Reading in a Foreign Language October 2021, Volume 33, No. 2
ISSN 1539-0578 pp. 238–246

Lexical Thresholds and Alleged Threats to Validity: A Storm in a Teacup?

Batia Laufer
University of Haifa
Israel

Coverages, thresholds and comprehension: A glimpse into the history

In the late 1980s I was working with teachers who believed that in order to understand a text it was enough to understand 80% of the text's word tokens. This belief was related to the popularity of cloze exercises and cloze tests where every fifth word, i.e., 20%, was deleted and yet the text could be fully reconstructed. It was clear to me that reading a real text did not resemble a cloze exercise, since unknown words were all lexical words, and since they were not evenly distributed, clusters of too many of them obscured the meaning of the text. I, therefore, set out to calculate the minimal text coverage, i.e., percentage of running words in a text the reader should understand in order to comprehend it reasonably well. I used the term 'lexical threshold' for this coverage, following Alderson (1984), who suggested that there was a 'second language threshold' that learners had to reach before they could transfer reading strategies from first language (L1) to second language (L2) reading. The result of my study (Laufer, 1989) was that 95% was the probabilistic lexical threshold above which a learner had a good chance of being a 'reader' as defined by a passing grade (55%) on a reading test in my university at that time.

My next task was to explore the second facet of lexical threshold, the number of words learners should know receptively in order to reach reasonable comprehension (Laufer, 1992). The results showed that a 3,000 vocabulary level would predict a reading score of 56%, a 4,000 vocabulary level would result in 63%, and a 5,000 level would yield a reading score of 70%. (Since the participants took vocabulary level and vocabulary size tests, the vocabulary results are in word families.) If a comprehension score of 70% was considered 'reasonable', the recommended vocabulary size would be 5,000 word families.

Two decades later (Laufer & Ravenhorst-Kalovski, 2010) I explored the relationship between text coverage, vocabulary size of the learners, and reading comprehension, by combining the three variables into one design. “Independent readers,” as defined by a national exam, had a vocabulary size of 6,000–8,000 word families and reached 98% of text coverage. “Readers with support” (learners who needed 56 academic hours before becoming independent readers) had a vocabulary size of 4,000–5,000 word families and reached 95% of text coverage. These results are similar to other studies with different learners and different texts, or corpora (Hu & Nation, 2000; Nation, 2006; Schmitt et al., 2011). Based on the combined evidence, the current suggestion is that an optimal coverage for reading of any text is 98% of word tokens and that the minimal coverage is 95%. In the case of authentic texts, the recommended vocabulary size is about 8,000–9,000 vocabulary families for reaching 98% coverage (assuming the proper nouns are familiar) and 5,000 for reaching 95% coverage.[1] Simplified texts will require a lower vocabulary size to reach the above coverages, depending

on the degree of simplification. Recently (Laufer, 2020), I found that 95% coverage could be reached with an initial knowledge of 90% words in the text and inferring an additional 5%. The findings regarding the lexical thresholds, in terms of coverages and learners' vocabulary size, can be used in planning lexical syllabi, grading reading texts in terms of lexical difficulty, matching particular texts to particular learners, dividing learners by lexical knowledge, and researching comprehension, inferencing and incidental acquisition.

Coverages, thresholds and comprehension: Incorrect assumptions?

The results of the thresholds studies cited here are based on data of over a thousand learners and millions of words of general and specialized corpora. Some later studies on vocabulary gains from written and oral texts used vocabulary size or levels tests to find out participants' lexical knowledge. The validity of these tests was established by research and published in refereed journals (Aviad-Levitzky et al., 2019; Beglar, 2010; Schmitt et al., 2001; Webb et al., 2017).

And yet, McLean (2021) argues that operationalization of the lexical difficulty of a text and the lexical level of learners suffers from limited validity, as it is based on incorrect assumptions and on convention rather than research findings. He claims that texts and learners' vocabulary knowledge are not matched correctly because (a) it is not clear at all what constitutes students' lexical mastery of particular vocabulary, (b) the word counting unit used when estimating the lexical difficulty of a text and the lexical ability of a learner is inappropriate, (c) the recognition formats of vocabulary tests are inappropriate for measuring the type of vocabulary knowledge needed for reading, and (d) the number of items selected to represent a 1,000-word band does not represent that band accurately. In the paper above he addresses the first two issues, the mastery level and the word counting unit. I will discuss them as well and offer a different, and, hopefully, more realistic perspective on these issues. For a critique of the last two points, see Webb (2021).

Matching texts to learners' level: A need for rigid mastery levels?

Texts can be matched to learners' vocabulary level by measuring the vocabulary knowledge of the learner and profiling the specific text. Table 1 shows the vocabulary profile of a mid-frequency reader that has been adapted to suit readers with a vocabulary of 4,000 words, A *Christmas Carol* by Charles Dickens (downloaded from Paul Nation's page of resources https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/readers/mid-frequency-graded-readers/A-Christmas-Carol-simplified-4000-level-version.pdf). The text was analyzed by Vocabulary Profile available at Tom Cobb's Lextutor (www.lextutor.ca). The left column indicates the 1,000 word frequency bands, the next 3 columns show the number of words at each band and their percent in brackets, and the last column—the cumulative percent of tokens, i.e. text coverage provided by each frequency level. The profile shows that 3,000 word families are necessary to reach 95% coverage and 5,000 to reach 98%.

Table 1

Vocabulary Profile of Christmas Carol (28,509 running words)

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token (%)
K-1	892 (39.8)	1729 (47.74)	24706 (86.7)	86.7
K-2	561 (25.1)	849 (23.44)	1919 (6.7)	93.4
K-3	274 (12.2)	351 (9.69)	554 (1.9)	95.3
		Coverage 95		
K-4	261 (11.7)	323 (8.92)	518 (1.8)	97.1
K-5	206 (9.2)	252 (6.96)	441 (1.5)	98.6
		Coverage 98		
K-6	18 (0.8)	18 (0.50)	106 (0.4)	99.0
K-7	9 (0.4)	9 (0.25)	23 (0.1)	99.1

Table 2 shows the scores on Vocabulary Levels Test (VLT) (Webb et al., 2017) of 10 English Foreign Language (EFL) College students from an English for Academic Purposes (EAP) class. Based on the National Psychometric Test of reading comprehension in English, they were placed in a group which required 112 academic hours of study to complete their English language requirements. At the end of the requirements the students are supposed to become independent readers of academic material. The VLT includes 150 items that represent 5,000 word families. It is separated into five levels, with 30 items at each frequency level that represent 1,000 word families. Thus, a total score of 150 on the VLT represents the knowledge of 5,000 word families, 120 represents the knowledge of 4,000 word families, 110 – of 3,666, 100 – of 3,333, 90 – of 3,000.

Table 2 shows the students' scores at each frequency level and the total test score.

Table 2

Vocabulary Levels Test scores (maximum level score =30)

Student	1st 1000	2nd 1000	3rd 1000	4th 1000	5th 1000	Total
1	29	27	25	26	16	122
2	30	25	23	24	21	122
3	29	25	27	24	18	121
4	30	25	21	15	14	103
5	29	27	14	14	15	98
6	28	27	13	26	5	97
7	30	26	23	24	20	122
8	30	26	24	16	13	109
9	30	26	26	16	23	121
10	30	27	23	18	21	119

The two tables show that none of the students reached the 98% text coverage that would require 5,000 word knowledge according to Table 1. However, the text can be read with support at 95% coverage, i.e., 3,000 word knowledge reflected in a VLT score of 90. If reading with support is what the teacher intends to do with the class, then all the students have the necessary total vocabulary score. In my research, I have often related the learner's total vocabulary score to text profiles. Other researchers established learner-text match by

setting mastery levels for individual vocabulary bands. Apparently, a figure of 26 out of 30, or 86.6%, has been used by most of them. McLean (2021) criticizes this figure suggesting that if readers need to reach a 95% or a 98% text threshold, the vocabulary mastery level in each band should be 95% or 98%. It is true that the mastery levels of 86.6% or 90% are not borne out by research. The question is whether setting mastery levels is at all necessary. If we set a mastery level at 95%, i.e., 29 out of 30, or 90%, i.e., 27 out of 30, then none of the students in Table 2 has the necessary knowledge of all three bands to read the text profiled in Table 1. Six students would not even be considered able to read texts at the 2,000 word level as their 2,000 band score is below 27. Even if we lowered the mastery level to 86.65%, i.e., 26 out of 30, then only student no. 9 would reach the necessary threshold, 3,000 word knowledge, for the text. And yet this student's total vocabulary, ~4,000 word families, is not different from students 1, 2, 3, 7, 10. If a 90% mastery level is insisted on, then an imaginary student with vocabulary test scores on the VLT of 27, 27, 27, 0, 0 would be a good match to the text. I have never seen scores like this on a vocabulary test. Most students know some words at most bands, and the scores usually decline with declining band frequency[2].

The text profile in Table 1 shows that knowledge of bands 4 and 5 may not be less important than of band 3 since the text coverages of these three bands are almost identical. The students in Table 2, like similar students in their group, read argumentative prose at the 3,000–4,000 word level in their EAP course where they were placed by their psychometric reading score. Had these students been given texts based on various mastery levels of individual word bands, the texts would not have matched their abilities.

In the case of advanced students with vocabulary sizes larger than 5,000 word families, setting mastery levels is not even possible. Since the Levels Test does not include bands of words less frequent than the 5th 1000, advanced students can be tested by size tests that assess the understanding of 14,000 or 20,000 word families. In these tests each frequency band is represented by 10 items. Ten items are not considered enough to distinguish partial from full mastery of vocabulary at a frequency level. What matters in a size test is the total score that indicates an approximate vocabulary size of the test-taker.

In Laufer and Aviad-Levitzky (2017), 116 students at four EFL reading levels were tested by the Vocabulary Size Test (VST) (Nation & Beglar, 2007). The students' vocabulary size ranged from ~2,000 to ~8,000. We examined whether these scores reflected students' classification by the national psychometric reading comprehension exam into the following groups requiring different numbers of instruction hours: non-readers (less than 2,000 words), basic readers (~3,500 words), pre-independent readers (~5,000 words) and independent readers (~8,000 words). Out of 116 participants, 10 were not classified correctly by the VST. In other words, matching the reading level to students' overall vocabulary size score was 91% accurate.

The data from this study and the examples of students' vocabulary in Table 2 suggest that the need to establish a definite mastery of an individual word band, 86.6, 88.8, or 90%, is unduly exaggerated. The overall vocabulary knowledge score, in the levels tests for low and intermediate level students, or size test for advanced students, may provide a more realistic and practical assessment of students' ability to comprehend specific texts than scores at separate frequency bands.

The lexical load of texts and the lexical knowledge of learners: An inappropriate word counting unit?

Partial morphological knowledge is not lack of knowledge

McLean (2021) argues that the word family, the counting unit of vocabulary tests and lexical text profiles, is inappropriate since not all learners possess the morphological knowledge necessary to understand all the possible derived forms of a familiar baseword. For example, if *center* and *develop* are known, but *decentralization* and *antidevelopment* are not, then learners' knowledge tested by a family-based test is overestimated and text difficulty profiled by a family-based profile is underestimated. Eight studies are cited as evidence for learners' incomplete morphological knowledge, because learners in these studies did not have a perfect knowledge of all the affixes, or affixed words. However, none of these studies showed that learners did not possess any morphological knowledge, even when the researchers' choice of test items looked rather unusual, e.g. *teachability, undevelopable, publishability*. Even low level learners, who scored 2,000–3,000 on a vocabulary test knew 50%–60% of derived words, while learners who knew 5,000 word families in Laufer et al. (2021) knew 85% of the derivatives and McLean's advanced students knew 83.9%.

Morphological knowledge is not the only measure of comprehension of derived words in texts

However, the main flaw in McLean's argument against word families lies in the unsubstantiated assumption that test results of morphological knowledge accurately reflect comprehension of derived words in texts. Perfect morphological knowledge implies that every possible derived word is understood if the baseword is familiar since the learner is supposed to know the meanings and grammatical functions of all the affixes, at least the 91 affixes from Bauer and Nation, (1993). Therefore, it does not matter whether the derived words are frequent, infrequent, or non-existent. (For example, *inacceptability* is understood if we know *accept* and the meaning and grammatical function of the affixes *in~, ~able, ~ity*).

But real texts include real, mostly frequent derived words, constructed with a limited number of affixes, particularly in simple texts. Laufer and Cobb (2020) found that derived words in graded readers were constructed with mostly 4 affixes, and in academic texts with 12 affixes, but only 3 appeared most frequently. Learners do not need perfect morphological knowledge to understand them.

Furthermore, many derived words are comprehended because they are more frequent than their corresponding base words. Some examples are *easy, healthy, government, conversation, explanation, basic*, *difference, beautiful, employment, careful, stranger, dirty, expensive, teacher.* I refer to them as “derived cores”, as they are the most frequent and useful members of the word family and have probably been learned holistically without awareness of their morphological composition.

An additional factor in comprehending derived words is text context that can help learners to infer meaning. For example, a student who knows *evaluate*, but not *evaluation* in isolation, may understand it in 'A thorough *evaluation* of the project is required'. Partial lexical knowledge, i.e., knowledge of the related base word, could often suffice to provide the necessary clues for comprehension. Preliminary results of a study on this topic show that this is indeed the case.

Morphological properties of a corpus are not morphological properties of texts

An example of how detached the "anti-word family" argument is from the real world is in the incompatibility between problems suggested in McLean (2021, Table 1) and the morphological makeup of texts that students read.

The table is based on Brown (2018), who examined a sample of 500 words representing 5,000-word families in the British National Corpus (BNC), and calculated that derived words constituted 13.4 % of the corpus. He then assumed that learners may not know derived words at all, or any of the derived words at different Bauer and Nation (1993) affix levels, and reduced the 95% and 98% coverages accordingly, by the proportion of all the derived words, or proportions of words at specific affix levels. This was done to show that lack of morphological knowledge reduces text coverage, and therefore family-based tests distort the information about students' true abilities to comprehend. However, this calculation was performed with imaginary students and imaginary texts. First, no study shows that students do not know all the derived words, or all the derived words at specific Bauer & Nation (1993) affix levels. Second, the overall percentage of derived words in the BNC does not reflect the distribution of derived words and affixes in real texts. Laufer and Cobb (2020) examined texts of several genres and difficulty levels and found that the average percentage of derived words was 7.78% in academic texts, 7.88% in newspaper articles, 5.04% in authentic novels, and 3.17% in graded readers.

As an example, I analyzed a graded text by Morpholex, a tool that analyzes texts into basewords, inflected and derived words. (The tool is available at https://www.lextutor.ca/morpho/).

The text is a story 'Mr. Harris and the night train'. It includes 1,723 words. 2,000 most frequent word families cover 95% of the text, and 3,000 – 98%. The analysis showed that derived words constituted 2.2% of text tokens. Table 3 shows the list of these words and the number of their occurrences in the text. The list shows what I have argued earlier. There is a small number of derived words, 15 different derived words that occur 38 times in 1,723 text words. There is a small number of frequent affixes, seven of the 15 derivatives are constructed with the affix *~ly*. Some words are 'derived cores', *easy, noisy, beautiful, famous*, possibly *Finnish*. *Carriage* (in *train carriage*) is probably learnt without even relating it to *carry*. It is rather unlikely that comprehension of these words in text context would require perfect morphological knowledge of all the members of word families.

Table 3

Derived Words in a Graded Reader

Derived word	Frequency in text
quickly	5
slowly	4
noisy	3
Finnish	2
carefully	2
unhappy	2
easy	2
loudly	1
quietly	1
suddenly	1
usually	1
beautiful	4
famous	2
carriage	7
midnight	1

I argue that word family is an appropriate counting unit as it does not invalidate text-reader compatibility. Understanding of derived words in texts is not reflected by tests of stemless affixes, or of infrequent derived words presented in isolation and in clueless sentences. Morphological knowledge is only one factor that affects comprehension of derived words. Other factors like derived coreness, frequency, text context, text level, text genre and student proficiency play parts as well. Besides, the number of derived words in texts is rather small. By the time learners are proficient enough to read academic texts with the largest (7.78 %) percentage of derivatives, they will have developed their vocabularies to at least 5,000 word families and therefore their morphological knowledge as well.

The claim that family-based tests result in such an overestimation of derivational knowledge that can quickly result in an incomprehensible text is an unnecessary and unsubstantiated exaggeration.

Concluding Remarks

It is rewarding to see that issues pertaining to lexical difficulty of texts, and learners' lexical knowledge necessary to comprehend them continue generating interest and discussion three decades after I started working on them. Scholarly activity involves revisiting and refining existing concepts and tools and applying different approaches to the same issue. However, older and newer approaches do not have to be either right or wrong, appropriate or inappropriate, valid or invalid. My discussion of 'lexical mastery levels' and word counting units is a case in point. It is easy to set a very high lexical mastery level when a text is composed of 1,000 or 2,000 most frequent words, and students possess limited vocabularies. It is much more problematic to do so with most text profiles (Table 1) and most students' vocabularies (Table 2). A total vocabulary size score is a practical alternative.

Researchers may use lemma profiles of texts and lemma-based tests with low level learners, if they wish. (Lemma-based tests for intermediate, let alone advanced, learners may be too long to be practical). But word family-based tests are appropriate for vocabulary assessment as well. They may not show how many and which derived words in a word family students know, but this does not mean they grossly overestimate the vocabulary that students can employ in comprehension. This is so because learners do possess morphological knowledge which improves and becomes quite good when they reach 5,000 word knowledge. Besides, morphological knowledge is only one factor that determines readers' comprehension of derived words in texts. Moreover, the percent of derived words in texts is rather low, particularly in low level texts read by low level learners.

Dire warnings against the family-based approach and exaggerations of its perils are unnecessary and counterproductive. Lemma and family approaches to text analysis and testing are based on different, not right or wrong, assumptions. Both approaches contribute to lexical research in different, not appropriate or inappropriate ways. They are complementary, not mutually exclusive.

Notes

1. According to Nation's (2006) figures, the former can sometimes be reached with 6,000 families and the latter with 3,000.
2. L1 speakers of Romance languages are an exception to the usual pattern of scores declining with frequency due to the large number of infrequent Latinate cognates (Cobb, 2000).

References

Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1–27). Longman.

Aviad-Levitzky, T., Laufer, B., & Goldstein, Z. (2019). The new computer adaptive test of size and strength (CATSS): Development and validation. *Language Assessment Quarterly*, *16*(3), 345–368. https://doi.org/10.1080/15434303.2019.1649409

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279. https://doi.org/10.1093/ijl/6.4.253

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, *27*, 101–118. https://doi.org/10.1177/0265532209340194

Cobb, T. (2000). One size fits all? Francophone learners and English vocabulary tests. *Canadian Modern Language Review, 57* (2), 295–324. https://doi.org/10.3138/cmlr.57.2.295

Hu, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403–430.

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint & P. Arnaud (Eds.), *Vocabulary and applied linguistics* (pp.126–132). Macmillan.

Laufer, B. (2020). Lexical coverages, inferencing unknown words and reading comprehension: How are they related? *TESOL Quarterly*, *54*(4), 1076−1085. https://doi.org/10.1002/tesq.3004

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*, 15−30.

Laufer, B., & Aviad-Levitzky, T. (2017). What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning recognition? *The Modern Language Journal, 101*(4), 729–741. https://doi.org/10.1111/modl.12431

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., Webb, S., Kim, S. K., & Yohanan, B. (2021). How well do learners know derived words in a second language? The effect of proficiency, word frequency and type of affix. *ITL-International Journal of Applied Linguistics*, Online-First Articles https://doi.org/10.1075/itl.20020.lau

McLean, S. (2021). The coverage comprehension model, its importance to pedagogy and research, and threats to the validity with which it is operationalized. *Reading in a Foreign Language*, *33*, 126–140.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63,* 59–82. http://dx.doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, *18*, 55–88. https://doi.org/10.1177/026553220101800103

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*, 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Webb, S. (2021). A different perspective on the limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition, 43*(2), 454–461.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics, 168*(1), 33–69. https://doi.org/10.1075/itl.168.1.02web

About the Author

Batia Laufer is Professor (Emerita) of applied linguistics in the Department of English Language and Literature at the University of Haifa, Israel. Her main research interests are vocabulary acquisition in additional languages, vocabulary testing, reading, lexicography, cross-linguistic influence, and language attrition. She has lectured, supervised research, and published widely on these topics.
Address for correspondence: Batia Laufer, Department of English Language and Literature, University of Haifa, Haifa, Israel. E-mail: batialau@research.haifa.ac.il