Studies in Second Language Acquisition, **43** (2021), 950–953
doi:10.1017/S0272263121000668

Critical Commentary

THE COMING PARADIGM SHIFT IN THE USE OF LEXICAL UNITS

Dale Brown *

Kanazawa University

Jeffrey Stewart

Tokyo University of Science

Tim Stoeckel

University of Niigata Prefecture

Stuart McLean

Momoyama Gakuin University

For any L2 research project or pedagogical purpose, determining what counts as a single lexical unit is a crucial issue. We would therefore like to thank *SSLA* for instigating this discussion of the issue, and Stuart Webb for his thoughts on this matter. Although proper perspective must always be maintained, we believe that in thinking about lexical units something of a paradigm shift (Kuhn, 1962) is currently underway in our field: At first, a practice is unquestioned, then conflicting evidence emerges (as has been the case recently on this issue, e.g., McLean, 2018; Stoeckel et al., 2020), alternate approaches become permissible, and eventually a new norm is established.

So just what is the emerging norm that appears nascent? Best practice, we would argue, is for researchers and educators to select a lexical unit that is appropriate, considering the evidence, for the learners and use in question. Webb suggests that thus far research has been limited and cautions against the overgeneralization of findings. However, while more evidence is always welcome, it is important to acknowledge that empirical evidence is available. As Brown et al.'s (2020) review found, to date the evidence is that (a) many learners (including those far beyond beginner levels) lack the knowledge necessary to deal with receptive uses of derivational forms of known headwords, and (b) derivational forms

* Correspondence concerning this article should be addressed to Dale Brown, Institute of Liberal Arts and Science, Kanazawa University, Kakuma-machi, Kanazawa-shi, Ishikawa-ken, 〒 920-1192 Japan. E-mail: dbrown@staff.kanazawa-u.ac.jp.

make up a significant portion of text (i.e., enough to impact comprehension). Moreover, there are indications that learners find inferring derivational forms encountered in context a challenge (Brown, 2013). Our position is that this evidence should be heeded: Decisions based on evidence (even if limited) are superior to decisions based on supposition.

For many L2 learners, therefore, for receptive as well as productive uses, the appropriate lexical unit is likely to be a smaller unit (perhaps the lemma or flemma); for young L2 learners and early beginners, word types are likely appropriate; and occasionally, for receptive uses with advanced learners, the word family may be reasonable. It is not therefore a dichotomous choice between lemmas/flemmas and families: Indeed, we have previously noted that "ideally, perhaps, a different unit might be adopted for each specific population of interest" (Brown et al., 2020, p. 2). However, it is the case, as Webb's article demonstrates, that only certain units tend to be given consideration (types, lemmas, flemmas, word families) because they have been conventionalized as named entities. Moreover, as we also observed previously, there is certainly value in the regular use of a small number of standard units for reasons of clarity, transparency, and comparability (see also Bauer & Nation, 1993, p. 264).

But what of the questions Webb raises about the utility and practicality of smaller lexical units? Regarding pedagogy, Webb suggests that larger lexical units have advantages in that viewing headwords alongside their inflections and derivations might accelerate lexical development. This is because learning morphologically related word forms may be easier than learning unrelated words and because acquisition of the inflectional and derivational systems may be facilitated. We agree. However, making use of larger lexical units in instruction makes sense precisely because learners need support in this area. That is, attention to individual family members is needed because knowledge of a word form is not a good indicator of knowledge of related forms, as Ward and Chuenjundaeng (2009) and McLean (2018) have shown. One issue at present is that different uses of larger lexical units are conflated: Lists of word families—useful as a pedagogical resource—are also used in testing and profiling (e.g., to match learners with lexically appropriate materials) without recognition of the problems raised (see Kremmel, 2016).

On wordlists, Webb argues that larger lexical units allow greater coverage to be achieved by a given number of units, observing, for example, that "the 1,000 most frequent word types, flemmas, and word families accounted for 76.46%, 80.97%, and 82.95% coverage of a 14-million-word corpus (Nation, 2016)" (Webb, 2021, p. 944). However, this additional theoretical coverage is illusory for a learner who lacks the ability to deal with the varied forms encompassed by the lexical unit. In and of itself the argument that a particular unit accounts for more coverage does not mean much. What matters is whether a lexical unit reflects the knowledge and abilities of learners (Gardner, 2007, p. 242).

Discussing assessment, Webb suggests that using a smaller lexical unit would mean a much larger number of items requiring assessment. Noting that the first 1,000 word families comprise 3,281 lemmas (Nation, 2016), Webb argues that in place of a 30-item test of those families, a 98-item test would be needed were lemmas to be used. However, there are two problems with this suggestion. First, from a statistical standpoint, this is incorrect. Vocabulary size and levels tests function as a poll of the proportion of words a learner knows within a given population of words (Gyllstad et al., 2020). In such polls, it is the sample size that determines the margin of error, not the size of the population being sampled (Smith, 2004). Therefore, tests with identical numbers of items to existing tests

https://doi.org/10.1017/S0272263121000668 Published online by Cambridge University Press

would not compromise accuracy of estimation. Second, Webb is mistaken in assuming that all the lemmas that comprise the first 1,000 families (in his example) would need to be covered by the test. In fact, many lemmas among the first 1,000 families are of low frequency and would be replaced by higher-frequency lemmas in a frequency-based lemmas list. Thus, switching from family- to lemma-based testing would not mean that all the resultant lemmas would be within the sampling pool. Accordingly, even if a sample of 30 items per 1,000-word band were desired, a lemma-based test need not contain many more items than a family-based test. This is demonstrated by the fact that a well-constructed lemma-based wordlist (e.g., Brezina & Gablasova, 2015) can provide coverage levels comparable to family-based lists.

We hope we have been able to address Webb's concerns in the space available to us. In closing, we note that precedent is a powerful force in research and pedagogy. We recognize that this shift in thinking about lexical units faces opposition from the inertia of many years of word-family usage by numerous practitioners. In much the same way less thorough statistical methodologies perpetuate in L2 research due to their long-held prior acceptability, the use of the word family remains common not just in areas where its use is well founded but also in others where it is not. We do not wish to denigrate such work. The word family has been the unit of choice of many seminal researchers, and their contributions have enhanced our understanding of L2 vocabulary. While rethinking the use of lexical units can be seen as a paradigm shift, ultimately it is but one of many changes in thinking in our field as practices improve. Scientific progress is about iteration and gradual improvement on what came before. Improvements in wordlists, in the accuracy of testing and in the precision of coverage estimates are just that, improvements. They are not a total negation of prior work. Our only wish is that future work makes use of a lexical unit that is appropriate, considering the available evidence, for the learners at hand.

REFERENCES

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

Brezina, V. & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, *36*, 1–22. https://doi.org/10.1093/applin/amt018

Brown, D. (2013). Types of words identified as unknown by L2 learners when reading. *System*, *41*, 1043–1055. https://doi.org/10.1016/j.system.2013.10.013

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.* Advance online publication. https://doi.org/10.1093/applin/amaa061

Gardner, D. (2007). Validating the construct of *Word* in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, *28*, 241–265. https://doi.org/10.1093/applin/amm010

Gyllstad, H., McLean, S., & Stewart, J. (2020). Using confidence intervals to determine adequate item sample sizes for vocabulary tests: An essential but overlooked practice. *Language Testing.* Advance online publication. https://doi.org/10.1177/0265532220979562

Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly*, *50*, 976–87. https://doi.org/10.1002/tesq.329

Kuhn, T. (1962). *The structure of scientific revolutions*. University of Chicago Press.

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 Publishing Company.

Smith, M. H. (2004). A sample/population size activity: Is it the sample size of the sample as a fraction of the population that matters? *Journal of Statistics Education*, *12*. https://doi.org/10.1080/10691898.2004.11910735

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

Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, *37*, 461–469. https://doi.org/10.1016/j.system.2009.01.004

Webb, S. (2021). The lemma dilemma: How should words be operationalized in research and pedagogy? *Studies in Second Language Acquisition*, *43*, 941–949. https://doi.org/10.1017/S0272263121000784