How L2 Working Memory Capacity for Japanese EFL Learners Are Related with Processing of Garden Path Sentences

Hiroshi NAKANISHI
Graduate School, Kwansei Gakuin University

Abstract

This study explores whether Japanese EFL learners, too, find the processing of Garden Path (GP) sentences difficult and whether the capacity of L2 Working Memory (WM) limits the L2 learners' ability to deploy semantic and syntactic cues in parsing sentences. The result showed that Japanese EFL learners find GP sentences difficult to understand and that the former use semantic information as clues to avoid the difficulty in processing GP sentences, regardless of WM capacity, which is inconsistent with the L1 study. These results may stem from the e-f score, which reflects the speed-accuracy trade-off in the high-and low-span groups. Therefore, we propose using the efficiency score (p-e score), which takes processing speed into account.

1. Introduction

A number of studies have demonstrated that WM, which is assumed to be a limited capacity system responsible for simultaneously processing and storing information in real time, plays an important role in reading comprehension. L1 studies that compared the comprehension behavior of high and low WM capacity participants have yielded useful data that help in specifying the manner in which WM constrains specific language processes (Miyake & Friedman, 1998). The language processes that have been shown to depend on WM capacity include the resolution of linguistic ambiguity (Miyake et al., 1994), the parsing of syntactically complex structures such as object-clause sentences (King & Just, 1991), and the processing of GP sentences (Macdonald et al., 1992; Just & Carpenter, 1992).

There is no doubt that L2 WM plays a critical role in the language comprehension of Japanese EFL learners, too. By using GP sentences, the present experiment explores the
relationship between L2 WM and the comprehension of syntactically ambiguous sentences. GP sentences have locally ambiguous syntactic structures and include a reanalysis of the entire sentence. An example of a GP structure is "The woman paid after the end of the month had worried the man." The word "paid" has the ambiguity of being in the past tense and in the past participle form; this ambiguity is not resolved until one reaches the word "had." If "paid" were to be considered as being in the past tense first, then one is required to reinterpret "paid" as being a past principle form.

In the L1 study, Macdonald et al. (1992) explored how readers with different WM capacities deal with GP sentences. In their experiment, they used GP sentences such as (a) and non-GP sentences such as (b), followed by true-false comprehension questions.

(a) The experienced soldiers warned about the dangers conducted the midnight raid
(b) The experienced soldiers who were warned about the dangers conducted the midnight raid.

The result was that the reading times (RTs) for the sentence-final words in GP sentences were longer than those in non-GP sentences. In addition, subjects made more errors in response to comprehension questions about GP sentences than in response to those about non-GP sentences. These results confirmed the difficulty of GP sentences. The researchers also observed that low-span readers made more errors in answering comprehension questions about GP sentences than in responses to those about non-GP sentences, while high-span readers had longer RTs for the words in GP sentences than in non-GP sentences. They argued that high-span readers can maintain multiple syntactic structures (i.e., a past tense and a past participle form), but low-span readers cannot have both structures available in WM, leading to more errors among low-span readers and slower processing times among high-span readers.

Just and Carpenter (1992) examined the individual differences of parsing in ambiguous sentences in terms of WM capacity. They used four types of sentences: (1) reduced relative clauses with +animate noun phrase (NP); (2) unreduced relative clauses –animate NP; (3) reduced relative clauses with +animate NP; and, (4) unreduced relative clauses with –animate NP. Examples of these sentences are as follows:

(1) The defendant examined by the lawyer turned out to be unreliable.
(2) The defendant that was examined by the lawyer turned out to be unreliable.
(3) The evidence examined by the lawyer turned out to be unreliable.
(4) The evidence that was examined by the lawyer turned out to be unreliable.

They reported the RTs using eye-movement monitoring techniques. The result is displayed in Figure 1.

Just and Carpenter found that there was a difference between high- and low-span readers in their first-pass fixation times on the "by phrase" of sentences with –animate NP. With regard to the reduced relative clause (types 1 and 3), the sentences with –animate NP
(type 3) decreased the first-pass fixation durations by 75 ms for the high-span readers, but not for the low-span readers. Likewise, with regard to the unreduced relative clauses (types 2 and 4), sentences with +animate NP (type 4) diminished the first-pass fixation durations by 69 ms for the high-span readers, but not for the low-span readers. This pattern produced a significant interaction of animacy and span group, \( F(1, 66) = 5.36 \) (\( p < .025 \)). The researchers contended that the ability to make use of semantic information during parsing depends on individual differences in WM capacity.

![Figure 1 The Result of Experiment by Just and Carpenter (1992)](image)

2. Purpose of the Study

The purposes of the present experiment were as follows:

1. To examine whether the processing of GP sentence is also difficult for Japanese EFL learners.
2. To explore whether the capacity of L2 WM limits L2 learners' ability to deploy semantic and syntactic cues in parsing sentences such as the type 2 and type 3 sentences.

3. Method

3.1 Participants

The participants of this experiment were 60 Japanese university students or graduate school students.

3.2 Procedure

All participants completed two tasks: (1) the Reading Span Test (RST) and (2) the sentence processing task. Each participant completed all the tasks on the computer monitor. The entire experiment took approximately 30 minutes.
3.2.1 The RST

The RST—originally developed by Daneman and Carpenter (1980)—has been used to measure a subject’s verbal WM capacity. In the original L1 test, the subjects are required to read aloud a set of unrelated sentences printed on cards and recall the final words of the sentences. In this process, their finite WM resources are consumed by having to process the sentences and remember the final words. The test measures the efficiency of both the processing and retaining of information. In the L1 case, subjects are supposed to read the sentences aloud in a rather automatic manner. Therefore, they can concentrate on comprehending the meaning of sentences while reading them aloud.

However, in the case of L2, reading aloud is performed in a less automatic manner. Therefore, the subjects must allocate their WM resources to reading aloud itself as well as to comprehending the sentences. The RST score is generally calculated by the number of final words remembered. Therefore, the subjects are considered to be trying to focus on memorizing the words by reading the sentences mechanically without comprehending the sentences. As a result, the scores measured by the original RST procedure do not reflect the efficiency of both the processing and retaining of information (Kadota, 2007).

In this experiment, we conducted a revised version of the RST for Japanese EFL learners by using a computer. In order to enable the subjects to concentrate on the comprehension of the sentences in the RST, we asked them to judge the truth or falsity of each sentence in the RST in Japanese.

The procedure of the RST in the present experiment was as follows. After the mark “+++++” was presented for one second, a sentence appeared on a computer monitor. The participants were required to press the spacebar immediately after they had silently read the sentence and to remember the final word of the sentence. The reading time, which is the time between the appearance of the sentence and the pressing of the spacebar, was recorded. After pressing the spacebar, the Japanese equivalent of the previous English sentence appeared on the monitor. The participants were required to judge whether the statement was true (B) or not (N) by pressing the corresponding key to record their choice. The reaction time, which is the time between the appearance of the statement and the pressing of the spacebar, was also recorded. Then, the next sentence appeared on the monitor following the “+++++” mark. The participants were asked to read the sentence while remembering the last word of the previous sentence. This procedure was repeated until they arrived at the instructions indicating the end of the session. Then, on an answer sheet, the participants were required to write down the final words of the sentences that had been presented. For example, under the four-sentence condition, after the participants had read the four sentences and answered each question, they encountered the instruction that indicated the end of the session. Then, they were required to write the four final words of the sentences on the answer paper. Sentences were presented to the participants under the conditions of two,
three, four, and five sentences. Each condition comprised three WM sessions. The participants were asked to silently read increasingly longer sets of sentences until they finished reading three sets of five sentences in the five-sentence condition. The sentences presented were eight to thirteen words long. The familiarity of English words among the sets for Japanese EFL learners (Yokokawa, 2006) was statistically the same on average ($F = 1.1445, ns.$).

3.2.2 The sentence processing task

Fifty-two sentences were used, including GP sentences, controls, and fillers. The thirty-two GP sentences and non-GP control sentences were analyzed excluding the filler sentences. The thirty-two sentences were categorized as follows: (1) reduced relative clauses with +NP; (2) unreduced relative clauses –animate NP; (3) reduced relative clauses with animate +NP; and, (4) unreduced relative clauses with –animate NP. Examples of these sentences are as follows:

(1) The woman paid after the end of the month had worried the man.
(2) The woman that was paid after the end of the month had worried the man.
(3) The bill paid after the end of the month had worried the man.
(4) The bill that was paid after the end of the month had worried the man.

The sentences without the “by” portion of the agentive phrase used in Ferreira and Clifton’s (1986) study were used because “by” is a conspicuous cue to easily identifying the first verb (e.g., paid) as being in the past principle form. In addition, we slightly modified the sentences used in Ferreira and Clifton’s study in order that the familiarity of the words become statistically identical among the four groups ($F = .0476, ns.$). The sentences we used were eleven to fifteen words in length.

Sentences were presented word-by-word in a computer-generated random order. When subjects pressed the spacebar to advance the display to the next word in a sentence, the letters of that word would replace the dashes and the letters of the previous word would revert to dashes. Each sentence was followed by a comprehension question written in Japanese. Participants were asked to as quickly as possible press either the “B” key if the statement corresponds with the experimental sentence or the “N” key if otherwise.

4. Results

The span tests were scored in two ways. The first method of scoring (score) was to count the total number of words that were correctly recalled. The maximum score was 42. The second method of scoring (error-free score: e-f score) was to count the number of words correctly recalled when the sentences presented were correctly processed by the participants. The maximum score from this method was also 42. Tables 1 and 2 present the descriptive
statistics of scores, e-f scores, mean reaction times (RTs), and solution times (STs) in the RST.

Table 1 Mean scores and Error-free Scores in the RST

<table>
<thead>
<tr>
<th>Number</th>
<th>Score</th>
<th>e-f score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>30.1</td>
<td>24.9</td>
</tr>
<tr>
<td>S.D.</td>
<td>6.4</td>
<td>6.1</td>
</tr>
</tbody>
</table>

(mark range: score, error-free score: 0–42)

Table 2 Mean RTs and STs per Sentence (msec.) in the RST

<table>
<thead>
<tr>
<th>Number</th>
<th>RTs</th>
<th>STs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>9784.5</td>
<td>3218.8</td>
</tr>
<tr>
<td>S.D.</td>
<td>3393.7</td>
<td>1141.9</td>
</tr>
</tbody>
</table>

Table 3 provides the descriptive statistics of scores, RTs per syllable for correct responses, and STs per syllable for correct responses. The RTs for “that was” in type 2 and type 4 sentences were excluded from the RTs in the sentences. Figures 2, 3, and 4 present the means of scores, RTs per syllable for correct responses, and STs per syllable for correct responses.

Table 3 Descriptive Statistics for all Participants in the Study

<table>
<thead>
<tr>
<th>Type</th>
<th>Scores</th>
<th>RTs per syllable for correct responses</th>
<th>STs per sentence for correct responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>4.8</td>
<td>5.6</td>
<td>7.0</td>
</tr>
<tr>
<td>S.D.</td>
<td>1.5</td>
<td>1.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean</td>
<td>558.0</td>
<td>541.2</td>
<td>531.8</td>
</tr>
<tr>
<td>S.D.</td>
<td>191.9</td>
<td>174.4</td>
<td>142.7</td>
</tr>
<tr>
<td>Mean</td>
<td>2671.7</td>
<td>2629.4</td>
<td>2403.9</td>
</tr>
<tr>
<td>S.D.</td>
<td>1507.9</td>
<td>969.4</td>
<td>781.6</td>
</tr>
<tr>
<td>Mean</td>
<td>964.8</td>
<td>964.8</td>
<td>964.8</td>
</tr>
<tr>
<td>S.D.</td>
<td>1507.9</td>
<td>969.4</td>
<td>781.6</td>
</tr>
</tbody>
</table>

Figure 2 Means of Scores

Figure 3 RTs per Syllable for Correct Responses
A significant main effect of sentence type was found in the mean scores, $F = 30.7173$ ($p < .01$). An analysis of multiple comparisons revealed that type 1 produced a significantly lower score than type 2 ($p < .05$), type 3 ($p < .01$), and type 4 ($p < .01$); further, type 2 produced a lower score than type 3 ($p < .01$), and type 4 ($p < .01$).

Furthermore, a significant main effect of sentence type was found in the RTs for correct responses, $F = 6.9323$ ($p < .01$). Type 4 was significantly faster than type 1 ($p < .01$), type 2 ($p < .01$), and type 3 ($p < .05$).

![Figure 4 STs per Syllable for Correct Responses](image)

There was no significant main effect of sentence type in the STs for correct responses, $F = 1.0377$.

As a result of the statistical analysis, it was revealed that (1) the processing of GP sentences is also difficult for Japanese EFL learners as well as for L1 native speakers, (2) the semantic and syntactic factors may prevent the processing difficulty for Japanese EFL learners, and (3) the semantic factor is more important in reducing the garden path phenomenon than the syntactic factor.

Then, we will explore L2 WM capacity in L2 parsing. Participants were divided into two groups in terms of WM capacity. The capacity of WM was assessed by using e-f scores. In order of ranking, participants with scores in the top half formed the high-span group and those with scores in the lower half were regarded as the low-span group. Table 4 displays the descriptive statistics for e-f high-span groups and low-span groups in the RST.

Table 4 Descriptive Statistics of Span Data for each Reading Span Group Divided by e-f Score

<table>
<thead>
<tr>
<th>Score</th>
<th>e-f score</th>
<th>RT/sentence</th>
<th>ST/sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Mean</td>
<td>35.0</td>
<td>30.0</td>
<td>11033.4</td>
</tr>
<tr>
<td>S.D.</td>
<td>4.3</td>
<td>4.1</td>
<td>3320.5</td>
</tr>
</tbody>
</table>
Table 5, and Figures 5, 6, and 7 below present the mean scores, RTs, and STs of types 1–4 sentences for high-span readers and low-span readers.

With regard to the scores, the crucial interaction between sentence type and span was not significant in our data, $F = 0.9282$ (ns.). There were significant main effects of sentence type, $F = 30.5743$, ($p < .01$), but not of span, $F = 0.1165$ (ns.). The results of multiple comparisons showed that the scores of type 3 were better than those of type 1 ($p < .01$), type 2 ($p < .01$), and type 4 ($p < .01$); further, type 4 produced more correct responses than type 1 ($p < .01$) for both span group readers.

Table 5 Mean Scores, RTs and STs for WM Span Group and Sentence Type

<table>
<thead>
<tr>
<th>Type</th>
<th>Scores</th>
<th>RTs per syllable for correct responses</th>
<th>STs per sentence for correct responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High span</td>
<td>Low span</td>
<td>High span</td>
</tr>
<tr>
<td>Type1</td>
<td>4.6</td>
<td>5.0</td>
<td>585.8</td>
</tr>
<tr>
<td>Type2</td>
<td>5.5</td>
<td>5.7</td>
<td>564.9</td>
</tr>
<tr>
<td>Type3</td>
<td>7.0</td>
<td>6.9</td>
<td>558.0</td>
</tr>
<tr>
<td>Type4</td>
<td>6.4</td>
<td>466.8</td>
<td>404.9</td>
</tr>
</tbody>
</table>

Figure 5 Mean Scores for WM Span Group and Sentence Type

Figure 6 Mean RTs for WM Span Group and Sentence Type
With regard to the RTs, there was no significant interaction between sentence type and span, $F = 0.4851 \text{ (ns.)}$. A significant main effect was found for sentence type, $F = 7.0188$, ($p < .01$) and span, $F = 6.8002$, ($p < .01$). The multiple comparisons revealed that type 4 had significantly shorter RTs than type 1 ($p < .01$) and type 2 ($p < .05$) for both group readers. It was also discovered that the RTs for high-span readers were slower than those for low-span readers in type 4 ($p < .05$).

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figure7.png}
\caption{Mean STs for WM Span Group and Sentence Type}
\end{figure}

With regard to the STs, neither the main effects nor interactions were significant for sentence type and span (sentence type: $F = 1.0381$, \textit{ns}; span: $F = 3.2801$, \textit{ns}.; interaction: $F = 0.2650$, \textit{ns}).

The statistical analysis revealed that (1) the low-span readers can understand the experimental sentences almost as well as the high-span readers can in terms of scores, RTs, and STs; further, (2) both span readers can make use of semantic information as clues to preventing the garden path effect.

5. Conclusions and Further studies

In general, the main findings of the present experiment can be summarized as follows: (1) GP sentences are more difficult for Japanese EFL learners to interpret regardless of their WM capacity. (2) Both high and low reading span readers make use of animate NP as a clue to avoiding the processing difficulty of GP sentences. (3) The low-span readers can process the sentences in a manner on par with the high-span readers. Results (2) and (3) of the present experiment was not consistent with that of Just and Carpenter (1992). Moreover, the result was contrary to our hypothesis. However, the result may have stemmed from RTs in the RST. Table 4 indicates that the high-span participants with longer RTs tended to achieve higher scores in the RST. In other words, the
high reading span group could obtain better scores due to slower processing speed. The results of the present study suggest that the RST score for L2 learners depends on the kind of strategy they adopt in processing the sentences in the RST.

This research suggests that the conventional method of scoring appears to be a slightly unrefined manner of measuring L2 WM capacity. A new index of the RST that takes processing speed into account is required. As a new and possible candidate, Nakanishi (2005) proposed the processing efficiency score (p-e score), which reflects the recalling accuracy of the final word, the processing accuracy, and the processing speed. We will analyze the results of the present experiment by using the p-e score.

Acknowledgement

I would like to thank Prof. Shuhei Kadota of Kwansei Gakuin University, for his thoughtful comments for this research.

6. References


Acknowledgement

I would like to thank Prof. Shuhei Kadota of Kwansei Gakuin University, for his thoughtful comments for this research.

6. References


Acknowledgement

I would like to thank Prof. Shuhei Kadota of Kwansei Gakuin University, for his thoughtful comments for this research.