



**UNIVERSITY OF LEEDS**

This is a repository copy of *Selecting lexical units in wordlist for EFL learners*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/179183/>

Version: Accepted Version

---

**Article:**

Dang, TNY [orcid.org/0000-0002-3189-7776](https://orcid.org/0000-0002-3189-7776) (2021) Selecting lexical units in wordlist for EFL learners. *Studies in Second Language Acquisition*, 43 (5). pp. 954-957. ISSN 0272-2631

<https://doi.org/10.1017/S0272263121000681>

---

This article has been published in a revised form in *Studies in Second Language Acquisition* <https://doi.org/10.1017/S0272263121000681>. This version is free to view and download for private research and study only. Not for re-distribution, re-sale or use in derivative works. © The Author(s), 2021. Published by Cambridge University Press

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

**Dang, T. N. Y.** (in press). Selecting lexical units in wordlist for EFL learners. *Studies in Second Language Acquisition*. Doi: [10.1017/S0272263121000681](https://doi.org/10.1017/S0272263121000681)

### **Selecting lexical units in wordlists for EFL learners**

Determining the lexical unit is at the heart of wordlist studies, whose findings in turn significantly influence research on testing, lexical profiling, and pedagogy. In this commentary, I will discuss issues in the selection of lexical units in wordlists for EFL learners by drawing on my experiences as a wordlist researcher and an EFL teacher and learner.

I agree with Webb (this volume) that various factors affect the decision for selection of the lexical unit. A key factor in wordlist studies is the list purpose, which in turn is influenced by the learning burden (i.e., how easy or difficult it is to learn a word). The assumption behind the selection of word types as the lexical unit is that all word types have a similar learning burden whether they share the same baseword or not. In contrast, larger lexical units take into account the relative difference in the learning burden of word types sharing the same baseword (e.g., *work, worked, worker*) and those not sharing the same baseword (e.g., *work, abduct, rumple*). The assumption behind the choice of lemmas as the lexical unit is that if learners know a baseword (*work*) and the inflectional system (e.g., *-ed*), and are able to guess the meaning of words from context, they would be able to recognize and infer the meaning of its inflected forms (e.g. *worked*) in the input even though teachers do not explicitly teach them these forms. Similarly, the choice of word families is based on the assumption that if learners know a baseword (*work*) and the inflectional and derivational systems (e.g., *-ed, -er*), and can guess the meaning of words from context, they would be able to infer the meaning of its inflected and derived forms (e.g., *worked, worker*) when they encounter them in the input without being explicitly taught by teachers. In brief, the choice of larger lexical units is based on the

assumption that the similarities in meanings and forms lighten the learning burden and allow learners to use learning strategies and inflectional and derivational knowledge to infer the meaning of unknown members of known word forms.

If a wordlist is used to develop tests measuring vocabulary knowledge or to determine the vocabulary load of texts, its lexical unit should be as precise as possible so that learners' current vocabulary knowledge and the difficulty of input would be accurately estimated (Brown et al., 2020; Stoeckel et al., 2020). If learners are yet to develop inflectional and derivational knowledge and vocabulary learning strategies (word parts, guessing from context), using lemmas and word families will underestimate the learning burden and overestimate learners' vocabulary knowledge. In contrast, if learners have already developed such kinds of knowledge, using word types and lemmas may overestimate the learning burden and underestimate learners' vocabulary knowledge.

If a wordlist is used to set vocabulary learning goals for learners, the lexical unit should be selected from a cost-effective perspective. In many EFL contexts, the learning and teaching time is limited, and learners' vocabulary growth rate is fairly slow (Webb & Chang, 2012). Adopting lemmas and word families seems to be more efficient than word types. Compared to explicit teaching of all word types, teaching basewords, core affixes, and vocabulary learning strategies (word parts, guessing from context) would mean that less instruction time is needed, the burden of learning new forms that share the same baseword with known forms is reduced, and learners could expand their vocabulary knowledge faster in a shorter period of time.

However, whether lemmas or word families are a more appropriate lexical unit for EFL learners is still a debate. Those in favor of lemmas (e.g., Brown et al., 2020) argue that choosing word

families may underestimate the learning burden of derived forms because EFL learners have insufficient derivational knowledge, which would negatively affect their comprehension of texts. It should be noted that while learners' insufficient derivational knowledge indicates that Level-6 word families is probably too large a lexical unit for EFL learners, it does not mean that we should completely dismiss the idea of word families. First, Laufer and Cobb's (2019) analysis of various kinds of texts that EFL learners are likely to read revealed that only knowledge of basewords, inflected forms, and derived forms of a small number of affixes is needed to reach 95% and 98% coverage of these texts. This indicates that it is unnecessary to know all derived forms of Level-6 word families for comprehension. Second, having incomplete derivational knowledge does not mean that learners do not know any derived forms nor that they are unable to acquire derivational knowledge. Research shows that even beginners know some derived forms and learners' vocabulary size positively correlates with their derivational knowledge (e.g., Mochizuki & Aizawa, 2000); importantly, word part training positively affects learners' vocabulary knowledge (e.g., Lin, 2019; Morin, 2003). Therefore, while EFL learners' insufficient derivational knowledge might be due to the difficulty in acquiring derived forms, it can also be the result of insufficient word part training. In fact, Ward and Chuenjundaeng's (2009) survey with Thai EFL students from the same cohort with those having incomplete derivational knowledge showed that although 84% of them had learned English suffixes at high-school, 78% were not sure about their knowledge, and only 18% acknowledged that their teachers paid considerable attention to English suffixes. Ward and Chuenjundaeng's informal conversation with high-school teachers and analysis of high-school textbooks revealed that most of these teachers used textbooks focusing on test-taking strategies, and the majority of these textbooks paid little attention to word part strategies. My experience as an EFL learner and

teacher in Vietnam also supports the idea that teachers do not pay much attention to word part training; if they do, instructions mainly focus on test-taking strategies to help students pass the national high-school graduation and university entrance exams. These tests use a multiple-choice format with very limited contextual information (Figure 1). To find the answer, students only need to know syntactical rules (article+ adjective + noun) and affix rules (*-ive* indicating an adjective) without having to know the meaning of the basewords (*attract*) and word part strategies. Consequently, in many classes, word part instructions mainly focus on training learners to mechanically apply syntactical and affix rules to answer multiple-choice questions. Learners may not see the importance of affixes and word part strategies in enhancing their vocabulary knowledge and text comprehension.

Question 12: The job gives you lots of chances to travel abroad; it's certainly a very \_\_\_\_\_ offer.  
A. attract                      B. attractively                      C. attractive                      D. attraction

*Figure 1. A test item in the Vietnamese National High School Graduation Exam 2020*

Therefore, from the pedagogical perspective, while Level-6 word families may be too large a lexical unit when setting vocabulary learning goals for EFL learners, it is reasonable to use word families which include the basewords, inflected forms, and derived forms of a small number of frequent affixes. The use of word family lists should combine with word part training, which focuses on learning the most frequent derivations and helps learners see the value of affixes and word part strategies in vocabulary development and comprehension.

Taken together, the selection of lexical units in wordlists for EFL learners should consider the list purposes, which in turn should take the learning burden into account. Several areas deserve attention in future research. First, evidence supporting large lexical units (lemmas, word

families) comes from research on reading. It is unclear whether the same results can be found with listening. Second, one possible reason for learners' incomplete derivational knowledge is the quality of word part training that they receive. Little research has examined the extent to which word part training is provided in real EFL classrooms nor how effective it is. Further research on word part training is needed. Last, Laufer and Cobb's (2019) findings of the core affixes are based on the analysis of a small number of texts. Whether similar findings can be found across a vast number of texts needs further investigation.

## 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*. <https://doi.org/10.1093/applin/amaa061>
- Laufer, B., & Cobb, T. (2019). How much knowledge of derived words is needed for reading. *Applied Linguistics*. <https://doi.org/10.1093/applin/amz051>
- Lin, M.-F. (2019). Developing EFL learners' morphological awareness: Instructional effect, teachability of affixes, and learners' perception. *IRAL*, 57(3), 289–325.
- Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, 28(2), 291–304.
- Morin, R. (2003). Derivational morphological analysis as a strategy for vocabulary acquisition in Spanish. *The Modern Language Journal*, 87(2), 200–221.

Stoeckel, T., McLean, S., & Nation, I. S. P. (2020). Limitation of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*.

<https://doi.org/10.1017/S027226312000025X>

Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, 37(3), 461–469.

Webb (this volume). The lemma dilemma: How should words be operationalized in research and pedagogy?

Webb, S. A., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal*, 43(1), 113–126.