

# Second Language Vocabulary Teaching and Learning: A Commentary on Four Studies for the JALT Vocabulary SIG

David Beglar  
*Temple University, Japan Campus*

## Abstract

Four papers by Brandon Kramer, Tohru Matsuo, Aaron Sponseller, Young Ae Kim, Suzuka Nishiyama, and Stuart McLean; Stuart Benson and Naheen Madarbakus-Ring; Michael McGuire and Jennifer Larson-Hall; and Atsushi Mizumoto, Geoffrey Pinchbeck, and Stuart McLean were presented in the morning session of the 2021 Vocabulary SIG Symposium held in Kyoto, Japan, on 04 December 2021. As discussant, it was my pleasure to comment on each of the papers. The four studies were investigations of the attrition of productive vocabulary knowledge; the lexical composition of two English for Academic Purposes (EAP) textbooks; the relationship between high-frequency multi-word sequences, speech rate, and listening perception; and the use of various word lists with the New Word Level Checker. After reviewing and summarising each study in this paper, I present suggestions for further developing the studies and expanding the research agenda of each study.

**Keywords:** vocabulary, attrition, productive knowledge, lexical coverage, multi-word sequences, speech rate, oral fluency, listening perception, word list

## 1 Introduction

Second language vocabulary acquisition (SLVA) has received a tremendous amount of attention from researchers in the past three decades for good reason. Vocabulary is widely seen as the heart of multiple models of first (L1) and second language (L2) acquisition. For instance, Cutler and Clifton (2000) placed lexis at the heart of their first language listening model, while Rost (2016), a leading L2 listening authority, stated, “Recognizing words in fluent speech is the basis of spoken language comprehension...” (p. 28). Levelt and his colleagues (Levelt, 1989; Levelt et al., 1999) proposed a modular L1 speech model in which lexis occupies a crucial position because lexical items govern syntactic processing. Lexis also plays the key role in bilingual speech production models (e.g., see Chapter 4 in Kormos, 2006). In addition, word knowledge plays the central role in theories of L1 reading comprehension, such as the reading systems framework (Perfetti & Stafura, 2014). This line of thinking has been echoed by second language reading specialists such as Grabe and Stoller (2020), who stated, “The most fundamental requirement for fluent reading comprehension is rapid and automatic word recognition” (p. 16).

In this paper, I review and comment on four papers: Measuring Attrition of L2 Productive Vocabulary Knowledge Over the Summer Vacation (Kramer, Matsuo, Sponseller, Kim, Nishiyama, & McLean), A Comparison of Textbook Vocabulary Load Analysis (Benson & Madarbakus-Ring), The Contribution of High-Frequency Multi-Word Sequences to Speech Rate and Listening Perception (McGuire & Larson-Hall), and Comparisons of Word Lists on New Word Level Checker (Mizumoto, Pinchbeck, & McLean).

## 2 The Four Studies

### **2.1 Measuring Attrition of L2 Productive Vocabulary Knowledge Over the Summer Vacation (Kramer, Matsuo, Sponseller, Kim, Nishiyama, & McLean)**

Relatively few second language vocabulary researchers have conducted studies of forgetting, and even studies of language attrition in the broader area of second language acquisition are rare; thus, little is known about this issue. For this reason, the study by Kramer et al. is a welcome addition to the SLVA literature. The primary purpose was to investigate the attrition of productive lexical knowledge over summer vacation by 81 first-year Japanese university students. The learners had taken part in a vocabulary instructional programme that included the use of a digital paired-associate learning application designed to develop both receptive and productive lexical knowledge and had also engaged in an extensive reading programme. The learners studied high-frequency vocabulary from the New General Service List (NGSL) (Browne et al., 2013) that covered the 501–1,507 word-frequency range and were asked to read 180,000 tokens of graded readers during the academic year. The participants' productive vocabulary knowledge of 60 randomly selected words from the targeted frequency range was assessed before and after summer break with an online vocabulary testing site, VocabLevelTest.org (McLean & Raine, 2019). The researchers' focus on productive vocabulary set a high bar, as learning productive vocabulary is more difficult than learning receptive vocabulary (Mondria & Weirisma, 2004, p. 79). A paired-samples *t*-test of the pre-test and post-test results showed no significant change in lexical knowledge ( $p = 0.90$ , Cohen's  $d = 0.01$ ). The researchers also conducted a multiple regression to investigate determinants of the changes in the productive test scores. Only vocabulary review with the digital application, which accounted for 4.90% of the variance, was a significant predictor. Extensive reading and overseas experience did not make significant contributions to the regression model.

Kramer et al. also used a more detailed approach to investigating possible attrition by using a five-part system for categorising lexical knowledge that ranged from *No demonstrated knowledge of the target word* to *Correct target word with correct spelling and derivational form*. The authors reported that 84.60% of the 12,115 total words tested did not change from pre-test to post-test, while the greatest amount of attrition at any level of the five-part system was 0.40%; thus, attrition was negligible. Taken together, the results indicated that the participants' productive vocabulary knowledge was generally stable over the summer break. This finding might have occurred for two reasons. First, the majority of the target items should have been known by the learners before entering the university; thus, the

instructional programme might have primarily served to consolidate rather than teach knowledge of the target words. Second, the digital application required productive practice; thus, it involved transfer-appropriate processing, which states that memory retrieval is easier when the cognitive processes engaged in when materials are encoded match those used during encoding (e.g., Graf & Ryan, 1990). In other words, the digital application matched the test format. Although admittedly difficult to determine, an effort needs to be made to operationalise the degree to which the learners used the receptive and productive parts of the digital application. This information could shed light on the effectiveness of the application and thereby provide usage guidelines for teachers and learners.

In addition, the fact that extensive reading was not a significant predictor in the regression model was reasonable given that the participants read little ( $M = 14,813$  words read) during summer vacation and it only provided receptive encounters with vocabulary. Even if overseas experience had been a significant predictor, it would have been uninterpretable without knowledge of the interactions the learners had while overseas. For instance, it is important to know whether the participants visited an English-speaking country, who they lived with (e.g., another Japanese student or an English-speaking family), how much English they were exposed to, and how much English they produced.

As the authors acknowledged, one problem with the study was the use of different words on the pre-test and post-test. This approach made direct comparisons impossible; instead, the researchers were forced to assume that the words on the two tests were approximately equal in difficulty. However, there was no indication that the words on the tests were controlled for issues such as frequency, concreteness, or part of speech. One way to make two or more tests directly comparable is to use Rasch item anchoring, as this technique allows for direct comparisons in person ability when a number of common items are included on each test form (see Chapter 5 in Bond et al., [2021] for further information).

This study raises interesting questions which, to the best of my knowledge, have not been investigated, and as is usual in many types of research, the most interesting questions revolve around the reasons why a phenomenon does or does not occur. For instance, does the degree to which a lexical item has been automatised influence its stability in long-term memory? The degree of automatisation can be assessed using reaction time measures, which can assess aspects of words such as orthographic form, phonological form, and semantic meaning (e.g., see Matsuo [2017] and Shimono [2019] for studies that included these types of reaction time measures). If automaticity is a determinant of stability in long-term memory, then it would be a candidate for becoming an instructional objective in foreign language curricula.

A second possibility concerns depth of lexical knowledge. If depth of knowledge is defined as knowledge of collocations and associations, then it means that a target word has been placed in a lexical network in the learners' mental lexicon. It is plausible that the development of this network consolidates knowledge of target words and provides multiple mental pathways to access them. A somewhat different version of this idea was proposed by Meara (2009, see Chapters 3 and 4), who suggested that receptive lexical knowledge develops

into more cognitively demanding productive knowledge as words are embedded into a lexical network. If it is shown that embedding target vocabulary in such a network improves long-term retention, then it would also be a candidate as an instructional objective.

Future researchers can also investigate part of speech and attrition. For instance, whereas adjective and adverbs generally have one form and nouns can only add two inflectional morphemes, plural *-s* and possessive *'s*, verbs are highly complex. Verbs can take four inflectional morphemes, irregular forms are commonplace and are formed in various ways, verb tenses combine to form a complex tense-aspect system, and many verbs convey abstract meanings, which raises yet another line of research: The possibility that attrition is related to the degree of concreteness of a word's meaning. This phenomenon, which is known as the concreteness effect (e.g., Pexman et al., 2007), states that concrete words are better remembered (Schwanenflugel et al., 1992), better recognised (Fliessbach et al., 2006), read faster (Schwanenflugel & Shoben, 1983), and acquired more quickly than more abstract words (Mestres-Missé et al., 2014).

## **2.2 A Comparison of Textbook Vocabulary Load Analysis (Benson & Madarbakus-Ring)**

Benson and Madarbakus-Ring conducted an investigation of the vocabulary load of an in-house textbook and a commercially produced textbook to determine the appropriateness of the two textbooks for first-year Japanese university students. The textbooks were analysed with Range (Heatley et al., 2002) and the New Word Level Checker (Mizumoto, 2012). Nation's British National Corpus/Corpus of Contemporary American English (BNC/COCA) base word lists, which were accompanied by his supplementary lists, were used as a general word list, and the New JACET8000 word list (JACET, 2016) was used as a context-specific word list, as it was created for tertiary-level Japanese learners of English. Using the conventional targets of 95% and 98% lexical coverage of a written text (Nation, 2013, pp. 206–208), the authors first analysed the inhouse textbook and reported that knowledge of the first 3,000 words on the BNC/COCA list and the supplementary list would enable learners to know 95% of the tokens in the textbook, while the first 5,000 words and the supplementary list were needed to meet the 98% criterion. The analysis of the commercial textbook produced similar results: Knowledge of the first 3,000 words was needed to meet the 95% criterion and knowledge of the first 4,000 words was necessary to meet the 98% criterion.

When the in-house and commercial textbooks were analysed with the New JACET8000 word list, they yielded the same results: The first 3,000 words provided 95% coverage, while knowledge of 8,000 words was needed to meet the 98% criterion.

One way to interpret the above results is to use the 3,715-word figure for written receptive vocabulary knowledge of Japanese university students provided by McLean et al. (2014). When matched with the results of this paper, average students would know slightly more than 95% of the tokens in both the in-house

and commercial textbooks, a figure I would argue is sufficient if the students are engaged in intensive reading. My position raises a question that has not been sufficiently addressed in the SLVA literature: Should the 95% and 98% criteria be applied equally to all types of reading, or should SLVA researchers distinguish different purposes for reading as reading specialists have done for decades? It is possible that different degrees of lexical coverage work effectively for different reading purposes. It is also important to acknowledge that not all unknown words are equal, as one might occur once in a text, while another occurs multiple times. Moreover, the helpfulness of the contexts in which they are embedded can vary greatly (see Nation [2013], pp. 363–364 for a list of 14 variables that influence an unknown word's guessability in a reading text).

Extensive reading and intensive reading usually differ in a number of important ways. Students typically engage in extensive reading without the benefit of pre-reading activities and with little or no teacher support. In such cases, the 98% criterion is defensible. However, intensive reading often involves pre-reading tasks, including vocabulary activities; the texts are far shorter than extensive reading texts; glossing can clarify the meaning of low-frequency or technical vocabulary (Nation, 2013, pp. 238–247), learners can use a dictionary (Nation, 2013, pp. 414–436), and learners have some ability to guess unknown words from context (Nation, 2013, pp. 348–388). In addition, it is generally agreed that some unknown vocabulary should be included in intensive reading texts given that without such lexical challenges, vocabulary growth would be unnecessarily limited. What this means where lexical coverage is concerned is unknown and therefore needs to be investigated, but there is a possibility that even 95% coverage is not required for intensive reading. Finally, vocabulary coverage figures might not provide an accurate picture of text comprehension because of what Hagoort (2005) called unification. During unification, word-level semantic and syntactic information is integrated into the developing context. Learners struggling with unification could have comprehension difficulties even when they know 100% of the tokens in a reading text because they fail to understand how individual tokens fit together to create meaning. Empirical research can shed light on this issue.

This study raises other interesting issues where the 95% and 98% criteria are concerned. It is possible that the two figures are interpreted too inflexibly in the field. For example, in Table 5 in Benson and Madarbakus-Ring's paper, knowledge of 8,000 words is needed to achieve 98% lexical coverage of the commercial textbook using the New JACET8000 list. However, knowledge of 4,000, 5,000, 6,000, and 7,000 words produced coverage figures of 97.01%, 97.40%, 97.65%, and 97.92%, respectively. Thus, knowledge of 4,000 words provides 97.01% coverage and knowledge of 8,000 words provides 98.15% coverage. This result raises two points. First, there is a massive difference between knowing 4,000 and 8,000 words in a foreign language. The McLean et al. (2014) study included 3,427 participants, yet the maximum vocabulary size estimate was 7,400 words; thus, while none of their participants appeared to have an 8,000-word receptive vocabulary, the mean of 3,715 was close to knowledge of 4,000 words. In contrast, the coverage difference between knowledge of 4,000 and 8,000 words in Table 5 was 1.14% ( $98.15 - 97.01 = 1.14$ ). This small difference raises an important question: Can the inflexible

use of the 95% and 98% coverage criteria lead researchers and classroom teachers astray in some instances?

Researchers can answer this question by investigating small differences in lexical coverage with learners at different proficiency levels using texts written at different levels of difficulty (e.g., general texts vs. academic texts) and that are used for different purposes (e.g., intensive reading, narrow reading, timed reading, and extensive reading).

Another issue to consider is that both word lists used in this study divide words into 1,000-word frequency bands. Although this division is conventional, it is imprecise and therefore of limited usefulness. If for instance, the texts were analysed using 500- or 100-word categories, it might be found that knowledge of 2,500 words rather than 3,000 words provides 98% coverage of a text. This 500-word difference is large in terms of percentage difference (i.e., 18%), and according to the result obtained from McLean et al. (2014) study, the Japanese learners acquired an average of 619 words per year ( $3,715 \text{ words} \div 6 \text{ years of study} = 619$ ) if they were first-year university students and less if they were in higher grades; thus, 500 words represents nearly 1 year of study. More detailed vocabulary coverage analyses are possible with *lexutor.com* and the Scale of Word Knowledge–Japanese (SEWK-J) list (Stuart McLean, personal communication, November 29, 2021).

Another concern with this type of analysis is that it assumes that knowledge of single words is an adequate predictor of text comprehension. There is an abundance of evidence that reading comprehension is based on multiple factors including knowledge of collocations (Matsuo, 2017), multi-word expressions (e.g., Siyanova-Chanturia & Van Lancker Sidtis, 2018), syntactic parsing (Jeon & Yamashita, 2014), background knowledge (Kintsch, 2012), and inferencing skills (van den Broek et al., 2015). Vocabulary researchers too rarely acknowledge that skilled reading is based on a number of distinct component skills, of which lexical knowledge is just one.

### **2.3 The Contribution of High-Frequency Multi-Word Sequences to Speech Rate and Listening Perception (McGuire & Larson-Hall)**

Relatively few studies have been focused on teaching vocabulary for productive use, so the study by McGuire and Larson-Hall addresses an important gap in the field. The researchers used Anki, a digital vocabulary learning application, to teach two types of multi-word sequences — bigrams such as *high probability* and trigrams such as *and then I*. They then investigated the degree to which knowledge of the multi-word sequences affected speech rate, one aspect of oral fluency, and performance on a listening dictation task. All 33 participants studied 10 collocations each week using the Anki app and described or listened to classmates describe three picture stories that incorporated the taught collocations. The experimental group ( $n = 16$ ) also practiced 10 of the multi-word sequences each week using Anki and watched a 5-minute introduction about them. The participants described three multi-panel pictures and talked about their ideal day on the speaking pre-test and post-test. Audio recordings were analysed with PRAAT (Boersma & Weenink, 2021), a computer program for analyzing, synthesizing, and

Table 1. Cohen's *d* Results for the Storytelling Task, Free Speech Task, and Dictation Task

	Pre-test	Post-test	Cohen's <i>d</i>
<b>Storytelling task</b>			
Control group	1.66	1.41	0.85
Experimental group	1.67	1.82	0.49*
<b>Free speech task</b>			
Control group	1.93	1.87	-0.18
Experimental group	1.84	2.05	0.50*
<b>Dictation task (total errors)</b>			
Control group	71.88	56.41	1.66
Experimental group	65.62	47.12	1.55
<b>Dictation task (MWS errors)</b>			
Control group	25.00	17.59	1.44
Experimental group	26.62	15.12	1.54

Note. CI = Confidence interval; \* = The 95% CI passed through zero; MWS = Multi-word sequence; Plonsky and Oswald (2014) recommend using the following values for interpreting Cohen's *d* for mean differences between groups: 0.40 = small, 0.70 = medium, 1.00 = large.

manipulating recorded speech samples, to ascertain speech rate, which was operationalised as the number of syllables produced divided by the duration of the speech sample. In the listening dictation task, the participants transcribed a naturalistic conversation containing 27 of the multi-word sequences while listening twice at normal speed. The pre-test and post-test results are shown in Table 1. As shown in the Table, the experimental group improved to a small degree on the speaking tests, as indicated by the Cohen's *d* effect sizes, while both groups made large improvements on the dictation tests. These findings indicated that receptive knowledge of the form of multi-word sequences was acquired better than productive knowledge of the sequences.

The overall results were promising, as they showed that L2 learners are capable of acquiring multi-word sequences for both receptive and productive use. Larger *n*-sizes would have produced more robust results and likely solved the statistical issues that arose with the 95% confidence intervals passing through zero for the experimental group. However, the authors are to be commended for using 95% confidence intervals—a practice that is all too rare in the field of SLVA—as they show the degree to which variables are measured precisely. In addition, a second experimental group would add more nuance to the study and make the results more informative. For instance, in the study, the participants were asked to engage in homework in which they counted the number of times they heard the reduced multi-word sequences. However, this type of focus on linguistic form can negate meaningful processing (VanPatten, 2020). Therefore, including an additional experimental group that receives input in which the focus is on meaning rather than form might provide more nuanced results.

An additional possibility is to assess oral fluency in a more detailed way. Oral fluency is often divided into three categories: breakdown fluency, speed fluency, and repair fluency. Breakdown fluency is assessed using silence and pausing

behaviour, and can be operationalised using silent pauses at AS-boundaries (AS = Analysis of speech unit; "...an independent clause or subclausal unit, together with any subordinate clause(s) associated with it" Foster et al., [2000], p. 365), silent pauses mid-clause, filled pauses at AS-boundaries, and filled pauses mid-clause (Pang & Skehan, 2021). Speed fluency can be assessed by calculating pruned (i.e., repetitions, restarts, and self-repairs are deleted) and unpruned speech rates. Moreover, speech rate can be divided into sub-components such as lexical retrieval speed, speed of articulation, and sentence building speed (De Jong et al., 2012). However, speech rate must be interpreted carefully, given that some native speakers of a language speak relatively slowly. This issue can be addressed by measuring research participants' L1 speech rates as suggested by Segalowitz (2010, p. 2). Repair fluency can be assessed using false starts/100 words, self-repairs/100 words, and repetitions/100 words. A final possibility is to gather qualitative judgements of oral fluency using human raters. In this way, it is possible to determine whether changes in oral fluency can be detected by the human ear. Human rater data can be profitably analysed using software such as FACETS (Linacre, 2021), which provides interval-level person ability estimates, fit statistics for all facets in the model (e.g., participants, raters, tasks, and evaluation criteria), and bias analyses.

In addition to a more detailed account of changes in oral fluency, when teaching vocabulary for productive purposes, it is useful to investigate changes in lexical complexity, which can be operationalised as lexical frequency (i.e., the number of tokens in particular word-frequency bands), lexical diversity (i.e., the ratio of the number of unique word types to the total number of tokens), and lexical density (i.e., the proportion of content words to the total number of tokens) (Révész et al., 2016). Lexical complexity is one form of linguistic development that would be reasonably expected to occur in a study such as this one (Baese-Berk et al., 2021).

A final possibility is to investigate the contribution of foreign language learning aptitude (e.g., Linck et al., 2013) to lexical acquisition in the context of an oral proficiency study such as this one. Two related constructs that could be investigated are working memory and rote learning ability (i.e., associative memory), as empirical research has suggested that they have direct relationships to vocabulary learning (e.g., Nagata et al., 1999).

## **2.4 Comparisons of Word Lists on New Word Level Checker (Mizumoto, Pinchbeck, & McLean)**

Analysing the lexical composition of listening and reading texts is a useful skill for foreign language instructors, materials developers, and curriculum specialists. For this reason, the paper by Mizumoto, Pinchbeck, and McLean provides useful data for each of the above audiences. The researchers presented an investigation of an online vocabulary profiling application, the New Word Level Checker (<https://nwic.pythonanywhere.com/>), in which they discussed the rationale for developing the application and compared the lexical coverage results of five word lists with two English-language examinations. The five word lists that make up the New Word Level Checker are the SEWK-J, the New JACET8000

Table 2. 95% and 98% Coverage Results for the Five Word Lists

Word list	Center and Kyotsu Tests		TOEIC Test	
	95% coverage	98% coverage	95% coverage	98% coverage
SEWK-J	4,000	—	5,000	10,000
New JACET8000	3,000	8,000	4,000–5,000	6,000
SVL 12000	4,000	9,000	5,000	8,000
NGSL	NGSL, NAWL, TOEIC	—	NGSL, NAWL, TOEIC	—
CEFR-J	B2	—	B2	—

Note. NGSL = New General Service List; NAWL = New Academic Word List; TOEIC = TOEIC (Test of English for International Communication) word list; CEFR-J = Common European Framework of Reference Japan.

(JACET, 2016), the Standard Vocabulary List 12000 (SVL12000), the NGSL (Browne et al., 2013), and CEFR-J (Tono, 2019). The authors described the word counting units and rules applied with each corpus, including how proper nouns and numerals were treated, lemmatisation, words with capital letters, contractions, hyphenated words, compounds, and multi-word units.

In the final part of the paper, the five word lists were used to analyse (a) the 2020 Center Test and 2021 Kyotsu Test as well as (b) a TOEIC test. The authors analysed the tests using the five word lists in order to investigate how much vocabulary knowledge was needed to reach the 95% and 98% text coverage criteria. The results are shown in Table 2. As indicated in the Table, using the SEWK-J, New JACET8000, and SVL 12000 word lists, knowledge of 3,000–4,000 words was needed to achieve the 95% criterion on the Center and Kyotsu examinations, while knowledge of 8,000–9,000 words was required to meet the 98% criterion. Importantly, the 98% criterion could not be determined using the SEWK-J, NGSL, and CEFR-J lists. The TOEIC test results showed that 95% coverage could be achieved with knowledge of 4,000–5,000 words. However, the results for the 98% criterion varied widely with the SEWK-J list indicating that knowledge of 10,000 words was required and the New JACET8000 list indicating that knowledge of 6,000 words was needed.

As noted above, in some instances, three of the lists were unable to determine 98% coverage due to limitations in the lists themselves. In addition, an anomalous result is that the SEWK-J list could not provide 98% coverage of the Center and Kyotsu tests even though it could for the more difficult TOEIC test. Furthermore, the results for the NGSL are not comparable to the other word lists, as it includes specialised lists such as the New Academic Word List and TOEIC list, which are made up of words from a variety of frequency levels.

The paper can be clarified and extended in a number of ways. First, the selection criteria for the five lists should be specified in order to justify why these lists are preferable to other better-known word lists such as the BNC and COCA. Second, as suggested by Laufer (2014), it would be helpful to show the amount of overlap among the five lists at the various word frequency levels, as this would clarify where they do and do not differ, and with the use of a more detailed analysis of individual words, how they differ. Third, the idea that proper nouns “...can be assumed to be

understood by learners” has not been properly researched and should therefore not be assumed, particularly where learners at lower proficiency levels are concerned.

The authors also used a multiple regression to estimate the probability of a word being known on the SEWK-J list by Japanese university students. The regression analysis could be improved in a number of ways, including checking the assumptions of the analysis, justifying the use of one vocabulary test to predict scores on another vocabulary test, using a vocabulary test that requires evidence of semantic knowledge rather than a yes-no test, ensuring that reaction time data are gathered with proper equipment and software, and treating loan word status in a nuanced way. A large-scale study of loanwords currently underway (Edelman, 2021) shows that knowledge of loanwords varies substantially among Japanese university students, and should not be treated in a monolithic way.

The authors cited Nation (personal communication) when stating that tests and lexical profilers should be based on what learners know rather than what they should know. However, justification is needed for this statement, as it cannot be generalised to all situations. Although this proposal might work when selecting a course textbook, there are good reasons for teaching learners words they do not know, but should know—a position the authors do not appear to support—particularly in educational contexts in which the learners have no specific need for the English language. In such cases, using frequency-based lists to determine learning targets often provides the best cost-benefit ratio (Nation, 2013, p. 38).

The researchers also concluded that all five word lists can be used for text analysis because they “...have sufficient coverage of English texts, up to 95%” despite mounting evidence that 98% is the more valid figure where unassisted reading is concerned. In addition, they stated that the CEFR-J list distinguishes part of speech and can be used for assessing learners’ productive vocabulary, “which in turn implies that CEFR-J could be used to measure the depth of vocabulary knowledge.” The reasoning behind this statement needs to be clarified, as the term *depth of vocabulary knowledge* has multiple meanings and it is unclear how a vocabulary list can be used to measure that construct, however it is defined.

Finally, the suggestions made above in connection to the Benson and Madarbakus-Ring paper also apply here: (a) Using 1,000-word frequency bands produces imprecise estimates and should therefore be abandoned, (b) assuming that knowledge of single words is an adequate predictor of text comprehension flies in the face of a large amount of L1 and L2 research, and (c) if learners are engaged in intensive reading, the 95% and 98% vocabulary coverage figures are potentially too strict. As noted above, each of these issues requires further study.

### 3 Conclusion

The four papers reviewed above provide a great deal of food for thought about how to move the field of SLVA forward. I enjoyed reading the papers and found them stimulating in a number of ways, as they touch upon a number of important issues where vocabulary learning and teaching are concerned: the stability of learners’ lexical knowledge, matching learners with textbooks, teaching productive vocabulary, and analysing the lexical composition of listening and reading texts. As researchers

shed further light on the learning processes involved in foreign language vocabulary acquisition, teachers and students will benefit greatly, as vocabulary knowledge is the linguistic foundation of second language acquisition.

## References

- Baese-Berk, M. M., Drake, S., Foster, K., Lee, D.-Y., Staggs, C., & Wright, J. M. (2021). Lexical diversity, lexical sophistication, and predictability for speech in multiple listening conditions. *Frontiers in Psychology, 12*, 1–19. <https://doi.org/10.3389/fpsyg.2021.661415>
- Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer (Version 6.1.44)* [Computer software]. <http://www.praat.org>
- Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). New York, NY: Routledge.
- Browne, C., Culligan, B., & Phillips, J. (2013). *The New General Service List*. <http://www.newgeneralservicelist.org>
- Cutler, A., & Clifton, Jr., C. (2000). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 124–166). Oxford, England: Oxford University Press.
- De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition, 34*, 5–34. <https://doi.org/10.1017/S0272263111000489>
- Edelman, C. (2021). Unpublished raw data. Temple University. *Cross-Linguistic Influences on English Loanword Learnability in the Japanese Context* [Unpublished doctoral dissertation]. Philadelphia, PA: Temple University.
- Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *Neuroimage, 32*, 1413–1421. <https://doi.org/10.1016/j.neuroimage.2006.06.007>
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). A unit for all reasons: The analysis of spoken interaction. *Applied Linguistics, 21*, 354–374. <https://doi.org/10.1093/applin/21.3.354>
- Grabe, W., & Stoller, F. L. (2020). *Teaching and researching reading* (3rd ed.). New York, NY: Routledge.
- Graf, P., & Ryan, L. (1990). Transfer-appropriate processing for implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 978–992. <https://doi.org/10.1037/0278-7393.16.6.978>
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9*, 416–423. <https://doi.org/10.1016/j.tics.2005.07.004>
- Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). *Range* [Computer software]. <https://www.wgtn.ac.nz/lals/paul-nations-resources>
- Japan Association of College English Teachers (JACET). (2016). *The New JACET list of 8000 basic words*. Tokyo, Japan: Kirihara Shoten.

- Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, 64(1), 166–212. <https://doi.org/10.1111/lang.12034>
- Kintsch, W. (2012). Psychological models of reading comprehension and their implications for assessment. In J. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 21–38). Lanham, MD: Rowman & Littlefield.
- Kormos, J. (2006). *Speech production and second language acquisition*. New York, NY: Routledge.
- Laufer, B. (2014). Vocabulary in a second language: Selection, acquisition, and testing: A commentary on four studies for JALT vocabulary SIG. *Vocabulary Learning and Instruction*, 3(2), 38–46. <https://doi.org/10.7820/vli.v03.2.laufer>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Science*, 22, 1–38. <https://doi.org/10.1017/S0140525X99001776>
- Linacre, M. (2021). *FACETS* [Computer software]. [www.winsteps.com](http://www.winsteps.com)
- Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., ... Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level proficiency. *Language Learning*, 63(3), 530–566. <https://doi.org/10.1111/lang.12011>
- Matsuo, T. (2017). *The roles of lexical size, depth, and automaticity of word recognition on reading comprehension*. [Unpublished doctoral dissertation]. Temple University.
- McLean, S., Hogg, N., & Kramer, B. (2014). Estimations of Japanese university learners' English vocabulary sizes using the vocabulary size test. *Vocabulary Learning and Instruction*, 3(2), 47–55. <https://doi.org/10.7820/vli.v03.2.mclean.et.al>
- McLean, S., & Raine, P. (2019). *VocabLevelTest.org* [Web application]. [www.vocableveltest.org](http://www.vocableveltest.org)
- Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam, The Netherlands: Benjamins.
- Mestres-Missé, A., Münte, T. F., & Rodriguez-Fornells, A. (2014). Mapping concrete and abstract meanings to new words using verbal contexts. *Second Language Research*, 30, 191–223. <https://doi.org/10.1177/0267658313512668>
- Mizumoto, A. (2012). *New Word Level Checker* [Web application]. <https://nwlc.pythonanywhere.com>
- Mondria, J.-A., & Weirisma, B. (2004). Receptive, productive, and receptive + productive L2 vocabulary learning: What difference does it make? In P. Boogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 79–100). Amsterdam, The Netherlands: Benjamins.
- Nagata, H., Aline, D., & Ellis, R. (1999). Modified input, language aptitude, and the acquisition of word meanings. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 133–150). Amsterdam, The Netherlands: Benjamins.

- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.
- Pang, F., & Skehan, P. (2021). Performance profiles on second language speaking tasks. *The Modern Language Journal*, 105(1), 371–390. <https://doi.org/10.1111/modl.12699>
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18, 22–37. <https://doi.org/10.1080/10888438.2013.827687>
- Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., & Goodyear, B. G. (2007). Neural correlates of concreteness in semantic categorization. *Journal of Cognitive Neuroscience*, 19, 1407–1419. <https://doi.org/10.1162/jocm.2007.19.8.1407>
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”: Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <http://doi.org/10.1111/lang.12079>
- Révész, A., Ekiert, M., & Torgensen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, 37(6), 828–848. <https://doi.org/10.1093/applin/amu069>
- Rost, M. (2016). *Teaching and researching listening* (3rd ed.). Routledge.
- Schwanenflugel, P. J., Akin, C., & Luh, W. M. (1992). Context availability and the recall of abstract and concrete words. *Memory and Cognition*, 20, 96–104. <https://doi.org/10.3758/bf032.08259>
- Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 82–102. <https://doi.org/10.1037/0278-7393.9.1.82>
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York, NY: Routledge.
- Shimono, T. (2019). *The dynamic cognitive processes of second language reading fluency*. [Unpublished doctoral dissertation]. Temple University.
- Siyanova-Chanturia, A., & Van Lancker Sidsis, D. (2018). What online processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language* (pp. 38–61). New York, NY: Routledge.
- Tono, Y. (2019). *CEFR-J Wordlist version 1.5*. [https://www.cefr-j.org/download.html#cefrj\\_wordlist](https://www.cefr-j.org/download.html#cefrj_wordlist)
- van den Broek, P., Beker, K., & Oudega, M. (2015). Inference generation in text comprehension: Automatic and strategic processes in the construction of a mental representation. In E. O’Brien, A. E. Cook, & R. Lorch (Eds.), *Inferences during reading* (pp. 94–121). Cambridge, England: Cambridge University Press.
- VanPatten, B. (2020). Input processing in adult L2 acquisition. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (pp. 105–127). New York, NY: Routledge.