Critical Commentary

WORD FAMILIES AND LEMMAS, NOT A REAL DILEMMA
INVESTIGATING LEXICAL UNITS

*Stuart Webb**

University of Western Ontario

INTRODUCTION

In my initial piece in this critical commentary, I briefly introduced the most frequently used lexical units and the different aspects of L2 research and pedagogy that are impacted by the choice of lexical unit. I argued that the selection of lexical units should vary according to research and pedagogical purpose, as well as learner variables (e.g., vocabulary size, morphological knowledge, and proficiency). In addition, I tried to highlight the lack of L2 studies on this topic, and a need for cautious interpretation of earlier findings. The other articles in this critical commentary provide different perspectives from which we might also consider lexical units, a lack of agreement on the topic, and suggestions for further research. In the following, I will briefly touch on key points from the other contributors. I will also look in detail at the research findings in support of using lemmas in pedagogy and research because Brown et al.'s (2021) recommendations are contrasted strongly by Laufer (2021), as well as by other contributors (Kremmel, 2021; Nation, 2021). I will conclude with an agenda for further research.

Nation (2021) suggests that comparing lemmas and flemmas with word families is not particularly useful; they represent different levels of derivational knowledge that develop along with proficiency. Each of these levels of knowledge fall within Bauer and Nation's (1993) classification of word families with lemmas categorized as Level 2 word families and word families categorized as Level 6. Bauer and Nation's (1993) word family levels and Sasao and Webb's (2017) word part levels provide two models that can be examined in relation to derivational knowledge. Laufer et al. (2021) found that each of these models was only partially accurate in predicting derivational knowledge. However, relatively few derivative forms were examined and further research is needed. Considering derivational knowledge to be moving along a continuum across levels is useful. The rate at which derivational knowledge is gained is likely affected by frequency of encounters, deliberate learning, and prior knowledge of the derivational system. Furthermore, in the early stages of lexical development the different levels of derivational knowledge will likely vary greatly across words (learners may gain knowledge of all members of a word family for

*Corresponding author. E-mail: swebb27@uwo.ca.

some words and few members for other words) with gains and losses in knowledge related in part to the amount of instruction and L2 exposure.

Dang (2021) discusses lexical units in relation to word lists. Dang suggests a flexible approach to selecting lexical units for word lists that is based on the list purpose. This has been reflected in lists that she has developed. The Essential Word List (Dang & Webb, 2016a) was developed for beginning learners and so versions were made up of either headwords or lemmas. In contrast, the Academic Spoken Word List (Dang et al., 2017) was developed for learners at a variety of levels and so versions were made up of either word families or flemmas. Similar to Nation (2021) and Laufer (2021), Dang (2021) notes that lists made up of word families do not necessarily have to include all of Bauer and Nation's (1993) Level 6 word family members. She suggests that including only core members that include very frequent affixes might make sense. This corresponds with the recently developed Nuclear Family List (Cobb & Laufer, 2021), which is made up of the most frequent word family members and excludes the least frequent members. The Nuclear Family List is useful because it makes the value of the items in each word family more transparent than earlier lists of word families. Another approach that would also be useful to consider is to provide frequency information for all word family members similar to the frequency information provided in learner dictionaries. This would still allow teachers and learners to gain awareness of all the different derivational forms of words, but also see which ones are most important for learning.

Gablasova and Brezina (2021) discuss lexical units from a corpus-based perspective. They argue for lemma as the lexical unit in corpus linguistics research because it provides greater precision and fewer assumptions about knowledge in comparison to larger units (flemmas and word families), and corpus linguistics tools can easily identify and count lemmas within corpora. However, they also report that corpus-based analysis of lemmas (and other larger units) is limited because items are identified according to their forms, and classification of homonyms (e.g., bank), polysemous words (e.g., cloud), and component words within formulaic language (e.g., make up) lead to a lack of precision. They suggest that using lexemes (lemma + sense disambiguation) as the lexical unit would further increase accuracy of corpus-based analyses. Similar to Brown et al. (2020), they report that lemmas may also have greater value for pedagogical applications than larger lexical units. However, the pedagogical value of lemmas appears rather opaque and neither discussion reveals how lemmas may have a positive impact on pedagogy. From a corpus-linguistics perspective, greater precision in identification of learner errors may provide value to teachers. However, from a pedagogical perspective the value of word families is more transparent. There are thousands of word families to learn but tens of thousands of lemmas to learn. It would seem much more efficient to learn words as larger lexical units. Development of morphological awareness through learning words in families may contribute to far-reaching gains in the lexical development of L2 learners of English through improving skill in inferring the meanings of novel, morphologically complex words, and using knowledge of word structure and context to comprehend spoken and written text containing derivatives (Goodwin & Ahn, 2010). It also seems unlikely that teachers will have classroom time to teach high-frequency lemmas such as *age*, *attack*, *attempt*, and *break* individually as nouns and verbs, and even less likely that lemmas such as *accept* and *acceptable*, *design* and *designer*, and *measure* and *measurement* that have relatively large variation in their frequencies within Brezina and Gablasova's (2015) New

General Service List would be taught at different times. Teachers and learners may also find greater value in measuring knowledge using word families because the benefit of determining whether unrelated words are known is likely more transparent than determining whether related words are known. However, as Kremmel (2021) rightly notes, this is speculation on my part, and there is a clear need for research investigating teacher and learner perspectives on the value of lexical units.

Kremmel (2021) expands on several of the points addressed in my introduction to this critical commentary and questions the degree to which we might be confident about using any one lexical unit for research and pedagogy. This makes sense as it allows for learners' development in derivational knowledge. He provides a balanced account of earlier research findings and discusses advantages and disadvantages of using lemmas and word families in pedagogy, lexical coverage, and testing. I agree with his conclusion that a great deal more research is needed in relation to the different contexts and purposes in which lexical units are used.

Laufer (2021) and Brown et al. (2021) provide contrasting perspectives on lexical units. Laufer (2021) provides a convincing argument why claims that word families are inappropriate as the lexical unit in tests are exaggerated. She is the researcher who is conducting the most research in this area with four recent studies, two of which compare L2 learner knowledge of headwords and their derivatives on different tests (Laufer et al., 2021; Snoder & Laufer, 2021), and two that look at the occurrence of derivatives within corpora (Cobb & Laufer, 2021; Laufer & Cobb, 2020). I agree with Laufer (2021, p. 968) that the claims made to reassess established vocabulary tests (Stewart et al., 2021; Stoeckel et al., 2021), lexical coverage and profiling studies (McLean, 2021), learning goals, word lists, and approaches to vocabulary teaching (Brown et al., 2020) are "unsubstantiated and inappropriate."

The premise of Brown et al.'s (2021) article is that there has long been a paradigm in which researchers use word families in research and pedagogy and that with the recent publication of studies that have investigated lexical units, as well as articles strongly advocating for the use of lemmas in research and pedagogy (Brown et al., 2021; Brown et al., 2020; McLean, 2021; Stewart et al., 2021; Stoeckel et al., 2021), researchers are moving to a new paradigm in which lemmas replace word families in research and pedagogy. I strongly disagree with this because I do not believe there has been a paradigm in which researchers view word families as the only appropriate unit in L2 research and pedagogy. Studies of intentional vocabulary learning tend to use word types (e.g., word cards, word lists) and lemmas (e.g., fill-in-the-blanks, sentence production). Reynolds and Wible's (2014) survey of L2 vocabulary learning studies revealed that although words are occasionally counted as word families, word types and lemmas were by far the most common lexical units. Although there are well-known lists of word families such as Coxhead's Academic Word List (2000) and Nation's (2012) British National Corpus/ Corpus of Contemporary American English word lists, there are also many lemma-based word lists (e.g., Davies, 2011–; Leech et al., 2001). There are also lemma-based tests (e.g., Peters et al., 2019), as well as family-based tests (e.g., Coxhead et al., 2015; Webb et al., 2017). To my knowledge, the only area to exclusively use word families as the lexical unit is lexical profiling, and as Gablasova and Brezina (2021) report, this is likely due in part to word family lists being developed for this purpose and provided together with lexical profilers. In addition, if we look at studies by specific researchers, we find variation in the

lexical unit across studies. For example, I have used lemmas (e.g., Webb, 2007) and word types (e.g., Pavia et al., 2019) as the lexical unit in studies of incidental vocabulary learning; word types as the lexical unit in studies of intentional vocabulary learning (e.g., Rogers et al., 2015); word types and lemmas to create different versions of the Essential Word List (Dang & Webb, 2016a); flemmas and word families to develop different versions of the Academic Spoken Word List (Dang et al., 2018); word types to evaluate different word lists (Dang & Webb, 2016b); word types (e.g., Nguyen & Webb, 2017); and word families (Webb et al., 2017) when developing tests of vocabulary knowledge, but only word families in lexical profiling studies (e.g., Webb, 2010). Similarly, over three decades, Batia Laufer has used lemmas in her empirical studies on various factors, for example, teaching conditions, dictionary types, and language used to explain new words, that affect intentional vocabulary learning (Laufer & Osimo, 1991; Laufer & Shmueli, 1997) and incidental vocabulary learning (Hill & Laufer, 2003; Laufer, 2000, 2003; Laufer & Girsai, 2008; Laufer & Hill, 2000; Laufer & Rozovzki-Roitblat, 2011, 2015). Posttests of all the experiments assessed the learning of the target lemmas. There was no expectation that learners, who were at an initial stage of learning target words, would understand their derived forms as well.

Rather than a paradigm of using word families in pedagogy and research, the choice of lexical unit in research has varied across research questions and pedagogical purposes. The promotion of a paradigm in which only one lexical unit should be used has only occurred recently, and then can only be found in recent articles by Brown, McLean, and colleagues repeatedly arguing that the lemma or flemma is the only appropriate choice (Brown et al., 2021; Brown et al., 2020; McLean, 2021; Stewart et al., 2021; Stoeckel et al., 2021).

EVIDENCE THAT L2 LEARNERS LACK KNOWLEDGE OF DERIVATIONS

Brown et al. (2021) argue that while there may be little evidence indicating L2 learners lack the knowledge necessary to deal with derivational forms, it is evidence nonetheless and should be followed. This seems reasonable. However, it is also very important to look carefully at this evidence. There are four studies that are suggested to provide evidence indicating that L2 learners lack knowledge of derivations (Brown, 2013; Kremmel & Schmitt, 2016; McLean, 2018; Ward & Chuenjundaeng, 2009). I will discuss the degree to which each of these studies provide evidence to support Brown et al.'s (2021) claims.

Brown et al. (2021, p. 951) cite Brown (2013) as indicating that L2 learners "find inferring derivational forms encountered in context a challenge." Brown (2013) asked low-intermediate Japanese EFL learners to circle unknown words that they encountered when reading several short articles. He then compared the degree to which these words were headwords, inflections, and derivatives. Brown found that more derivatives at the 1000 (unknown derivatives = 40) and 3000 (unknown derivatives = 70) word levels were unknown than headwords (21 and 57). However, he also found that participants marked more headwords as unknown at the 2000, 4000, and 5000 levels (unknown headwords = 45, 113, 103, respectively) than derivatives (unknown derivatives = 29, 16, 72, respectively). Moreover, the overall number of unknown headwords (339) was much higher than for derivatives (227) and inflections (224). Together, this provides little if any

support that L2 learners lack knowledge or have difficulty inferring derivatives in comparison to headwords and inflections.

Kremmel and Schmitt (2016) measured participants' knowledge of derivations by the degree to which they could correctly spell headwords (e.g., accurate) after being provided with three derivatives (e.g., inaccuracy, accurately, accuracies). Scores on this test of derivational knowledge were then compared to test scores of form-meaning knowledge. Kremmel and Schmitt found that participants' ability to score correctly on their derivation test overlapped with measures of form-meaning knowledge for 63–67% of the target items, which led them to suggest that receptive tests of form-meaning connection may overestimate receptive knowledge of word family members. There are, however, several problems with this conclusion. First, the extent to which the test of derivational knowledge measured derivational knowledge is not clear. It was a productive test of orthography that involved assessing the degree to which test takers could successfully produce the written forms of headwords. The degree to which participants had receptive knowledge of the form-meaning connections of derivations or even the receptive and productive knowledge of the orthography of derivations was not measured. Although this test may provide useful information about test takers spelling of the written forms of headwords, there is little that can be inferred about receptive derivational knowledge.

There are two studies that have shown that L2 learners were unable to translate the meanings of base forms and affixes of all derivative test items. Ward and Chuenjundaeng (2009) found that low-level EFL learners' knowledge of Academic Word List (AWL) headwords and their derived forms varied. For some items, there was greater knowledge of the headword, while for others there was greater knowledge of the derivative. The degree to which headwords and their corresponding derivatives were successfully translated was not explicitly examined. However, Brown et al. (2020) reported that Ward and Chuenjundaeng's (2009) participants were unable to translate 49% of the derivatives for AWL headwords that they could translate. They also found that these learners could not translate a slightly smaller percentage of headwords (46%) for the derivatives that they could translate. This is perhaps the most convincing evidence that L2 learners lack knowledge of derivatives. However, the slight difference in receptive knowledge between headwords and derivatives for low-level learners provides relatively little support for the many claims made in relation to this topic.

McLean (2018) investigated knowledge of 14 headwords (e.g., standard, adjust) and their derivatives (e.g., nonstandard, standardize, substandard, standardization). Target items were presented in contexts with minimal information (e.g., The stereo is now **usable**.) and participants needed to translate the underlined target items. Responses scored as correct needed to demonstrate knowledge of both the base forms and any included affixes. McLean's results showed that Japanese L2 learners could translate a significantly greater proportion of headwords than their derivatives. Because McLean's (2018) study of receptive derivational knowledge seems to have spurred the recent discussions of lexical units, it is useful to consider the internal and external validity of the study to better understand the degree to which its findings should be generalized to other contexts.

There are several things to consider about McLean's (2018) research design. First, the primary conclusion of the study is that Japanese L2 learners lack receptive knowledge of derivatives. However, to make claims about a lack of knowledge, there should be a

baseline for comparison. This is particularly true in this case because there is a wealth of research indicating that L1 speakers struggle to demonstrate knowledge of derivatives (e.g., Derwing & Baker, 1979; McCutchen & Stull, 2015; Nagy et al., 1993; Tyler & Nagy, 1990; Wysocki & Jenkins, 1987). Because the meaning recall test used in McLean's (2018) study was quite demanding, L1 speakers are also likely to score lower on derivatives than headwords included in the study. For example, L1 speakers may have the same difficulty as L2 learners at providing the meanings of both the base forms and included affixes for the following target items: *encirclement, publishability, collectivization, standardization, maladjusted, countermove, centralized, teacherly, developmentally, antidevelopment, teachability, acceptability, maintainability.* Without an L1 speaker baseline, we do not know whether L2 learners' scores on the test are similar or lower than L1 speakers. Iwaizumi and Webb (2021a) found that L2 productive knowledge of derivation on a decontextualized form recall test increases with vocabulary knowledge and there was no statistically significant difference between the scores of L1 learners and advanced L2 learners. However, this likely depends to some degree on the test format. Iwaizumi and Webb (2021b) found that L1 learners were able to score significantly higher than advanced L2 learners in the production of derivatives in a contextualized form recall test. Thus, the degree to which the results from one test format may apply to another is a very important consideration when generalizing beyond a data set.

A second aspect of McLean's (2018) research design worth considering is the ecological validity of the contexts in which target items were encountered as well as the degree of difficulty in inferring the meanings of headwords, inflections, and derivations from those contexts. Derivations are considered to be relatively easy to infer from context when their base forms are known. The similarity in form and meaning between headwords (*use*), inflections (*uses, used, using*), and derivations (*useful, useless, user*) may facilitate comprehension of different word family members in comparison to words with unrelated forms and meanings (*help, end, useless*) (Bauer & Nation, 1993). The overlap in form and meaning should reduce learning burden in the same way that research indicates that it does for cognates (e.g., Rogers et al., 2015). For example, Peters and Webb (2018) found that the odds of successfully learning cognates incidentally through viewing television were 2.5 and 8 times higher than noncognates in two experiments. The degree to which similarity between forms and meanings of word family members is recognized (as well as the L1 and L2 forms of cognates) is likely dependent to some degree in the sentences and surrounding contexts in which they are encountered. The individual sentences in which derivatives were encountered in McLean's (2018) study (e.g., His <u>countermove</u> was expected. The <u>collectivization</u> of farming was very fast. She is very <u>teacherly</u>. The group is <u>antidevelopment.</u> The use of the car was <u>nonstandard</u>. The house has limited <u>maintainability</u>. <u>Standardization</u> is common in farming.) were likely much more difficult to infer the meanings of target items than longer passages and contexts that L1 and L2 learners typically encounter. If we are aiming to determine whether L2 learners can infer the meanings of derivations when they are encountered, it makes little sense to present these words in contexts with little ecological validity. Moreover, the meanings of headwords and inflections in the sentences from the test often appear to be more transparent than those of derivatives (e.g., Pandas <u>move</u> slowly. The pandas are <u>moving</u> slowly. The panda <u>moved</u> slowly. The panda *has* <u>moved</u> from the tree to the window.).

A third aspect of McLean's (2018) research design that is worth considering further is the scoring of responses. Research has consistently indicated that recall tests are less sensitive to vocabulary knowledge than recognition tests (e.g., Laufer & Goldstein, 2004; Nagy et al., 1985). This is because test takers may often produce responses that demonstrate knowledge but do not meet the scoring criteria. For example, responses such as *working*, *okay*, and *repaired* would all make perfect sense within the context of *The stereo is now **usable.*** Although this might also occur with headwords and inflections, because derivations required a rather complex response to be scored as correct (knowledge of the meanings of base forms and affixes) this may have occurred more often with derivatives. In addition, although it is reasonable to score responses for derivatives that demonstrated knowledge of only the meanings of the base forms (e.g., use) as incorrect, it is not clear whether that indicates that participants did not have knowledge of the affixes (e.g., -able) or were unable to understand the context. In paper-and-pencil recall tests where partial knowledge is demonstrated such as in this study, there is no way to determine whether participants did or did not have knowledge of the items. Thus, responses might be scored as incorrect when participants do have knowledge of the target item. Schmitt and colleagues have often used interviews with follow-up prompts to more accurately gauge knowledge in recall tests to avoid these problems (e.g., Pellicer-Sánchez & Schmitt, 2010; Schmitt, 1998).

If we are to heed this evidence that L2 learners lack receptive knowledge of derivations as Brown et al. (2021) recommend, we need to make the following inferences from McLean's (2018) findings: (1) infer that the results were valid and reliable, (2) infer that the results are relevant to encounters with derivative forms in meaningful contexts, (3) infer that the participant sample can be generalized to other participant samples, (4) infer that the test results can be generalized to other test formats, (5) infer that results for the 14 target words can be generalized to other target words, and (6) infer that L1 learners would score perfectly on the test. In addition, this evidence has also been used by Brown, McLean, Stoeckel, and colleagues to reach the following conclusions: (1) the use of vocabulary learning goals and word lists using word families should be reconsidered (Brown et al., 2020), (2) approaches to teaching vocabulary may need to be reconsidered (Brown et al., 2020), (3) tests such as the Vocabulary Levels Test and Vocabulary Size Test overestimate vocabulary knowledge and should be replaced with lemma-based tests (Stewart et al., 2021; Stoeckel et al. 2021), and (4) lexical profiling and lexical coverage research has limited validity (McLean, 2021). Because of the absence of any research indicating that the use of lemmas is more beneficial than word families for (1) vocabulary learning goals and word lists, (2) approaches to teaching vocabulary, (3) the Vocabulary Levels Test and Vocabulary Size Test, and (4) lexical profiling and lexical coverage, the claims made by Brown, McLean, Stoeckel and colleagues are not warranted (Brown et al., 2021; Brown et al., 2020; McLean, 2021; Stewart et al., 2021; Stoeckel et al., 2021). This is not an endorsement for word families as the lexical unit for use in all aspects of pedagogy and research for all L2 learners; the appropriacy of lexical units is likely to vary across research and practical purposes. Lemmas are likely to have similar or greater appropriacy for some aspects of pedagogy and research. However, we should be cautious about making such claims about the different lexical units.

SETTING A RESEARCH AGENDA

The most useful area for research on lexical units is examining comprehension and use of lemmas, flemmas, lexemes, and derivations in meaningful contexts. The reason for this is that discussion of other research topics (testing, coverage) should be based on the degree to which L2 learners of varying proficiency levels and vocabulary knowledge can understand and produce word types, lemmas, and families. Investigating the extent to which L2 derivative forms of known headwords in learning materials such as graded readers and course books are encountered and understood would be a useful starting point. Studies could examine the degree to which learner variables (proficiency, prior vocabulary knowledge, L1, learning context), text variables (text type, text length, frequency of encounters, contextual clues), lexical variables (derivation frequency, word family frequency, word length, part of speech), and test format (meaning recognition, form recognition, meaning recall, form recall, cued recall, sentence production) moderate knowledge.

It would also be of great benefit to investigate the learning of derived forms and the development of morphological awareness (the ability to understand and manipulate word parts) through different types of instruction. In the introduction to this critical commentary, I suggested that one of the benefits of larger lexical units to pedagogy is that presenting headwords together with their inflections and derivations may provide a shortcut to lexical development. However, as Dang (2021) notes, teachers may not explicitly focus on instruction designed to promote gains in derivational knowledge and morphological awareness. The value of developing morphological awareness is clear; lexical development involves understanding and using thousands of derived and inflected forms. Although there are many studies examining the benefits of morphological interventions for L1 learners (e.g., Goodwin & Ahn, 2010, 2013), this line of research has been neglected within L2 studies. Research examining the degree to which different instructional interventions facilitate gains in derivational knowledge, morphological awareness, and vocabulary breadth may spur improvement in this aspect of L2 pedagogy.

A third key area for research is to examine teacher and student perceptions and uses of learning materials and activities, and tests that use word types, lemmas, flemmas, and families. The development of lists of word families such as Nation's (2012) British National Corpus/Corpus of Contemporary American English word lists and Coxhead's (2000) Academic Word List was based on the pedagogical value of presenting headwords together with their inflected and derived forms to aid learning. However, there is little information available about the degree to which words are taught and learned as types, lemmas, and families, or whether teachers and learners perceive lists of word families to be useful. In addition, there would be great value in looking at the types of tests that teachers and students value for measuring vocabulary knowledge. In my introduction to this critical commentary, I suggested that an advantage to administering tests that use word families as the lexical unit is that they measure knowledge of distinct words (e.g., *happy*, *sad*, *love* rather than morphologically related words, e.g., *happy*, *happiness*, *unhappy*). This was also based in part on my experiences as a language teacher and learner. My perception is that both teachers and learners would find much greater value in measuring knowledge of word families than lemmas, together with tests that assess knowledge of word parts, because this provides them with more useful diagnostic

information; determining knowledge of distinct words and affixes may provide a more useful measure of lexical development than measuring knowledge of morphologically related words. Understanding the needs of teachers and learners should guide the use and development of tools such as word lists and tests, and so there is great value in researching this area.

Much of the recent discussion of lexical units (e.g., Brown et al., 2020; Kremmel, 2016; McLean, 2018; Stewart et al., 2021; Stoeckel et al., 2021; Webb, 2021) has focused on assessment and the degree to which tests that use lemmas and word families may accurately evaluate vocabulary knowledge. Laufer and colleagues (Laufer et al., 2021; Snoder & Laufer, 2021) have begun to look at this topic by comparing the results of tests that employ headwords or their derivatives as target items, as well as the extent to which different factors (proficiency, derivation frequency, word family frequency, affix type) moderate scores. This is a useful approach to determining the degree to which lemma- and family-based tests may affect estimations of vocabulary size and levels. It would also be useful to examine the degree to which the scores of different measures of derivational knowledge are correlated with those on tests with different lexical units. The degree to which knowledge of derivations and lemma- and family-based tests correlate should provide some indication of the degree to which tests of vocabulary size and levels tap into morphological knowledge. We might expect lemma-based tests to have a higher correlation with derivational knowledge because part of the value of using lemmas as the lexical unit is that they may provide a better indication of whether morphologically complex words are known. However, because vocabulary size and levels tests have relatively few test items, it may be that tests with both lexical units provide a similar correlation with knowledge of derivations. Investigating the relationships between scores on a test of affix knowledge and lemma- and family-based tests might further clarify the degree to which tests that use different lexical units are effective in gauging knowledge of the morphological system.

Another avenue for research is the development of tests designed to measure derivational knowledge. Tests that measure the degree to which learners can recognize and produce morphologically complex words might be useful diagnostic tools to assess lexical development more accurately in addition to existing tests of vocabulary size (e.g., Aviad-Levitzky et al., 2019; Coxhead et al., 2015), levels (Webb et al., 2017), and word part knowledge (Sasao & Webb, 2017). Tests of derivational knowledge using meaning recognition formats similar to those used to measure knowledge of form-meaning connection would provide a useful comparison to these aspects of receptive vocabulary knowledge. Productive tests that assess the degree to which test takers can produce different derivative forms of target items might provide a useful indication of the degree to which learners can use words in writing. Schmitt and Zimmerman (2002) and Iwaizumi and Webb (2021a, 2021b) provide examples of how this might be done.

A final direction for further research is the development of lemma-based tests of form-meaning connection. This is a high-risk/high-reward topic because test development and initial validation takes a relatively long time, and lemma-based tests may have less face validity and provide similar results to existing tests of form-meaning connection. However, lemma-based tests could add precision for investigating some research questions such as the lexical development of those in the initial stages of L2 learning. Moreover, lemma-based tests could be designed to investigate the development of derivational

knowledge through examining recognition and production of derivations by affix type, as well as providing greater insight into how different learner (e.g., proficiency, L1, learning context, exposure to L2 input), lexical (derivation frequency, word family frequency, word length, part of speech, affix type), and test (meaning recognition, form recognition, meaning recall, form recall, cued recall, sentence production) variables moderate knowledge.

CONCLUSION

This critical commentary has highlighted a lack of agreement about the value of different lexical units for research and pedagogy. This is because lemmas and word families are not dichotomous options for which only one is appropriate for a given purpose. In all likelihood, types, lexemes, lemmas, flemmas, and word families will each have advantages and disadvantages for a given purpose, and there might occasionally be little difference between the benefits of the options. To some degree I have highlighted the value of word families in this article. This is not because I believe word families to be the only appropriate option, but because articles advocating the value of lemmas (Brown et al., 2020; McLean, 2021; Stewart et al., 2021; Stoeckel et al., 2021) have exaggerated the internal and external validity of earlier findings, have not adequately addressed the complexity of the issue, and have not addressed the need for further research. I hope that this critical commentary will provide useful justification for future studies investigating the relative values of lexical units to research and pedagogy, and that it will stimulate further research.

REFERENCES

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*, 345–368.

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*, 253–279.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, *36*, 1–22.

Brown, D. (2013). Types of words identified as unknown by L2 learners when reading. *System*, *41*, 1043–1055.

Brown, D., Stoeckel, T., & McLean, S. (2021). The coming paradigm shift in the use of lexical units. *Studies in Second Language Acquisition*, *43*, 950–953.

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). The Nuclear Word Family List: A list of the most frequent family members, including base and affixed words. *Language Learning*, *71*, 834–871.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213–238.

Coxhead, A., Nation, P., & Sim, D. (2015). Measuring the vocabulary size of native speakers of English in New Zealand secondary schools. *New Zealand Journal of Educational Studies*, *50*, 121–135.

Dang, T. N. Y. (2021). Selecting lexical units in wordlists for EFL learners. *Studies in Second Language Acquisition*, *43*, 954–957.

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The Academic Spoken Word List. *Language Learning*, *67*, 959–997.

Dang, T. N. Y., & Webb, S. (2016a). Making an essential word list for beginners. In I. S. P. Nation, *Making and Using Word Lists for Language Learning and Testing* (pp. 153–167, 188–195). John Benjamins.

Dang, T. N. Y., & Webb, S. (2016b). Evaluating lists of high frequency vocabulary. *ITL–International Journal of Applied Linguistics*, *167*, 132–158.

Davies, Mark. (2011–). *Most frequent 100,000 word forms in English (based on data from the COCA corpus).* https://www.wordfrequency.info/

Derwing, B. L., & Baker, W. J. (1979). Recent research on the acquisition of English morphology. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 209–223). Cambridge University Press.

Gablasova, D., & Brezina, V. (2021). Words that matter in L2 research and pedagogy: A corpus-linguistics perspective. *Studies in Second Language Acquisition*, *43*, 958–961.

Goodwin, A. P., & Ahn, S. (2010). A meta-analysis of morphological interventions: Effects on literacy achievement of children with literacy difficulties. *Annals of Dyslexia*, *60*, 183–208.

Goodwin, A. P., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. *Scientific Studies of Reading*, *17*, 257–285.

Hill, M. M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics*, *41*, 87–106.

Iwaizumi, E., & Webb, S. (2021a). Measuring L1 and L2 productive derivational knowledge: How many derivatives can L1 and L2 learners with differing vocabulary levels produce? *TESOL Quarterly.* Advance online publication. https://doi.org/10.1002/tesq.3035

Iwaizumi, E., & Webb, S. (2021b). To what extent does productive derivational knowledge of adult L1 speakers and L2 learners at two educational levels differ? [Manuscript submitted for publication]. Faculty of Education, University of Western Ontario.

Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly*, *50*, 976–987.

Kremmel, B. (2021). Selling the (word) family silver? A response to Webb's "lemma dilemma." *Studies in Second Language Acquisition*, *43*, 962–964.

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly*, *13*, 377–392.

Laufer, B. (2000). Electronic dictionaries and incidental vocabulary acquisition: Does technology make a difference? In U. Heid, S. Evert, E. Lehmann, C. Rohrer (Eds.). *EURALEX, Proceedings of the 9th EURALEX International Congress* (pp. 849–854). Institutfür Maschinelle Sprachverarbeitung, Universität Stuttgart.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? *Canadian Modern Language Review*, *59*, 565–585,

Laufer, B. (2021). Lemmas, flemmas, word families and common sense. *Studies in Second Language Acquisition*, *43*, 965–968.

Laufer, B., & Cobb, T. (2020). How much knowledge of derived words is needed for reading? *Applied Linguistics*, *41*, 971–998.

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, *29*, 694–716.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptive-ness. *Language Learning*, *54*, 399–436.

Laufer, B., & Hill, M. M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning and Technology*, *3*, 58–76.

Laufer, B., & Osimo, H. (1991). Facilitating vocabulary retention: The second hand cloze. *System*, *19*, 217–224.

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research*, *15*, 391–411.

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*, *19*, 687–711.

Laufer, B., & Shmueli, K. (1997). Memorizing new words: does teaching have anything to do with it? *RELC Journal*, *28*, 89–108.

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

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus.* Longman.

McCutchen, D., & Stull, S. (2015). Morphological awareness and children's writing: Accuracy, error, and invention. *Reading and Writing*, *28*, 271–289.

McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, *39*, 823–845.

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.

Nagy, W. E., Diakidoy, I. A. N., & Anderson, R. C. (1993). The acquisition of morphology: Learning the contribution of suffixes to the meanings of derivatives. *Journal of Reading Behavior*, *25*, 155–170.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, *20*, 233–253.

Nation, I. S. P. (2012). The BNC/COCA word family lists. http://www.victoria.ac.nz/lals/about/staff/paul-nation

Nation, P. (2021). Thoughts on word families. *Studies in Second Language Acquisition*, *43*, 969–972.

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, *21*, 298–320.

Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning from listening to L2 songs. *Studies in Second Language Acquisition*, *41*, 745–768.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*, 31–55.

Peters, E., Velghe, T., & Van Rompaey, T. (2019). The VocabLab tests: The development of an English and French vocabulary test. *ITL–International Journal of Applied Linguistics*, *170*, 53–78.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through watching a single episode of L2 television. *Studies in Second Language Acquisition*, *40*, 551–577.

Reynolds, B. L., & Wible, D. (2014). Frequency in incidental vocabulary acquisition research: An undefined concept and some consequences. *TESOL Quarterly*, *48*, 843–861.

Rogers, J., Webb, S., & Nakata, T. (2015). Do the cognancy characteristics of loanwords make them more easily learned than noncognates? *Language Teaching Research*, *19*, 9–27.

Sasao, Y., & Webb, S. (2017). The word part levels test. *Language Teaching Research*, *21*, 12–30.

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, *48*, 281–317.

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, *36*, 145–171.

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.

Stewart, J., Stoeckel, T., McLean, S., Nation, P., & Pinchbeck, G. G. (2021). What the research shows about written receptive vocabulary testing: A reply to Webb. *Studies in Second Language Acquisition*, *43*, 462–471.

Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, *43*, 181–203.

Tyler, A., & Nagy, W. (1990). Use of derivational morphology during reading. *Cognition*, *36*, 17–34.

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

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*, 46–65.

Webb, S. (2010). Pre-learning low frequency vocabulary in second language television programs. *Language Teaching Research*, *14*, 501–515.

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*, 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*, 34–70.

Wysocki, K., & Jenkins, J. (1987). Deriving word meanings through morphological generalization. *Reading Research Quarterly*, *22*, 66–81.