on reading compared to natural reading. Second, because the response task was conducted after reading the sentence disconfirming the inference (e.g., Suddenly, the ball dropped...), it provided little information about how participants processed the sentence. Consequently, it is difficult to determine what kinds of comprehension processes are related to the difficulty of revising predictive inferences. For instance, given that L2 readers are more likely than L1 readers to dismiss inconsistencies included in the passages (Morishima, 2013), it is possible that the participants in Nahatame’s study dismissed inconsistencies between drawn inferences and the disconfirming context, or that insufficient attention was paid to the inconsistencies. Alternatively, even though participants were aware of inconsistencies, they might have more difficulty integrating the disconfirming context with prior context. Long and Chong (2001) suggested such difficulty among poor L1 readers when the text included inconsistencies. However, examining these possibilities requires the use of measures that can tap into real-time comprehension processes when encountering the sentence disconfirming the inferences.

To overcome these limitations of Nahatame’s (2014) study, it will be helpful to adopt eye-tracking (i.e., recoding participants’ eye movements during reading). The next section reviews previous eye-tracking studies of significance to the present study.

1.2 Use of Eye-Tracking in Reading Research

Eye-tracking has been widely used in the study of L1 reading, and a growing number of L2 studies are also adopting this methodology (see Roberts & Siyanova, 2013, for a review). Eye-tracking is useful in reading research for the following reasons. First, it allows participants to read a passage in a more natural situation by presenting an entire passage and not giving participants extra requirements (e.g., frequently pressing a button). Second, it discriminates between effects on early comprehension processes (e.g., lexical access) and late comprehension processes (e.g., information reanalysis, discourse integration) during reading. Third, it captures participants’ regressions or look backs to the previously processed context.

Previous L2 studies have often used eye-tracking to investigate learners’ lexical processing and syntactic parsing. For example, Kadota and Kuramoto (2006) used eye-tracking to examine Japanese EFL learners’ processing of garden-path sentences. Nevertheless, some researchers suggest that eye-tracking is applicable to the study of processing larger text units, such as discourse (Hyöniä, Lorch, & Rinck, 2003; Rayner, Chace, Slattery, & Ashby, 2006). Rayner et al. (2006) noted that, “the time is ripe for more comprehension studies to use eye movement data to understand discourse processing” (p. 252). Indeed, some studies have used eye-tracking to
investigate discourse processing in L1 reading (e.g., Rayner et al., 2006; Rinck, Gamez, Diaz, & De Vega, 2003; van der Schoot, Reijntjes, & van Lieshout, 2012). These studies used similar eye movement measures to lexical processing or syntactic parsing research (e.g., first-pass and second-pass reading times), but most defined the clause or sentence as the unit of eye movement analysis, rather than a single word or phrase.

Some studies have explored how L1 readers deal with inconsistent statements during discourse processing. For instance, van der Schoot et al. (2012) measured the eye movements of L1 children while reading a narrative passage containing inconsistent character information. The passage included the target sentence describing a character’s action (e.g., Peter ordered a hamburger), which was either consistent or inconsistent with previously introduced character information (e.g., Peter always wanted to eat junk food or Peter had been a strict vegetarian). The results showed that inconsistencies increased the duration of first-pass fixations on target sentences as well as total fixation durations at the end of the sentence. Although Rinck et al. (2003) conducted a similar experiment on the processing of temporally inconsistent information (Mark arrived before Claudia vs. Claudia was already waiting for Mark when he arrived), eye movements on target sentences were not affected by inconsistencies in the text; instead, inconsistencies caused more and longer regressions towards the previously stated information.

The present study examined Japanese EFL readers’ eye movements using a similar methodology to these studies on inconsistency processing. Specifically, it analyzed eye movements on the sentence disconfirming predictive inferences drawn from the prior context (see Scoring and Data Analysis for further explanations).

1.3 Hypothesis and Research Questions of This Study

As described above, Nahatame’s (2014) methodological limitations can be overcome by the use of the eye-tracking method. Specifically, it can examine learners’ comprehension processes in a more natural reading setting. Furthermore, it allows us to investigate the real-time processing of sentences that disconfirm predictive inferences. It can also identify what comprehension processes are more related to difficulty in revising inferences by discriminating early and late processes in text comprehension.

The purpose of the present study was to confirm and extend Nahatame’s (2014) findings. First, to refine the previous finding that learners have difficulty eliminating disconfirmed inferences from text memory, their text memory was examined using the same recognition task used by Nahatame. Second, eye movement data during reading were collected to examine comprehension processes when readers encounter sentences that disconfirm predictive inferences. Furthermore, comprehension processes were examined in terms of learners’ L2 reading proficiency. Consequently, the following hypothesis (H) and research questions (RQs) were addressed:

H  Japanese EFL readers have difficulty eliminating disconfirmed predictive inferences from text memory.
RQ1 How does the disconfirmation of predictive inferences affect Japanese EFL readers’ text comprehension processes?

RQ2 Does L2 reading proficiency influence the extent to which predictive inference disconfirmation affects Japanese EFL readers’ text comprehension processes?

2. Method

2.1 Participants
The participants were 26 Japanese university students (16 male and 10 female; aged 18–22 years, $M = 19.96, SD = 1.46$). All participants were native Japanese speakers with normal or corrected-to-normal visual acuity. In addition, students had studied English as a foreign language for more than six years as part of Japanese formal education, and they were assumed to have intermediate-level English proficiency. The data from eight participants was excluded from analyses because of inaccurate eye-movement recordings. Therefore, the following analyses were based on 18 participants.

2.2 Materials
2.2.1 Passages
Twelve short narratives called disconfirming passages were adopted from Nahatame (2014). Each passage consisted of five sentences and included approximately 70 words. The fourth sentence strongly induces a single, specific predictive inference (i.e., target inference), but this inference is disconfirmed by the fifth sentence. For the sample passage in Table 1, the fourth sentence strongly suggests that the boy will hit the ball in the next situation, but the fifth sentence indicates that he actually did not hit the ball. Nahatame’s (2014) response and recognition tasks showed that Japanese university students with intermediate-level English proficiency were likely to draw target inferences after reading the fourth sentence.

| Table 1 |
| Sample of Experimental Passages |
| The boys’ high school baseball team was having tests for the spring season. The coach decided to test the boys’ baseball skills before he did anything else. The first batter to step up to the plate was a new boy on the team. As the pitcher released the ball, the boy raised his bat and the ball went directly towards him. (Disconfirming) / without knowing it was a forkball. (Control) Suddenly, the ball dropped in front of the bat and fell in the catcher’s mitt. |

The present study constructed new control passages corresponding to each disconfirming passage. The control passages were created by rewriting the fourth sentence of each disconfirming passage so that it did not induce the target inference and naturally precedes the content of the fifth
sentence (see Table 1 for an example). The plausibility of the fourth sentences of the control passages was confirmed and improved in a small pilot study with 12 Japanese graduate students majoring in applied linguistics. Furthermore, the fourth sentences of the control passages were not significantly different than those of the disconfirming passages in terms of the sentence length (i.e., number of words), lexical frequency (JACET 8000 level), and semantic relatedness to the fifth sentences, all ts < 1.77 (see Table 2 for the descriptive statistics).

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Number of words</th>
<th>Lexical frequency</th>
<th>Semantic relatedness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Disconfirming</td>
<td>15.67</td>
<td>3.55</td>
<td>1.48</td>
</tr>
<tr>
<td>Control</td>
<td>16.17</td>
<td>3.46</td>
<td>1.45</td>
</tr>
</tbody>
</table>

Note. Lexical frequency is the mean JACET 8000 level of words included in the sentence. Semantic relatedness ranges from 0 to 1 and was calculated using Latent Semantic Analysis.

In addition to the experimental passages, six filler passages were used to prevent participants from becoming aware of the purpose of the experiment. The filler passages were similar in length to the experimental passages, but neither elicited nor disconfirmed any specific inferences. All experimental and filler passages were paired with simple yes-no comprehension questions to encourage careful reading. These questions concerned a piece of information explicitly stated in the text before the fifth sentences (e.g., Was the first batter a new boy on the team?).

Two material sets were constructed, resulting from the counterbalance of passage types. Each list included six disconfirming, six control, and six filler passages. This ensured that each participant read an equal number of experimental passages of each type, and that each type of experimental passage was presented to one half of the participants.

2.2.2 Target Sentences for the Recognition Task

Target sentences for the post-reading recognition task were also adopted from Nahatame (2014). Each text was paired with three types of target sentences: explicit, inference, and inconsistent. An explicit sentence described an event explicitly mentioned in the passage (e.g., The coach decided to test the boys’ baseball skills). An inference sentence described a future event that were initially predicted but later disconfirmed in the disconfirming passage (e.g., The boy hit the ball). An inconsistent sentence described an event not mentioned or suggested in the passage (e.g., The boy cleaned his baseball glove). Target sentences were also written for filler passages, but each filler passage was paired with two explicit sentences and one inconsistent sentence to balance the number of yes-no responses in the task.
Note that all of the target sentences were written in Japanese to avoid the effect of participants’ surface text memory about word forms and sentence structures on recognition judgments (Muramoto, 2000; Nahatame, 2014). Furthermore, similar to previous studies, each target sentence was paired with a 4-point scale of recognition judgment confidence (1 = low, 2 = relatively low, 3 = relatively high, and 4 = high).

2.2.3 L2 Reading Proficiency Test

To examine participants’ L2 reading proficiency, an English reading test was constructed based on the reading subsection of the Eiken test (Eiken Foundation of Japan). The Eiken test has been widely used to measure Japanese EFL learners’ reading proficiency. Considering the participants’ estimated proficiency, two passages were chosen from retired copies of Grade Pre-1, two passages from Grade 2, and one passage from Grade Pre-2. Each passage included three to five multiple-choice questions and there were 20 items in total.

2.3 Apparatus and Procedure

The L2 reading proficiency test was conducted 30 minutes prior to the main experimental session. In the main session, participants sat approximately 40 cm from the computer screen displaying the texts. At this distance, two letter spaces equaled 1° of the visual angle. The participants’ head position was fixed by means of a chinrest. Using the SuperLab 4.5 program, each text was displayed in its entirety on the computer screen. To make analysis of eye movements more accurate, each sentence started on a new line, and line spacing was inserted between sentences. Eye movements during reading were recorded with an eye-tracker (EMR 9) manufactured by nac Image Technology Inc. (Tokyo, Japan).

After the eye-tracker was adjusted for optimal tracking, participants’ eye movements were calibrated with a standard 9-point grid. In the subsequent experimental trials, participants were given one of the two material sets. Each trial began with “Ready” displayed on screen. Participants were asked to push the button on the Response Pad RB-730 to indicate they were ready to read a passage. Pushing the button caused the passage to appear on the screen. Participants then silently read the passage in a self-paced manner. They were instructed to press the button again when they finished the reading. Subsequently, the current passage was replaced with the comprehension question. Participants responded to the question using a pair of yes-no buttons, and received feedback about accuracy. This trial was repeated for each of the 18 passages presented in random order. Before starting the experiment, participants completed practice trials for three passages.

After reading all passages, participants received the booklet that presented target sentences for the recognition task. The booklet included all three types of target sentences for each text they had just read. Participants were instructed to: (a) judge whether each sentence was described in the text they had read, and (b) rate how confident they were of their judgments on a 4-point scale.
2.4 Scoring and Data Analysis

The data from the sentence recognition task was analyzed in the same way as in Muramoto (2000) and Nahatame (2014). The recognition scores, ranging from 0 to 6, were calculated from participants’ responses to the target sentences and their confidence ratings (see Table 3). When participants recognized a target sentence as written in the text, the scores ranged from 3–6, depending on their confidence rating. In contrast, when participants recognized a target sentence as not written in the text, the scores ranged from 0–3 contingent upon their confidence rating. If participants did not make the target inferences or eliminated the drawn inferences from text memory, inference sentences should be recognized as not written in the text with high confidence, resulting in scores close to 0.

Table 3

<table>
<thead>
<tr>
<th>Response</th>
<th>Confidence rating</th>
<th>Recognition scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>This sentence was written in the text.</td>
<td>4 (high)</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>3 (relatively high)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2 (relatively low)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1 (low)</td>
<td>3</td>
</tr>
<tr>
<td>This sentence was not written in the text.</td>
<td>1 (low)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2 (relatively low)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3 (relatively high)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4 (high)</td>
<td>0</td>
</tr>
</tbody>
</table>

Regarding the eye movement data, those from the selected eye (usually the right eye) were analyzed. All trials where recording problems occurred were omitted from analysis (1.39% of all trials). This study analyzed fixation times on the fifth sentences of experimental passages (these sentences were common between passage types and defined as target sentences). Prior to analysis, short fixation durations of less than 100 ms were excluded, along with long fixation durations beyond three standard deviations from each participant’s mean in a condition (1.63% of the data set). In the analysis, individual eye fixations within a target sentence were ignored, but their durations were summed to yield the fixation durations for the sentence (Hyōnä et al., 2003).

As in previous studies (e.g., Rinck et al., 2003; Hyōnä et al., 2003), this study analyzed first-pass and second-pass reading times on target sentences. First-pass reading time is defined as the sum of the duration of all fixations on the target sentence until finishing reading the sentence or before looking back to prior sentences. This eye fixation measure reflects the initial processing of the sentence, and is indicative of the readers who noticed inconsistencies between drawn inferences and the sentence being processed. If readers consciously noticed the inconsistencies, such awareness should make initial processing of the target sentence more difficult, resulting in an
increase of first-pass reading times (van der Schoot et al., 2012). On the other hand, if readers dismissed or did not fully notice the inconsistencies, little effect should appear on this measure.

Second-pass reading time was defined as the duration of additional fixations on the target sentence that occurred after looking back to one or more preceding sentences. This eye fixation measure reflects later comprehension processes, such as the integration of the target sentence with comprehension of preceding context (i.e., developing representations). Thus, if readers have more difficulty in integrating the target sentences, some effects are assumed to appear on this measure.

In addition to these eye fixation measures, sentence wrap-up time was also analyzed as the total fixation times on the final two words of the target sentence (van der Schoot et al., 2012). This was because second-pass reading on the target sentences did not frequently occur in this study (see Results and Discussion). The processes during reading the end of the sentence (i.e., sentence wrap-up) include not only integrating words with the sentence being processed, but also integrating the sentence with the preceding context (Just & Carpenter, 1980). Thus, this eye fixation measure also indicates the discourse integration difficulty of target sentences.

3. Results and Discussion

3.1 L2 Reading Proficiency Test

The reliability of the L2 reading proficiency test was sufficient (Cronbach’s α = .81) after excluding three low discriminability items. Before the main analysis, participants were classified into either a higher (n = 8) or lower reading proficiency group (n = 10) based on a median split of the test scores. The mean scores were significantly higher for the higher proficiency group (M = 15.00, SD = 1.67, Max = 17, Min = 12) than the lower proficiency group (M = 8.60, SD = 1.84, Max = 11, Min = 6), t(16) = 7.76, p < .001, d = 3.68.

3.2 Sentence Recognition Task

Table 4 shows the mean recognition scores for the sentence recognition task. A 3 (Sentence: explicit, inference, inconsistent) × 2 (Passage: disconfirming, control) × 2 (Proficiency: higher, lower) analysis of variance (ANOVA) was conducted on the mean scores, with Sentence and Passage as within-participants variables and Proficiency as a between-participants variable. The results indicated significant main effects of Sentence, F(2, 32) = 228.20, p < .001, ηp² = .93, and Passage, F(1, 16) = 11.97, p = .003, ηp² = .43. In addition, there was a significant interaction between Sentence and Passage, F(2, 32) = 8.48, p = .001, ηp² = .35. Any other potential main or interaction effects were not significant or marginally significant (all Fs < 1).

The subsequent analysis of the Sentence × Passage interaction revealed a significant difference between disconfirming and control passages for inference sentences (p < .001, d = 1.10), with higher scores for disconfirming than control passages. There was no significant difference between passages for explicit (p = .492, d = 0.23) and inconsistent sentences (p = .527, d = 0.15).
These results suggest that disconfirming and control passages differed only in the likelihood of target inference generation: Participants were less likely to generate target inferences when reading control compared to disconfirming passages. This supports the claim that the control passages worked as intended in the present study.

Table 4

<table>
<thead>
<tr>
<th></th>
<th>Explicit</th>
<th></th>
<th>Inference</th>
<th></th>
<th>Inconsistent</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dis</td>
<td>Con</td>
<td>Dis</td>
<td>Con</td>
<td>Dis</td>
<td>Con</td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Higher</td>
<td>5.13</td>
<td>0.59</td>
<td>2.21</td>
<td>1.06</td>
<td>0.92</td>
<td>0.67</td>
</tr>
<tr>
<td>Lower</td>
<td>5.17</td>
<td>0.49</td>
<td>2.53</td>
<td>1.07</td>
<td>0.73</td>
<td>0.38</td>
</tr>
</tbody>
</table>

*Note.* The scores ranged from 0 to 6. Dis = disconfirming passages; Con = control passages.

More importantly, mean recognition score of the inference sentences for the disconfirming passages ($M = 2.39$, $SD = 1.05$) was similar to that of Nahatame (2014) ($M = 2.62$, $SD = 1.20$), $t(52) = 0.69$, $p = .491$, $d = 0.20$. This score suggests that participants in the present study correctly recognized the sentences describing the disconfirmed inferences (e.g., *The boy hit the ball*) as not written in the text, but the confidence of their judgments were low to relatively low (see Table 3). Thus, these results support the hypothesis that Japanese EFL learners have some difficulty eliminating the memory trace of initially activated, but later disconfirmed, predictive inferences. The following section discusses what comprehension processes are related to such difficulty based on eye movement data during reading.

3.3 Eye Movement Data

3.3.1 First-Pass Reading Times

Table 5 shows mean first-pass reading time of target sentences for each passage type and proficiency group. A 2 (Passage: disconfirming, control) $\times$ 2 (Proficiency: higher, lower) mixed ANOVA indicated significant main effects of Passage, $F(1, 16) = 12.78$, $p = .003$, $\eta_p^2 = .44$, and Proficiency, $F(1, 16) = 5.77$, $p = .029$, $\eta_p^2 = .27$. Specifically, mean reading time was significantly longer for disconfirming than control passages. Furthermore, mean reading time was significantly shorter for higher than lower proficiency learners. There was no significant interaction effect between Passage and Proficiency, $F(1, 16) = 0.26$, $p = .620$, $\eta_p^2 = .02$.

As already noted, disconfirming and control passages differed only in the generation of target inferences. Therefore, the longer first-pass reading time for the disconfirming passages is attributed to the generation of the target inferences during reading the fourth sentences, which made the initial processing of the subsequent sentences more difficult. This suggests that
participants noticed inconsistencies between drawn inferences and target sentences during the immediate processing of the sentence, and such awareness resulted in more and longer fixations on target sentences during the first-pass reading.

Table 5
First-Pass Reading Times (in Milliseconds) of Target Sentences

<table>
<thead>
<tr>
<th></th>
<th>Disconfirming</th>
<th></th>
<th>Control</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Higher</td>
<td>6,455</td>
<td>2,339</td>
<td>5,421</td>
<td>1,498</td>
</tr>
<tr>
<td>Lower</td>
<td>8,370</td>
<td>1,359</td>
<td>6,996</td>
<td>1,513</td>
</tr>
</tbody>
</table>

3.3.2 Second-Pass Reading Times

Mean second-pass reading time of target sentences are shown in Table 6. A 2 (Passage: disconfirming, control) × 2 (Proficiency: higher, lower) mixed ANOVA found neither significant main effects of Passage, $F(1, 16) = 0.00, p = .992, \eta_p^2 = .00$, nor Proficiency, $F(1, 16) = 0.66, p = .427, \eta_p^2 = .04$. There was also no significant interaction effect between Passage and Proficiency, $F(1, 16) = 0.33, p = .574, \eta_p^2 = .02$. The lack of significant difference between disconfirming and control passages can be interpreted as showing that the difficulty of integrating target sentences into developing text representations was not affected by whether the sentences disconfirmed the predictive inferences drawn from the prior context.

Table 6
Second-Pass Reading Times (in Milliseconds) of Target Sentences

<table>
<thead>
<tr>
<th></th>
<th>Disconfirming</th>
<th></th>
<th>Control</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Higher</td>
<td>291</td>
<td>428</td>
<td>439</td>
<td>748</td>
</tr>
<tr>
<td>Lower</td>
<td>694</td>
<td>945</td>
<td>616</td>
<td>1,221</td>
</tr>
</tbody>
</table>

However, it should be noted that the mean second-pass reading time was quite small, regardless of passage types and L2 proficiency, which corresponded to only a few fixations. This suggests that participants did not frequently reread the target sentences after regressions towards the prior sentences. In addition, the large standard deviations suggest that second-pass reading times greatly varied between individual participants. Indeed, the overall probability of second-pass reading on target sentences was less than 20%, and no second-pass reading on target sentences occurred throughout all trials for seven of 18 participants. Therefore, it is difficult to interpret the results of second-pass reading times based on such a limited number of fixations and participants. Consequently, this study also analyzed the sentence wrap-up times as another measure reflecting
discourse integration difficulty of target sentences (see the following section).

There are some possible reasons why rereading of the target sentences did not frequently occur. One is that, instead of looking back to prior context, readers could rely on their text memory to access prior text information because each passage consisted of only five simple sentences. Another possibility is that readers stayed on the target sentence, having felt that the existing text information was not sufficient to revise the drawn inferences and hoping to find further information in the subsequent context.

3.3.3 Sentence Wrap-Up Times

Table 7 shows mean wrap-up times of target sentences. A 2 (Passage: disconfirming, control) × 2 (Proficiency: higher, lower) mixed ANOVA indicated significant main effects of Passage, $F(1, 16) = 13.71, p = .002, \eta_p^2 = .46$, and Proficiency, $F(1, 16) = 9.50, p = .007, \eta_p^2 = .37$. More importantly, there was a significant interaction effect between Passage and Proficiency, $F(1, 16) = 15.50, p = .001, \eta_p^2 = .49$. The subsequent analysis indicated that sentence wrap-up times were not significantly different between disconfirming and control passages among the higher proficiency group ($p = .877, d = 0.05$), whereas the lower proficiency group demonstrated significantly longer times for disconfirming than control passages ($p < .001, d = 1.17$). These results suggest that lower proficiency learners had considerable difficulty integrating the target sentences into their developing text representations when reading disconfirming passages.

<table>
<thead>
<tr>
<th>Wrap-Up Times (in Milliseconds) of Target Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disconfirming</td>
</tr>
<tr>
<td>$M$</td>
</tr>
<tr>
<td>Higher</td>
</tr>
<tr>
<td>Lower</td>
</tr>
</tbody>
</table>

Although higher proficiency learners showed no significant difference in wrap-up times between passage types, this result should be interpreted with care. The results of sentence recognition suggest that the disconfirmed inferences were not completely deleted from text memory, even for higher proficiency learners. Furthermore, Nahatame's (2014) response time analysis suggested that even higher (generally, upper-intermediate) proficiency learners failed to suppress the activation of drawn inferences immediately after reading the disconfirming sentence. Therefore, it is possible that in the present study higher proficiency learners also experienced some difficulty integrating disconfirming sentences, although the difficulty was not large enough to be reflected in eye movement data. Nevertheless, the results of wrap-up times at least suggest that lower proficiency learners experienced much greater difficulty than higher proficiency learners in integrating the disconfirming sentences into developing text representations.
3.4 Comprehension Questions

A 2 (Passage: disconfirming, control) × 2 (Proficiency: higher, lower) mixed ANOVA was conducted on the correct answer rates of comprehension questions (see Table 8). The results indicated a significant main effect of Passage, $F(1, 16) = 4.87, p = .042, \eta_p^2 = .23$, but there were neither a significant main effect of Proficiency, $F(1, 16) = 0.52, p = .482, \eta_p^2 = .03$, nor an interaction between Passage and Proficiency, $F(1, 16) = 0.99, p = .757, \eta_p^2 = .01$.

Table 8

<table>
<thead>
<tr>
<th></th>
<th>Disconfirming</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Higher</td>
<td>87.08</td>
<td>19.55</td>
</tr>
<tr>
<td>Lower</td>
<td>81.67</td>
<td>18.34</td>
</tr>
</tbody>
</table>

As shown in Table 8, performance on comprehension questions was poorer for the disconfirming than the control passages. Similar results were obtained by Nahatame (2012), who showed that disconfirmation of drawn predictive inferences had some negative impacts on text information recall. Given the results of eye movement data analysis, the poor comprehension of disconfirming passages can be attributed to the additional demand of processing and integrating target sentences. Consequently, participants might have few cognitive resources for maintaining the previously processed text information until the time when they answered comprehension questions. Although this result does not directly relate to the present RQs, it provides an important suggestion for teaching and instruction of EFL reading (see Limitations and Implications).

4. Conclusion

This study investigated Japanese EFL readers’ text comprehension when they make predictive inferences and the inferences are disconfirmed by the subsequent context. The results of sentence recognition performance supported the proposed hypothesis: Consistent with Nahatame (2014), the participants in the present study had some difficulty eliminating memory trace of disconfirmed predictive inferences after reading.

Regarding the RQs, the analysis of eye movement data revealed that the disconfirmation of drawn predictive inferences (i.e., inconsistencies between the inferences and the subsequent context) affected some aspects of EFL learners’ real-time text comprehension processes. First, as indicated by longer first-pass reading times, drawing on predictive inferences that are inconsistent with the following sentences caused extra demand of initial processing of the sentences, regardless of L2 proficiency. This suggests that both higher and lower proficiency learners immediately noticed inconsistencies between drawn inferences and the sentence being processed. Second, as
revealed by longer sentence wrap-up times, such inconsistencies also caused difficulty with integrating the sentence disconfirming the inferences into developing text representations, especially among lower proficiency learners. Taken together, the present study suggests that difficulty of revising predictive inferences in memory does not result from insufficient awareness of inconsistencies between drawn inferences and the disconfirming context during reading. Instead, it is more likely to be related to difficult integration of the disconfirming context into developing text representations, which is more demanding for less proficient readers.

In contrast, inconsistencies between inferences and the text did not affect eye fixations for reanalyzing the sentences disconfirming the inferences: There was no significant difference in second-pass reading times between passage types. As noted above, in this study, participants were highly unlikely to reread the target sentence. On the one hand, this supports the validity of the results of Nahatame’s (2014) sentence-by-sentence experiment because participants generally processed the text sequentially, even when the text was presented in its entirety. However, on the other hand, this relates to one of the limitations of the present study as noted below.

5. Limitations and Implications

One important limitation of this study is that the passages included only a single sentence that disconfirmed the drawn inferences. The present study needed to adopt the same materials used in Nahatame (2014) to confirm and extend the findings of that study. However, because there were no successive sentences that followed the target sentence, second-pass reading times reported in the present study are confined to forward fixations from prior sentences. Consequently, regressive or backward fixations from successive sentences to target sentences were not examined. Therefore, future research should examine how frequently and for how long learners regress to the disconfirming sentence from the subsequent context. Similarly, it is necessary to examine how the amount of disconfirming context affects learners’ processing and memory of inferences disconfirmed by the context.

Another limitation of this study is the small number of participants. Although some results showed distinctive differences between higher and lower proficiency learners, higher proficiency learners demonstrated dissociation between the results of eye movement analysis (i.e., wrap-up times) and sentence recognition. Thus, the effect of L2 proficiency on processing of the disconfirming context and memory of inferences disconfirmed by the context also requires further investigation involving a larger number of participants with more variations in L2 proficiency.

Finally, some educational implications are suggested from the present findings. First, the present study suggests that when learners make predictions that are immediately disconfirmed by the following context, they tend to construct less accurate text comprehension. Specifically, learners will have difficulty eliminating the disconfirmed predictions from text memory. Furthermore, as indicated by the results of comprehension questions, the comprehension of
explicit text information may be impaired by the disconfirmation of the predictions. Therefore, teachers should recognize these risks of students making incorrect predictions in reading.

Second, using eye-tracking allowed for the identification of the difficulty less proficient learners may have in on-line reading. Specifically, they have difficulty with the integration of the context disconfirming predictions into developing text representations, which may be more important for achieving comprehension revision. Thus, it is important to focus on and develop the ability to flexibly integrate unexpected text information with current comprehension during text processing. Although more evidence is needed to provide implications for specific instruction, the present study emphasizes such flexible text comprehension processes as one aspect that characterizes proficient and effective reading.

Note

1. The eye movement data analysis yielded consistent results when fixation durations divided by the number of syllables or characters in the target sentence were analyzed. In addition, similar results were obtained, regardless of whether the analysis included or excluded the trials where participants made errors on comprehension questions.

Acknowledgements

I would like to thank Prof. Yuji Ushiro for his enormous support and insightful comments on this paper. This research was supported by a Grant-in-Aid for JSPS Fellows (No. 24 • 1337).

References


