Studies in Second Language Acquisition, **43** (2021), 965–968
doi:10.1017/S0272263121000656

Critical Commentary

LEMMAS, FLEMMAS, WORD FAMILIES, AND COMMON SENSE

Batia Laufer
University of Haifa

Webb (this volume) points out that discussions of lexical units have presented lemmas/flemmas and word families as dichotomous options of which one is more appropriate than the other. This is indeed the approach of lemma/flemma proponents. The basic "antifamily" argument is that comprehension of *center* does not necessarily mean the learner knows the related lemmas *centrist*, *central*, *centralization*, *centralize*, *centralized*, *centralism*, *centralist*, *centrality*, *centrally*, *centeredness*, and *centric.* Therefore, it is argued that family-based tests overestimate vocabulary knowledge, and word families are unsuitable counting units for vocabulary tests, teaching materials, and research. I will question the logic of the assumptions underlying this one-sided approach and argue that test results of morphological knowledge do not reflect learners' comprehension of derived words in texts.

LEARNERS IN ONE AREA OF THE WORLD DO NOT REPRESENT ALL LEARNERS

There are very few studies that examined comprehension of base words and related derived forms (including identical forms with different parts of speech), thus investigating whether knowledge of base words extended to derived words. (Studies that require participants to supply the target items, i.e., demonstrate productive knowledge, or test general affix knowledge without comparing base and derived words are less relevant to our argument.)

The participants of the relevant studies were mostly of low and intermediate EFL levels. Ward and Chuenjundaeng's (2009) Thai students knew 25%–50% of the base words from the Academic Word List. Only 17 Japanese learners in McLean (2018) knew 5,000 word families. Others knew 3,000 (n = 176), or less than 2,000 (n = 84). Almost all Stoeckel et al. (2020), participants were at A2 or B1 CEFR level. These learners knew on average 50%–60% of derived test items. The authors state that "for receptive purposes, most L2 English learners … lack the morphological knowledge to make the word family a suitable unit" (Brown et al., 2020, p. 5). And yet low-intermediate Asian learners do not represent most learners and all proficiency levels. It is plausible that morphologically different L1s

Correspondence concerning this article should be sent to Batia Laufer, Department of English Language and Literature, University of Haifa, Haifa, Israel 3491077. E-mail: batialau@research.haifa.ac.il

will affect morphological awareness and learning differently. More research is necessary with different populations, different educational contexts, and a range of proficiencies. Recent studies by Laufer et al. (2021) and Snoder and Laufer (2021) show that L1 speakers of Hebrew and Swedish who scored ~5,000 on a vocabulary size test have almost identical knowledge of base words and derived words. Mochizuki and Aizawa's (2000) Japanese learners of a similar vocabulary size understood 77% of affixes in nonwords. This is less than perfect, but this is not a lack of morphological knowledge. These studies indicate that knowledge of derivations is likely to increase with vocabulary size to a point at which it is similar to that of base words. If learners after years of English in some educational contexts fail to see the relationship between *agree* and *agreement*, let alone, *work (n)* and *work (v)*, the remedy is not to dismiss word families, but to provide effective teaching.

WORDS IN ISOLATION ARE NOT WORDS IN CONTEXT

Almost all the studies on base words and related derived words present the target items in isolation, or in isolated sentences that do not give away the meaning. Furthermore, too often the derived test items are infrequent, as frequency figures in COCA show (in parentheses). Examples from McLean (2018) are *antidevelopment* (4), *publishability* (7), *mistaught* (5), *teacherly* (50), and *undevelopable* (12). Such methodology is appropriate for investigating learners' morphological knowledge in its own right. However, one assumption behind comprehension of derived words in texts is that the surrounding context can help learners to infer meaning. I have always expressed concerns about overreliance on context, mainly because context may not provide the necessary clues for unfamiliar words. However, knowledge of the meaning of a base word *is* a clue for understanding the related derived word. (A study on comprehension of derived words in context is in preparation, and preliminary results suggest that base words as clues are helpful.)

DERIVED WORDS IN TESTS ARE NOT DERIVED WORDS IN TEXTS

While the studies of morphological knowledge include infrequent derived words or nonwords with affixes, the derived words in learner texts are mostly frequent words constructed with a small number of affixes, particularly in simple texts. Laufer and Cobb (2020) showed that graded readers included derived words with mostly four affixes. Derived words in academic texts tended to be constructed with 12 affixes, but only 3 appeared most frequently.

Furthermore, many derived words are more frequent than the corresponding base words and are, therefore, learned earlier. Some examples from Cobb and Laufer's (2021) first 1,000-word family list are *easy*, *healthy*, *government*, *conversation*, *explanation*, *basic*, *difference*, *beautiful*, *employment*, *security*, *careful*, *stranger*, *suddenly*, *dirty*, *expensive*, and many more, that constitute more than 40% of the derived forms in the list. I refer to them as "derived cores" as they are the most frequent and useful words of the family. Concerns about morphological knowledge tend to overlook the presence of derived cores that are prevalent in English and are learned holistically without learners' awareness of their morphological makeup.

https://doi.org/10.1017/S0272263121000656 Published online by Cambridge University Press

DERIVED WORDS IN CORPORA ARE NOT DERIVED WORDS IN LEARNER TEXTS

The lemma proponents calculate how lack of morphological knowledge may impact text coverage. They cite Brown (2018), who examined a sample of 500 words representing 5,000-word families in the British National Corpus. Derived words constituted 13.4% of the corpus. Then they argue that "if the first 5,000 word families provided 95% coverage of a particular text, the actual proportion of known tokens would be just 82.3% (i.e., 95% × [100 – 13.4%]) for learners unable to comprehend derivational forms" (Brown et al., 2020, p. 4). However, no study shows that learners are unfamiliar with all the derived words, particularly if they know 5,000 words. Moreover, there is no support for the assumption that the proportion of derived words in a large corpus is the same as in texts that students read, and that derived words are distributed equally in texts of different difficulties. In a comprehensive study of the proportions of derived words in different text types, Laufer and Cobb (2020) 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, and that the number of different affixes that make up the derived words in texts is small. Thus, it is possible to reach 95% of text coverage with three or four derivational affixes in academic and newspaper texts, one affix (ly) in novels, and none at all in graded readers. This implies that, contrary to the claims made in Brown et al. (2020), reaching the lexical thresholds for reading does not require knowledge of most of the derived words in a word family because a small number of frequent affixes will provide the necessary coverage together with the base words and inflections.

NOT EXTENDED, BUT NUCLEAR WORD FAMILIES

A possible solution to the family and lemma-based counting principles was proposed by Cobb and Laufer (2021). They produced a Nuclear Family List (NFL) of the 3,000 most frequent families. The NFL includes frequent and useful family members that are most often encountered in the input (*apply*, *application*), and excludes many of the infrequent derived words (*misapplication*, *inapplicable*). It includes 5,610 lemmas and 22 frequent affixes, as opposed to 9,132 lemmas and 81 affixes in the 3,000 BNC/COCA lists. The nuclear family might be a more useful word counting unit for basic and intermediate learners than lemmas or word families.

CONCLUSION

The objection to family-based tests rests on the assumption that because learners may not be familiar with all the word family members of test items, the tests overestimate the vocabulary that learners can employ in comprehension tasks. But text analysis shows that such full knowledge is not necessary. It is particularly unnecessary for low-level learners who read simple texts. The simpler the text, the simpler and smaller the proportion of derived words. Besides, many of the derived words are derived cores, frequent and useful items that are learned holistically, and, therefore, comprehended regardless of affix knowledge or lack thereof. Furthermore, if the derived word is unfamiliar, knowledge of the base word meaning and text context can facilitate its comprehension. Finally, morphological knowledge develops with lexical and general language proficiency, and,

https://doi.org/10.1017/S0272263121000656 Published online by Cambridge University Press

like other language areas, could be affected by learners' L1 and by teaching. Hence, tests of morphological knowledge do not reflect the employability of derived words in real language tasks, and do not provide evidence against word family as a counting unit. The call of lemma proponents to reevaluate tests, coverage studies, curriculum goals, word lists, text profiling, and approaches to vocabulary teaching (Brown et al., 2020) is at best unsubstantiated and inappropriate.

REFERENCES

Brown, D. (2018). Examining the word family through word lists. *Vocabulary Learning and Instruction*, *7*, 51–65. http://vli-journal.org/wp/vli-v07-1-2187-2759/

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

Cobb, T., & Laufer, B. (2021). A Nuclear word family list: The most frequent family members—base words and affixed words. *Language Learning*, *71*, 834–871 https://doi.org/10.1111/lang.12452

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, K. S., & 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*, *172*, 229–258. https://doi.org/10.1075/itl.20020.lau

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

Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, *28*, 291–304. https://doi.org/10.1016/S0346-251X(00)00013-0

Snoder, P. & Laufer, B. (2021). EFL learners' receptive knowledge of derived words: The case of Swedish adolescents [Manuscript submitted for publication]. Department of Language Education, Stockholm University and Department of English Language and Literature, University of Haifa.

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/S0346-251X(00)00013-0