The new index score of the L2 reading span test : The relationship between L2 working memory span and the processing of garden path sentences

The new index score of the L2 reading span test : The relationship between L2 working memory span and the processing of garden path sentences

Hiroshi Nakanishi

Abstract

This paper explores whether or not a new scoring method (processing efficiency score) of the L2 Reading Span Test (RST), which takes processing speed into account, predicts the performance of the garden path sentences. In Nakanishi (2007a) study, participants were required to take the RST and perform sentence processing tasks including garden path sentences. However, the scoring method of the RST, which was adopted in Nakanishi (2007a), did not predict the performances related to the garden path sentences for Japanese EFL learners. This is probably because high span readers in Nakanishi's study spent a considerable amount of time processing the sentences of the RST and memorizing the final words of the sentences, which boosted their score on the test. The present study reanalyzed the data of Nakanishi (2007a), using the processing efficiency score. The result showed that the high span readers assessed by the processing efficiency score performed better in the processing of garden path sentences task than did the low span readers. Therefore, the current paper proposes a processing efficiency score of the RST as an excellent predictor of L2 reading performance.

Key words

working memory, reading span test, processing efficiency score, garden path sentences

1. Introduction

Working Memory (WM) is conceived of as a cognitive system responsible not only for the storage of information, but also for the simultaneous processing of information. Much of the research has shown that the WM capacity is an excellent predictor of reading comprehension ability (Daneman & Carpenter, 1980 ; Daneman & Merikle, 1996). In addition, studies comparing the language comprehension behavior of higher and lower WM capacity participants have yielded useful data that help specify the manner in which WM constrains specific language processes (Miyake & Friedman, 1998) such as the resolution of linguistic ambiguity (Miyake *et al.*, 1994) and the parsing of syntactically complex structures such as object-clause sentences (King & Just, 1991) and garden-path (GP) sentences (Just & Carpenter, 1992).

Above all, Just and Carpenter (1992) explored the relationship between WM capacity and parsing in ambiguous sentences. They administered two tasks ; (1) Reading Span Test (RST) and (2) Sentence processing task. (1) In the RST, which is originally developed by Daneman and Carpenter (1980), participants are required to read aloud sets of sentences printed on cards (the processing requirement) while trying to remember the final words in the sentences for later recall (the storage requirement). In their view, WM resources are shared with both processing and storage functions. Therefore, the test measures the efficiency of both the processing and retaining of information. According to their scores, the participants were divided into two groups (high- and low-span). (2) In sentence processing task, participants were required to read 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, 4) unreduced relative clauses with <−animate> NP. Examples of the 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.

The time needed to read the sentences was measured using the eye-movement monitoring techniques. The reaction times (RTs) for four types of sentences by WM group are summarized in Figure 1. The analysis of a two-way analysis of variance (ANOVA) produced a significant interaction of animacy and span group (F (1,66) = 5.36, $p < .025$). This suggests that only the participants with large WM capacity have the ability to make use of semantic information during parsing.

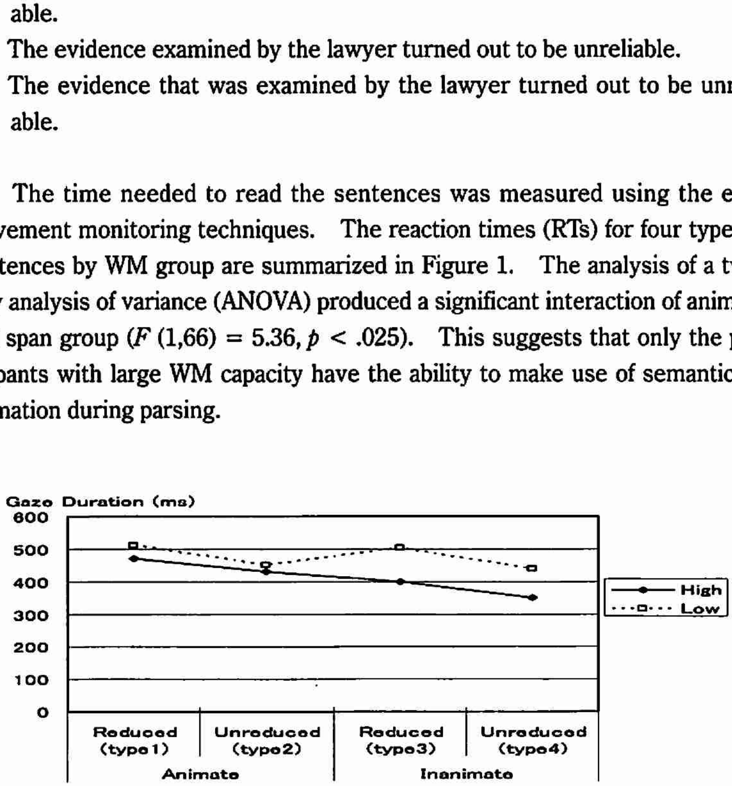


Figure 1 Just and Carpenter's (1992) result

Since L2 learners are also supposed to utilize the same cognitive language processing system, it is natural to assume that WM plays a crucial role in L2 language comprehension as well as L1. Nakanishi (2007a) explored this issue by investigating the affect of L2 WM capacity on the comprehension of syntactically ambiguous sentences, using GP sentences. Nakanishi (2007a) administered a computer-based RST and a sentence processing task which consisted of GP sentences and filler sentences.

The RST procedure adopted in Nakanishi (2007a) was illustrated in the following section (see 2.3 Procedure (1) The reading span test). The RST, originally developed by Daneman and Carpenter (1980), has been used to measure a participant's verbal WM capacity. In the original test, a participant is asked to read increasingly longer sets of sentences printed on cards and remember the final words of the sentences. The RST score is generally calculated based on the number of final words remembered. In Nakanishi (2007a), there are two main revisions to the original version of the RST, in particular, these relate to the procedure and the scoring method. First, the participants are required to read sentence silently and then answer a comprehension question. Second, the score is calculated as the number of correctly recalled words when the sentences presented were correctly processed by the participants.

The result of the study (Nakanishi, 2007a) was surprising in that the low span readers were able to perform the sentence processing task as well as the high span readers, which is inconsistent with previous L1 studies. We suggested that the result might stem from reaction times in the RST. In other words, high span readers in the study may have obtained better scores due to slow processing since they spent much more time memorizing the final words of the sentences of the RST. Thus, the RST scores in Nakanishi (2007a) might just reflect a trade-off between the reaction times (RTs) and the number of sentence-final words readers could recall. Therefore, if we adopt processing efficiency score (p-e score) which takes processing speed into account, the p-e score may be a better predictor of language comprehension ability (Nakanishi, 2005). The formula which represents the p-e score with a slight modification is presented in Table 1 as follows : First, calculate the processing speed (syllables per minute, spm) for each sentence. Second, if the sentence is correctly understood, multiply the processing speed (spm) by 1, or if it is not correctly understood, multiply the processing speed (spm) by 0. Next, multiply the entire score by 1 if the participant recalls the final word correctly, or multiply by 0 if it is recalled incorrectly. We calculate the each score for all 42 sentences using this formula and finally we total the score.

Table 1 The formula for the processing efficiency score

The processing efficiency score of each sentence = ∑[processing speed for each sentence (syllables per minute) × processing accuracy (0 or 1) × final word recall (0 or 1)]

2. Method and Procedure

This paper aims to test the validity of the 'p-e score' using the data from the Nakanishi (2007a) study. First, this section will review the procedure of the experiment presented in Nakanishi (2007a).

2.1. Participants

The participants for this experiment were 60 Japanese university students or graduate school students learning English as a foreign language.

2.2. Method

All participants completed 2 tasks : (1) the reading span test (RST), (2) the sentence processing task. The tasks were administered to the participant on a computer monitor. The entire experiment took approximately 30 minutes.

2.3. Procedure

(1) The reading span test

After a fixation marker was presented for one second on the computer monitor, the marker was replaced with a sentence. The participants were required to push the space button immediately after they read the sentence silently and to remember the sentence-final word. The reading time, which the participants needed to push the space button from the emergence of the sentence, was recorded. After pushing the button, the Japanese equivalent of the previous English sentence appeared on the monitor. The participants were asked to judge whether it was true (B) or not (N) and then to push the corresponding button. The reaction time which the participants needed to push the space button from the emergence of the question was also recorded. Then, the next sentence appeared on the monitor following the fixation mark. The participants were asked to read the sentence while remembering the final word from the former sentence. This procedure was repeated until they saw the instruction indicating the end of the session. The participants were then required to write down the final words of the sentences that had been presented on the answer sheet.

The sentences were presented in increasing set size that consisted of two to five sentences. There were three sessions of each set size, which added up to 42 recall words. The length of the sentences ranged from nine to thirteen words. The sentences were selected from the Osaka and Osaka's (1992) and Harrington and Sawyer's (1992) L2 study. However, sentences were modified so that the familiarity of the words (Yokokawa, 2006) among the sets would be statistically similar on average (F = 1.1445, *ns*.).

(2) The sentence processing task

Fifty-two sentences were used including 32 garden-path (GP) and non-GP control sentences and 20 filler sentences. The 32 sentences were categorized as follows : 1) reduced relative clauses with <+animate> NP 2) unreduced relative clauses with <+animate> NP 3) reduced relative clauses with <−animate> NP 4) unreduced relative clauses with <−animate> NP. Examples of the 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, selected from Ferreira and Clifton's (1986), were also slightly modified so that the familiarity of the words would be statistically similar across the 4 groups (F = .0476, *ns*.). The sentences were eleven to fifteen words in length. Sentences were presented word-by-word in a computer-generated random order. The procedure was as follows : When participants pushed the space button to advance the display to the next word in a sentence, the letters of that word would appear in the place of 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 quickly press either the 'B' key if the statement corresponds with the experimental sentence or the 'N' key if not.

3. Results

The reading span data was scored using the p-e method described above. Table 2 below shows the p-e score, reaction times (RTs) for correct response, and solution times (STs) for correct response for the high- and low-span groups assessed by p-e scores in the RST.

Table 2 Descriptive statistics of reading span data for reading span group divided in terms of the p-e score

High span

	p-e score	RTs (msec.)	STs (msec.)
Number	30	30	30
Mean	3,037.9	8,461.5	3,019.3
S.D.	723.0	2,482.5	958.3

Low span

	p-e score	RTs (msec.)	STs (msec.)
Number	30	30	30
Mean	1,775.0	11,107.6	3,418.3
S.D.	347.2	3,694.1	1,285.4

Table 3 and Figure 2 below stand for the mean scores of the sentence processing task in type 1-4 sentences for the high- and low-span groups divided according to p-e score.

An ANOVA analysis confirmed that there were significant main effects of sentence type, (F = 31.335, p < .01), and of span, (F = 7.918, p < .01). The results of Bonferoni's multiple comparisons further revealed that there were significant differences between type 1 and type 3, between type 1 and type 4, and between type 2 and type 3 in both span readers (i.e., those with p values of <.01). It was also discovered that high span readers got significantly better scores than low span readers in type 3 (p < .01).

Table 3 Mean scores of the sentence processing task for WM span group and sentence type

	type1	type2	type3	type4
High span	5.0	5.9	7.3	6.6
Low span	4.6	5.3	6.6	6.3

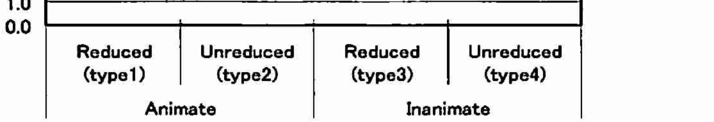


Figure 2 Mean scores of the sentence processing task for WM span group and sentence type

In the following Table 4 and Figure 3 are shown the mean RTs (msec.) per syllable for correct response in type 1-4 sentences for the high- and low-span groups divided according to p-e score.

Table 4 Mean RTs of the sentence processing task for WM span group and sentence type

	type1	type2	type3	type4
High span	498.4	480.3	487.9	397.6
Low span	617.6	602.1	575.7	474.1

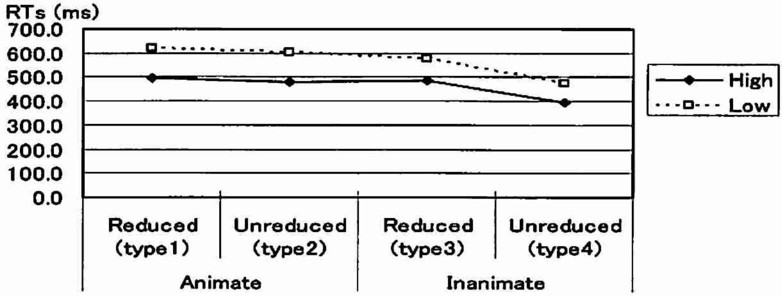


Figure 3 Mean RTs of the sentence processing task for WM span group and sentence type

According to an ANOVA analysis, there were significant main effects of sentence type, (F = 7.6012, p < .01) and of span group, (F = 25.805, p < .01). The results of Bonferoni's multiple comparison showed that, while there were significant differences between type 1 and type 4 (p < .01), between type 2 and type 4 (p < .05) and between type 3 and type 4 (p < .01) for high span readers, for low span readers the RTs in type 1 were significantly slower than those in type 4 (p < .01). It was also found that RTs for high span readers were shorter than those for low span readers in all the types of sentences (p < .01).

Table 5 and Figure 4 below represent the mean STs (msec.) per sentence for correct response in type 1-4 sentences for the high- and low-span groups divided according to p-e score.

Table 5 Mean STs of the sentence processing task for WM span group and sentence type

	type1	type2	type3	type4
High span	2,529.8	2,525.6	2,200.5	2,495.4
Low span	2,813.6	2,733.1	2,607.3	2,957.5

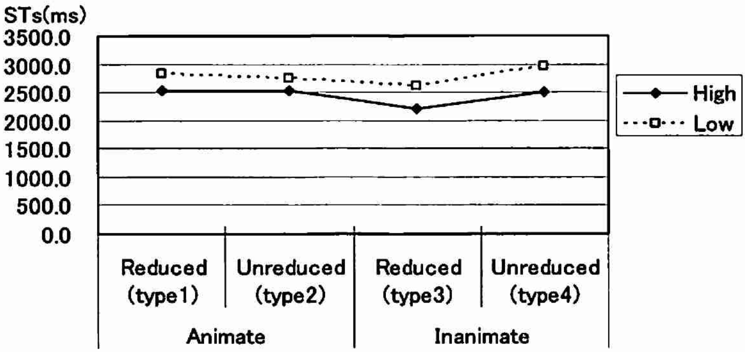


Figure 4 Mean STs of the sentence processing task for WM span group and sentence type

An ANOVA analysis revealed that there was no significant main effects of sentence type, (F = 6.046, *ns*), but significant main effects of span, (F = 6.046, $p < .05$). The results of Bonferoni's multiple comparison revealed that high span readers could answer the comprehension question significantly faster than low span readers in type 3 and type 4 ($p < .05$).

As the result of statistical analysis, it was discovered that high span readers assessed by p-e score tended to get better scores, read sentence faster, and solve the questions faster than low span readers. This finding is inconsistent with the results of Nakanishi's (2007) analysis. In addition, both span readers exhibited similar patterns of processing ; Both span readers make use of <−animate> NP cue as a clue to avoid the processing difficulty.

RTs were also analyzed by region across WM span group. Regions were divided into the three areas of the disambiguating phrase as follows :

The woman paid after the end of the month

$_{\alpha}$had /$_{\beta}$worried /$_{\gamma}$the man.

Region α (e.g. had) contains the initial word of the verb, which disambiguates the sentence, Region β (e.g. worried) contains the rest parts of the verb phrase (VP). Region γ (e.g. the man) contains the remaining parts of the sentence. Table 6 and Figure 5 below show the mean RTs per syllable by sentence types and regions for all participants.

Table 6 Mean RTs by sentence types and regions

	type1	type2	type3	type4
		RTs for region α		
mean	985.5	787.9	717.4	655.6
S.D.	441.5	282.6	309.1	288.8
		RTs for region β		
mean	532.1	478.5	494.6	462.9
S.D.	225.0	136.3	158.1	182.4
		RTs for region γ		
mean	476	435	425	414
S.D.	123	88	106	96

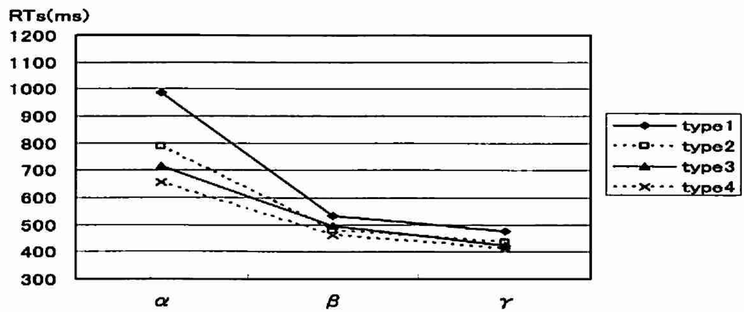
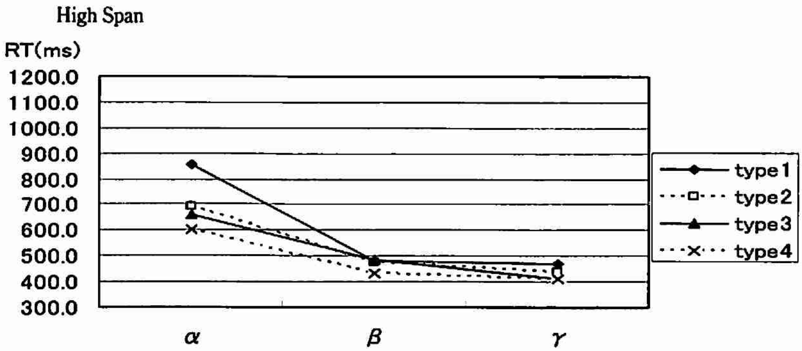


Figure 5 Mean RTs by sentence types and regions

An ANOVA analysis indicated a significant interaction between sentence types and regions ($F = 5.255, p < .01$). As a significant interaction was indicated, the simple main effects were further analyzed. The analysis disclosed that RTs for the region α in all types of sentences were significantly longer than those for the region β and γ in all types of sentences ($p < .01$). It was also found that RTs in type 1 for the region α were significantly longer than those in any other sentence type for the region α ($p < .01$).

Table 7 Mean RTs by sentence type and region for high and low span group

	type1	type2	type3	type4
		RTs for region α		
High span	856.6	693.4	657.5	603.0
Low span	1,114.4	882.5	777.3	708.0
		RTs for region β		
High span	481.2	476.6	486.9	432.0
Low span	582.9	480.3	502.3	493.7
		RTs for region γ		
High span	472	436	413	412
Low span	480	433	436	416



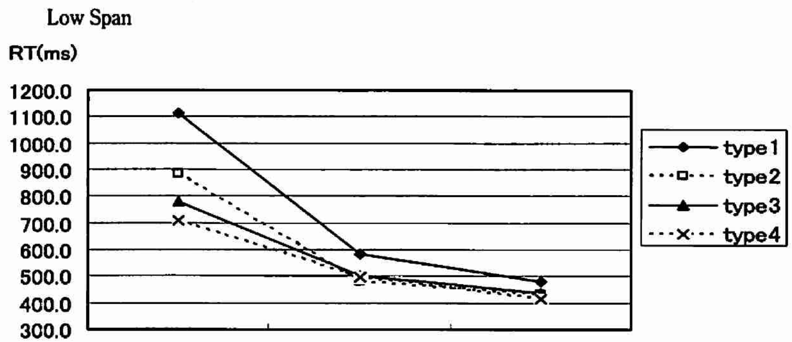


Figure 6　Mean RTs by sentence type and region for high and low span group

In Table 7 and Figure 6 are shown RTs per syllables by regions and sentence types for high and low span readers.

In the case of high span readers, RTs produced a significant interaction of sentence type and region ($F = 3.1369, p < .01$). The simple main effect analysis showed that RTs for the region α in all types of sentences were significantly longer than those for the region β and γ in all types of sentences ($p < .01$). It was also found that RTs in the type 1 for the region α were significantly longer than those in any other sentence type for the region α ($p < .01$).

Similarly, in the case of low span readers, there was a significant interaction between sentence types and regions ($F = 2.993, p < .01$). The result of the simple main effect analysis revealed that RTs for the region α were longer than any other region in all types of sentences ($p < .01$). It was also discovered that RTs in type 1 were longer than any other sentence type for the region α ($p < .01$).

As a whole, the major conclusions of the present experiment can be summarized as follows :

(1) The high span group divided by p-e score can process sentences more efficiently than low span group.
(2) Semantic information (i.e. the animacy of the subject NP) reduces the processing difficulty of the GP sentences for both high and low span readers.
(3) Both high and low span readers spend longer time reading region α (the initial word of the verb) than any other area especially in GP sentences. This suggests that region α is the most capacity demanding.

4. General Discussion

The present study asserts that the new index score (p-e score) for measuring L2 WM capacity can predict language processing efficiency for Japanese EFL learners more accurately than the score used in Nakanishi (2007a).

In the original version of the RST (Daneman & Carpenter, 1980), participants are required to read aloud a set of unrelated sentences printed on cards, while remembering the last word of each sentence for later recall. The test measures the efficiency of both the processing and retaining of language information. In the case of L1 studies, participants are supposed to read the sentences aloud in a rather automatic manner. It is hypothesized that reading aloud naturally accompanies reading comprehension with little WM resources. As a result, WM resources are allocated to the processing of sentences and the retaining of the final words in the performance of the RST task. The RST score is generally calculated based on the number of final words to be remembered.

However, in the case of L2 studies, reading aloud is performed in a less automatic manner. Therefore, the participants must allocate their WM resources to reading aloud itself, using many WM resources. In addition, for L2 learners, reading aloud is not necessarily accompanied by sentence comprehension. Consequently, WM resources are allocated to reading aloud and memorizing the final words of the sentences in the performance of the RST. As a result, the scores obtained by the original RST procedure for L2 learners do not necessarily reflect the efficiency of both the processing and retaining of language input (Kadota, 2007).

Accordingly, Nakanishi (2007a) argues that the RST for L2 learners should include tasks that direct participants' attention to the comprehension of sentences, such as sentence comprehension tasks and grammatical judgment tasks. In Nakanishi (2007a), the participants were required to judge whether or not the Japanese equivalent of the previous English sentence is correct. Additionally, the scoring method was also revised. The method of scoring involves counting the number of correctly recalled words when the sentences presented are correctly processed by the participants, instead of just counting the total number of words recalled.

However, the scoring method in Nakanishi (2007a) could not predict the performance of garden path sentence. This is because the high span readers of Nakanishi's (2007a) study spent much more time processing the sentences than did the low span readers, in order to remember the final words, which boosted their RST score. Kadota (2007) points out that the total number of final words to be remembered (original RST score) depends on the strategies that the L2 learners adapted in the performance of the RST. For example, as was illustrated in Nakanishi's (2007a) study, participants tend to spend a considerable amount of time intentionally memorizing the final words in order to memorize the target words in the performance of the RST.

Therefore, the RST score for L2 learners should include the time needed for language processing in the RST, such as the p-e score, as is proposed in the present paper. By taking RTs measure into consideration, the WM index for predicting language comprehension would be much upgraded, because reading behaviors are usually performed under the condition of time-dependent circumstance.

5. Conclusions and Further Study

The present study aimed to test whether the processing efficiency score (p-e score) of the reading span test (RST) can be a better predictor of the comprehension of syntactically complex sentences such as garden path (GP) sentences than the score used in Nakanishi (2007a). The results demonstrate that the p-e score can predict performance in language processing tasks better than that used in Nakanishi (2007a). Thus, the p-e score, which takes reaction times (RTs) of the RST into consideration, is a more effective measure of WM capacity and predictors of their reading performance. In other words, a potential determinant factor of WM capacity could be how efficiently the participants process sentences of the RST. However, the RST used in Nakanishi (2007a) is a revised version, which differs from the original in that participants are always posed comprehension questions about the prior sentence. Therefore, we cannot compare the results of the present analysis with the original version of the score used in L1 studies. Further research must be conducted to replicate the current study with an original version of the RST.

6. Acknowledgements

This paper is based on my Ph.D. thesis (Nakanishi, 2007b) submitted to Kwansei Gakuin University, in March, 2007. I would like to express my sincere gratitude to Professor Shuhei Kadota for his insightful advices, and I would like to thank the anonymous reviewers for their valuable suggestions for improvement of this paper.

References

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension : A meta-analysis. *Psychonomic Bulletin and Review, 3*, 422-433.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25-38.

Just, M. A., & Carpenter, P. A. (1992). Individual differences in working memory. *Psychological Review, 99*, 122-149.

Kadota, S. (2007). *Science of shadowing and reading aloud, [shadowing to ondoku no kagaku]*. Tokyo : Cosmopier.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing : The role of working memory. *Journal of Memory and Language, 30*, 580-602.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24*, 56-98.

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity : Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language, 33*, 175-202.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency : working memory as language aptitude. In Healy, A. F., & Bourne, L. E. (Eds.), *Foreign language learning : Psycholinguistics Studies on Training and Retention*, (pp. 339-364). New Jersey : Lawrence Erlbaum Association.

Nakanishi, H. (2005). How memory spans for reading, listening and spatial thinking are related with reading and listening comprehension for Japanese EFL learners. *Language Communication and Culture, 3*, Kwansei Gakuin University Graduate School, The Society of Language, Communication, and Culture.

Nakanishi, H. (2007a). How L2 working memory capacity for Japanese EFL learners are related with processing of garden path sentences. *ARELE, 18*, 191-200.

Nakanishi, H. (2007b). *Insights into second language working memory : An empirical study for Japanese learners of English*. Doctoral Dissertation Submitted to the Graduate School of Society of Language, Communication, and Culture, Kwansei Gakuin University.

Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society, 30*, 287-289.

Yokokawa, H. (Ed). (2006). *Familiarity of English words for Japanese EFL learners—Second language database for education and research. [Nihonjin eigo gakushusha no eitango shinmitsudo moji hen—kyoiku kenkyu no tameno dainigengo database]*. Tokyo : Kuroshio Pub.