Depth of Processing in Japanese EFL Learners’ Reading Comprehension

Akari KAI
Graduate School, University of Tsukuba / The Japan Society for the Promotion of Science

Abstract

Semantic analysis for the establishment of mental representation is not necessarily complete and people have enormous difficulty in detecting the anomaly in sentences such as, “They were trying to decide where to bury the surviving dead.” Two experiments were carried out to examine depth of reading processing among Japanese EFL readers. In Experiment 1, the language effects (L1 vs. L2) on anomaly detection were investigated. In Experiment 2, whether or not anomaly detection occurs online was examined. The results of the two experiments showed that many readers had a tendency to process text in a shallow or incomplete manner, frequently failing to detect anomalies. This tendency was more remarkable in the L2 condition. Furthermore, participants’ reading time revealed that an anomaly can be detected online, although their reaction to the anomaly might be delayed. Failures to detect contradictions are accounted for by assuming that non-detectors believe too strongly in the global text interpretations they create.

1. Introduction

An important goal of studies of reading comprehension is to specify the cognitive processes involved in building a representation of the text. One component of this involves investigating how readers maintain both local and global coherence of a text. Traditional models of reading comprehension (e.g., Just & Carpenter, 1980; Kintsch & van Dijk, 1978) assume that the local meaning of a phrase is computed prior to incorporating it into the more global representation of a text. This is called the local-to-global processing assumption.

However, recent work on discourse processing reveals growing evidence that discourse representations may vary in the detail they encode and may sometimes be underspecified (Barton & Sanford, 1993; Ferreira, Bailey, & Ferraro, 2002; Sturt, Sanford, Stewart, & Dawydiak, 2004). Moreover, some studies have suggested that the reading process may not support the local-to-global assumption (Barton & Sanford, 1993; Hannon & Daneman, 2004). Barton and Sanford (1993), for instance, showed that young adult readers often failed to detect a semantic anomaly embedded in a text. The researchers used an incidental anomaly detection
task by giving readers the following short text that ended with one of four anomalous terms:

There was a tourist flight traveling from Vienna to Barcelona. On the last leg of the journey, it developed engine trouble. Over the Pyrenees, the pilot started to lose control. The plane eventually crashed right on the border. Wreckage was strewn equally in France and Spain. The authorities were trying to decide where to bury the surviving / injured/ surviving injured / surviving dead.

Then they analyzed whether or not their readers detected the anomalous terms. Their results showed high levels of detection failures, and the detection rates differed across the four anomalous terms (59% for survivors, 7% for injured, 65% for surviving injured, and 23% for surviving dead).

Barton and Sanford (1993) used their findings to argue that their very low detection rates for the internally incoherent anomalous noun phrase (hereafter, NP), surviving dead, challenged the commonly held assumption of local-to-global processing. This was because if they processed the text in a local-to-global manner, they should have detected the internally incoherent anomalous NP immediately. However, their results showed that local semantic coherence was not always established prior to incorporation into the global text representation. Furthermore, they argued that the proclivity toward shallow semantic processing is particularly evident when a word or phrase has a good semantic fit with the preceding theme of the passage. In the case of Barton and Sanford’s plane crash passage, because the term survivors fitted well with the global plane crash theme, readers may have failed to include the is alive feature of survivors in their representation, and consequently failed to notice the anomaly between survivors and bury. However since Barton and Sanford did not investigate the comprehenders’ mental representation, the interpretations these readers made after encountering an anomaly remain to be specified.

The detection of inconsistencies in texts has been widely documented by other researchers studying comprehension monitoring. Individual and age differences in the detection of text contradictions (Daneman, Hannon, & Barton, 2006), as well as the effects of anomaly type on detection (Bohan & Sanford, 2008; Stewart, Kidd, & Haigh, 2009) have been found by a number of researchers. Hannon and Daneman (2004), for instance, replicated Barton and Sanford’s finding that young adult readers of all skill levels are susceptible to shallow semantic processing, but they also showed that less-skilled readers are much more susceptible. Furthermore, they found that it was only the less-skilled readers who often failed to compute the local meaning of an NP prior to integrating it into their more global representation of the text.

Although research using the anomaly detection paradigm was conducted in the first language (L1) reading area with the aim of revealing the process of reading comprehension, no such research so far seems to have been conducted in the second language (L2) research area.
Moreover, if L1 readers’ comprehension skill impacts on shallow semantic processing as Hannon and Daneman suggest, what about L2 readers? Since L2 readers are considered to have less reading ability compared to L1 readers, their semantic processing may be much shallower than those of L1 readers. The present study attempted to replicate and extend some of Hannon and Daneman’s findings with respect to how reading comprehension skill in terms of first and second language skills impacts on shallow semantic processing.

In the two experiments to follow, the depth of processing among Japanese EFL readers was investigated. In Experiment 1, by using Barton and Sanford’s anomaly detection paradigm, the effects of language (L1 vs. L2) on anomaly detection during comprehension of a short text were investigated. In Experiment 2, the time course was carefully examined with the aim of revealing the process of anomaly detection. Additionally, across the two experiments, a free written recall task was introduced to reveal the mental representation that comprehenders had built up. It was hoped that these complementary methodologies would help assess the nature of reading processes among Japanese EFL learners, as well as the mechanism of anomaly detection (or failures). Three research questions were set as follows: (1) Do success rates of anomaly detection differ across language conditions (L1 vs. L2)?; (2) Do Japanese EFL learners read a text in a local-to-global fashion?; (3) Do Japanese EFL learners detect anomalies online?

2. Experiment 1

The objective of Experiment 1 was to investigate the effects of language on anomaly detection. Barton and Sanford’s (1993) incidental anomaly detection task was used to investigate differences in the text processing styles of L1 and L2 young adult readers.

2.1 Participants and Materials

A total of 59 Japanese university students with diverse majors participated in this study. Their English proficiency was estimated at between 125 and 345 based on the score for the reading section of the TOEIC Practice Test. Because of the limited number of participants, and with the numerous experimental conditions, the current study did not take into consideration the participants’ proficiency levels as a factor in the following analysis.

The reading material used in this study was from Barton and Sanford (1993). Wording was slightly modified so that there would be no difficult words for the participants. Although the previous study used four anomalous terms (e.g., survivors, injured, surviving injured, and surviving dead), the present study used only two anomalous NPs (i.e., surviving injured and surviving dead) in two language conditions. This was done to shed more light on the examination of the local-to-global processing assumption. It was expected that if readers failed to detect the locally incoherent anomalous NP (surviving dead), the findings would not support the local-to-global processing assumption. For the Japanese version, the original text was
carefully translated into Japanese by the investigator and then checked later by another person. The anomalous terms for the Japanese version (i.e., seizonshya no itai for surviving dead, and seizonshya no kanjiya for surviving injured) were decided based on the results of a pilot study.

2.2 Procedure

The experiment took place in regular English classes. Participants read Barton and Snaford’s (1993) plane crash passage with one of the four anomalous endings (the locally incoherent NPs: surviving dead, seizonshya no itai, and the locally coherent NPs; surviving injured, seizonshya no kanjiya). Assignment to passage version was random with the constraint that there should be equal numbers of participants in each of the four conditions. The experimental design was a between-subjects design. Half of the participants read the Japanese (L1) version and half read the English (L2) version. Then in each condition, half read the locally incoherent anomalous NP, and half read the locally coherent anomalous NP. For the current study, a booklet containing the experimental material, a comprehension question, a free written recall test, and a questionnaire were made and distributed to each participant. Participants were asked to write all their responses in their native language, Japanese.

As in the previous studies (Barton & Sanford, 1993; Hannon & Daneman, 2004), after reading and responding to the experimental passage, the participants underwent a structured debriefing questionnaire about the passage. This structured debriefing was included to identify participants who had detected the anomaly but had made cooperative written responses because they assumed that it had been an unintentional error in the text. Participants were first asked whether they had noticed anything odd about the passage. After that, participants were shown the anomalous term that had appeared in their version of the passage and were asked whether they had noticed it and whether this NP made the passage odd in any way. Each task was printed on one side of the paper and participants were explicitly prohibited to turn back to the previous pages of the material. Furthermore, participants were given no warning about potential anomalies prior to reading the passage.

2.3 Results and Discussions

2.3.1 Recall Production

Table 1 summarizes the results of recall production. The scoring was done by using idea units, and one point was given when the participants correctly recalled an idea unit. The results of a 2 (NP: locally coherent, locally incoherent) x 2 (Language: Japanese, English) ANOVA revealed no interaction between NP and language conditions, $F (1, 55) = .24, p = .626$, no significant main effect of NP, $F (1, 55) = 1.23, p = .273$ and no main effect of language, $F (1, 55) = 3.11, p = .083$, on the amount of recall production. These results suggest that the recall performance was not significantly different across the four experimental conditions and that all of the experimental conditions were around the same difficulty level.
Table 1 Descriptive Statistics of Recall Production

<table>
<thead>
<tr>
<th></th>
<th>Locally Coherent NP</th>
<th>Locally Incoherent NP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>Japanese (L1)</td>
<td>15</td>
<td>9.15</td>
</tr>
<tr>
<td>English (L2)</td>
<td>15</td>
<td>7.53</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>8.33</td>
</tr>
</tbody>
</table>

_Note._ Full score = 13

2.3.2 Anomaly Detection

The participants were classified as detectors or non-detectors based on their written responses to the question posed at the end of the passage (e.g., "What should the governments do?"). as well as their answers in the questionnaire. As in previous studies (Barton & Sanford, 1993; Hannon & Daneman, 2004), participants who explicitly mentioned the anomaly in their written responses and in the questionnaire were classified as detectors; participants who failed to mention the anomaly in their written responses but pointed out the anomaly when asked in the questionnaire were also classified as detectors; participants who explicitly failed to mention the anomaly in both their written responses and when asked about the oddness of the passage were classified as non-detectors.

Table 2 shows the percentage of anomalies detected for the four experimental conditions. The top two columns show the results of the Japanese version, and the bottom two columns show those of the English version.

Table 2 Percentage of Anomalies Detected as a Function of Anomaly Type in Experiment 1

<table>
<thead>
<tr>
<th>Language</th>
<th>Anomalous Term</th>
<th>Type of Anomaly</th>
<th>N of Detectors (%)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>seizonshya no kanjya</td>
<td>coherent NP</td>
<td>67</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>seizonshya no itai</td>
<td>incoherent NP</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>English</td>
<td>surviving injured</td>
<td>coherent NP</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>surviving dead</td>
<td>incoherent NP</td>
<td>0</td>
<td>15</td>
</tr>
</tbody>
</table>

Consistent with Barton and Sanford’s (1993) findings, the results of the current study showed high levels of anomaly detection failure, indicative of shallow semantic processing. Moreover, the pattern of the results showed different detection rates across the experimental conditions. As for the NP condition, locally incoherent NPs produced lower detection rates than locally coherent NPs. This tendency became much stronger in the L2 condition. Whereas only 20% of the L2 readers detected the anomaly in the locally coherent NP condition, no one in the locally incoherent NP condition detected the anomaly.

In order to confirm the findings described above, Fisher’s exact test was conducted. Table 3 shows the cross table. Since the number of items in each category differed greatly, Fisher’s exact test, which is very robust, was considered to be more suitable than Chi-square tests.
Table 3 *The Cross Table for Anomaly Detection*

<table>
<thead>
<tr>
<th></th>
<th>NP Locally Coherent</th>
<th>Non-detector</th>
<th>Non-detector</th>
<th>Japanese (L1)</th>
<th>English (L2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>AR</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Detector</td>
<td>12</td>
<td>22.0</td>
<td>2.2</td>
<td>5</td>
<td>8.5</td>
</tr>
<tr>
<td>Non-detector</td>
<td>17</td>
<td>28.8</td>
<td>-2.2</td>
<td>24</td>
<td>40.7</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>50.8</td>
<td></td>
<td>29</td>
<td>49.2</td>
</tr>
</tbody>
</table>

*Note.* AR = Adjusted Residuals

Fisher’s Exact test confirmed the description of the results insofar as there was a significant main effect of NP types, Adjusted Residual = 2.2, *p* = .047, which was attributable to the fact that the locally coherent NP was detected more frequently than the locally incoherent NP. Furthermore, it was revealed that those who read L1 text detected anomalous terms more frequently than those who read L2 text, Adjusted Residual = 3.5, *p* = .001. The low detection rates for the locally incoherent NP suggests that the local semantics of the phrase (e.g., *surviving dead*) are not computed prior to their incorporation into the more global representation of the text. In sum, these findings do not support the local-to-global processing assumption. Additionally, the results of the Japanese version, in terms of the NP type effects, showed a similar pattern to the previous studies (e.g., Barton & Sanford, 1993), although the detection rates for the English version were much lower than the previous studies.

### 2.3.3 Mental Representation

The results of the current study showed high levels of anomaly detection failure as described above. Then, what kind of interpretation had those readers made? In order to investigate the comprehenders’ built up mental representation after anomaly detection failure, a qualitative analysis was performed on non-detectors’ recall production of the target sentence (*they were trying to decide where to bury the surviving dead/injured*). A particular concern in this analysis was the possibility that participants might interpret anomalous NPs (e.g., *surviving dead, surviving injured*) as meaning *intact dead bodies*, or something similar.

The results revealed different tendencies across language conditions. As for the Japanese version, nearly 80% of the non-detectors interpreted *seizosnya no itai* (surviving dead) or *seizosnya no kanziya* (surviving injured) as meaning *dead bodies*. Thus, these readers failed to include the meaning of *seizosnya* (survivors) in their recall, and its core meaning (i.e., the *is alive* feature) was not recovered in their interpretation.

In the English version, however, fewer people showed this tendency, and only 30% of the non-detectors discarded the core meaning of *surviving*. Moreover, an additional 15% of the non-detectors made another interpretation. Indeed these readers did include the core meaning of *surviving* in their recall, but misinterpreted the remaining part of the target sentence. Specifically, they interpreted “bury the surviving injured/dead” part as “missing survivors are
buried (under something).” The remaining readers in the English condition either did not recall the target part (44%) or gave unclassifiable responses (11%).

In sum, the qualitative analysis of written recall indicated that non-detectors believed too strongly in the global text interpretations they created and they tended to make up alternative interpretations for the target anomalous terms which better suited their interpretation. The interpretations those readers made differed across the language conditions and showed that readers frequently fail to retrieve the meaning of a word in a text, or they do retrieve the meaning but change their interpretations of other parts.

3. Experiment 2

The objective of Experiment 2 was to reveal the time course of anomaly detection.

3.1 Participants, Apparatus, and Materials

A total of 48 subjects participated in this study. The experiment used a computer running SuperLab software to record participants’ responses. The same material as Experiment 1 was used with the following modifications. The experimental material contained one of three possible target NPs: (a) an internally coherent anomalous NP (i.e., surviving injured); (b) an internally incoherent anomalous NP (i.e., surviving dead); and (c) a non-anomalous control NP (i.e., unfortunate dead), and participants read only the English version. Furthermore, after the target NP, the following words (i.e., from the plane crash) were added so that the target NP did not occur in the sentence-final position. This was done to avoid sentence wrap-up effects (Gurzman & Klin, 2000).

3.2 Procedure

The passage was presented one word at a time on a computer screen and readers controlled the rate of presentation of each word by pressing a button. As each word appeared, the previous word disappeared. At the end of the text, the question “What should the governments do?” appeared on the screen, preceded by the prompt **QUESTION**. Participants were asked to write their solution to the problem on the piece of paper provided. Participants were instructed to read for comprehension and were given no warning about potential anomalies prior to reading the passage. For the classification of detectors and non-detectors, the same method as Experiment 1 was used.

3.3 Results and Discussions

3.3.1 Recall Production

Table 4 summarizes the results of recall production. The results of a one-way ANOVA revealed no significant main effect of NP on the amount of recall production, $F(2, 45) = 2.96, p$
= .062. These results imply that the recall performance was not significantly different across the experimental conditions.

Table 4 Descriptive Statistics of Recall

<table>
<thead>
<tr>
<th>NP type</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coherent NP</td>
<td>16</td>
<td>8.25</td>
<td>3.40</td>
</tr>
<tr>
<td>Incoherent NP</td>
<td>16</td>
<td>10.56</td>
<td>2.16</td>
</tr>
<tr>
<td>Control</td>
<td>16</td>
<td>9.19</td>
<td>2.40</td>
</tr>
</tbody>
</table>

Note. Full Score = 13

3.3.2 Anomaly Detection

Table 5 shows the percentage of anomalies detected for the two anomalous terms for 32 readers who read either internally in/coherent anomalous NPs. Consistent with the findings of previous studies as well as those of Experiment 1, the results of the current study showed high levels of anomaly detection failure. Furthermore, as in Experiment 1, anomaly detection rates differed across NP conditions. Fisher’s Exact tests confirmed that there was a significant main effect of NP types, Adjusted Residual = 2.4, p = .043, which was attributable to the fact that surviving injured was detected more frequently than surviving dead.

Table 5 Percentage of Anomalies Detected as a Function of NP Type

<table>
<thead>
<tr>
<th>Anomalous Term</th>
<th>Type of Anomaly</th>
<th>Number of Detectors (%)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>surviving injured</td>
<td>coherent NP</td>
<td>31</td>
<td>16</td>
</tr>
<tr>
<td>surviving dead</td>
<td>incoherent NP</td>
<td>0</td>
<td>16</td>
</tr>
</tbody>
</table>

3.3.3 Online Detection of Anomalous Terms

In order to investigate whether participants in the current study detected anomalies online, reading times in the critical region (i.e., bury the surviving injured from the plane crash) were examined. Figure 1 shows the mean response times of 16 readers (detectors vs. non-detectors) in the internally coherent NP condition. Since the participants in the present study detected anomalies only in the internally coherent NP condition, the analysis of the reaction time was restricted to this condition.

As can be seen in Figure 1, the detectors seem to have taken a longer time than non-detectors to read the critical region. In particular, there seems to be a large time difference in injured, from, and crash parts. For the statistical analysis, t tests were performed to reveal the reading time differences between detectors and non-detectors.

Contradictory to the prediction, however, the statistical analysis showed that there was no reading time difference in the anomalous NP itself, t (14) = .013, p = .990 for surviving, and t (14) = 1.11, p = .285 for injured, respectively. In addition, the sentence-final position, crash, did
not show a significant time difference, 
\( t(14) = 1.37, p = .193. \)

However, one part of the critical region showed a significant time difference between detectors and non-detectors; this was the position of "from", immediately after the anomalous NP, \( t(14) = 3.21, p = .006. \) It was revealed that if they noticed the refutation, participants took 929 ms longer to read the word which was located immediately after the anomalous NP. These results indicate that comprehenders can detect anomalies online, although their response might be delayed.

### 3.3.4 Mental Representation

As in Experiment 1, non-detectors’ built-up representations in terms of their written recall production were examined. While 30% of the readers excluded the meaning of surviving from their recall, an additional 15% of the readers did include its core meaning (is alive feature) in their recall, but they misinterpreted other parts (e.g., missing survivors are buried under something). The remaining readers gave unclassifiable responses (16%) or did not recall the part (40%). These results indicate that if readers hold on too strongly to their initially formed macroproposition (in this case, the plane crash theme), they will make up their own interpretation and consequently they will misunderstand what the text actually says.

### 4. Conclusion

The new contribution of the current study was to show how Japanese EFL learners used shallow semantic processing. The results of the current study revealed both quantitative and qualitative differences in the text processing styles of L1 versus L2 readers. Although readers of both languages were susceptible to anomaly detection failures, L2 readers were significantly more susceptible and frequently failed to spontaneously detect an anomalous term. This finding indicates that L2 readers have a tendency to engage in partial semantic processing, producing mental representations of the text that are underspecified and incomplete. Furthermore, the qualitative analysis of comprehenders’ mental representation confirmed this finding. The analysis revealed that non-detectors misinterpreted not only the anomalous part itself but also other parts of the target sentence. Their interpretations were more suited to their initially formed macropropositions rather than the accurate meaning of the original text.

Another contribution of the current study was to shed light on the time course of anomaly
detection of L2 readers. It was revealed that although readers can detect anomalies online, their response may be delayed. These results indicate that detectors did not pause longer on the anomalous NP itself; however they did spend longer looking at the word which immediately followed the anomalous NP. The results of the current study provided strong support for the view that L2 readers are not scrupulous about processing and do not integrate every word into their mental representation of the text. Considering this issue, educators may need to look at the balance in comprehension processes and encourage readers to pay attention to the text more evenly, not just to a few selected pieces as nondetectors did.

Acknowledgment

I would like to thank Professor Yuji Ushiro for his comments and suggestions for this research.

References


