

Vocabulary Education & Research Bulletin

Summer 2022

Volume 11, Issue 1

This first issue of 2022 contains two very nice articles that we hope will be of interest to our Vocabulary community. Michael and I have been happy to work together on Verb as editors and are always delighted to get good submissions for the journal.

The first, by **John Collins**, touches on the somewhat controversial topic of including an “I don't know” (IDK) option in multiple choice tests, specifically the VST. Collins finds that 30% of his large population of test takers (N=526) did not use the IDK at all, even though they were instructed to use it and probably needed to use it. The short paper is accompanied by three useful graphics and a quick summary of the literature and I'm sure you will find it a strong argument for avoiding the IDK option (which Collins recommends).

The second, by our own **Michael McGuire** (handled by me, Jenifer, of course!) provides a categorized list of the most common trigram multi-word sequences (MWS) that exist in the Open American National Corpus. McGuire created it as a way to help students identify very rapid sequences of words that are currently not studied or found in any teaching materials. Such sequences may be less salient to English language learners but if they can be understood they may improve listening comprehension. Actually, we are currently using the list in our own classrooms. We are gaining more participants for our study that examines whether focusing on these MWSs may help students improve their listening ability. This paper establishes why and how Michael made his list.

We hope to see colleagues at the upcoming Pan-SIG conference in Nagano and please keep sending us your papers!

Michael McGuire & Jenifer Larson-Hall, VERB editors

Table of Contents

Short Articles

John B. Collins	3
<i>Creating One Problem While Addressing Another: 'I Don't Know' Options in the VST</i>	
Michael P. McGuire	9
<i>A Categorized List of Phonologically Reduced High-Frequency Spoken Multi-Word Sequences for Classroom Use</i>	
SIG News	18
VERB Call for Papers	21
VLI Call for Papers	22

Short Paper

Creating One Problem While Addressing Another: ‘I Don’t Know’ Options in the VST

John B. Collins john.buchanan.collins@gmail.com

Background

For teachers and other professionals involved in language program coordination and development, understanding our students’ levels of vocabulary knowledge is vital. As observed by Alderson (2005), “From the point of view of the diagnosis of language strengths and weaknesses, a measure of vocabulary size... [is] of considerable value in its own right, [and]...a quick and reliable placement procedure for more detailed diagnoses of different aspects of language ability” (p. 88). Of the many tools available to measure vocabulary knowledge, one of the most widely used is the Vocabulary Size Test (Nation & Beglar, 2007), which tests passive vocabulary knowledge by way of multiple-choice questions (MCQs), the focus of the current article.

An obvious weakness inherent in the MCQ design is its susceptibility to successful responses due to guessing. As described by Bachman and Palmer (1996), since “the probability of getting the correct answer by chance alone depends on the number of options from which... to choose” (p. 229), it follows that increasing the number of options would reduce the impact of guessing and subsequent score inflation. Creating such a test, or indeed making alterations to existing ones, would likely be prohibitively time consuming for classroom teachers and program coordinators since it would require considerable development, piloting, and refinement – one of the criticisms of MCQs made by Wesche and Paribakt (1996, as cited in Read, 2000). Not addressing this issue, however, would be to ignore the previously identified matter of potential score inflation due to blind guessing (Stewart, 2014, McLean, Kramer & Stewart, 2015).

A more practical and less time-consuming alternative is to allow for an “I don’t know” (IDK) option which, in combination with a discretionary penalty system for incorrect answers, test-takers would select when they would otherwise take a blind guess. Zhang (2013) found that offering a penalty system for incorrect guesses not only discouraged random guessing, resulting in more test-takers using the IDK option, but also resulted in lower overall scores. Likewise, Bennett, and Stoeckel (2012) found that rewording test instructions to indicate a penalty system increased test-takers willingness to use the IDK. Studies involving post-test recall interviews have demonstrated that test-

takers opted for an IDK only when they lacked sufficient knowledge (Lucovich, 2014a, 2014b). Limitations of these studies, however, include the very small data set (one and two subjects respectively) and the test subjects' educational level (both studies involved highly educated university graduates). Offering an IDK option, however, is not without its own problems, most concerning of which arises from the fact that "the tendency to guess...varies greatly from one individual test taker to the next. The 'cautious' type of test taker will typically omit most, if not all, responses to items of which they are unsure, while the 'risk taker' may not omit any responses at all" (Bachman & Palmer, 1996, p. 204). The addition of an IDK option also raises concerns about test validity due to the introduction of a secondary construct, namely willingness to use the IDK. In their study involving computer-generated data simulating a group of 1,000 test-takers, Stoeckel, Bennett, and McLean (2016) demonstrated that correlations between actual ability and VST scores improved when 93% of test-takers used the IDK. However, this correlation was lower when IDK use fell below 93%. They concluded that differential use of the IDK not only undermines test validity, but also renders the results of "little research or pedagogic value" (p.973). Indeed, they not only suggest that the addition of an IDK option without a penalty system for incorrect responses be avoided, but also that the IDK should be avoided altogether and other approaches be considered including retrospective score correction for incorrect responses.

Aim and Methods

The data presented here were gathered at a small private university in Kyushu, Japan, and utilized the Vocabulary Size Test (VST) developed by Nation and Beglar (2007). Based on the British National Corpus and comprised of 140 questions, the VST was developed as a "reliable, accurate, and comprehensive measure of a learners' vocabulary size" (Nation & Beglar, 2007, p. 9) and typically requires 40 minutes to complete. By administering only the first 100 questions, the test can be completed in as little as 30 minutes (Nation, 2012). A total of 526 first-year undergraduate students across 25 classes completed the first 100 questions of the VST using a mark sheet and were instructed to select the IDK option when they had no idea of the correct answer. Subjects were informed that the results would have no impact on their grades, and that the test was being administered for the purposes of improving the English program. Subjects' test scores were not returned, and nor did they count towards any assessment component. No penalty system was used for incorrect responses. The test was administered during the last 30 minutes of a regular class and proctored by an English teacher. Data were gathered as part of a needs analysis in order to more clearly identify and delineate student English ability levels, which were essentially unknown due to the absence of data from standardized tests.

Interviews with university faculty and administrative staff indicated, however, that student ability levels varied considerably. The data have been analyzed here in light of a troubling observation made in relation to the above discussion. More precisely, the issue of variance in willingness to use the IDK option is examined through the following research questions:

RQ1: How many test-takers selected the IDK option at least once when offered the choice?

RQ2: What degree of variance occurred in test-takers' use of the IDK option among those who used it?

RQ3: How did using the IDK affect overall test scores?

Results

Figure 1 shows the number of test-takers who selected the IDK option at least once during the test. Of the 526 test-takers, 376 (71.48%) selected the IDK option at least once. The remaining 150 (28.52%), did not.

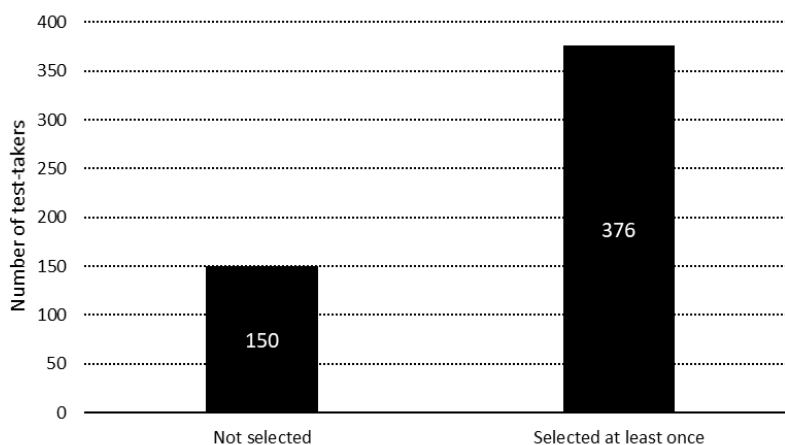


Figure 1. Number of test-takers that selected the IDK option at least once

Figure 2 shows the distribution of test-takers' use of the IDK option among those who used it. Other than the sharp drop-off identified at the 91–95 and 96–100 range, the use of the IDK option appears evenly spread, suggesting either considerable variance in test-takers' use of the IDK option or variance in student ability.

Figure 3 shows the variance in test scores between the two groups. The average score of those who did not use the IDK option was 47.93 while the average score for those who did use the option was 30.94, equating to a difference of 16.98 correct responses. Following the instructions provided by Nation (2012) this figure is multiplied by 100 to produce the test taker's approximate vocabulary size, which in this case results in a difference between the two groups of 1,698 words.

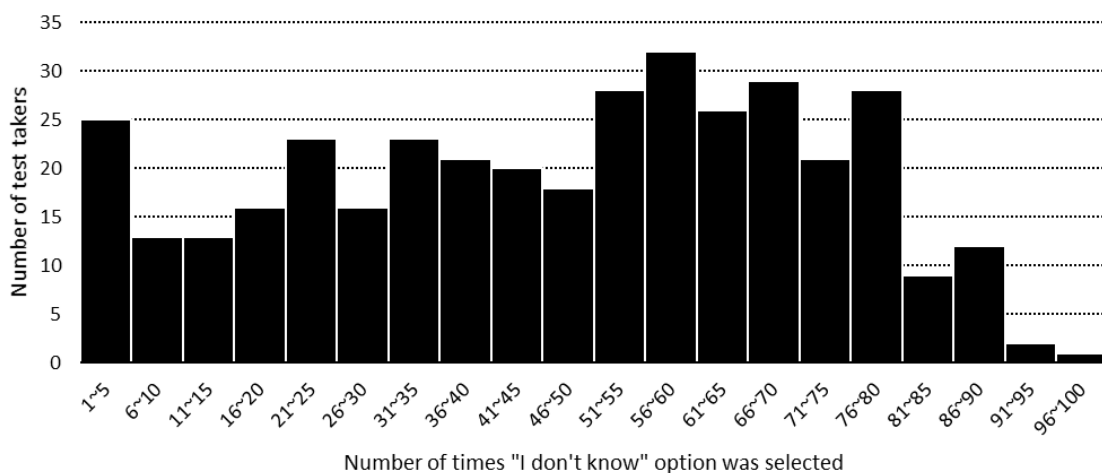


Figure 2. Distribution of test-takers' use of the IDK option among those who used it

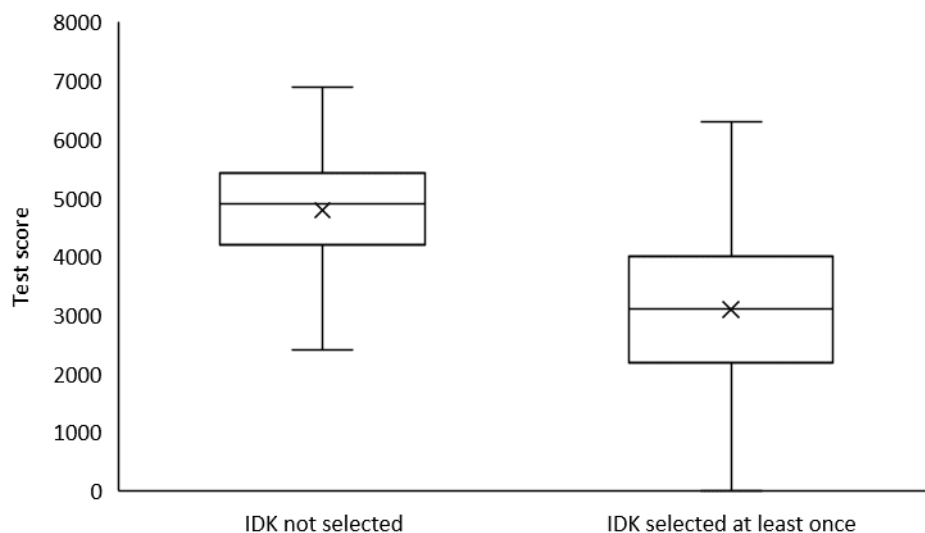


Figure 3: Box plots of test scores for the two groups

Conclusion and Future Directions

Although providing the IDK option was intended to reduce score inflation due to random guessing and thus provide a clearer picture of students' vocabulary knowledge, the data presented here indicate that a considerable portion (28.52%) of test-takers chose not to use it, despite being instructed to do so. This could be due to any number of factors, including the lack of a penalty system for incorrect answers, failure to follow instructions, or simply that, due to its zero-stakes nature, many did not take the test seriously—a potential limitation of the current study. Furthermore, only 71.48% of test-takers opted for the IDK, well below the 93% threshold suggested by Stoeckel, Bennett, and McLean

(2016), which raises serious doubts about the validity of any inferences made regarding student vocabulary sizes. Considerable variance was also observed among those who did use it, indicating either one of two things: 1) variance in vocabulary knowledge (the primary construct), or 2) variance in willingness to use the IDK (the secondary construct). However, in the absence of independent data regarding student ability, it is impossible to ascertain which construct accounts for this variance – a further limitation of the current study.

In order for educators to gain reliable and valid insights into students' vocabulary sizes, the issue of score inflation due to guessing needs to be addressed. As demonstrated here, however, attempting to achieve this by using the VST with an IDK option is problematic. Based on previous findings (Zhang, 2013; Bennett & Stoeckel, 2012), it could be argued that a penalty system could raise IDK uptake and reduce variance in willingness to use the IDK. It remains to be seen, however, if variance could be reduced to an acceptable level or if IDK uptake could be raised to above the 93% threshold. For these reasons, while providing an IDK option may offer a seemingly simple and attractive solution to the issue of blind guessing, with or without a penalty system, it may in fact create more problems than it addresses. For this reason, retrospective score correction for incorrect responses may offer a more straightforward and robust approach.

References

- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. Continuum.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Bennett, P., & Stoeckel, T. (2012). Variations in format and willingness to skip items in a multiple-choice vocabulary test. *Vocabulary Education and Research Bulletin*, 1(2), 2–3.
- Lucovich, D. (2014a). Adding “I don’t know” to the Vocabulary Size Test. In N. Sonda & A. Krause (Eds.), *JALT2013 Conference Proceedings* (pp. 683–691). Tokyo: JALT.
- Lucovich, D. (2014b). Test taking and DK use on the vocabulary size test. *Vocabulary Learning and Instruction*, 3(2), 69–77. <https://doi.org/10.7820/vli.v03.2.lucovich>
- McLean, S., Kramer, B., & Stewart, J. (2015). An empirical examination of the effect of guessing on vocabulary size test scores. *Vocabulary Learning and Instruction*, 4(1), 26–35. <https://doi.org/10.7820/vli.v04.1.mclean.et.al>

- Nation, I. S. P. (2012). The Vocabulary Size Test: information and specifications. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/VocabularySize-Test-information-and-specifications.pdf>
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13. Retrieved from jalt-publications.org/files/pdf/the_language_teacher/07_2007tlt.pdf
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511732942>
- Stoeckel, T., Bennett, P., & McLean, S. (2016). Is “I don’t know” a viable answer choice on the vocabulary size test? *TESOL Quarterly*, 50(4), 965–975. <https://doi.org/10.1002/tesq.325>
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13–40. <https://doi.org/10.3138/cmlr.53.1.13>
- Zhang, X. (2013). The I don't know option in the Vocabulary Size Test. *TESOL Quarterly*, 47(4), 790–811. <https://doi.org/10.1002/tesq.98>

Short Paper

A Categorized List of Phonologically Reduced High-Frequency Spoken Multi-Word Sequences for Classroom Use

Michael P. McGuire mmcguire@mail.doshisha.ac.jp

Introduction

The use of corpus frequency to prioritize language instruction has become customary, as seen in popular corpus-based vocabulary lists such as the Oxford 3000 (Hornby, 2020), the New General Service List (NGSL; Browne et al., 2013), the Academic Vocabulary List (AVL; Gardner & Davies, 2014). High-frequency words hold a higher relative value for language learners because not knowing them can impair understanding and communication (Nation & Webb, 2017). At the same time, corpus analysis has revealed the prevalence of multi-word sequences (MWSs) such as collocations and formulaic sequences in both written and spoken language. Even so, learners' knowledge of such MWSs is lacking (Rogers, 2016). This has led to a push for increased attention being given to MWSs in language instruction.

MWSs and Spoken Fluency

MWSs play a central role in fluent speech. Spoken fluency refers to “the productive processes involved in the planning and delivery of speech” (Schmidt, 1992, p. 92). Fluency in speech is possible when these processes become automatic and function “easily and efficiently” (Lennon, 1990, p. 391). The cognitive processing load of fluent speech is reduced by the recurrent use of prefabricated, or formulaic, MWSs (Wood, 2010) that are stored and recalled as single units (Wray, 2002). Rather than building every utterance one word at a time as many language learners need to, fluent speakers assemble a substantial amount of their speech from MWSs. In fact, their occurrence is so pervasive that they make up one-third to one-half of language (Conklin & Schmitt, 2008).

The cognitive load of fluent speech is reduced by the automation of MWSs through repetition (Bybee, 2002a). Disfluency, on the other hand, is a sign of cognitive overload, which suggests that a speaker is not yet proficient in the process of planning and delivering speech at the same time. “Nonfluent speech is effortful and requires a great deal of attention, so that nonfluent speakers exhibit many hesitations and other

manifestations of groping for words and attempting to combine them into utterances” (Schmidt, 1992, p. 93). This suggests that language learners could benefit from focused study and practice of MWSs. In fact, a number of studies have demonstrated that focused practice of MWSs leads to improved quantitative fluency measures in ESL students (McGuire & Larson-Hall, 2017, 2021; Thomson, 2017; Wood, 2009).

MWSs and Listening Perception

Language learners may also face challenges in listening perception due to considerable phonological reduction that occurs in high-frequency MWSs (Bybee, 2002b). This is the result of “the automatization of neuro-motor sequences which comes about with repetition” (Bybee & Hopper, 2001, p. 11) which “involves the reduction of the magnitude of articulatory gestures and the increased overlap of these gestures” (Bybee & Hopper, 2001, p. 11). The higher the frequency of an MWS, the more likely it will be to feature phonological reduction; in fact, some unique reductions and assimilations are only found in high-frequency phrases, such as the palatalization in *did you* or *don't you* (Bybee, 2002b).

The highest frequency MWSs in English are composed primarily of function words (articles, prepositions, pronouns, conjunctions, etc.), which are unstressed and therefore greatly reduce and link together in between stressed content words (Avery & Ehrlich, 1992). This presents a serious challenge for language learners, as Avery and Ehrlich point out: “the fact that function words in English are generally unstressed and reduced makes them almost unrecognizable to beginning ESL students” (p. 82). Henrichsen (1984) saw significantly lower accuracy in 65 L2 students’ transcriptions of English speech recordings when they featured phonological reductions compared to transcriptions of recordings without reduction. Ito (2006) found that students struggle more with phonological reductions than with lexical reductions such as contractions when tested on listening perception. Wong et al. (2017) found that “the ability to perceive phonologically reduced forms was a significant predictor of comprehension of native English connected speech” (p.23).

All of this suggests that attending to the phonological reductions during focused listening and speaking practice of such highly frequent MWSs could help improve students’ spoken fluency, listening perception, and listening comprehension. If this is the case, classroom practice can be greatly supported by a concrete corpus-based list of the highest-frequency multi-word sequences in spoken English.

Such MWS learning materials are scarce, especially in comparison with single word vocabulary lists (Nesselhauf, 2005). While a number of corpus-based MWS lists have been developed, most have focused on bigram collocations (Ackermann & Chen,

2013; Lei & Liu, 2018; Shin, 2006; Shin & Nation, 2007). As for n-gram or conogram MWSs of three words or longer, Simpson-Vlach and Ellis (2010) created the Academic Formula List (AFL), which includes 607 academic MWSs categorized by pragmatic function. The Phrasal Expression List, or PHRASE List (Martinez & Schmitt, 2012), compiles 505 figurative phrasal expressions identified for their potential difficulty for language learners. More recently, Rogers et al. (2021) created a significantly larger list of over 5,000 academic multi-word units.

All the above-mentioned lists serve different pedagogical purposes and used distinct selection criteria. This paper presents a list of high-frequency multi-word sequences that has been categorized specifically for classroom emphasis on phonological reductions.

Reduced MWS List

This list is intended to be used over the course of a semester in a university level listening and speaking course, so rather than a long single list, it takes the form of ten categories of MWSs that allows for focus on different phonological reductions in weekly lessons. As mentioned earlier, the highest-frequency MWSs are composed almost entirely of short function words, which allows for less psycholinguistic salience than idioms or fixed expressions. Thus, rather than memorization, the objective is to give students targeted exposure to and practice with phonological reductions in their most frequent MWS contexts.

The list does not attempt to represent all phonological reductions that occur in English but instead presents a pedagogically useful collection of the reductions that occur in the most common MWSs. It is hoped that after focused practice with this list, students will be better able to perceive these reductions more accurately while listening to fluent speech and make use of them in their own speech. McGuire and Larson-Hall (2021) provide a clear example of a semester-long course that uses this list.

The initial list was created using the spoken section of the Open American National Corpus (Ide & Macleod, 2001). This free and open corpus was chosen due to its large size and its content of conversational spoken American English. Future versions of the MWS list may be refined using the larger Corpus of Contemporary American English (Davies, 2008)

The first step of creating the list was to generate a list of n-grams (3, 4, and 5-grams) sorted by frequency. These raw results contained many similar or overlapping MWSs and therefore required some filtering to make the list more pedagogically functional. The following steps were taken to refine the list:

1. N-grams containing disfluencies such as hesitation markers (*uh* as in *uh you know*) and word repetition (such as in *I I do, yeah yeah I, and and*

uh) were removed. While such disfluencies do occur frequently in fluent speech, they are often the result of planning problems (Clark & Wasow, 1998) and are not pedagogically useful.

2. The highly frequent contraction *don't* was counted as a single item rather than two items (*do* and *n't*) in order to select MWSs with greater saliency. For example, *I don't know* was kept, while *I don't* and *don't know* were removed.

The resulting list was then categorized for weekly lessons to allow a focus on specific phonological reductions. Ten categories were created based on recurrent keywords or bigrams found in the top 100 MWSs: *of/the*, *I don't*, *you know*, *I think*, *that*, *and*, *have*, *was*, *do you*, and *to*. Each category was then filled with the ten most frequent MWSs containing those keyword(s).

Twelve MWSs featured keywords from multiple categories and were thus repeated in the list. This way each phonological reduction could be addressed separately in the classroom, and MWSs that contain multiple reductions could be revisited. For example, the MWS *do you have* appears in the *do you* category and the *have* category, which focus on different reductions. The resulting list of 100 MWSs (88 unique MWSs with 12 appearing in multiple categories) divided into ten categories can be found in the Appendix.

References

- Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235–247. <https://doi.org/10.1016/j.jeap.2013.08.002>
- Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford University Press.
- Browne, C., Culligan, B., & Phillips, J. (2013). *The new general service list*. Retrieved from <http://www.newgeneralservicelist.org>
- Bybee, J. (2002a). Sequentiality as the basis of constituent structure. In T. Givón & B. Malle (Eds.), *The evolution of language out of pre-language* (pp. 109–134). John Benjamins Publishing Company.
- Bybee, J. (2002b). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, 24(2), 215–221. <https://doi.org/10.1017/S0272263102002061>
- Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Benjamins.

- Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37(3), 201–242. <https://doi.org/10.1006/cogp.1998.0693>
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72–89. <https://doi.org/10.1093/applin/amm022>
- Davies, M. (2008). *The corpus of contemporary American English (COCA): 560 million words, 1990-present*. <https://www.english-corpora.org/coca/>
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. <https://doi.org/10.1093/applin/amt015>
- Henrichsen, L. E. (1984). Sandhi-variation: A filter of input for learners of ESL. *Language Learning*, 34(3), 103–123. <https://doi.org/10.1111/j.1467-1770.1984.tb00343.x>
- Hornby, A. S. (2020). *Oxford advanced learner's dictionary*. Oxford University Press.
- Ide, N., & Macleod, C. (2001). The American National Corpus: A standardized resource of American English. *Proceedings of Corpus Linguistics*, 3, 1–7.
- Lei, L., & Liu, D. (2018). The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics*, 23(2), 216–243. <https://doi.org/10.1075/ijcl.16135.lei>
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40(3), 387–417. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299–320. <https://doi.org/10.1093/applin/ams010>
- McGuire, M., & Larson-Hall, J. (2017). Teaching formulaic sequences in the classroom: Effects on spoken fluency. *TESL Canada Journal*, 34(3), 1–25. <https://doi.org/10.18806/tesl.v34i3.1271>
- McGuire, M., & Larson-Hall, J. (2021). The contribution of high-frequency multi-word sequences to speech rate and listening perception among EFL learners. *Vocabulary Learning and Instruction*, 10(2), 18–29. <https://doi.org/10.7820/vli.v10.2.mcguire.larson-hall>
- Nation, P., & Webb, S. (2017). *How vocabulary is learned*. Oxford University Press.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins. <https://doi.org/10.1075/scl.14>
- Rogers, J. M. (2016). *What are the collocational exemplars of high-frequency English vocabulary? On identifying multi-word units most representative of high-frequency lemmatized concgrams*. (Unpublished doctoral dissertation). University of Southern Queensland. Retrieved from https://eprints.usq.edu.au/32849/2/Rogers_2016_whole.pdf

- Rogers, J., Müller, A., Daulton, F. E., Dickinson, P., Florescu, C., Reid, G., & Stoeckel, T. (2021). The creation and application of a large-scale corpus-based academic multi-word unit list. *English for Specific Purposes*, 62, 142–157. <https://doi.org/10.1016/j.esp.2021.01.001>
- Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14(4), 357–385. <https://doi.org/10.1017/S02722263100011189>
- Shin, D. (2006). *A collocation inventory for beginners*. (Unpublished doctoral dissertation). Victoria University of Wellington. Retrieved from <http://researcharchive.vuw.ac.nz/handle/10063/553>
- Shin, D., & Nation, P. (2007). Beyond single words: The most frequent collocations in spoken English. *ELT Journal*, 62(4), 339–348. <https://doi.org/10.1093/elt/ccm091>
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487–512. <https://doi.org/10.1093/applin/amp058>
- Thomson, H. (2017). Building speaking fluency with multiword expressions. *TESL Canada Journal*, 34(3), 26–53. <https://doi.org/10.18806/tesl.v34i3.1272>
- Wong, S. W. L., Mok, P. P. K., Chung, K. K.-H., Leung, V. W. H., Bishop, D. V. M., & Chow, B. W.-Y. (2017). Perception of native English reduced forms in Chinese learners: Its role in listening comprehension and its phonological correlates. *TESOL Quarterly*, 51(1), 7–31. <https://doi.org/10.1002/tesq.273>
- Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *Canadian Journal of Applied Linguistics*, 12(1), 20.
- Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. Bloomsbury Publishing.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Appendix

The Reduced Multi-Word Sequence List

Category	Example Phonological Reduction(s)	MWS	Frequency (Occurrences per million words)
of / the	[ʌv] → [ə] [ði] → [ðə]	a lot of	1854
		one of the	449
		some of the	329
		a couple of	235
		that kind of*	222
		a lot of people	191
		kind of a	189
		what kind of	170
		out of the	157
		lot of the	157
I don't	linking [t] to next sound unreleased stop [t]	I don't know	1549
		I don't think	470
		I don't have*	190
		I don't like	106
		I don't really	101
		I don't want	74
		I don't even	72
		I don't see	51
		I don't remember	47
		I don't get	35
you know	[ju] → [jə] linking [w] to next sound	you know I	973
		you know it	627
		you know and	480
		you know the	454
		you know you	428
		you know they	399
		you know it's	358
		you know that*	342
		you know we	293
		you know if	226

I think	linking [k] to next sound unreleased stop [k]	I think that*	741
		I think it	570
		and I think*	408
		I think it's	300
		I think that's*	271
		I think they	250
		I think the	205
		yeah I think	204
		but I think	188
		well I think	173
that	[ðæt] → [ðət]	I think that*	741
		you know that*	342
		that you know	319
		I think that's*	271
		things like that	256
		something like that	243
		that kind of*	222
		to do that*	177
		I know that	166
		that would be	149
and	[ænd] → [ən], [n]	and you know	523
		and I think*	408
		and it was*	289
		and I don't	223
		and then I	175
		and things like that	154
		and I was*	148
		and all that	144
		and so I	143
		and then you	137
have	[hæv] → [əv], [v]	you have to	359
		I have a	291
		do you have*	255
		we have a	206
		I don't have*	190
		you have a	163

		to have a*	156
		have a lot	138
		I have to*	137
		they have a	134
was	[wʌz] → [wəz]	and it was*	289
		when I was	275
		it was a	264
		it was just	163
		and I was*	148
		I was in	137
		and that was	113
		there was a	106
		was kind of	103
		that was a	100
do you	[du ju] → [də jə]	what do you	266
		do you have*	255
		do you think	251
		do you do	129
		how do you	104
		do you like	96
		do you know	93
		do you feel	64
		or do you	49
		well do you	48
to	[tuw] → [tə]	to do it	214
	[tə] → [də]	I used to	201
		be able to	192
		to be a	191
		to go to	190
		to do that*	177
		to have a*	156
		I like to	138
		I have to*	137
		talking to you	133

**MWSs that appear in more than one category*

SIG News:

JALT
VOCAB SIG
2022
SYMPOSIUM
on Vocabulary & Learning



October 29, 2022

Tokyo International University

Kawagoe Campus - Building 2 - Room 232

¥1500 General Admission - ¥1000 JALT Members

Strand 1:
Vocabulary Learning
10:30am - 1:00pm

Discussant: Junko Yamashita
Nagoya University

Speaker 1: Kevin Mueller
Tokyo International University

Speaker 2: Suwako Uehara
The University of Electro-Communications

Speaker 3: Adam Dabrowski
The University of Electro-Communications

Speaker 4: Kimberly Klassen
Tokyo Woman's Christian University

Poster Session and Lunch: 1:00pm - 2:30pm

Strand 2:
Advances in
Vocabulary Research
2:30pm - 5:00pm

Discussant: Kristopher Kyle
University of Oregon

Speaker 1: Masaki Eguchi
University of Oregon

Speaker 2: Ali H. Al-Hoorie
Saudi TESOL Association

Speaker 3: Joseph P. Vitta
Kyushu University

Speaker 4: Derek E. Canning
Seigakuin University

Conference Chairs: Jeffrey Stewart, Joseph P. Vitta, Kevin Mueller, Christopher Nicklin



2022 Vocabulary Symposium

The 2022 Vocabulary Symposium is set for **October 29, 2022**, at **Tokyo International University's Kawagoe Campus**. Stay tuned for further details but the time has come to submit your proposals for poster presentations. We look forward to seeing you at the event. Description and application procedures are as follows:

Call for Poster Session Presentations

The Vocabulary SIG is excited to announce that there will be a poster session at its 2022 symposium which features Kristopher Kyle (University of Oregon) and Junko Yamashita (Nagoya University) as discussants with speakers presenting on a wide range of topics spanning from classroom vocabulary learning to lexical/linguistic complexity. As a component of the event, there will be a 90-minute poster session in the middle of the day. To that end, we would like to invite poster session proposals.

- Proposal should feature L2 vocabulary learning and/or instruction.
- Proposals can be present completed research or works in progress.
- Proposals can straddle the practitioner-research divide in presenting reflections, evidence-supported teacher tips, and/or action research.

How to apply?

Please complete the following Google form by July 15, 2022:

https://docs.google.com/forms/d/1GwfpSA6ss_B8a0nHlsm-fV3msN3lBHD9dfXP3MyJog4/viewform

Acceptance notifications will be sent out by mid-August.

Poster presenters are expected to attend the conference (1500 yen entrance fee; 1000 yen if JALT member). A presenter can only be first author on one poster.

Vocab SIG Research Grants

The JALT Vocab SIG is pleased to announce that the two recipients of the 2022 Research Grant have been decided through a double-blind peer review. We look forward to hearing more from them in the coming years.

- Eric Shepherd Martin “An Investigation of Japanese EFL Learners’ High-Frequency Written and Spoken Receptive Word Knowledge”
- Jeffrey Martin “Effects of Lexical Coverage and Genre on L2 Listening Comprehension”

Finally, a big thank you to our peer reviewers who not only helped with the selection process, but also took the time and energy to give a lot of useful feedback to all applicants.

VERB Call for Papers

The VERB welcomes submissions related to vocabulary research and education.

Short papers are peer reviewed and may require rewriting and resubmission for acceptance. They must not exceed 1500 words, excluding references, tables, and titles. Short papers fall into the categories of completed research, ongoing research, and teaching and learning in practice.

Other submissions encouraged are classroom activities related to vocabulary, book reviews, opinion pieces, and event reports and commentary. All submissions are expected to adhere to APA 7th edition formatting guidelines.

Summer Issue Deadline: March 15th each year

Winter Issue Deadline: September 15th each year

For submissions and all correspondence: <jaltvocabsig.verb@gmail.com>

Latest information: <https://jaltvocab.weebly.com/publications.html>

The following are guidelines for short paper submissions (please include these sections):

Completed research:

- * Background
- * Aims
- * Methods
- * Results
- * Conclusions
- * Future directions

Ongoing research:

- * Background
- * Aims
- * Methods
- * Sample
- * (Preliminary) Results
- * (Preliminary) Conclusions
- * Future directions

Teaching and learning in practice:

- * Theoretical framework
 - * Teaching context
 - * Procedure
 - * (Preliminary) Results
 - * (Preliminary) Conclusions
 - * Future directions
-

**If you are thinking about submitting, but your article doesn't fit into one of the above categories, please email us at the above address and let us know what you would like to submit, and we can discuss the possibility with you.

***Vocabulary Learning & Instruction* Call for Papers**

The Vocabulary SIG's *Vocabulary Learning and Instruction* (VLI) journal is calling for submissions for an upcoming issue. Submissions will be published online upon acceptance and combined into an issue later in the year.

VLI accepts long-form research papers (2000-7000 words) and brief reports, summaries, and commentaries (2000-3000 words) related to vocabulary acquisition, pedagogy, assessment, and lexical networks.

As an open journal, content is indexed on Google Scholar and made freely available on the internet without paywalls. Authors are free to also make their work available on sites such as academia.edu and ResearchGate.

All submissions are subject to a 2-step peer-review process:

A) Editors review manuscripts to ensure basic requirements are met, and that the work is of sufficient quality to merit external review. This process typically takes 1-2 weeks, at which point authors are informed of the outcome.

B) Submissions which meet these requirements are sent out for blind peer review by 2-3 experts in the field. This process takes approximately 1-2 months. Following external review, authors are sent copies of external reviewers' comments and notified of decisions (accept, accept pending changes, revise, and resubmit, or reject).

Please see <http://vli-journal.org/submissions.html> for details

The VERB salutes the cooperation and hard work of our reviewers:

VERB Reviewers: Phil Bennett, Thuy Bui, David Coulson, Tomoko Ishii, Magda Kitano, Brandon Kramer, Jenifer Larson-Hall, Mimi Masson, Stuart McClean, Michael McGuire, Atsushi Mizumoto, Ian Munby, John Racine, James Rogers, Rachel Ruegg, Jeff Stewart, Raymond Stubbe, and Yuka Yamamoto.