

Receptive and productive knowledge of frequent and infrequent vocabulary

MOCHIZUKI, Masamichi

Research Background

A variety of models have been proposed to describe vocabulary knowledge: the breadth and depth model (Nagy & Herman, 1987); the size and dimension model (Meara, 1996); the partial/precise, depth, and receptive and productive model (Henriksen, 1999); and the word knowledge framework model (Nation, 2001). These models differ in that some include the distinction between receptive and productive aspects of vocabulary knowledge (Henriksen, 1999; Nation, 2001), while others do not (Nagy & Herman, 1987; Meara, 1996).

Whether they are a component of a lexical knowledge model or not, the distinction between receptive and productive vocabulary might be important to assess one's vocabulary knowledge, especially that of a second language (L2) learner. L2 learners tend to have irritating experiences of not being able to make themselves understood in the L2 because they lack productive vocabulary. They may understand what is said or written in an L2 with the help of their receptive L2 vocabulary but may have difficulty expressing themselves in an L2 due to their limited

productive vocabulary. Folse (2004) relates his own awkward experience of not being able to buy flour in Japan just because he did not know the Japanese word for it. It is commonly assumed that receptive knowledge precedes productive knowledge and the former size is larger than the latter. Griffin and Harley (1996) show that it takes more time and effort to learn vocabulary productively than receptively. However, it is difficult to clearly draw a line between receptive and productive vocabularies. Melka (1997) suggests that each person has different degrees of familiarity for different aspects of word knowledge and argues: “A crucial factor would be to establish at what point familiarity is such that one could say that knowledge is no longer receptive, but is productive, or at which point receptive knowledge can be converted into productive knowledge. The question is clearly not easy to answer” (p.86).

One attempt to find the dividing point between receptive and productive vocabulary is to measure the strength of word knowledge in four modes (Laufer & Goldstein, 2004). Although Laufer and Goldstein do not use the terms receptive and productive, what they attempted to show was to describe the degree of L2 learners' vocabulary knowledge strength. They proposed four modes of vocabulary knowledge strength along the two axes of active/passive and recall/recognition. Active recall vocabulary knowledge enables a learner to produce an L2 word on an L1 cue, active recognition knowledge helps a learner to recognize an L2 word on an L1 cue, passive recall knowledge enables a learner to produce an L1 word on an L2 cue, and passive recognition knowledge helps a learner to recognize an L1 word on an L2 cue. Laufer and Goldstein hypothesized that active recall is the strongest form of vocabulary knowledge, passive recognition is the weakest, and passive recall and active recognition lie somewhere between the two extremes. They tested the hypotheses by giving the tests of four modes to 435 high school and university students. Thirty words in each of the 2,000-, 3,000-, 5,000-word and Academic Word List levels were tested in the four modes. The 2,000 word level test was given to 52

ninth and 82 tenth graders; The 3,000-, 5,000-word, and Academic Word List level tests were given to 124 eleventh and 133 twelfth graders; 44 university students took the 5,000-word level test. They took the active recall mode first, then either the passive recall or the active recognition modes, and finally the passive recognition mode. Laufer and Goldstein found that there was a hierarchy of difficulty in the four modes: active recall was the most difficult, followed by passive recall, then active recognition, and passive recognition the least difficult. The hierarchy was not dependent on vocabulary frequency (Figure 1).

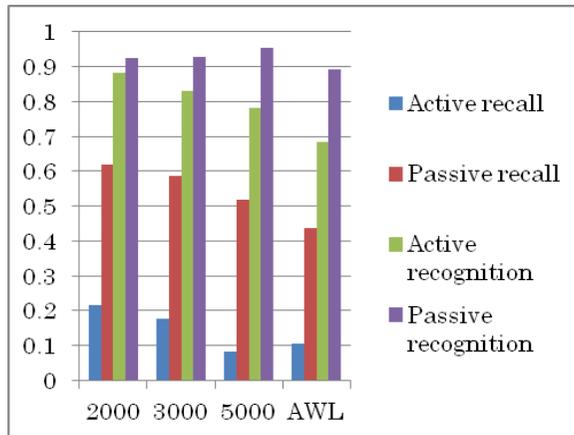


Figure 1. Facility Values of the Four Modes (created with the data from Laufer & Goldstein, 2004).

The findings of Laufer and Goldstein (2004) suggest that the difference between active and passive recognition knowledge is smaller than that between active and passive recall knowledge. The difference between the latter two looks quite large as shown in Figure 1. Active recall vocabulary knowledge seems to correspond to productive vocabulary and passive recall to receptive vocabulary. Thus, Laufer and Goldstein imply that learners have much larger receptive vocabulary sizes than productive vocabulary sizes.

In addition to Laufer and Goldstein (2004), there are a number of studies that address the difference between receptive and productive vocabulary sizes (Fan, 2000; Laufer, 1998; Waring, 1997; Webb, 2008). These studies, except for Webb (2008), employed Nation's Vocabulary Levels Tests (Nation, 1990; 2001) as a receptive vocabulary test and the Productive Levels Tests (Laufer & Nation, 1999) as a productive vocabulary test. They report differing ratios between the receptive and the productive vocabulary sizes of their participants.

Webb (2008) points out the drawback of these studies exploring the difference between the two different vocabularies. He argues that the use of the receptive and productive Vocabulary Levels Tests biased the estimates of receptive vocabulary sizes for four reasons. First, test-takers have a 17% chance of making a correct guess without any knowledge of a target word in the receptive VLT, while they have no chance of correctly guessing in the productive VLT. Thus, they are likely to score higher in the receptive test than in the productive one. Second, test-takers are tested on knowledge of form and meaning of target words in the receptive VLT, whereas they are also required to demonstrate knowledge of grammatical functions of those words in the productive VLT (it is possible to avoid this by ignoring grammatical functions in marking, though). This makes the productive test more demanding. Third, the receptive test employs a recognition format, while the productive test uses a recall format. The difference in formats is likely to produce an inaccurate comparison. Fourth, Webb claims that the test that gives part of the spelling of a target word as a cue might not measure productive knowledge but receptive knowledge. He argues that the productive VLT is a stricter test of receptive knowledge based on the assertion of Melka (1997); "it is, then, clear that the presence of partial information is often sufficient to recognize a word" (p.87) and the assumption that the cued recall test measures only productive knowledge of orthography. Thus, Webb maintains that the findings of the previous

studies comparing the receptive and productive vocabulary sizes are misleading.

Webb (2008) made an attempt to compare the receptive and productive vocabulary sizes in a valid way. He chose 60 words from each of the three frequency bands of the COBUILD dictionary: the 701st to 1,900th, 1,901st to 3,400th, and the 3,401st to 6,600th most frequent words in English. These 180 words were tested both receptively and productively in two versions of the test. In the receptive test, test-takers were given an L2 target word and were to write its meaning in their L1, Japanese. In the productive test, they were given the L1 recall of an L2 target word and were to spell the L2 word. Two versions of tests were created. Version A tests the receptive knowledge of 90 words and the productive knowledge of the other 90. Version B tests the same words in the other modes. Eighty-three Japanese university students participated in the study. Half of the participants took version A and the other half version B. They were tested on 90 words either receptively or productively and 180 words in total. When the results of the two versions were combined, all the 180 target words were tested both receptively and productively. The productive test responses were marked in two ways: strict and sensitive markings: the former only accepted correctly spelt words, while the latter accepted minor spelling mistakes and gave a full point to misspelt words.

The results of Webb (2008) showed that the receptive vocabulary sizes were larger in each of the three frequency bands than the productive vocabulary sizes in either marking method. The ratios of productive to receptive vocabulary sizes remained constant in the sensitive marking: 95% in frequency band 1, 91% in band 2, and 94% in band 3 in the sensitive marking. On the other hand, the ratios declined as the frequency levels lowered in the strict marking: they were 88%, 73%, and 65% (Figure 2). Webb ascribed to three factors the small difference between the receptive and productive vocabulary sizes in the sensitive marking. One is

the limitation of the study that measured only the form and meaning relationship in the productive test. The gap would have been larger if other aspects of word knowledge such as collocation and syntax had been addressed as well. The second factor Webb referred to was the EFL situation in which the participants studied English. He argued that the Japanese learners receive more explicit vocabulary instruction than ESL learners, who, like L1 learners, tend to learn vocabulary incidentally by receiving input. Explicit instruction might bring more productive vocabulary learning than incidental learning does. The third factor was rote learning Japanese learners often employ. The rote vocabulary learning with L2 to L1 translation and vice versa might help learners develop receptive and productive vocabulary to a similar degree. Webb suggests that “vocabulary instruction in Japan might offer a more balanced approach to learning vocabulary” (p.90).

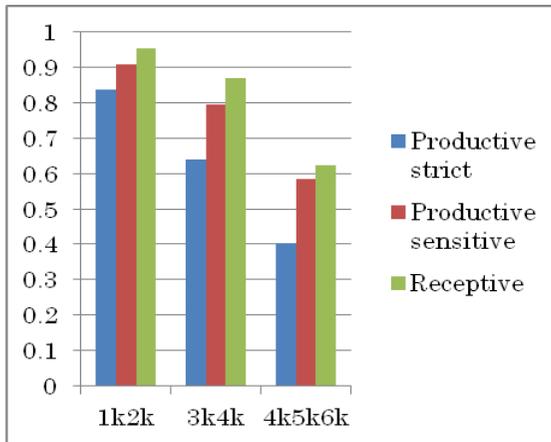


Figure 2. Facility Values of Receptive and Productive Vocabulary Tests
(created with the data from Webb, 2008).

The findings of the previous research provide an interesting issue for future studies. That is related to the difference between receptive and productive vocabulary. Laufer and Goldstein (2004) showed that there was a hierarchy

of strengths of vocabulary knowledge irrespective of vocabulary frequency, while Webb (2008) argued that his Japanese participants showed almost the same sizes for receptive and productive vocabularies in the sensitive marking. The findings of Laufer and Goldstein's might be interpreted as suggesting that L2 learners learn vocabulary knowledge in a fixed order no matter how frequently words are used. That is, learners learn L2 vocabulary by first recognizing an L1 word when presented with an L2 word, then by recognizing an L2 word when presented with an L1 word, next by recalling an L1 word on an L2 word cue, and finally by recalling an L2 word on an L1 word cue. On the other hand, Webb's findings might be interpreted as suggesting that Japanese learners learn both receptive and productive knowledge of L2 words to similar degrees when a sensitive marking method is adopted. Thus, it would be intriguing to find whether Japanese EFL learners learn receptive and productive vocabularies to similar degrees or they learn some aspects of vocabulary knowledge better than others. This study addresses the following two research questions:

RQ1: Is there a hierarchy of vocabulary knowledge strengths when EFL learners are tested English words in four modes: L2 recall, L1 recall, L2 recognition, and L1 recognition?

RQ2: Are receptive and productive vocabularies learned in a similar way irrespective of their frequency levels?

Method

Participants

Eighty-eight university students majoring in English participated in the study. They had studied English at least for six years before and they were intermediate-level learners. They took the four modes of a vocabulary test as part of their course.

Vocabulary tests

The vocabulary test employed in the study was a computer program called

J8VST (Mochizuki, 2007). It tests 125 words in four modes: English (L2) recall, Japanese (L1) recall, L2 recognition, and L1 recognition, each of which corresponds to active recall, passive recall, active recognition, and passive recognition respectively in Laufer and Goldstein's (2004) terms. The 125 words can be divided into five one-thousand-word bands, each of which contains 25 words. That means 25 words from the most frequently used 1,000 words in The *JACET List of 8,000 Basic Words* (JACET Basic Words Revision Committee, 2003), another 25 words from the second most frequent 1,000 words, another 25 words from the third most frequent 1,000 words, and so forth. The test estimates a test-taker's vocabulary size up to 5,000 words in each mode.

When a test-taker starts the computer program, the instruction screen appears and tells the test-taker how to take the four modes of the test. In the L2 recall mode, test-takers type a target word from the cue of an L1 word and a hint of the first letter of the L2 target word. The first letter is given to restrict possible answers to a target word. For example, the Japanese word “genkyu suru” can be translated into *refer to, tell, allude to*, in addition to the target word *mention*. In order to restrict the correct answer to *mention*, the first letter *m* is given (See Figures 3.1 and 3.2).

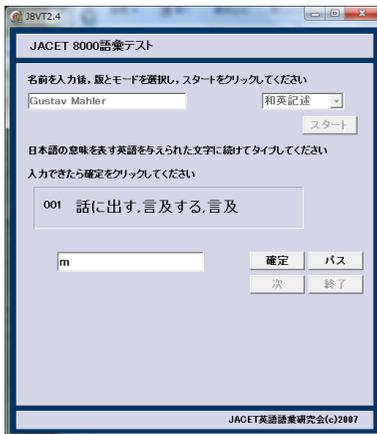


Figure 3.1. L2 Recall.



Figure 3.2. L2 Recall Answered.

Receptive and productive knowledge of frequent and infrequent vocabulary
(MOCHIZUKI, Masamichi)

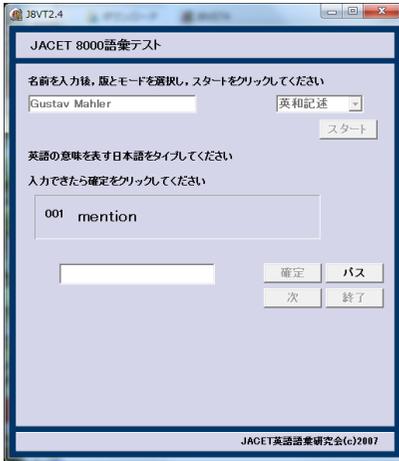


Figure 3.3. L1 Recall.



Figure 3.4. L1 Recall Answered.

In the L1 recall mode, test-takers enter an L1 word that corresponds to an L2 target word (Figures 3.3 and 3.4). In the L2 recognition mode test-takers choose an L2 word out of five options on an L1 word cue (Figure 3.5) and in the L1 recognition mode an L1 word out of five options on an L2 word cue (Figure 3.6).

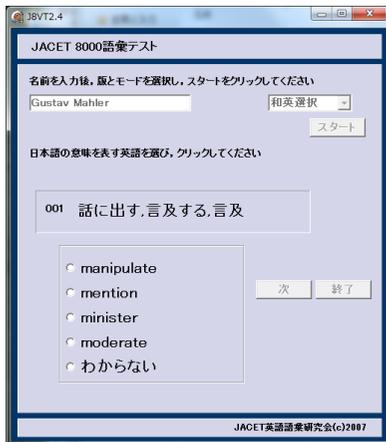


Figure 3.5. L2 Recognition.

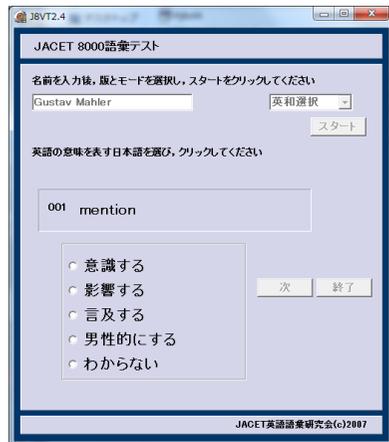


Figure 3.6. L1 Recognition.

The participants took the four modes of the test in the order of L2 recall, L1 recall, L2 recognition, and L1 recognition modes. They took the four modes in the computer laboratory in two 90-minute sessions. At the end of each mode they got the feedback of their vocabulary sizes of that mode and a text file was automatically created that recorded their responses to 125 test items.

Results

Out of 88 participants, three failed to complete one or two modes of the test and so were excluded from further analyses. Table 1 shows the means and standard deviations of the four modes of the vocabulary size test. On average, the participants recalled 2010 words on L1 cues and 2838 words on L2 cues, while they recognized 3902 words on L1 cues and 4271 words on L2 cues.

Table 1

Means and Standard Deviations of the Four Modes

N=85	L2 Recall	L1 Recall	L2 Recognition	L1 Recognition
<i>Mean</i>	50.25	70.94	97.56	106.78
<i>S.D.</i>	14.01	14.88	10.1	9.46

A one-way analysis of variance (ANOVA) was conducted to find a significant difference among the four mode means: $F(3)=369.161, p < .001$. A Bonferroni analysis found there were significant differences between any pair of modes ($p < .001$).

Table 2 shows the means of the four modes in the five vocabulary frequency bands. The means decreased as the vocabulary tested got infrequent as is seen in Figure 4. A one-way ANOVA was conducted for each of the five vocabulary frequency bands and found that there was a significant difference in means in each of the five bands: $F(3)=62.455$ for

the first 1,000-word band; $F(3)=229.372$ for the second 1,000-word band; $F(3)=324.658$ for the third band; $F(3)=269.672$ for the fourth band; $F(3)=292.496$ for the fifth band. Bonferroni analyses found that there were significant differences between the recall and the recognition modes ($p < .001$) but no significant differences between the two recall modes and between the two recognition modes in the first band. In the other four frequency bands, significant differences were found among the four modes at the $p < .001$ level, except for the difference between the L2 and the L1 recognition modes in the second band ($p < .05$).

Table 2

Means of the Four Modes in Five Vocabulary Bands

	1k	2k	3k	4k	5k
L2 Recall	21.6	11.9	7.6	4.6	4.5
L1 Recall	22.0	18.1	13.1	10.8	7.0
L2 Recognition	24.1	22.0	20.8	16.7	13.9
L1 Recognition	24.2	23.3	22.7	18.8	17.8

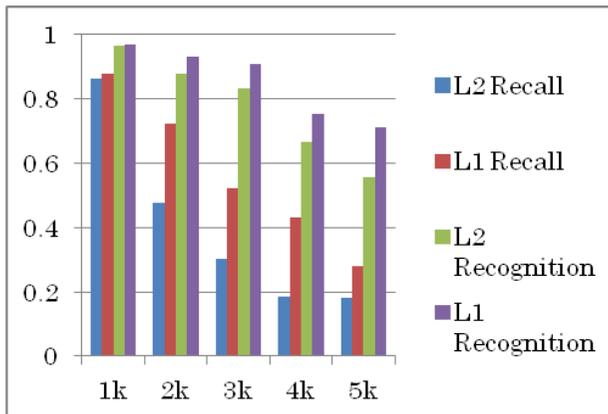


Figure 4. Mean Percentages of the Four Modes in Five Vocabulary Bands.

Discussion

The results present answers to the research questions. The significant difference among the means of the four modes implies a hierarchy of vocabulary knowledge strength: L1 recognition, L2 recognition, L1 recall, and L2 recall in the order of increasing difficulty. This supports the findings of Laufer and Goldstein (2004) and is consistent with the established fact that recall is psychologically more demanding than recognition (e.g. Tulving & Watkins, 1973; Griffin & Harley, 1996). It is natural for L2 learners to have difficulty producing L2 words because they contain phonemes and phonetic structures that are not used in their L1. So it is more difficult for learners to produce L2 words than L1 words. It may be argued that L2 learners learn L2 vocabulary first by recognizing L1 words when shown L2 words, second by recognizing L2 words when shown L1 words, third by recalling L1 words when given L2 words, and finally by recalling L2 words when given L1 words. This presumes that production of L2 words on L1 cues is more difficult than that of L1 words on L2 stimuli. This contradicts Webb's (2008) argument that his Japanese participants learned receptive and productive vocabularies to similar degrees because of the explicit vocabulary instruction they received. It is true that most Japanese EFL learners learn L2 vocabulary by L2 to L1 and L2 to L1 translations but they still follow more or less fixed stages of vocabulary learning: L1 recognition to L2 recall.

The second research question examined if receptive and productive vocabularies are learned in a similar way irrespective of their frequency levels. The results show that there were significant differences among the four modes of the vocabulary test except for the first 1,000 word level. This suggests that receptive and productive vocabularies are learned in a similar way, i.e. from L1 recognition to L2 recall, at least at the second, third, fourth and fifth 1,000 word levels. This also supports the findings of Laufer and Goldstein (2004). It may suggest that cognitive loads of recall or recognition of L1 or L2 words are similar regardless of their frequency.

Whether L2 words are frequently used or not, learners learn L2 words in four stages. In the first stage they recognize the L1 translation for an L2 word. In the second stage they recognize the L2 word for an L1 word. In the third stage they recall the L1 word for an L2 word. Finally, in the fourth stage they recall the L2 word for an L1 word. The result of this study suggests that L2 learners follow the four stages in vocabulary learning except for the most frequently used 1,000 word level.

The fact that the order was not observed in the first 1,000 word level may be accounted for by the proficiency level of the participants. They were all English-major university students who were at an intermediate level. They had learned the first 1,000 word vocabulary to such a degree that they showed no difference between receptive and productive vocabularies in the recognition or recall modes. They were able to recognize or recall receptive and productive vocabularies to the same degree at the first 1,000 word level, although they had more difficulty in recall than recognition. It may be hypothesized that less proficient learners would show the same learning order at the first 1,000 word level.

This study does not resolve the contradiction between Laufer and Goldstein's (2004) and Webb's (2008) findings. Laufer and Goldstein found an order of vocabulary knowledge strengths, which suggests learners learn receptive and productive aspects of L2 vocabulary in a fixed order. Their findings are more or less supported by this study. On the other hand, Webb found his Japanese EFL learners were able to produce L1 and L2 words on their counterpart cues to similar degrees when a sensitive marking was adopted. This suggests the learners had learned receptive and productive vocabularies to similar degrees. This should be addressed in future research.

This study has three major limitations. First, the participants took all the four modes in succession: L2 recall, L1 recall, L2 recognition and L1

recognition. So there might have been a practice effect in the L1 recall, and L2 and L1 recognition modes. Because the results show the difficulty order is exactly the same as the order of the test administrations, it would be necessary to replicate the study in a design that excludes practice effect in order to confirm the findings. Second, the test of L2 recall mode may not be testing productive vocabulary in a strict sense. The L2 recall mode of the vocabulary test, J8VST, gives an L1 word as a stimulus and the first letter of the target L2 word as a hint in order to exclude possible correct responses whose meanings correspond to the L1 stimulus word. However, some researchers like Webb (2008) may not regard a test giving the first letter of a target word as a test of productive vocabulary in a strict sense. Third, the L1 recall mode may not give possible correct responses a point. Because the J8VST was a computer program of vocabulary test, all answer keys had been installed in the program. In the L1 recall mode, answer keys were programmed in Chinese characters and Hiragana. So, for example, for the L2 cue word *nation*, L1 responses such as 国家, 国民, 国, 民族, こっか, こくみん, くに, and みんぞく were registered as answer keys. If test-takers wrote one of these responses, they were awarded with a point for a correct response. However, if test-takers wrote one of these responses in katakana letters such as コツカ, it was marked as incorrect. The study has these limitations and so care must be taken when we interpret the results.

Conclusion

This study addressed two issues related to receptive and productive vocabularies. First, it found that Japanese EFL learners showed differing degree of knowledge in the four modes of the vocabulary test: L2 recall the most difficult, L1 recall, L2 recognition, and L1 recognition the easiest. This supports Laufer and Goldstein's (2004) hierarchy of vocabulary knowledge strengths. Second, the study found that the difficulty order was the same in four of the five frequency bands. This implies that EFL learners learn L2 words in a fixed order irrespective of their frequency: L1

recognition, L2 recognition, L1 recall, and L2 recall.

Acknowledgement

The study is supported by the Grants-in-Aid for Scientific Research (Research No.22320110). I am grateful to Professor UEMURA Toshihiko for his cooperation in data collection.

References

- Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, 31, 105-119.
- Folse, K. S. (2004). *Vocabulary myths: Applying second language research to classroom teaching*. The University of Michigan Press.
- Griffin, G. f., & Harley, T. A. (1996). List learning of second language vocabulary. *Applied Psycholinguistics*, 17, 443-460.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21, 303-317.
- JACET Basic Words Revision Committee. (2003). *JACET List of 8000 Basic Words*. The Tokyo: Japan Association of College English Teachers.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 19, 255-271.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: size, strength, and computer adaptiveness. *Language Learning*, 54, 399-436.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16, 33-51.
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp.35-53). Cambridge: Cambridge University Press.
- Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In Schmitt, N., and McCarthy, M. (eds.), *Vocabulary: Description*,

- acquisition and pedagogy*. Cambridge University Press, pp. 84-102.
- Mochizuki, M. (2007). Development of a vocabulary test based on JACET 8000. In K. Aizawa (Ed.), *Construction of a vocabulary list for Japanese learners of English and development of a system for analyzing educational materials based on large-scale corpora* (Kaken Report: Research Number 1632076), pp.134-145.
- Nagy, W.E., Herman, P. A., & Anderson, R.C. (1985). Learning words from context. *Reading Research Quarterly*, 20, 233-253.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Tulving, E., & Watkins, M.J. (1973). Continuity between recognition and recall. *American Journal of Psychology*, 86, 352-373.
- Waring, R. (1997). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata*, 1, 53-68.
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30, 79-95.