How L2 Working Memory Capacity in Japanese EFL Learners Is Related to the Processing of Filler-Gap Sentences

Hiroshi NAKANISHI
Kobe University (part-time lecturer)

Abstract

By using filler-gap sentences, the present study aimed to explore how sentence-internal structural factors (i.e., syntactic features of the filler and the length of the filler-gap distance) induce L2 working memory (WM) cost. We used six types of experimental sentences: sentences with three types of syntactic features (i.e., subject and object WH-questions and whether questions) and two versions of the filler-gap distance (i.e., long and short versions). The results showed that Japanese EFL learners as well as L1 readers find object-relative sentences more difficult to process than they do other syntactic types of sentences. On the other hand, the short distance between the filler and its gap does not improve language processing performance, which is not in line with the results of the L1 studies. It was also found that the participants with a large WM capacity performed better than those with a small WM capacity. This suggests that for Japanese EFL learners, in the processing of filler-gap sentences, syntactic information maintained in WM is a more important factor than the distance over which syntactic information is held in WM.

1. Introduction

Working memory (WM) is assumed to be a limited capacity system that is responsible for the simultaneous storage and processing of information during language comprehension (Just & Carpenter, 1992). L1 studies that compare the comprehension behavior of participants with large and small WM capacities have shown how WM constrains specific language processes, including the parsing of syntactically complex structures such as object-clause sentences (King & Just, 1991), garden path (GP) sentences (Just & Carpenter, 1992), and filler-gap sentences (Fiebach et al., 2002).

There is no doubt that L2 WM also plays a crucial role in language comprehension for Japanese EFL learners. Nakanishi (2007) showed that participants with a large WM processed GP sentences better than those with a small WM capacity. The data supports the involvement of
L2 WM resources during the parsing of L2 sentences, which is consistent with the results of L1 studies. However, how sentence-internal structure induces L2 WM cost is still uncertain. To answer this question, the present study explores the relationship between L2 WM and the comprehension of filler-gap sentences.

Prior to discussing the experiment, let us consider one of the filler-gap dependencies. In an English declarative sentence such as “The little boy read the book in school,” the noun phrase (NP) the book appears in the direct object position, following the verb read. On the other hand, there is also the WH-question, for example, “Which book did the little boy read in school?” wherein the corresponding NP which book appears in the sentence-initial position, and no NP appears in the postverbal object position of read. The WH-element which book is called the filler, and the postverbal object position is considered the gap. The following two processes are necessary to comprehend such types of sentences: (1) the filler must be maintained in WM until the gap is encountered and then, (2) the filler must be integrated at the gap position. These two processes are considered to incur WM cost (Yasunaga & Sakamoto, 2005). The former is called syntactic memory cost and the latter is called syntactic integration cost (Fiebach et al., 2002).

In their L1 study, King and Just (1991) showed that the processing of syntactic information (syntactic integration) affects WM resources. The participants were divided into small and large WM groups and read subject-relative and object-relative sentences (e.g., The reporter that attacked the senator admitted the error/ The reporter that the senator attacked admitted the error). The results indicated that reading the verbs of an object-relative sentence took more time for readers with a small WM capacity for language, $F = 10.34, p < .01$, and their resulting comprehension is less accurate, $F = 4.99, p < .05$.

Furthermore, Fiebach et al. (2002) explored how syntactic memory costs and syntactic integration costs affect WM resources. In this study, German participants processed German subject and object WH-questions and whether-questions with either a short or a long distance between the filler and its gap. The distance between the WH-filler / whether and the second NP was varied by inserting either one (short sentence condition) or two (long sentence condition) prepositional phrases. The behavioral data indicated that subjects took more time to answer the long questions than they took to answer the short questions, $F = 13.31, p = .001$ and that the number of correct responses was greater for the whether-questions than for the WH-questions, $F = 14.28, p < .0001$. In addition, participants with a small WM capacity tended to make more errors than those with a large WM capacity, $F=3.65, p = .07$. The event-related brain potentials (ERPs) data showed a sustained left anterior negativity in the region of the filler-gap dependency for the long object WH-questions compared to the subject WH-questions and a late positivity at the second NP position for the short and long object WH-questions. Furthermore, the sustained negativity was stronger and more broadly distributed for individuals in low WM capacity. A sustained left anterior negativity is supposed to reflect the WM processes required for maintaining the filler in memory, whereas a positive-going ERP effect is assumed to indicate the WM
processes required for integrating the stored filler element into the gap.

The combined results of L1 studies revealed that both the length of the filler-gap distance and syntactic integration impose a burden on WM resources. It was also found that the participants with a small WM capacity are affected by the demands of WM costs; consequently, their language processing performance is worse than that of their counterparts with a large WM capacity.

2. Purpose of the Present Study

The purposes of the present experiment were as follows:
(1) To examine whether or not syntactic information (i.e., subject and object WH-questions and whether-questions) induces WM costs
(2) To explore whether or not the length of the filler-gap distance (i.e., short and long sentence conditions) elicits WM costs
(3) To investigate if participants with a large WM capacity performed better than those with a small WM capacity

3. Method

3.1 Participants

The participants for this experiment were 62 Japanese university students or graduate school students.

3.2 Procedure

The subjects completed two tasks: (1) the Reading Span Test (RST) and (2) a sentence processing task. All the tasks were completed on the computer monitor. The entire experiment took approximately 40 minutes.

3.2.1 The RST

The RST, originally developed by Daneman and Carpenter (1980), is a commonly used assessment of a subject's verbal WM capacity. In the original version of L1 RST, the subjects were asked to read aloud a set of unrelated sentences presented on cards and then recall the final words of the sentences. In this process, the subjects' finite WM resources are consumed in processing the sentences and in remembering the final words. This test measures the efficiency in both the processing and storage of information. However, we should be careful in adapting this procedure to the L2 RST for Japanese EFL learners (Nakanishi, 2007). In the case of the L1 test, the participants are supposed to read the sentences aloud in a rather automatic manner, which allows them to concentrate on sentence comprehension. However, in the case of the L2 test, reading aloud is performed in a less automatic manner. Therefore, the subjects must allocate
their WM resources to reading aloud itself and to comprehending the sentences. Furthermore, the subjects are trying to memorize the final words of the sentences, as the total number of final words remembered is calculated as the RST score. As a result, the scores measured by the original RST procedure do not reflect the efficiency of both the processing and retaining of information for Japanese EFL learners (Kadota, 2007).

For these reasons, in this experiment, we adopted a revised version of the RST. This revised version was different from the original version with regard to the following aspects: (1) in the revised version, the RST sentences are presented on a computer monitor; (2) following each RST sentence, the subjects have to perform an insert verification task wherein they have to determine whether the sentence is true in Japanese; and (3) the processing efficiency score (p-e score), which reflects the total number of final words to be remembered, processing accuracy, and processing speed, is considered. The procedure of administering the RST was the same as that in Nakanishi (2007).

3.2.2 The sentence processing task

Three types of indirect questions (i.e., subject and object WH-questions and whether-questions) with two types of filler-gap distance conditions (i.e., short and long distances) as experiment sentences (Table 1) and fillers were used in the present experiment.

Table 1 Sample sentences from the sentence processing task

<table>
<thead>
<tr>
<th>(1) Short subject WH-question</th>
</tr>
</thead>
<tbody>
<tr>
<td>The secretary checked who on Sunday called the client.</td>
</tr>
<tr>
<td>(2) Short object WH-question</td>
</tr>
<tr>
<td>The secretary checked who on Sunday the client called.</td>
</tr>
<tr>
<td>(3) Short whether-question</td>
</tr>
<tr>
<td>The secretary checked whether on Sunday the client called.</td>
</tr>
<tr>
<td>(4) Long subject WH-question</td>
</tr>
<tr>
<td>The secretary checked who on Sunday evening after the work called the client.</td>
</tr>
<tr>
<td>(5) Long object WH-question</td>
</tr>
<tr>
<td>The secretary checked who on Sunday evening after the work the client called.</td>
</tr>
<tr>
<td>(6) Long whether-question</td>
</tr>
<tr>
<td>The secretary checked whether on Sunday evening after the work the client called.</td>
</tr>
</tbody>
</table>

A total of 68 sentences, including 8 sentences for each sentence type and 20 sentences for fillers were used. Part of the translated versions of the experimental sentences used in Fiebach et al. (2002) was used in this experiment. The words used in the experimental sentences were selected so that the familiarity of the English words for Japanese EFL learners (Yokokawa, 2006)
among the sets was, on average, statistically the same, $F = .347, ns$.

Sentences were presented word-by-word in a computer-generated random order. When subjects pressed the space button 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 press the “B” key if the statement corresponded with the experimental sentence and the “N” key if it did not, as soon as possible.

4. Results and Discussion

Table 2 provides the descriptive statistics of the scores, reaction times (RTs; msec) per syllable for correct responses, and solution times (STs; msec) per syllable for the correct responses in the short- and long-distance conditions. The maximum score was 24. The RTs for the short-distance sentences were significantly longer than those for the long-distance sentences, $t = 3.110, p < .01$; however, the scores and STs between the short-distance and long-distance conditions were not significantly different, $t = 1.0496, ns$ and $t = -1.229, ns$, respectively.

Table 3 presents the descriptive statistics for the scores, RTs, and STs in the subject and object WH-question and whether-question conditions. The maximum score was 16. There was a significant main effect of syntactic features in the score data, $F = 45.0, p < .01$ and in the ST data, $F = 3.8, p < .05$. Analysis of multiple comparisons revealed that the scores in the WH-subject condition were higher than those in the WH-object ($p < .01$) and whether-question conditions ($p < .05$) and that the RTs in the WH-object condition were longer than those in the whether-question condition ($p < .05$).

<table>
<thead>
<tr>
<th>Table 2 Scores, RTs, and STs in the short- and long-distance conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short-distance</strong></td>
</tr>
<tr>
<td>scores</td>
</tr>
<tr>
<td>average</td>
</tr>
<tr>
<td>S.D.</td>
</tr>
</tbody>
</table>

(mark range: score 0–24)

<table>
<thead>
<tr>
<th>Table 3 Scores, RTs, and STs in the subject and object WH- and whether-question conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WH-subject</strong></td>
</tr>
<tr>
<td>scores</td>
</tr>
<tr>
<td>average</td>
</tr>
<tr>
<td>S.D.</td>
</tr>
</tbody>
</table>

(mark range: score 0–16)

Figure 1 provides the scores for individuals with high and low reading spans in the long-
and short-distance conditions. There was a significant main effect of WM, $F = 11.380, p < .01$ in the score data. The results of multiple comparisons showed that the scores of high span readers were significantly higher than those of low span readers in both the short- and long-distance conditions, $p < .01$ and $p < .05$, respectively. Figure 2 shows the scores for individuals with high and low reading spans in the WH-subject, WH-object, and whether-question conditions. There were significant main effects of WM, $F = 14.219, p < .01$ and of syntactic features, $F = 46.641, p < .01$. The results of multiple comparisons revealed that the scores in the WH-object condition were significantly lower than those in the WH-subject and whether-question conditions for the high and low reading span readers ($p < .01$) and that the scores in the WH-subject condition were higher than those in the whether-question condition for low span readers ($p = .01$). In addition, the high span readers obtained higher scores than the low span readers ($p < .01$).

![Figure 1 WM × distance (scores)](image1)

Figure 1 WM × distance (scores)

![Figure 2 WM × filler (scores)](image2)

Figure 2 WM × filler (scores)

Figure 3 shows the RTs for individuals with high and low reading spans in the short- and long-distance conditions. There were significant main effects of WM, $F = 49.424, p < .01$ and of distance, $F = 24.396, p < .01$ in the RT data. Multiple comparisons revealed that the RTs for the high span readers were significantly shorter than those of the low span readers in the short- and long-distance conditions ($p < .01$). It was also found that both the high and low span readers took significantly shorter in the long-distance condition than in the short-distance condition, $p < .05$ and $p < .01$, respectively. Figure 4 presents the RTs for the high and low span readers for the WH-subject, WH-object, and whether-question sentences. There was a significant main effect of WM, $F = 46.240, p < .01$. The results of multiple comparisons revealed that the RTs of the high span readers were shorter than those of the low span readers, regardless of the syntactic condition.

Figure 5 represents the STs for the participants with high and low reading spans in the filler-gap distance condition. There were no significant main effects of distance and syntax (distance: $F = 2.645, ns$; syntax: $F = 1.349, ns$). The STs for the high and low span participants in the three types of filler conditions are shown in Figure 6. It was found that there was a significant main effect of syntactic type, $F = 3.869, p < .05$. Analysis of multiple comparison revealed that participants with high reading spans answered the WH-object questions faster than
those with a low reading span, \( p < .05 \).

Figures 2, 4, and 6 show that the scores and STs reflect processing difficulty according to syntax (i.e., WH-subject, WH-object, and whether), but the RTs do not. The RTs are supposed to reflect on-line processing. In this stage, the subjects may combine word with word incrementally in their minds. On the other hand, the scores and STs are assumed to reflect off-line processing. In this stage, they may understand the entire sentence. RTs, which reflect on-line processing, still do not reveal the differences in syntax.

Subsequent analysis focused on the RTs for certain regions for all participants. The RTs indicate the times per syllable for correct responses. The analyses were conducted by dividing each sentence into six regions as indicated by the “/” character in the sample sentence:

The secretary / checked / who / on Sunday / called / the client.

The first region (e.g., the secretary) contains \{{[the]} + {[initial noun]}\}, the second region (e.g., checked) contains \{{[initial verb]}\}, the third region (e.g., who) contains \{{[WH-filler]}\}, the fourth region (e.g., on Sunday) contains \{{[1 vs 2 prepositional phrases]}\}, the fifth region (e.g., called) contains \{{[second verb]}\} in the WH-subject questions or \{{[second noun]}\} in the WH-object questions and whether-questions, and the sixth region (e.g., the client) contains the remaining parts of the sentences. Figure 7 shows the mean RTs per syllable by sentence type and region for all
participants.

The RTs revealed a reliable interaction of sentence type and region, $F = 14.303, p < .01$. The analysis of the simple main effect of sentence type and region disclosed that the RTs for the region *who* in the WH-questions were significantly longer than those in the whether-questions ($p < .01$). The result suggests that the region *who* may impose a burden on the WM of the participants because they have to consider the two possibilities of *who* (i.e., subject or object). Then it was also found that the RTs for the region *called* in the WH-subject sentences were significantly longer than those in the other types of sentences ($p < .01$) and that the RTs for the region *the client* in the WH-subject sentences were shorter than those in the other types of sentences ($p < .01$). The result suggests that in WH-subject sentences, integrating the filler into its gap induces WM cost for the participants.

![Figure 7 RTs for all the participants](image)

Figures 8 and 9 given below, present the RTs per syllable by region and sentence type for the high and low span readers. There was a significant interaction between sentence type and region for the high span readers, $F = 9.726, p < .01$ as well as for the low span readers, $F = 6.560, p < .01$.

With regard to the high span readers, the simple main effect analysis revealed that there was a difference between the RTs of the WH-subject sentences and the other types of sentences ($p < .01$) for the region *who*. It was also disclosed that at the region *called*, the RTs for the short WH-subject questions were significantly longer than those for the short whether-questions ($p < .05$), the RTs for the long WH-subject questions were significantly longer than those for the short whether-questions ($p < .01$) and long whether-questions ($p < .05$), and at the region *the client*, the RTs for the WH-subject questions were significantly shorter than those for the other types of questions ($p < .01$).

With regard to the low span readers, the simple main effect analysis revealed that for the region *who*, the RTs for the WH-questions were significantly longer than those for the whether-questions. It was also disclosed that the RTs for the region *called* in the short WH-subject questions were significantly longer than those in the whether-questions and long-object questions ($p < .01$) and those in the short-whether questions ($p < .05$); further, the RTs
for the region *the client* in the WH-subject questions were significantly shorter than those for the other types of sentences (*p* < .01), and the RTs for the short-object questions were longer than those for the long-object questions (*p* < .01) and short-whether questions (*p* < .05).

As a result of the statistical analysis, it was revealed that (1) the RTs in the long-distance condition were shorter than those in the short-distance condition; (2) the scores for the WH-object questions were lower than those for the WH-subject and whether-questions, and the STs for the WH-object questions were longer than those for the whether-questions; (3) the RTs for the high span readers were shorter than those for the low span readers, regardless of syntactic type; (4) the high span readers obtained better scores in the case of the whether-questions and had shorter STs for the WH-object questions than did the low span readers; and (5) the RTs for the filler region (i.e., *who* region) for the WH-questions were longer than those for the whether-questions, and the RTs for the gap region (i.e., *called* region) for the WH-subject sentences were longer than those for the other types of sentences.

5. Conclusions and Further Studies

The main findings of the present experiment can be summarized as follows:

1. Syntactic processing has a greater influence on WM resources than does the filler-gap distance.
2. Participants with a large WM capacity can perform better than those with a small WM capacity.
3. The WM cost at the filler and gap positions is greater than it is at other regions in the processing of the WH-subject questions.
An unexpected finding was that the RTs per syllable in the long filler-gap distance were shorter than those in the short distance. This result stems from the fact that the participants in the present study process the region of adverb phrases faster than they do other regions (Figures 7, 8, and 9). The reasons for faster processing in this region may be as follows.

1. The words used in the adverb phrases in this experiment were more familiar to the Japanese EFL learners (Yokokawa, 2006) than those used in the other phrases.
2. The comprehension questions used in this experiment were presented in Japanese. The questions asked only about the relationship between agent and patient, excluding adverb phrases.
3. The filler-gap distance in this study was varied by inserting prepositional phrases. If we manipulate the length between the filler-gap, using other adverb phrases (i.e., relative-clause sentences), the processing speed in the adverb area may be slower.

In future studies, it is necessary to control these variables and investigate the factor of sentence-internal structure that increases WM cost for Japanese EFL learners.

Acknowledgement

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

6. References


門田修平. (2007). 『シャド－イングと音読の科学』コスモビア。


