Second Language Vocabulary Research: 2004

Yuji Tanaka

This is a review article on second language vocabulary research. Articles published in leading international research journals in 2004 are the scope of this investigation. The present review comprises the following key themes: testing vocabulary size and strength, the avoidance of phrasal verbs in English, homography in the Academic Word List, gesture and speech in the vocabulary explanations of one ESL teacher, validating a measure of written lexical retrieval in a second language, the effect of second language reading proficiency on lexical inferencing, and semantic transfer in second language vocabulary learning.

Testing Vocabulary Size and Strength

Laufer and Goldstein (2004)[1] described a bilingual computerized test of vocabulary size and strength, and conducted a paper-and-pencil study with a view to examining whether their hypothesized hierarchy of vocabulary learning difficulty was valid.[2] (The hierarchy is to be shown in the paragraphs that follow.)

The term *vocabulary size* is familiar to those involved in second language vocabulary research, and what it means is the number of words a person knows. In contrast, the term *vocabulary strength* is not so often used as the term *vocabulary size,* and I think that this term (i.e., *vocabulary strength*) needs to be explained.

Vocabulary strength comprises four degrees of knowledge assumed to constitute a hierarchy of difficulty: passive recognition (the easiest among the four), active recognition, passive recall, and active recall (the hardest) (pp. 404-408).

These four degrees of knowledge (i.e., passive recognition, active recognition, passive recall, and active recall) are called *degrees of "strength" of knowledge* in Laufer and Goldstein's (2004, p. 408) article.[3]

Regarding these degrees of "strength" of knowledge, the two researchers state that "if[4] active knowledge is more difficult to achieve than passive knowledge, and if recall is more difficult than recognition, then the most advanced degree of knowledge is reflected in active recall and the least advanced knowledge is passive recognition" (p. 408).

As for the remaining two (i.e., passive recall and active recognition), Laufer and Goldstein (2004) state that they intuitively felt that passive recall would require more knowledge than active recognition but that, before their validation study, they refrained from hypothesizing which of these degrees of strength was more difficult to attain (p. 408).

This study comprised 391 high school students (52 9th graders, 82 10th graders, 124 11th graders, and 133 12th graders) and 44 university students learning English as a foreign[5] language (N = 435). Specifically, the participants consisted of 278 native speakers of Hebrew, 140 native speakers of Arabic, and 17 native speakers of Russian (pp. 415-416).

The results indicated that the hypothesized hierarchy was present at all word

frequency levels examined in Laufer and Goldstein's (2004) study. (The researchers examined 2,000-word level, 3,000-word level, 5,000-word level, and the Academic Word List level.) At all of these frequency levels, "the hierarchy was the same: active recall yielded the lowest scores, and then in ascending order, passive recall, active recognition, and passive recognition" (p. 417).[6]

The central argument of Laufer and Goldstein's (2004) article is that "the main component of vocabulary knowledge, which is the knowledge of form-meaning relationship, can be construed as a hierarchy of four degrees of strength: active recall, passive recall, active recognition, and passive recognition" (p. 421). The overall results lent support to this hypothesis.

The Avoidance of Phrasal Verbs in English

Liao and Fukuya (2004) investigated the avoidance of English phrasal verbs by Chinese learners of English.

After giving descriptions of "avoidance" (pp. 194-196) and "phrasal verbs in English" (pp. 196-197) and reviewing previous studies on "avoidance of phrasal verbs in English" (pp. 197-201), Liao and Fukuya (2004) presented the following three research questions (p. 201): (a) Do Chinese learners avoid using phrasal verbs?; (b) Does their avoidance, if any, reflect differences in the semantic nature of phrasal-verb types (figurative vs. literal)?; and (c) Does[7] their avoidance, if any, reflect the ways their performance is measured?

This study comprised 15 native speakers of English, 30 advanced Chinese learners of English, and 40 intermediate Chinese learners of English. Specifically, the native speakers of English were undergraduate students at the University of Hawai'i[8] at Mānoa. As for the advanced learners, they were graduate students at the same

university. Their Test of English as a Foreign Language (TOEFL) scores were all above 600. These students were randomly assigned to one of three groups, with each group given one of the three tests (multiple-choice, translation, or recall) (pp. 201-202).[9]

Regarding the intermediate learners of English, 10 were graduate students at the University of Hawai‘i[10] at Mānoa and the remaining 30 were college students in China. The TOEFL scores of the graduate students ranged from 500 to 590. In respect of the college students in China, all of them had studied English for a minimum of 8 years (6 years in secondary school and 2 in college) and had passed Band 6 of the College English Test (a national standardized English test for college students in China), which is roughly equivalent to 500-600 on the TOEFL. These 40 intermediate learners were combined into one group on the grounds that the 10 graduate students at the university had been in the United States for only a short period of time.[11] Of the 10 learners from the university, five took the multiple-choice test and five took the translation test. The 30 learners from China were randomly assigned to one of three groups, with each group given one of the aforementioned three tests (i.e., multiple-choice, translation, or recall) (p. 202).[12]

First, the researchers asked the 15 native speakers to take the multiple-choice test with a view to identifying 15 cases of native-speaker preference for a phrasal verb over a semantically equivalent one-word verb in a short dialogue. Specifically, the 15 pairs of phrasal and one-word verbs used in Liao and Fukuya's (2004) study are the following (p. 222): (a) literal phrasal verbs and their one-word equivalents (*get up, rise; go away, leave; take away, remove;* and *come in, enter*) and (b) figurative phrasal verbs and their one-word equivalents (*show up, appear; brush up on, review [improve];*[13] *let down, disappoint; go off, explode; hold on, wait; put out, extinguish; make up, invent; give in, surrender; turn down, refuse; show off, boast;* and *run into,*

meet). Next, after identifying the 15 pairs, the researchers conducted a study with the intention of investigating whether and to what extent these Chinese learners of English would avoid using phrasal verbs (p. 203).

In what follows, a concise description of the results will be given. First, let me begin with the results pertaining to the aforementioned research question (a). In Liao and Fukuya's (2004) article, learners' avoidance of phrasal verbs is defined as "a usage that is lower than that of the native speakers at a statistically significant level" (p. 209). "On the multiple-choice test, the advanced learners used phrasal verbs 75% of the time and one-word verbs 21% of the time, whereas the intermediate learners used them 45% and 43% of the time, respectively. The native speakers, on the other hand, used phrasal verbs 84% of the time and one-word verbs 16% of the time" (p. 209). Additionally, the researchers analyzed the obtained data using a 3 x 2 analysis of variance (ANOVA), with Proficiency Level (native speakers vs. advanced learners vs. intermediate learners) serving as a between-participant variable and Phrasal-Verb Type (figurative vs. literal) serving as a within-participant variable. (They call this ANOVA "Analysis I.") The results of Analysis I showed that "the intermediate learners produced phrasal verbs much less frequently than both the advanced learners and the native speakers, which means that the intermediate learners avoided using phrasal verbs and preferred the one-word verbs" (pp. 209-210). Moreover, "although[14] the advanced learners did not perform very differently from the native speakers, they also showed a slight tendency to use phrasal verbs less than the native speakers" (p. 210). Furthermore, the obtained data were also analyzed using a 2 x 3 x 2 ANOVA, with Proficiency Level (advanced learners vs. intermediate learners) and Test Type (multiple-choice vs. translation vs. recall) serving as between-participant variables, and Phrasal-Verb Type (figurative vs. literal) serving as a within-participant variable. (This ANOVA is called "Analysis II.") The results of Analysis II demonstrated that

“on all three elicitation tests, the advanced learners used significantly more phrasal verbs than the intermediate learners did” (p. 210).

Second, let me move on to the results concerning the research question (b). (To reiterate, the question [b] investigated whether the avoidance of phrasal verbs by Chinese learners would reflect differences in the semantic nature of phrasal-verb types [i.e., figurative vs. literal].) Liao and Fukuya (2004) summarized the main points of the results as follows:

> Analysis I found phrasal-verb type to have been statistically significant on the multiple-choice test, with the mean for literal phrasal verbs higher than that for figurative phrasal verbs. There was no interaction between the phrasal-verb types and the three groups, which means that all three groups favored more literal phrasal-verb production than figurative phrasal-verb production on the multiple-choice test. Analysis II also found that learners favored more literal phrasal-verb production than figurative phrasal-verb production. (More precisely, significant results were obtained only on the translation test.) Analysis II found no interaction between group and phrasal-verb type, which means that learners of both proficiency levels performed in a similar way, using figurative phrasal verbs less often than literal ones. Finally, there was no difference between the advanced and intermediate learners in terms of using figurative phrasal verbs less than literal ones. (p. 214)

Third, in respect of the research question (c), which examined “the test effect on the participants’ avoidance of phrasal verbs” (p. 215), the results are summarized in the following statement by Liao and Fukuya (2004):

> Analysis II on the test effect revealed that there was an interaction between test type and phrasal-verb type and that this interaction was found only on the translation test. This means that it was only on the translation test that the

> Chinese learners (both advanced and intermediate) showed the tendency to use figurative phrasal verbs less often than literal ones. Hence, inherent L2 complexity played a role in the intermediate Chinese learners' avoidance of phrasal verbs when they took the translation test, which made neither the phrasal verbs nor their one-word equivalents available. This is also a reason that the advanced Chinese learners used literal phrasal verbs more often than figurative ones. (p. 216)

Learners' phrasal-verb avoidance behavior is undoubtedly one of the issues second language vocabulary research should deal with, and I think that this article (i.e., Liao & Fukuya, 2004) is a valuable addition to the existing literature on the avoidance of phrasal verbs.

Homography in the Academic Word List

Wang Ming-Tzu and Nation's (2004) study investigated the words in the Academic Word List (AWL; Coxhead, 2000). Specifically, the two researchers scrutinized all the 570 items in the AWL with the intention of ascertaining whether the existence of unrelated meanings for the same word form (i.e., the existence of homographs) resulted in the inclusion of words which would not be in the AWL if their clearly different meanings had been distinguished (p. 291).

After outlining Coxhead's (2000) AWL and giving a concise description of such terms as *polyseme, polysemy, homograph, homography,* and *homonym* (pp. 291-296), the researchers presented the following research questions (p. 296): (a) What proportion of the word families in the AWL contains homographs that occur frequently in the Academic Corpus?;[15] (b) If these homographs are counted as separate word families, what word families in the AWL would not meet Coxhead's criteria for

inclusion in the AWL?;[16] and (c) If[17] a word family contains homographs, to what degree are the different homographs limited to one particular discipline or larger subject area?

The results of this study showed that only a small proportion of the word families in the AWL contained homographs and that in almost all cases, one of the members of a pair or group of homographs was much more frequent and widely used than the other(s). Furthermore, the analysis also revealed that three word families (i.e., *intelligence, offset,* and *panel*) had to be deleted from the AWL; this is because none of their homographs separately met the criteria for inclusion in the AWL (p. 291, pp. 305-309).

Specifically, the findings of this study indicated that of the 570 word families in the AWL, 60 contained homographs. (Appendix of Wang Ming-Tzu and Nation's [2004, pp. 311-313] article shows these 60 word families.) More specifically, of these 60 word families containing homographs, 21 had homographs that were frequent enough in the Academic Corpus,[18] and "of these 21, 6 had members which met the criteria for being additional AWL families, and 3 of the 21 had to be excluded from the AWL because neither homograph in each family met the criteria for inclusion. For the remaining 12 of the 21, no change was required in the AWL because the homographs were so infrequent. So for the 570 original word families, there were only 9 changes (6 additions and 3 deletions)" (pp. 305-306).[19]

Gesture and Speech in the Vocabulary Explanations of One ESL Teacher

Lazaraton (2004) investigated gesture and speech in the vocabulary explanations of one teacher of English as a second language (ESL) by employing

microanalysis in conjunction with conversation analysis (CA) (Atkinson & Heritage, 1984). Microanalysis has, according to Lazaraton (2004), "recently been embraced by applied linguists interested in validating oral language tests (Lazaraton, 2002), in understanding SLA (Markee, 2000), and in teaching oral skills (Riggenbach, 1999) and involves segmenting data into turns of talk and analyzing single cases to build a collection of instances that exemplifies some interactional phenomenon" (pp. 93-94).

The focus of Lazaraton's (2004) article is "at the intersection of two understudied areas in SLA: nonverbal behavior and unplanned vocabulary explanations. That is, in addition to examining the gestures used by one ESL teacher, this study also places the locus of interest on unplanned vocabulary explanations in the L2 context" (pp. 88-89).

After reviewing previous studies on "nonverbal behavior in the L2 context" (pp. 84-88) and "vocabulary explanations" (pp. 89-90), Lazaraton (2004) gave a description of her study that explored the role of teacher gestures in unplanned vocabulary explanations.

The data for this study consisted of three 50-min videotaped classes conducted by one ESL teacher. This teacher, a female master of arts (MA) ESL graduate student in her late 20s from Japan, had 5 years of experience teaching English as a foreign language in a private language school in Japan and 1 year of U.S. ESL experience as a teaching assistant in an intensive English program (IEP) at the time of the taping. Three of her IEP grammar classes were taped. The teaching points for these classes were (a) relative clauses and gerunds, (b) past progressive verb tense, and (c) mass/count nouns and quantifiers. A total of 23 students participated in one or more of the three videotaped classes. There was a nearly even mix of males and females, from their late teens to their early 30s. The videotapes were transcribed by employing both microanalysis and conversation analysis to represent the speech, gesture, and

other nonverbal behavior that accompanied unplanned vocabulary explanations (pp. 91-94).

In the analysis of the data obtained from the videotapes, "the[20] gesture classification system of McNeill (1992), which delineates different types of hand movements (iconics, metaphorics, deictics, beats), was used to understand the role the gestures played in these explanations" (p. 79).

The findings revealed that the aforementioned teacher used a variety of gestures in giving vocabulary explanations. Specifically, Lazaraton (2004) gave a detailed description of the following lexical items (pp. 94-106): *mislay, swear, sweep, dig, feed, tear, weave, argue, tiptoe, a majority of,* and *hypothesis.* After examining the obtained data, Lazaraton (2004) states as follows:

> The data in this study point to the potential significance of gestural input to L2 learners. In each of the data fragments analyzed, the verbal input from the teacher was only part of what the learners were presented with during these unplanned vocabulary explanations by the teacher. Overlooking the input provided by the nonverbal channel in L2 classrooms, or in other forms of face-to-face L2 interaction, cannot be justified on the grounds that "it doesn't matter," unless it can be shown *empirically* to be of no consequence. (pp. 106-107)

Lastly, I would like to say that I am in agreement with Lazaraton's (2004) comment that "nonverbal[21] behavior is certainly a fundamental aspect of L2 teaching" (p. 109). As she puts it, "further research is needed to understand the parameters of nonverbal behavior in the L2 classroom, its status as input, and its role in learner output and uptake of the L2" (p. 109).

Validating a Measure of Written Lexical Retrieval in a Second Language

Snellings, van Gelderen, and de Glopper (2004) aimed to validate a measure of written lexical retrieval in a second language.

In contrast to picture naming tasks, the above measure, called Written Productive Translation Task (WPTT), is not restricted to concrete nouns and verbs that can be depicted by pictures. In addition, by the use of this measure, we can test not only single words but also combinations of words. Besides, the WPTT can be used in a classroom setting because individual testing is not required (p. 174, p. 178).

Drawing on Messick's (1989) validity framework, Snellings et al.[22] (2004) investigated whether the WPTT was an appropriate measure of the construct of written lexical retrieval (p. 174, pp. 178-180).

This study comprised 109 Dutch learners of English. The ages of these participants ranged from 14 to 15; they were 9th graders (p. 181). Specifically, of the initial 109 participants, four students were deleted from the analysis because they failed to meet a criterion set by the three researchers; in short, these students' data were regarded as outliers. Additionally, another eight missed a test session. Therefore, in the final analysis, the total number of the participants became 97 (pp. 186-187).

With the intention of examining how the WPTT relates to such constructs as lexical access and orthographic encoding and whether the scores of the WPTT resemble those of a picture naming task, which is commonly used to measure the construct of lexical retrieval, Snellings et al. (2004) administered the following tasks in addition to the WPTT: (a) a picture naming task (a measure of lexical retrieval), (b) a lexical decision task (a measure of lexical access), and (c) an orthographic encoding task, which is said to measure the subprocess of written lexical retrieval. Additionally,

with a view to eliminating the effect of typing skill differences, (d) a typing task was also administered, and each participant's typing speed was measured (pp. 181-185).

The results of this study showed that the internal consistency of the WPTT was high with Cronbach's alpha of .95 (p. 189). Moreover, the results of correlational analyses indicated that the correlations (Pearson product-moment correlations) between the WPTT and the aforementioned tasks (i.e., [a], [b], [c], and [d]) were .71 (the correlation between the WPTT and the picture naming task), .43 (that between the WPTT and the lexical decision task), .53 (that between the WPTT and the orthographic encoding task), and .60 (that between the WPTT and the typing task), respectively; all of the above correlations were statistically significant, $p < .01$, one-tailed (pp. 189-190).

It should be also noted that after the effect of typing skill differences was eliminated, the correlations between the residual[23] WPTT (WPTT') and the above tasks (i.e., [a], [b], and [c]) became .59 (the correlation between the WPTT' and the residual picture naming task [PN']), .31 (that between the WPTT' and the residual lexical decision task [LDT']), and .45 (that between the WPTT' and the residual orthographic encoding task [OE']), respectively; all of these correlations were statistically significant, $p < .01$, one-tailed (p. 191).

Following the correlational analyses, the researchers conducted multiple regression analysis with a view to investigating what proportion of variance of the WPTT' scores could be explained by the following three independent variables (p. 192): PN', LDT', and OE'.

The results of the multiple regression analysis indicated that the following two independent variables contributed to the prediction of the dependent variable (i.e., the WPTT'): the PN' (semipartial correlation squared, .19) and the OE' (semipartial correlation squared, .05). Taken together, these two variables explained 41% of the

variance of the WPTT', R^2 = .41. Regarding the LDT', the analysis showed that this independent variable did not contribute to the prediction of the dependent variable (i.e., the WPTT') (pp. 192-193).

I think that in addition to the empirical investigation summarized above, pages 174-180 of Snellings et al.'s (2004) article are informative because they provide an excellent summary of research on lexical access and retrieval.

The Effect of Second Language Reading Proficiency on Lexical Inferencing

Bengeleil and Paribakht (2004) reported on an introspective study that investigated the effect of second language (L2) reading proficiency on L2 lexical inferencing[24] while reading an English expository text.

The participants of this study were 10 intermediate and 7 advanced Arabic medical students learning English as a foreign language in Libya. Their ages ranged from 22 to 25 years old, and the above two reading proficiency levels (i.e., intermediate or advanced) were determined by the result of a reading comprehension section of the Canadian Test of English for Scholars and Trainees (CanTEST)[25] (pp. 228-229).

The main results of this study can be summarized as follows. First, in lexical inferencing, both the intermediate and the advanced learners used almost the same knowledge sources and contextual cues, with the exception of word association, which was used by the intermediate learners only (pp. 230-235).

Second, both groups made use of single as well as multiple (two to five) knowledge sources in lexical inferencing; however, "the intermediate group used multiple sources more often and showed more variation in the combinations of

knowledge sources and contextual cues . . . than the advanced group" (p. 236).

Third, the results also indicated that intralingual L2-based knowledge sources were used far more often than interlingual L1-based knowledge sources, and that among the three knowledge sources (i.e., word-level, sentence-level, and discourse-level knowledge sources), sentence-level knowledge source was most often used by both groups (p. 236).

Fourth, it was also shown that the advanced learners had a higher percentage of correct and partially correct answers and a lower percentage of wrong inferences than the intermediate learners (p. 236).

Lastly, the obtained data also indicated that although there were gains in the participants' knowledge of the target words immediately after inferencing and over time, these gains were not statistically significant either within or across groups, but that there was a significant rate of learning and retention over time for the combined sample (i.e., $N = n_{\text{intermediate}} + n_{\text{advanced}} = 17$) (pp. 237-239).

Semantic Transfer in Second Language Vocabulary Learning

Jiang's (2004) article investigated semantic transfer in second language vocabulary learning.

This study (i.e., Jiang, 2004) is a replication of Jiang's (2002) article. The researcher intended to replicate the findings reported in Jiang (2002) by employing different ESL learners (i.e., Korean learners of English [$n = 15$]); in contrast, the participants of Jiang's (2002) study were Chinese learners of English ($n_{\text{Experiment 1}} = 25$, $n_{\text{Experiment 2}} = 18$). To be precise, both studies (i.e., Jiang, 2002, 2004) included native speakers of English as well. Specifically, 27 and 18 native speakers of English participated in Experiment 1 and Experiment 2 of Jiang's (2002) study, respectively;

15 native speakers of English participated in Jiang's (2004) study.

The participants of this study (i.e., Jiang, 2004) were, as noted in the preceding paragraph, 15 Korean learners of English as a second language (ESL) (5 females and 10 males) and 15 native speakers of English. The Korean ESL learners were 14 graduate students and an undergraduate student studying at Auburn University, the United States of America. The ages of the Korean ESL learners ranged from 20 to 37, with a mean of 30.8 (*SD* = 4.6). Their TOEFL scores ranged from 550 to 595, with a mean of 571.7 (*SD* = 18). The above learners ranged in the years of residence in the United States from 0.7 to 8.0, with a mean of 3.4 (*SD* = 1.9). As for the native speakers of English, they were students and staff members at Auburn University. At the time of testing, none of them had prior knowledge of Korean (pp. 421-422).

As for experimental materials used in this study, they consisted of (a) 64 English word pairs that were related in meaning and (b) 64 English word pairs that were unrelated in meaning. Specifically, the above (a) was composed of (c) 32 English word pairs that shared the same Korean translation equivalent and (d) 32 English word pairs whose two members were related in meaning but had two different Korean translation equivalents. In Jiang's (2004) article, the above (c) is called *same-translation pairs*[26] and the above (d) *different-translation pairs* (pp. 421-423).

The participants (*N* = 30) performed a semantic judgment task in which they were asked to judge as quickly and accurately as possible whether or not two English words in each pair were related in meaning. They were tested individually with the aid of a computer, and their reaction times and error rates were recorded. All the participants had to respond to the same 128 items (i.e., the above [a] plus [b]), but the presentation order was randomized and thus different for each participant (pp. 421-423).

This study employed a 2 x 2 x 2 mixed design, with Participant Group (native

speakers of English vs. Korean ESL learners) serving as a between-participant variable and Word Relatedness (related pairs vs. unrelated pairs) and Translation Status (same-translation pairs[27] vs. different-translation pairs) serving as within-participant variables (p. 423).

The findings of this study demonstrated that the Korean ESL learners responded to the same-translation pairs 113 milliseconds (ms) faster than to the different-translation pairs. Specifically, their average reaction times for the same-translation pairs and different-translation pairs were 1549 ms and 1662 ms, respectively. This difference was statistically significant. Additionally, it was also shown that they made significantly more errors on the different-translation pairs than on the same-translation pairs. (Their average error rates for the same-translation pairs and different-translation pairs were 7.7% and 13.3%, respectively.) As for the native speakers of English, their average reaction times for the same-translation pairs and different-translation pairs were 1058 ms and 1074 ms, respectively; their average error rates for the same-translation pairs and different-translation pairs were 8.7% and 8.1%, respectively. That is, they responded to the same-translation pairs 16 ms faster than to the different-translation pairs, but this difference was not statistically significant; nor was the difference in error rates (pp. 423-425).

Based on the results of this study, Jiang (2004) explored pedagogical implications for second language vocabulary teaching (pp. 425-428). Specifically, the researcher discussed "a broader definition of lexical competence, the use of the L1 as a means of semantization, and ways to help learners overcome semantic fossilization" (p. 425).

In my judgment, this article (i.e., Jiang, 2004) is one of the basic readings for those interested in the same-translation effect in second language semantic judgment.

Conclusion

In this article second language vocabulary research published in leading international research journals in 2004 was reviewed. In addition to the articles examined in the preceding sections, the following papers were also published in 2004: for example, Barcroft (2004a),[28] Barcroft (2004b),[29] Berkel (2004),[30] Biber, Conrad, and Cortes (2004),[31] Chung and Nation (2004),[32] Collentine (2004),[33] Cotterill (2004),[34] Durán, Malvern, Richards, and Chipere (2004),[35] Gardner (2004),[36] Grant and Bauer (2004),[37] Laufer, Elder, Hill, and Congdon (2004),[38] Meara (2004),[39] Morris and Cobb (2004),[40] Murphy and Kandil (2004),[41] Nassaji (2004),[42] Pulido (2004),[43] Qian and Schedl (2004),[44] Rott (2004),[45] and Smith (2004).[46]

This is the fifth attempt to tackle the task of reviewing second language vocabulary research (Tanaka, 2008, 2009, 2010, 2011). Specifically, Tanaka (2008) examined articles published in 2006, Tanaka (2009) focused on articles published in 2007, Tanaka (2010) dealt with articles published in 2008, and Tanaka (2011) investigated articles published in 2005. I hope that the present review, together with the above four (i.e., Tanaka, 2008, 2009, 2010, 2011), will be of help to those involved in second language vocabulary research.

Notes

[1] It should be noted that this study (i.e., Laufer & Goldstein, 2004) is a follow-up to Laufer, Elder, Hill, and Congdon's (2004) article.

[2] To validate their hypothesized hierarchy of vocabulary learning difficulty, Laufer and Goldstein (2004) employed a paper-and-pencil format. For the use of this format, see pages 411-

412 of Laufer and Goldstein's (2004) article.

[3] Laufer and Goldstein (2004) also use the term *vocabulary strength* on page 410 of their article (i.e., Laufer & Goldstein, 2004).

[4] It should be noted that in the original source, the first letter of this word (i.e., if) was "I" (uppercase letter). In accordance with section 3.37 of "*Publication Manual of the American Psychological Association,* 2001, p. 119," the first letter of the first word was changed from "I" (uppercase letter) to "i" (lowercase letter).

[5] For the learning environment (i.e., ESL or EFL), see "lines 24-25 of page 423" and "lines 10-11 of page 424" of Laufer and Goldstein's (2004) article.

[6] Regarding this (i.e., the hierarchy of four degrees of strength), Table 7 of Laufer and Goldstein's (2004, p. 418) article is informative.

[7] In the present review, there are several places in which sentences beginning with such words as (a), (b), and (c) are enumerated after a colon. It should be noted that in such cases, sentence-initial words begin with an uppercase letter (even if they are preceded by the conjunction *and*).

[8] This use of ' is not a typographical error.

[9] "Of the 30 advanced learners, 10 took the multiple-choice test, 10 the translation test, and 10 the recall test" (Liao & Fukuya, 2004, pp. 202-203).

[10] See Note 8.

[11] Specifically, at the time of the data collection, of the 10 graduate students, "7 of them had been in the United States for about 5 months, and 3 for about 9 months" (Liao & Fukuya, 2004, p. 202).

[12] "Of the 40 intermediate learners, 15 took the multiple-choice test, 15 the translation test, and 10 the recall test" (Liao & Fukuya, 2004, p. 203).

[13] It should be noted that on page 222 of Liao and Fukuya's (2004) article, *review* is shown as the one-word equivalent for *brush up on* but that on page 223, a different word *improve* is written as the equivalent.

[14] It should be noted that in the original source, the first letter of this word (i.e., although) was "A" (uppercase letter). In accordance with section 3.37 of "*Publication Manual of the American Psychological Association,* 2001, p. 119," the first letter of the first word was changed from "A" (uppercase letter) to "a" (lowercase letter).

[15] Coxhead's (2000) Academic Word List was created "from a corpus of 3,600,000 running words, called the Academic Corpus, from 28 subject areas equally distributed over the four divisions of Commerce, Law, Science, and Arts" (Wang Ming-Tzu & Nation, 2004, p. 292).

[16] Coxhead's criteria for inclusion in the Academic Word List were that a word family had to (a) occur at least 100 times in the Academic Corpus, (b) occur in all four divisions of Commerce, Law, Science, and Arts with a frequency of at least 10 in each division, and (c) occur in at least 15 of the 28 subject areas (Wang Ming-Tzu & Nation, 2004, p. 292).

[17] See Note 7.

[18] Table 6 of Wang Ming-Tzu and Nation's (2004, pp. 307-308) article shows distribution of primary and other meanings of these 21 word families. Specifically, the 21 items are the following (pp. 307-308): *consist, issue, credit, normal, correspond, volume, attribute, project, decline, generation, objective, abstract, brief, intelligent, appreciate, offset, tense, induce, accommodate, converse,* and *panel.*

[19] For the six additional word families, see Wang Ming-Tzu and Nation (2004, p. 305). Specifically, they are the following (p. 305): *consist, issue, volume, attribute, objective,* and *abstract.*

[20] It should be noted that in the original source, the first letter of this word (i.e., the) was "T" (uppercase letter). In accordance with section 3.37 of "*Publication Manual of the American Psychological Association,* 2001, p. 119," the first letter of the first word was changed from "T" (uppercase letter) to "t" (lowercase letter).

[21] It should be noted that in the original source, the first letter of this word (i.e., nonverbal) was "N" (uppercase letter). In accordance with section 3.37 of "*Publication Manual of the American Psychological Association,* 2001, p. 119," the first letter of the first word was changed from "N" (uppercase letter) to "n" (lowercase letter).

[22] This is not a typographical error. In the present review, "et al." is not italicized (*Publication Manual of the American Psychological Association,* 2001, p. 102).

[23] It should be noted that the term *residual* was used in Snellings, van Gelderen, and de Glopper's (2004) article to refer to a variable from which the effect of typing skill differences had been eliminated.

[24] This word (i.e., inferencing) is one of the technical words used in second language vocabulary research. In other words, this is not a typographical error. Lexical inferencing means guessing the meaning of an unfamiliar word.

[25] The CanTEST is "a standard test used to assess academic English language proficiency. It was developed and validated at the Second Language Institute, University of Ottawa, and is currently used by a number of institutions in Canada and worldwide for student admission purposes" (Bengeleil & Paribakht, 2004, p. 242).

[26] It should be noted that *the* was not used before this phrase in accordance with the way

Jiang (2004) used this term.

[27] See Note 26.

[28] Does writing new words in sentences affect second language vocabulary learning? This is the main research question addressed in Barcroft's (2004a) study.

[29] Barcroft's (2004b) article considers the importance of vocabulary in second language acquisition (p. 201), presents an overview of major strands of research on second language vocabulary acquisition (pp. 201-203), and discusses five principles for effective second language vocabulary instruction (pp. 203-206).

[30] Berkel's (2004) article pertains to the acquisition of English spellings by Dutch learners of English. Specifically, this study addressed the following two research questions: How do Dutch learners of English manage to acquire the English spelling system? What spelling knowledge do they actually acquire?

[31] Biber, Conrad, and Cortes (2004) investigated the use of multi-word sequences in university classroom teaching and textbooks.

Biber et al. (2004) took a frequency-driven approach to the identification of multi-word sequences referred to as "lexical bundles." These three researchers (i.e., Biber et al., 2004) compared the lexical bundles in university classroom teaching and textbooks with those found in Biber, Johansson, Leech, Conrad, and Finegan's (1999, pp. 990-1024) study on conversation and academic prose. (NB: It should be noted that in the present review, "et al." is not italicized [*Publication Manual of the American Psychological Association,* 2001, p. 102].)

[32] Chung and Nation's (2004) article pertains to identifying technical vocabulary. Specifically, these two researchers compared four different approaches to identifying technical vocabulary (i.e., [a] using a rating scale [pp. 252-253], [b] using a technical dictionary [pp. 253-256], [c] using clues provided in the text [pp. 256-257], and [d] using a computer-based approach [pp. 257-260]).

[33] Collentine (2004) investigated the effects of learning contexts on morphosyntactic and lexical development. Specifically, he compared the effects of (a) learning Spanish as a second language in a study abroad program in Alicante, Spain (n = 26) versus (b) learning it as a foreign language in a formal university classroom at the University of Colorado at Boulder (n = 20).

[34] "Much of the existing literature on the language of the courtroom . . . has focused on," according to Cotterill (2004), "the study of lawyers' use of interrogative forms such as closed yes/no and tag questions as a means by which witness testimony is controlled. In contrast, relatively little analytical attention has been devoted to *lexical* aspects of direct and cross-examination and the role of lexis in creating these subtle nuances of meaning for the jury"

(p. 518).

Drawing on a 5-million word corpus of rape/sexual assault and domestic violence trials held in the late 1990s in the UK, Cotterill's (2004) article examined lawyers' control of witness testimony through lexical negotiation.

The researcher (i.e., Cotterill) used "a combination of corpus linguistics and discourse analysis to study the use of strategic lexical choices in the verbal construction and representation of the crime, its perpetrator and its victim in a variety of rape/sexual assault and domestic violence cases" (p. 520).

[35] Durán, Malvern, Richards, and Chipere (2004) discussed issues in measuring lexical diversity and outlined an approach based on mathematical modelling that produced a measure, *D*.

What are typical values and ranges of *D* in children acquiring their first language between 18 months and 5 years, and how do these compare with values obtained from other corpora? Does *D* show a consistent and significant developmental trend across this age range? How well does *D* correlate with other measures of language production and comprehension at each age? To what extent does *D* vary according to how word types are defined? These questions were addressed in Durán, Malvern, Richards, and Chipere's (2004, p. 226) article.

[36] Gardner's (2004) study analyzed "the lexical differences between narrative and expository reading materials used in upper-elementary education (10- and 11-year-old children)" (p. 1) and explored "how these differences could affect children's potential vocabulary acquisition through reading" (p. 1).

[37] Grant and Bauer (2004) aimed to lay down new criteria for defining idioms.

[38] Laufer, Elder, Hill, and Congdon (2004) described the process of developing and validating a vocabulary test designed to measure vocabulary size and strength. The test is called Computer Adaptive Test of Size and Strength (CATSS). For CATSS, see also Laufer and Goldstein (2004).

[39] Meara (2004) simulated vocabulary attrition using a random autonomous Boolean network model. Specifically, he worked with "lexicons of 2,500 words——a figure which is small enough to be manageable, but large enough to be plausible in terms of what we know about actual lexicons" (p. 139).

In the concluding section of his article, Meara (2004) mentions that the models described in his paper "are not intended as theories of attrition which can be supported . . . by data from human experiments. Rather, they offer a way of exploring what the important features of lexicons might be, and how these features interact with each other in the process of attrition" (p.

154). Furthermore, with regard to the significance of “formalizing our assumptions by means of simulations” (p. 155), Meara (2004) states as follows:

> I hope I have shown that working with models . . . forces us to be much more explicit about our assumptions, and that interesting research questions can grow out of this process. In other modes of research, we can often take these assumptions for granted, and fail to appreciate their wider implications. The assumption highlighted in this paper is that lexicons are above all *networks* of words, and that in order to really understand how lexicons work, we need to understand how the processes we are interested in operate in a network, rather than at the level of individual words. (p. 154)

[40] Morris and Cobb (2004) examined whether vocabulary profiles could serve as predictors of academic performance in undergraduate Teaching English as a Second Language (TESL) programs.

The participants of this study were 122 Canadian TESL trainees registered in either a 4-year Bachelor of Education or a 1-year Certificate program in a Quebec university (p. 80).

[41] Murphy and Kandil (2004) investigated word-level stress patterns of the words contained in the Academic Word List (AWL; Coxhead, 2000).

Specifically, these two researchers examined the AWL’s 525 polysyllabic headwords and 2,454 polysyllabic sublist items, or 2,979 polysyllabic academic words in total. It should be noted that they excluded 64 monosyllabic words contained in the AWL inventories because their focus was on word-level stress patterns of polysyllabic words (p. 65).

[42] Nassaji (2004) investigated the relationship among ESL learners’ depth of vocabulary knowledge, their lexical inferencing strategy use, and their success in deriving the meaning of an unknown word from context.

The participants of this study were 21 adult intermediate ESL learners from different language backgrounds, including Chinese, Spanish, Persian, Portuguese, and Arabic (p. 113).

[43] Pulido (2004) investigated the relationship between second language (L2) text comprehension and L2 incidental vocabulary acquisition. The effect of topic familiarity on the above relationship was also examined.

This study comprised 99 adult learners of Spanish as an L2; all of them were native speakers of English. These learners were recruited from the following three university course levels: beginning (2nd-semester elementary language), n = 43; intermediate (5th-semester composition), n = 39; and advanced (8th-semester literature), n = 17 (p. 480).

[44] Qian and Schedl (2004) evaluated an in-depth vocabulary knowledge measure for assessing reading performance; the measure is called the Depth of Vocabulary Knowledge

measure or the Depth of Vocabulary Knowledge test. For the details of this test, see Qian and Schedl (2004, pp. 37-38, pp. 50-52).

The above test (i.e., the Depth of Vocabulary Knowledge test) was developed by the two researchers (i.e., Qian and Schedl) and a team of TOEFL test-developers at the Educational Testing Service; in the test development, they referred to the format of Read's (1993, 1998) word associates test (Qian & Schedl, 2004, p. 37).

[45] Rott (2004) sought to develop a better understanding of the effects of output tasks on second language vocabulary acquisition and text comprehension. Specifically, she compared the effects of (a) a cued-output task, (b) a self-selected-output task, and (c) a reading-only control task (i.e., un-enhanced control reading).

[46] Smith (2004) was interested in the relationship between computer-mediated negotiated interaction and second language lexical acquisition, and reported on a study conducted with 24 intermediate learners of English as a second language (14 females and 10 males) taken from an intensive English language program at a large Midwestern university (p. 373).

References

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Atkinson, J. M., & Heritage, J. (Eds.). (1984). *Structures of social action: Studies in conversation analysis.* Cambridge, England: Cambridge University Press.

Barcroft, J. (2004a). Effects of sentence writing in second language lexical acquisition. *Second Language Research, 20,* 303-334.

Barcroft, J. (2004b). Second language vocabulary acquisition: A lexical input processing approach. *Foreign Language Annals, 37,* 200-208.

Bengeleil, N. F., & Paribakht, T. S. (2004). L2 reading proficiency and lexical inferencing by university EFL learners. *The Canadian Modern Language Review, 61,* 225-249.

Berkel, A. V. (2004). Learning to spell in English as a second language. *IRAL, 42,* 239-257.

Biber, D., Conrad, S., & Cortes, V. (2004). *If you look at . . .* : Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25,* 371-405. (NB: It should be noted that the title of this article [i.e., *If you look at . . .*] should be italicized.)

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Essex, England: Pearson Education.

Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System, 32,* 251-263.

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26,* 227-248.

Cotterill, J. (2004). Collocation, connotation, and courtroom semantics: Lawyers' control of witness testimony through lexical negotiation. *Applied Linguistics, 25,* 513-537.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34,* 213-238.

Durán, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics, 25,* 220-242.

Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found in children's narrative and expository reading materials. *Applied Linguistics, 25,* 1-37.

Grant, L., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics, 25,* 38-61.

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 24,* 617-637.

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal, 88,* 416-432.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing, 21,* 202-226.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54,* 399-436.

Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests.* Cambridge, England: Cambridge University Press.

Lazaraton, A. (2004). Gesture and speech in the vocabulary explanations of one ESL teacher: A microanalytic inquiry. *Language Learning, 54,* 79-117.

Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning, 54,* 193-226.

Markee, N. (2000). *Conversation analysis.* Mahwah, NJ: Erlbaum.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

Meara, P. (2004). Modelling vocabulary loss. *Applied Linguistics, 25,* 137-155.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan.

Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System, 32,* 75-87.

Murphy, J., & Kandil, M. (2004). Word-level stress patterns in the Academic Word List. *System, 32,* 61-74.

Nassaji, H. (2004). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Canadian Modern Language Review, 61,* 107-134. (NB: This article was reprinted in "*The Modern Language Journal,* Vol. 90, No. 3, pp. 387-401, [2006]" as an exchange article between *The Modern Language Journal* and *The Canadian Modern Language Review.*)

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning, 54,* 469-523. (NB: This article was reprinted in "*Language Learning,* Vol. 57, Supplement 1, pp. 155-199, [2007].")

Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21,* 28-52.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10,* 355-371.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. J. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th Language Testing Research Colloquium, Long Beach* (pp. 41-60). Mahwah, NJ: Erlbaum.

Riggenbach, H. (1999). *Discourse analysis in the language classroom: Vol. 1. The spoken language.* Ann Arbor, MI: University of Michigan Press.

Rott, S. (2004). A comparison of output interventions and un-enhanced reading conditions on vocabulary acquisition and text comprehension. *The Canadian Modern Language Review, 61,* 169-202.

Smith, B. (2004). Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition, 26,* 365-398.

Snellings, P., van Gelderen, A., & de Glopper, K. (2004). Validating a test of second language written lexical retrieval: A new measure of fluency in written language production. *Language Testing, 21,* 174-201.

Tanaka, Y. (2008). Second language vocabulary research: 2006. *Administration, 15*(1 & 2), 77-103.

Tanaka, Y. (2009). Second language vocabulary research: 2007. *Administration, 16*(1), 55-72.

Tanaka, Y. (2010). Second language vocabulary research: 2008. *Administration, 17*(1 & 2), 37-57.

Tanaka, Y. (2011). Second language vocabulary research: 2005. *Administration, 17*(3 & 4), 61-

81.

Wang Ming-Tzu, K., & Nation, P. (2004). Word meaning in academic English: Homography in the Academic Word List. *Applied Linguistics, 25,* 291-314.