DOI: 10.1002/tesj.622

WILEY

Introducing Knowledge-based Vocabulary Lists (KVL)

Norbert Schmitt[1] | **Karen Dunn**[2] | **Barry O'Sullivan**[2] | **Laurence Anthony**[3] | **Benjamin Kremmel**[4]

[1]University of Nottingham

[2]British Council

[3]Waseda University

[4]University of Innsbruck

Correspondence
norbert.schmitt@nottingham.ac.uk

Funding information
British Council

1 | INTRODUCTION

Vocabulary is probably the key type of knowledge necessary for any language use, because if words to express concepts are not known, all syntactic and discourse knowledge is of little use. This is true for both first languages (L1) and second languages (L2). First language vocabulary (at least in its spoken form) is mainly learned incidentally from the massive exposure which children receive, and much second language vocabulary can be learned this way as well (see the other articles in this *TESOL Journal* special issue). But in classrooms and materials, second language vocabulary is largely taught in an explicit manner. This means that second language pedagogy must have a way of prioritizing which words to teach from among the multitude available. (English is estimated to have about 10 million different individual words (Brysbaert, Stevens, Mandera, & Keuleers, 2016).

For the past hundred years or so, word lists have been used to inform this prioritization. These lists have largely been based on word frequency, that is, how often various words appear in (typically L1) discourse. These word lists have been used to identify the words that learners are likely to encounter, thus likely to need, and likely to know. However, frequency has its limitations when predicting which words might be known by L2 learners, particularly at the level of being able to *produce* those words (i.e., being able to recall and spell them accurately) (see Schmitt, Dunn, O'Sullivan, Anthony, & Kremmel, in press, for evidence of this). It is true that learners typically know more words in high-frequency bands than lower frequency bands; for example, learners will typically know more words in the first thousand frequency band (1K) than in the second thousand frequency band (2K). But while frequency ranking works relatively well for *bands* (at least up until about the 5K band), it does not work so well for

individual *words* (e.g., we know that *pencil* and *orange* will likely be some of the early words learned by L2 learners, but appear as relatively lower frequency vocabulary on frequency lists).

This suggests the need for word lists based on direct tests of learner knowledge, rather than on frequency, which merely predicts imperfectly what learners might have learned from their language exposure. This *Current Issues* article introduces the *Knowledge-based Vocabulary Lists* (KVL) created to fill this need. The lists should be useful for pedagogical purposes in which it is beneficial to know whether learners are likely to be able to produce and correctly spell the words they know.

2 | DEVELOPING THE KNOWLEDGE-BASED VOCABULARY LISTS

In order to build the KVL, we needed to test which words learners know. Most previous lists have used the counting unit of *word families* (stem + inflections + derivatives: *nation* + *nations* + *national, nationalize, nationally, international*, etc.). However, research has shown that L2 learners typically do not know all of these family members, especially the derivatives (e.g., *nationalize, nationally*). Nevertheless, most learners do know the inflected forms (e.g., McLean, 2018). This makes the counting unit of *lemma* (stem + inflections only: *look* + *looked, looking, looks*; *boy* + *boys*) more appropriate for the KVL, due to learners' typically limited morphological knowledge.

Lemmas also take account of word class. Thus, the noun *look* is a separate lemma from the verb *look*, as is the noun *dance* from the verb *dance*. In these cases, the meaning is essentially the same: the former being an activity, and the latter entailing engagement in that activity. But often different word classes have different meanings. For example, *abandon* (verb) means to "leave something completely and finally," whereas *abandon* (noun) has the unrelated meaning of "a complete surrender to natural impulses without restraint or moderation." Knowing one of these meanings does not imply knowing the other. Thus, by using the counting unit of lemmas, meanings like these were tested separately, and if both uses were known well enough, the words were inserted as two separate entries onto the KVL lists.[1]

We needed to test learners' knowledge of the lemmas, but we wanted a test which demonstrated relatively good mastery of the words, ideally productive knowledge. Common test formats like multiple-choice or matching demonstrate receptive mastery at best, and they are prone to learners using guessing or test-taking strategies to obtain correct answers (Gyllstad, Vilkaitė, & Schmitt, 2015). We chose the following form-recall format for several reasons: being able to accurately produce the spelling of a word indicates a relatively robust level of knowledge of that word (Laufer & Goldstein, 2004), guessing is very difficult with this format, and the format could be computerized and scored automatically.[2] The format indicated the meaning of the target lemma with an L1 translation and

[1]Another possible counting unit (which we did not use) is the *flemma*. In it, different word classes with the same spelling are combined into a single unit. While some research suggests it can be useful when working with L2 learners (e.g., Brown, Stoeckel, McLean, & Stewart, 2020; McLean, 2018; Pinchbeck, McLean, Brown, & Kramer, 2016), other research indicates that lemmas might be preferable (Stoeckel, Ishii, & Bennett, 2020).

[2]The format also had limitations. One was that some lemmas have alternative spellings in British and American English (*centre/center, colour/color, apologise/apologize*). As a result, we were unable to produce knowledge rankings for 145 lemmas with dual spellings, and so these are excluded from the KVL lists. A list of these excluded lemmas is provided at the end of the User Manual. Another limitation is that a lemma needed to have a meaning in order to prompt the spelling answer. Thus, function/grammar words (e.g. *the*, *a*, *they*) were not tested, with the exception of a few that carried a clear meaning (e.g., *above* = being higher or greater than something else). A third limitation is that the respondents were given the first letter of the lemma, and the number of letters required. This means that test was not the same as having to recall the spelling of a lemma in the real world, where such prompts are usually not available.

example sentence. The answers were constrained by giving the first letter of the lemma and number of letters necessary. Here is an example item for Spanish speakers for the lemma *house* (English = I live in a large house that has three bedrooms):

casa Vivo en una **casa** grande que tiene tres dormitorios.

h __ __ __ __

We needed to decide on the size of the KVL, but there proved to be little research on the vocabulary size requirements for productive use (writing and speaking) to guide us. Almost all of the research has been on receptive use (reading and listening), and is almost solely denominated in word families, not lemmas. For example, 2,000–3,000 word families appear to be enough to understand daily chat like informal storytelling (van Zeeland & Schmitt, 2013). The most frequent 3,000–5,000 families should allow initial access to authentic reading materials (Laufer & Ravenhorst-Kalovski, 2010). The most frequent 5,000 families should provide most of the lexical resources needed to understand movies (Webb & Rodgers, 2009a) and television programs (Webb & Rodgers, 2009b). Thus, the most frequent 5,000 word families are enough to do many things in English. While word families are a more expansive counting unit than lemmas (also including derivatives), Laufer and Cobb (2020) have shown that only a few derivative affixes are very common, and so many derivative members of a word family will not occur very often in reading. This makes the text coverage provided by lemmas and flemmas not so different compared to that provided by word families. This was demonstrated by Nation (2016) who showed that the most frequent 3,000 flemmas provided 87.77% coverage of a 14-million-word spoken and written corpus, while the most frequent 3,000 word families provided only slightly more (90.38%). Thus, we felt able to use the word family research as guidance in setting our KVL list size in lemmas. Ultimately, we settled on a 5,000-lemma list as a reasonable target. More lemmas would have been better, but it was only practical to test enough lemmas to create a 5,000-lemma list. Also, in order to discover the best-known 5,000 lemmas, we needed to "overshoot" and test more of the likely candidate lemmas to see which were the best known. Ultimately, we tested a pool of 7,679 lemmas for each of three languages, with the intention that the best-known 5,000 of these would be retained in the list.

We know that a learner's L1 affects how they learn words, with L1–L2 similarities facilitating learning and L1–L2 differences increasing difficulty (Otwinowska & Szewczyk, 2017). Therefore, we tested learners from three diverse languages: German (English is a Germanic language), Spanish (a cognate Romance language), and Chinese (an unrelated language). German learners answered, on average, 37 test items from the candidate lemma pool, Spanish respondents answered 32, and Chinese learners 25. We discovered that the results from the three language groups were substantially different, and so we created bespoke lists for each of these languages instead of one combined, universal KVL list.

Test results were gathered via internet crowdsourcing on a British Council website, as they were sponsoring the research. The data were gathered from late 2018 to mid-2020. In total, over 3.3. million test item responses were collected from over 100,000 respondents.

3 | ANALYZING THE KNOWLEDGE-BASED VOCABULARY DATA

Perhaps the most important finding from our analyses of the data was the surprising disconnect between knowledge rankings and frequency rankings. The Kendall's tau correlations for these rankings

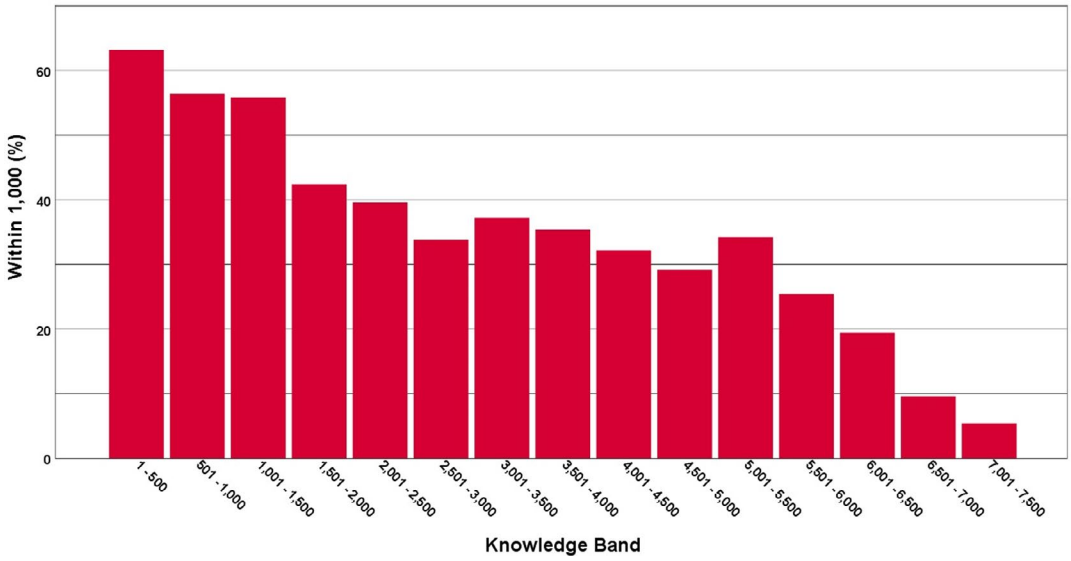

FIGURE 1 Percentage of frequency ranks within ±1,000 of knowledge ranks (Chinese results)

were .226 for the Spanish respondents, .224 for the German respondents, and .402 for the Chinese respondents (all $p < .001$).[3] Overall, these correlations indicate a very modest correspondence, given that frequency has always been taken as a reasonable proxy to indicate learner knowledge. It seems that for productive knowledge (or more precisely the ability to spell an L2 word correctly once its meaning is known), frequency rankings largely fail to predict knowledge rankings. (Note that frequency is somewhat better at predicting receptive knowledge; see Brysbaert, Keuleers, & Mandera, 2020).

This appears to be the case for at least two reasons. The first is cognates. *Cognates* are words that have a similar form in two languages, often the result of coming from a common ancestor word. Loanwords also have similar forms. This similarity obviously makes words easier in the L2 (e.g., English *generation*, *zombie*, *infirm*, *ecology* versus Spanish *generación*, *el zombie*, *enfermo*, *ecología*).

The second reason is that frequency data are collected from corpora (large databases of language), and these corpora often do not represent the kind of exposure L2 learners typically receive very well. For example, for German respondents, *reading* was extremely well-known, placed in the best-known 100 lemmas (#96). This is unsurprising as reading is usually a key feature of L2 instruction and learners will become familiar with this lemma very early. Yet the frequency ranking is only #1,586, based on the reference corpus we used (the *Corpus of Contemporary American English (COCA)*, Davies, 2008–). Thus, the corpus-based frequency ranking does not match actual German learner knowledge very well.

Figure 1 illustrates the knowledge–frequency disconnect that we found for the Chinese learners, showing that relatively few (often not much more than one-third) of the lemmas had frequency rankings that were within even ±1,000 of the KVL knowledge rankings. Note however, that the match is somewhat better at higher knowledge rankings than lower ones. (The Spanish and German learners

[3]These correlations are based on COCA frequency data. Other corpora based on oral language may provide somewhat closer correspondences. However, when we looked at frequency information from the SUBTlex corpus (compiled from subtitles from American films and television series), we found that, although somewhat better, they still did not accurately predict KVL knowledge rankings. See Schmitt et al. (in press) for details.

produced similar profiles.) Clearly, if practitioners working with learners from one of the three language backgrounds included wish to know the likelihood of their learners knowing particular lemmas to a productive level, the empirically based KVL lists will be more informative than frequency-based lists. (The full report of the KVL development and analyses is available in Schmitt et al., in press.)

4 | EXAMPLES OF THE SPANISH, GERMAN, AND CHINESE KVL LISTS

The complete KVL lists consist of 5,000 lemmas, but to give an idea of what they look like, Table 1 presents the first 50 lemmas for the three language groups we tested. Several points are noticeable. First, there are some surprisingly well-known lemmas in these "top 50" lists, for example *German*, *elite*, *atomic*, and *oil*. This is likely due to the teaching syllabi, materials, and emphases used in the different countries. This highlights that fact that although incidental learning may well be reasonably predicted by frequency and repetition (see the other articles in this issue), the learning of vocabulary which is explicitly taught will always depend on what is in the materials. Second, some of the lemmas which are surprisingly well known can be explained by cognateness (e.g., *federal*, *generation*, *clinic*).

Third, the three lists are largely different, with only a few common words across the three lists (e.g., *kiss*). However, this is somewhat misleading. If we take larger 500-lemma groupings, we see that there is about 50% overlap between the three language groups at the 1–500 knowledge level (Table 2). This steadily decreases until there is not much more than a 10% overlap at the 4,501–5,000 knowledge level. If we take the three complete 5,000-lemma lists, there is about 80% overlap. This shows that while the best-known 5,000 lemmas are substantially the same between the three language groups, the knowledge ranking of each individual lemma is typically different between the groups. Take *republic* for example. The Spanish knowledge ranking is 157 (known by 98% of Spanish respondents), a high score facilitated by the Spanish equivalent being a very similarly spelled cognate (*república*). The German equivalent is also a cognate (*republik*), but the German knowledge ranking is a lower 1,176 (93%). In the non-cognate language of Chinese, it unsurprisingly has the lowest ranking of 2,202 (79%). This shows that the L1 plays an important part in determining how well each lemma is likely to be known.

5 | FINDING THE KVL LISTS

The KVL lists will only be available online on the British Council KVL website https://www.britishcouncil.org/exam/aptis/aptis-expertise/knowledge-based-vocabulary-lists-kvl. This is because the KVL may be updated in the future as the researchers receive additional test data from users which could lead to revisions. The revisions need to be done on a single authorized site, or else a range of different versions may proliferate on different websites, some revised and others not. The main KVL lists for each language are presented in Excel spreadsheets named *KVL-Spanish*, *KVL-German*, and *KVL-Chinese*. Each of these spreadsheets contains several columns of basic information. These include the following:

- *Lemma*: the lemma being described
- *Word Class*: the part of speech of the lemma (e.g., noun, verb, adjective)
- *Knowledge Rank Order*: the order of likelihood of the lemma being known to a form-recall level of mastery.

TABLE 1 The 50 best-known[a] lemmas for Spanish, German, and Chinese learners

Knowledge ranking	Spanish respondents	German respondents	Chinese respondents
1	dance	bomb	car
2	jet	English	cat
3	city	eat	just
4	zoo	hungry	boy
5	orange	kiss	open
6	sun	ice	both
7	German	dance	dream
8	book	fit	story
9	one	sun	kiss (verb)
10	two	day	kiss (noun)
11	radio	burger	food
12	no	better	no
13	human	not	time
14	hour	blog	dog
15	winner	wind	keep
16	change	cat	tree
17	hero	camel	plan
18	family	west	hi
19	day	elite	heart
20	acid	jeans	home
21	black	Korean	name
22	atomic	idol	fish
23	boy	drive	north
24	federal	orange	cry
25	my	generation	must
26	shoe	dear	page
27	beach	king	far
28	love	kiss	open
29	ice	ball	map
30	key	star	kid
31	clinic	lamp	ready
32	play	football	care
33	west	egg	sleep
34	religion	person	drink
35	car	zoo	class
36	atom	one	shy
37	hot	art	two
38	fruit	win	write

(Continues)

TABLE 1 (Continued)

Knowledge ranking	Spanish respondents	German respondents	Chinese respondents
39	red	six	oil
40	body	water	table
41	air	invest	shopping
42	drive	law	music
43	table	war	ball
44	class	song	land
45	moon	dad	house
46	bird	bingo	team
47	please	bag	girl
48	bye	text	bit
49	expert	call	win
50	apple	birth	life

[a]All lemmas were known by at least 97% of respondents.

- *Frequency Rank Order*: the frequency rank of the lemma based on frequency counts of the *Corpus of Contemporary American English* (COCA) in 2018.

The website contains a *User Manual*, which should be read before using the KVL lists. There is also another set of spreadsheets available which are meant for researchers and contain more technical information about the lemmas' characteristics (e.g., percentage of respondents answering the test item correctly), and the design of the test items on which the knowledge rankings are based.

6 | USING THE KVL LISTS WITH SPANISH, GERMAN, AND CHINESE LEARNERS OF ENGLISH

The KVL lists will often be used in conjunction with frequency lists, and so it is important to understand the most advantageous use of each. Frequency lists report how often words appear in written and/or spoken discourse, which makes them good sources for information about which words learners are likely to come across in English, and need to know to operate in English. Thus, they are particularly useful in regard to the receptive skills of reading and listening.

The KVL lists were compiled based on the direct testing of a large sample of English language learners and show the likelihood of these kinds of learners knowing individual lemmas to the level of being able to spell them accurately. Thus, the KVL lists should be useful when it is advantageous to understand which lemmas learners are likely to know to a productive written level.

There are a range of possible applications for this knowledge-based information. We list a few possibilities to illustrate the uses of the KVL lists, but there will be many more.

- One main application will be informing the likelihood of learners achieving spelling mastery of vocabulary. Current lists used to predict the knowledge/difficulty of words are usually based on frequency or on receptive measures of vocabulary knowledge. But the analyses of the KVL data (Schmitt et al., in press) show that receptive knowledge (i.e., just recognizing words in text) does

TABLE 2 Percentage of common lemmas within various knowledge ranking levels

Knowledge Band	Spanish ↔ German (%)	Spanish ↔ Chinese (%)	German ↔ Chinese (%)
1–500	48.6[a]	50.4	48.2
501–1,000	25.7	23.2	24.1
1,001–1,500	18.3	17.4	18.3
1,501–2,000	15.2	13.9	12.8
2,001–2,500	12.4	11.8	13.0
2,501–3,000	14.2	12.9	12.3
3,001–3,500	11.2	13.2	9.8
3,501–4,000	15.0	12.8	8.2
4,001–4,500	12.0	11.4	13.1
4,501–5,000	13.1	12.9	10.9
1–5,000	84.7	82.4	83.3

[a]Percentage of lemmas that were common between the two languages at the specified knowledge band (e.g., 48.6% of the lemmas were shared between the 500 lemmas appearing in the Spanish 1–500 knowledge band and the 500 lemmas appearing in the German 1–500 knowledge band).

not predict productive knowledge very well (i.e., being able to recall and spell words correctly). The KVL provides much better predictions about the sequence in which learners achieve spelling control over a range of lemmas. This information should be useful for writing teachers, and for test developers who assess writing ability.

- In selecting reading materials, it is often useful to grade the readings to match the abilities of learners. This is currently done by frequency profiles. But frequency is only a crude proxy for knowledge. Using the KVL lists should give a better idea of whether learners know the words in particular texts or not. While the KVL lists are based on spelling tests, research shows that if words are mastered to a spelling level, learners can typically also understand the words when they see them (Laufer & Goldstein, 2004). This makes the KVL lists potentially informative for reading-based applications.
- The sequencing of the KVL lists should provide a baseline for understanding which lemmas learners know. For example, if learners know many lemmas at the 1,000–1,500 level, it can be inferred that they will probably also know most of the other lemmas in that band, and also in the 1–1,000 range as well. While frequency lists also do this to some extent, the KVL lists are customized to each of the three language groups (Spanish, German, and Chinese speakers), and so take account of words which are relatively easy for each group due to cognateness. Frequency lists do not take account of cognateness, and so the KVL lists are a much better representation of learner knowledge than frequency lists.
- In testing, we often attempt to measure or discover which words learners have learned, both from incidental learning outside the classroom and from explicit study within it. To do this, we first need to build a pool of which words learners *might* know. Frequency lists have typically been drawn upon to build this pool. However, as frequency does not predict knowledge of individual words very well, target words drawn from frequency lists will not match learner knowledge very closely, which makes for inefficient and potentially misleading tests. Drawing on the KVL lists for pools of test words should give test developers a better chance of targeting the words on their tests to the level

of their test takers. (See Gibson & Stewart, 2014 for one example of how a vocabulary list created in a similar way to the KVL was used to provide vocabulary size estimates for Japanese university students.)

- In psycholinguistic experiments, a range of factors affect the processing of vocabulary. This makes it crucial that target words are selected which are controlled in terms of the word characteristics that make the words easier or more difficult to process. Frequency has been shown to be a robust word characteristic which affects processing. However, frequency does not account for cognateness, and so target words may be far easier or more difficult than frequency might suggest for particular language groups. The KVL list provides psycholinguists a valuable alternative source of information about potential word knowledge/difficulty to use in building their experiments.

The *User Manual* has more information on these and other potential uses of the KVL lists. The current KVL lists cover Spanish, German, and Chinese learners of English, but there is also interest in creating bespoke KVL lists for other languages. It will be interesting to see how these efforts progress.

7 | THE AUTHORS

Norbert Schmitt is emeritus professor of applied linguistics at the University of Nottingham. He has researched second language vocabulary issues for over 25 years, and is interested in all aspects of vocabulary description, acquisition, and use. He continues to write, lecture, and consult internationally on these topics.

Karen Dunn is a senior researcher in the British Council's Assessment Research Group. Her work includes evaluating the functioning of new and established tests across the Aptis suite, and contributing to a range of British Council research projects, including those in conjunction with international organisations, such as the Rwandan Ministry of Education and the Pratham-ASER Centre in India.

Barry O'Sullivan is Head of Assessment Research and Development, the British Council, London and has undertaken research across many areas on language testing and assessment (e.g., the development of the sociocognitive model of test development and validation). He was the founding president of the UK Association of Language Testing and Assessment (UKALTA).

Laurence Anthony is a professor of applied linguistics at Waseda University, Japan. His main research interests are in corpus linguistics, educational technology, natural language processing, and English for specific purposes (ESP). He received the National Prize of the Japan Association for English Corpus Studies (JAECS) in 2012.

Benjamin Kremmel is the Head of the Language Testing Research Group at the University of Innsbruck, Austria, where he teaches and researches language learning, teaching, and assessment. He holds a PhD in applied linguistics from the University of Nottingham, UK. His research interests include vocabulary assessment, diagnostic assessment, and language assessment literacy.

ACKNOWLEDGEMENTS

This research was supported by the British Council. We also wish to thank the many researchers who helped at various stages of the project, who are listed in full in Schmitt et al. (in press).

ORCID

Norbert Schmitt https://orcid.org/0000-0001-8590-5471
Karen Dunn https://orcid.org/0000-0002-7499-9895
Barry O'Sullivan https://orcid.org/0000-0002-2256-8209

Laurence Anthony https://orcid.org/0000-0002-8249-2824
Benjamin Kremmel https://orcid.org/0000-0002-2536-0091

REFERENCES

Brown, D., Stoeckel, T., McLean, S., & Stewart, J. (2020). The most appropriate lexical unit for L2 vocabulary research and pedagogy: A brief review of the evidence. *Applied Linguistics*, amaa061. https://doi.org/10.1093/applin/amaa061

Brysbaert, M., Keuleers, E., & Mandera, P. (2020). Which words do English non-native speakers know? New supernational levels based on yes/no decision. *Second Language Research*, *37*, 207–231. https://doi.org/10.1177/0267658320934526

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, F. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*, 1116. https://doi.org/10.3389/fpsyg.2016.01116

Davies, M. (2008–). *The corpus of contemporary American English (COCA): 560 million words, 1990–present*. Available online at https://www.english-corpora.org/coca/

Gibson, A., & Stewart, J. (2014). Estimating learners' vocabulary size under item response theory. *Vocabulary Learning and Instruction*, *3*(2), 78–84. https://doi.org/10.7820/vli.v03.2.gibson.stewart

Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *International Journal of Applied Linguistics*, *166*, 276–303. https://doi.org/10.1075/itl.166.2.04gyl

Laufer, B., & Cobb, T. (2020). How much knowledge of derived words is needed for reading? *Applied Linguistics*, *41*, 971–998. https://doi.org/10.1093/applin/amz051

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*, 399–436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30.

McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, *39*, 823–845. https://doi.org/10.1093/applin/amw050

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins.

Otwinowska, A., & Szewczyk, J. M. (2017). The more similar the better? Factors in learning cognates, false cognates and non-cognate words. *International Journal of Bilingual Education and Bilingualism*, *22*, 974–991. https://doi.org/10.1080/13670050.2017.1325834

Pinchbeck, G. G., McLean, S., Brown, D., & Kramer, B. (2016, September). Aspects of word list development: Reference corpora and the definition of "word" [working title]. In *Revisiting the word family: What is an appropriate lexical unit for Japanese EFL learners?* Colloquium conducted at the Vocab@Tokyo conference, Tokyo, Japan.

Schmitt, N., Dunn, K., O'Sullivan, B., Anthony, L., & Kremmel, B. (in press). *Knowledge-based vocabulary lists*. British Council Monographs on Modern Language Testing. Equinox Press.

Stoeckel, T., Ishii, T., & Bennett, P. (2020). Is the lemma more appropriate than the flemma as a word counting unit? *Applied Linguistics*, *41*, 601–606. https://doi.org/10.1093/applin/amy059

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*, 457–479. https://doi.org/10.1093/applin/ams074

Webb, S., & Rodgers, P. H. M. (2009a). The lexical coverage of movies. *Applied Linguistics*, *30*, 407–427. https://doi.org/10.1093/applin/amp010

Webb, S., & Rodgers, P. H. M. (2009b). Vocabulary demands of television programs. *Language Learning*, *59*, 335–366.

How to cite this article: Schmitt, N., Dunn, K., O'Sullivan, B., Anthony, L., & Kremmel, B. Introducing Knowledge-based Vocabulary Lists (KVL). *TESOL J.* 2021;12:e622. https://doi.org/10.1002/tesj.622