

# A Review of McLean’s Evidence for the Adoption of the *Flemma* as an Appropriate Word Counting Unit

Yajie LI

Graduate School of Integrated Arts and Sciences

Hiroshima University

Simon FRASER

Institute for Foreign Language Research and Education

Hiroshima University

Jon CLENTON

Graduate School of Integrated Arts and Sciences

Hiroshima University

## INTRODUCTION

Studies show that vocabulary knowledge contributes to language proficiency (Milton, 2013; Qian & Lin, 2019). Such studies also indicate that morphological knowledge (i.e., of inflected, derived, and base forms) is an essential part of vocabulary knowledge. Once we consider knowledge of morphology to be an integral aspect of vocabulary knowledge, we need to acknowledge the *word family* (Bauer & Nation, 1993) as a counting unit, with different morphological levels (see Appendix 1). The creation of the word family levels also implies proficiency related knowledge (Nation, 2021), if, for instance, low-level learners have not acquired comprehensive word family knowledge because of their language proficiency. A more practical comparison might exist between *lemmas* and word families, with the lemma as a part of the word family levels. Bauer and Nation’s second word family level can be taken as the lemma level (a headword and its inflected forms), with a *flemma* consisting of a headword and its inflected forms comprising different parts of speech.

There has been support in the recent literature for the lemma or *flemma* as a suitable word unit for English language learners (Brown et al., 2022; McLean, 2018; Stoeckel et al., 2020). These studies call attention to the widespread acceptance of word family levels as a word unit in language learning, teaching, and vocabulary tests, and suggest that any estimation of word part knowledge depends on how “words” are counted. McLean’s (2018) article questions the validity of using word families as a word counting unit. His article, reviewed below, considers a potential gap in word family-based research, and focuses on participants’ ability to comprehend inflected and derived forms. Investigating a cohort of L1 Japanese learners at different proficiency levels, his findings highlight a lack of knowledge of all participants regarding inflectional and derivational forms. McLean concludes that the *flemma* is a more appropriate word unit than the word family, at least for this specific group of language learners. In this review article, we examine and evaluate his evidence for this claim.

## Explanations and Definitions of Key Terms

There are a number of ways in which words can be counted, and these can vary according to the pur-

pose of the count as well as researcher preference. What follows is a brief introduction to the key terms. Word families comprise seven different levels of affixes based on eight level-ordering criteria, and contain a headword plus its inflected and derived forms (Bauer & Nation, 1993). McLean’s article uses the term “WF6” to refer to the word family levels to level 6 based on Bauer and Nation’s word family criteria. Lemmas comprise the base word and its inflected forms with the same part of speech (POS); these include plural, third person singular, present tense, past tense, past participle, *-ing*, comparative, superlative, and possessive forms. The difference between lemmas and flemmas is that flemmas treat the words in their inflected forms with different POS as the same word. In Table 1 below, we summarize the key terms, taking the word *abstract* as an example.

**TABLE 1. Key Terms and Definitions**

Key Term	Definition
Word unit	The lexical unit used in the analysis and counting of words in vocabulary research. The most common word units include <i>tokens</i> (the running words in a text), <i>word types</i> , <i>word families</i> , <i>lemmas</i> , <i>flemmas</i> , and <i>multiword units</i> .
Word type	Each unique word occurring in a text is counted as a different word.
Word family	Seven different levels proposed by Bauer and Nation (1993) (see Appendix A for more information). Word families consist of a headword with its inflected forms and most derived forms. If we use the word family as a word unit, the inflected forms of “abstract” ( <i>abstract</i> , <i>abstracts</i> , <i>abstracting</i> , <i>abstracted</i> ) and derived forms ( <i>abstractedly</i> , <i>abstractly</i> , <i>abstractness</i> , <i>abstraction</i> , <i>abstractions</i> ) are counted as the same word. A word family count assumes that learners have the knowledge of inflected and derived forms of the words.
WF6	The term used in McLean (2018) to refer to word families excluding level 7 of Bauer and Nation (1993).
Lemma	A headword with its inflected forms of the same part of speech (POS). If we use the lemma as a word unit, the adjective <i>abstract</i> , noun <i>abstract/ abstracts</i> , and verb <i>abstract/ abstracts/ abstracted/ abstracting</i> are counted as three different words. A lemma count assumes that learners have knowledge of inflected forms but do not have POS knowledge.
Flemma	Like the lemma, but does not distinguish the POS of words. If we use the flemma as a word unit, the adjective ( <i>abstract</i> ), noun ( <i>abstract/ abstracts</i> ), and verb ( <i>abstract/ abstracts/ abstracted/ abstracting</i> ) are all counted as one word. A flemma count assumes that learners have the knowledge of inflected forms and can distinguish POS.

## A SUMMARY OF MCLEAN (2018)

McLean's (2018) article is important because it addresses the need to evaluate word unit choice in second language research and teaching. McLean criticizes the common practice of accepting the use of word families, as categorized by Bauer and Nation (1993). He argues that learners may not be able to recognize all the word family members belonging to the same base word, resulting in an overestimation of learner vocabulary knowledge. McLean reports on an investigation in which he assessed 279 L1 Japanese L2 English language learning participants from three different proficiency levels, as determined by the New Vocabulary Levels Test (NVLT, McLean & Kramer, 2015). He investigates whether the use of the flemma, rather than Bauer and Nation's word family, provides a better estimate of participant vocabulary knowledge at three different proficiency levels. McLean concludes that the flemma is a more appropriate unit for estimating the vocabulary knowledge of language learners than the word family.

### Issues Relating to Levels of the Word Family

McLean (2018) begins by outlining the importance of inflected and derived forms of words in vocabulary learning. He follows Bauer and Nation's (1993) categorization of word families into seven levels (*each form is a different word* + level 2–level 6). Referring in his paper to a word family as WF6 (Bauer and Nation's affix criteria), McLean highlights two major issues. The first relates to how Bauer and Nation (1993) define *word family*:

From the point of view of reading, a word family consists of a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately. (p. 253)

With this claim, he defines the word family from a reading perspective, and the learning of new words implies the learning of new word families. The notion of WF6 knowledge stems from the idea that learners who know one form of the word family are likely to know other forms that belong to the same base word (Webb, 2010). McLean (2018), however, disagrees with this view, especially for low-proficiency language learners, who may not know the meaning of words that are derived forms of a newly learnt word, such that knowledge of *marmelize* may not imply a knowledge of the word *marmelization*. McLean also contends that because the WF6 standard emanates from an L1 English speaker knowledge standard, it does not follow that English language learners with a different L1 such as Japanese will share the same knowledge. He argues that the use of the flemma or lemma may be more appropriate than WF6.

### Evidence Concerning the Use of the Lemma or Flemma

McLean suggests that the lemma or flemma may be more appropriate in assessing L2 language learners than the conventionally accepted WF6 standard, and he presents five major studies to support this. Schmitt (2010) states that the lemma is a better choice of word counting unit. The other four studies (Schmitt & Meara, 1997; Mochizuki & Aizawa, 2000; Ward & Chuenjundaeng, 2009; Sasao & Webb, 2015) provide evidence that the flemma is a more appropriate word unit than WF6. The participant learners in these studies demonstrate that they have limited knowledge of the inflected and derived forms of word knowledge. Learners with different proficiency levels show differently acquired knowledge of inflected and derived forms.

Although they provide evidence in support of his argument, McLean observes that the four studies are not without their shortcomings. First, there is a lack of reliable data collected in these studies; second, two of the studies (Schmitt & Meara, 1997; Ward & Chuenjundaeng, 2009) only test knowledge of suffixes, and do not include validation of prefix assessment; third, the studies do not test knowledge of multiple affixes for the same base word; and fourth, the multiple-choice test format used might overestimate learner morphological knowledge.

To question whether word families are appropriate word units in estimating word knowledge, McLean called for an investigation of the morphological knowledge of WF6 words with multiple affixes. His study attempts to answer the following questions:

- 1) Are there significant differences between L1 Japanese learners' ability to comprehend the base form and their ability to comprehend the inflectional and derivational forms of the same word family?
- 2) Are there significant differences between L1 Japanese learners' ability to comprehend the base form and their ability to comprehend the inflected forms of the same word family?
- 3) Can use of the flemma overestimate or underestimate learners' ability to understand the inflected forms and derived forms?

### McLean's Study

The participants in McLean's investigation were L1 Japanese undergraduates ( $N=279$ ). They were required to complete the New Vocabulary Levels Test (NVLTL, McLean & Kramer, 2015), using a bilingual Japanese version, within 30 minutes. The participants were asked to complete 24 multiple-choice items for each of the first five 1,000-word bands, based on the BNC/COCA (British National Corpus / Corpus of Contemporary American English) word bands. McLean then divided the participants into three groups based on the scores from the NVLT, which included a beginner group ( $n=85$ ), an intermediate group ( $n=177$ ), and an advanced group ( $n=17$ ). An ANOVA (one-way analysis of variance) showed that there were significant differences between the groups ( $p<0.001$ ).

To measure participant vocabulary knowledge of the inflected and derived forms of English words, McLean used a comprehension test which presented high frequency words from the first 2,000 word families of the British National Corpus (BNC). Twelve words (*use*, *move*, *collect*, *center*, *teach*, *accept*, *maintain*, *develop*, *standard*, *circle*, *adjust*, and *publish*) were selected for the test because these words include many inflected and derived forms, according to Bauer and Nation's (1993) word family criteria. For the comprehension test, McLean used 100 sentences in which both inflected forms and derived forms were embedded. The participants were required to translate the underlined L2 English words into their corresponding L1, Japanese, within 30 minutes. The inflected and derived forms of the same base word were presented in the same sets (see the following example sentences for the word *use*, included in the test).

1. He is useless. = \_\_\_\_\_
2. How do you use this? = \_\_\_\_\_
3. He used the computer yesterday. = \_\_\_\_\_
4. The computer is now usable = \_\_\_\_\_
5. He is using the computer. = \_\_\_\_\_

6. He has used the computer all day. = \_\_\_\_\_
7. He is a user. = \_\_\_\_\_
8. Computers are very useful. = \_\_\_\_\_
9. The usage of this word is common in law. = \_\_\_\_\_
10. Please reuse the paper. = \_\_\_\_\_
11. The bag is reusable. = \_\_\_\_\_

Both the NVLT and the sentence comprehension test were scored; multiple raters were used to score the comprehension test, with an inter-rater reliability of over 0.91 using Kappa analysis. In response to each research question, McLean used Cochran's Q test to analyze the data, and treated the base form, inflected forms, and derived forms as repeated measures. The dichotomous data acquired significant differences between the base form, inflected form, and derived form. For his third research question, McLean used McNemar chi-square to investigate whether the adoption of the flemma was appropriate for his participants.

### Findings

Regarding his first research question, McLean hypothesized that if there were significant differences found in Cochran's Q analysis, this would indicate that participants differed in their ability to understand the base words, the inflected forms, and the derived forms of the words, and so would not support the adoption of WF6 as an appropriate word-counting unit. However, if no significant difference was found between the base form, the inflected forms, and the derived forms, the adoption of WF6 as a word counting unit would be appropriate. The Cochran's Q analysis indicated a significant difference regarding the number of correct responses to the base word and other members of the same WF6. The large effect size in the study also showed that the participants differed considerably in their ability to understand the base forms and WF6 forms.

In response to his second research question, McLean hypothesized that participants had the same ability to understand the base and inflected forms. The results showed that the eight tested words (*use*, *move*, *collect*, *teach*, *accept*, *maintain*, *adjust*, and *publish*) showed no significant difference in the flemmas. Only three words (*center*, *develop*, and *circle*) indicated significant differences, but the effect size was small. These results showed that participants have the same ability to understand the base word and its corresponding inflected forms, indicating that the flemma is an appropriate word unit for the L1 Japanese participants.

In relation to the third research question, the results (see Table 2) showed that using the flemma as a word unit underestimated participant derived knowledge of *-er* for the three tested words *use*, *teach*, and *publish*, but not for *develop* and *adjust*. Regarding inflected knowledge, using flemmas would overestimate the tested words *center* (with *-ed*, *-ing*, and *have -ed*), *develop* (with *-ing* adjective), and *circle* (with *-ed*, *-ing*, and *have -ed*). The effect size values were minimal with the existing significant differences and were therefore tolerated. As with the advanced-level participants ( $n=17$ ), using the flemma as a word unit would not overestimate knowledge but could underestimate 19 derived forms that had been tested.

TABLE 2. The Significance and Effect Size of Differences in the Number of Participants Who Comprehend Base Forms and the Number of Participants Who Comprehend Associated Inflected Forms and Derivational Forms

word form	use	move	collect	center	teach	accept	maintain	develop	standard	circle (verb)	adjust	publish
-ed	0.01	0.0	0.0	0.08*	0.01	0.01	0.01	0.03		0.04*	0.01	0.01
-ing	0.0	0.01	0.02	0.12*	0.01	0.02	0.03	0.03		0.07*	0.02	0.01
-ing adjective								0.09*				
have -ed	0.03	0.02	0.03	0.13*	0.03	0.03	0.03	0.03		0.07*	0.01	0.01
-er	0.02				0.01			0.16*			0.1*	0.01
-able	0.2*		0.2*								0.26*	
-less	0.41*											
...	...	...	...	...	...	...	...	...	...	...	...	...

*Note.* Effect sizes ( $\Phi$ ) of differences between the number of participants who comprehend base words and associated inflected or derived forms.

\*Significant difference. Alpha values for comparisons established by using the Bonferroni adjustment.

McLean's results show that the participants in his study differed in their ability to comprehend the base forms and WF6, demonstrating that using WF6 as a word counting unit is inappropriate for measuring language ability. Using the flemma as the word counting unit only slightly overestimated the beginner and intermediate group participants. McLean, therefore, suggests that the flemma is an appropriate word counting unit for L1 Japanese learners.

McLean highlights the problems of current vocabulary tests that are based on word family counts such as the Vocabulary Size Test (VST, Nation & Beglar, 2007) and the Vocabulary Levels Test (VLT, Nation, 1983), and he offers an alternative in order to build reliable vocabulary knowledge measures. The vocabulary tests based on word families can overestimate participants' vocabulary knowledge, because learners can guess the meanings of the words when selecting from multiple-choice phrases with a knowledge of the base form of the words. McLean proposes building an accurate measurement of language learners' vocabulary knowledge in three ways: (a) knowing the participants' inflectional and derivational knowledge levels before conducting vocabulary tests and delivering vocabulary lists; (b) using derived forms mainly known by participants in the research (i.e., L1 Japanese learners); and (c) adopting a flemma counting unit as a practical solution for L1 Japanese participants.

### **A CRITIQUE OF MCLEAN (2018)**

McLean's article is a valuable contribution to the research relating to inflected and derived vocabulary knowledge. His results show that participants, especially those at beginner and intermediate levels, have very limited knowledge of derivational forms. To avoid using an inappropriate word-counting unit that overestimates participant morphological knowledge of vocabulary, researchers should be wary of using WF6. McLean proposes the adoption of the flemma as a practical solution. While his article offers an innovative way of understanding the vocabulary knowledge of L1 Japanese participants, suggesting the flemma as an appropriate word unit to assess word part knowledge, it is not without its weaknesses. These include the number of words selected in the study, the suitability of using the flemma for assessing language skills other than reading, and the fact that McLean only included learners with a single L1, Japanese.

#### **Number of Words Selected**

The number of words presented in the study is likely to be problematic. McLean tested 12 English words from the first 2,000 of the BNC on the basis that the low-proficiency Japanese participants could understand these 12 English words. These words had 100 different inflected and derived forms in total, in accordance with the word family levels suggested by Bauer and Nation (1993). Using low-frequency English words ( $n=12$ ) to judge the morphology knowledge of participants with different language proficiency levels raises concerns, because these 12 English words cannot comprehensively represent participant knowledge of inflected and derived forms of English words. Even the participants identified as being at the same proficiency levels demonstrated discrepancies or individual differences in their inflectional and derivational knowledge of English words belonging to the same or different frequency bands. Including more English words at a wide range of frequency levels is worth investigating in future studies. Two recent studies carried out by Iwaizumi and Webb (2021, 2022) suggest that learners' derived vocabulary knowledge is associated with their proficiency levels and vocabulary size.

### **Suitability of the Flemma for Assessing All Language Skills**

A problem remains with using the flemma as a word unit for second language research and teaching. McLean's article focuses on participant understanding of morphological knowledge mainly of a single receptive language skill (reading), not for other language skills (listening, speaking, and writing). Brown et al. (2022) suggest that for second language studies a smaller word unit, the lemma or flemma, should be adopted based on a review of the previous morphology studies. The lemma comprises the base word and its inflected forms of the same part of speech (POS) in the English language, including plural, third person singular, present tense, past tense, past participle, *-ing*, comparative, superlative, and possessive forms. The only difference between the lemma and the flemma is that the flemma treats the words in their inflected forms with different POS as the same word. In other words, a flemma as a word unit can include more members than a lemma. However, a lack of sufficient empirical research in this area sheds significant doubt on McLean's suggestion. In addition, morphological knowledge is linked to language proficiency, vocabulary size, and other related factors. Participants with the same proficiency levels may have different derivational knowledge of individual words. Morphological knowledge may to some extent depend on different language skills. The word unit (flemma) recommended in McLean's study for reading may not be suitable for the other language skills (i.e., listening, speaking, and writing). A recent paper (Maw et al., 2022), for instance, suggests that two different word counts (flemma and lemma) might present different interpretations of writing proficiency. What, therefore, needs determining is whether word count units might vary once we consider the four skills and their assessment.

### **Language Background of Participants**

McLean's article only tested L1 Japanese participants, and we may not necessarily expect the same results for language learners from other language backgrounds. We might see a different set of results if we conduct a replication study for different L1 background populations than those reported in McLean (2018). As Nation and Bauer (2022) state in their article on morphological awareness:

English and many other languages, including Japanese, have words that are made up of meaningful parts and these parts systematically contribute to the meaning of words. (p. 1)

Crucially, McLean fails to report on participants from other language backgrounds, such as L1 Chinese learners of English. The Chinese and Japanese languages are similar in that there are no singular/plural changes for proper nouns and personal pronouns. However, the two languages differ in that while derivational changes in Japanese are similar to those in English, this is not the case with Chinese.

Were we to use the same method to test the L1 Chinese participants as those reported in McLean's article by translating a single word from English to Chinese, it would not detect the morphological changes because no meaningful parts can be added to the changes to the Chinese language, unlike Japanese or English. Thus, testing participants with a wide range of language backgrounds might indicate that no single measure is universally appropriate.



## PEDAGOGICAL IMPLICATIONS

In his article, McLean indicates that his participants had insufficient vocabulary knowledge in understanding word families, especially for the derived forms. This implies that in teaching practice, even though language learners have been taught these words, word part knowledge is not emphasized in classroom activities. Dang (2021), for instance, suggests that there is a lack of training in the knowledge of word parts, and that through learning the most frequent derived forms, learners can recognize the important role that an understanding of morphology plays in increasing their overall vocabulary knowledge. If language instructors can emphasize the significance of teaching word part knowledge, it may help learners to improve their language skills. In support, Webb (2021) suggests that “presenting headwords together with their inflections and derivations may provide a shortcut to lexical development” (p. 942). Such a claim also emphasizes that building morphological knowledge can improve vocabulary knowledge learning in practice. However, we need further research on how exactly morphological knowledge influences vocabulary learning.

While it is important that both language instructors and language learners are made aware of the importance of gaining word part knowledge, questions remain regarding precisely what knowledge is necessary. In response, Nation and Bauer (2022) suggest that affixes should be learned in relation to learners’ vocabulary size: They claim that level 3 affixes should be studied by learners who know the first 1,000 flemmas, level 4 affixes should be studied when learners have acquired the first 2,000 to 3,000 words, and level 5 affixes and beyond are appropriate for learners who know at least the first 4,000 word families. Nation and Bauer’s (2022) claim offers a solution in teaching practice, that language instructors should also consider learner vocabulary size when they present the knowledge of word families in their teaching activities.

## CONCLUSION

McLean’s article contributes significantly to vocabulary research and our understanding of vocabulary learning. It investigates an important issue: the most appropriate word-counting unit for language learners that current vocabulary researchers should focus on. The article evaluates L1 Japanese participants from three different proficiency levels (beginner, intermediate, and advanced) by giving them an English-to-Japanese word translation comprehension test that uses sentences for context. The findings show that L1 Japanese participants had limited knowledge of the word family; using word families as a word unit for L1 Japanese participants, would, therefore, overestimate their knowledge of word parts. However, McLean’s study indicates that use of a flemma count might only slightly overestimate participant knowledge of the tested words (e.g., *center*, *circle* as verbs, *develop* with *-ing* forming an adjective). On this basis, McLean suggests that the appropriate word unit is the flemma. We should, though, bear in mind the problems we have highlighted in our evaluation related to suggesting the flemma as an appropriate unit in teaching and research. Other problems with McLean’s study include the small number of words tested and the use of a comprehension test for participants with different L1 backgrounds.

Importantly, McLean’s article offers possible implications for language learning and teaching. His paper stresses the need for an emphasis on morphological knowledge because such knowledge relates to learners’ vocabulary size and language proficiency level. Research, however, suggests that there is a lack of word part knowledge training in teaching practice (Dang, 2021). Meanwhile, the way in which word units

are processed in research is far from perfect based on current processing tools, and innovation is needed in lexical processing methods (Gablasova & Brezina, 2021). Further research might consider how a focus on morphology that includes derivational forms could be incorporated into language teaching.

## REFERENCES

- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253–279. <https://doi.org/10.1093/ijl/6.4.253>
- Brown, D., Stoeckel, T., Mclean, S., & Stewart, J. (2022). The most appropriate lexical unit for L2 vocabulary research and pedagogy: A brief review of the evidence. *Applied Linguistics*, 43(3), 596–602. <https://doi.org/10.1093/applin/amaa061>
- Dang, T. N. Y. (2021). Selecting lexical units in wordlists for EFL learners. *Studies in Second Language Acquisition*, 43(5), 954–957. <https://doi.org/10.1017/S0272263121000681>
- Gablasova, D., & Brezina, V. (2021). Words that matter in L2 research and pedagogy: A corpus-linguistics perspective. *Studies in Second Language Acquisition*, 43(5), 958–961.
- Iwaizumi, E., & Webb, S. (2021). To what extent does productive derivational knowledge of adult L1 speakers and L2 learners at two educational levels differ? *TESOL Journal*, 12(4), e640. <https://doi.org/10.1002/tesj.640>
- Iwaizumi, E., & Webb, S. (2022). To what extent do learner- and word-related variables affect production of derivatives? *Language Learning*. <https://doi.org/10.1111/lang.12524>
- Maw, T. M. M., Clenton, J., & Higginbotham, G. (2022). Investigating whether a flemma count is a more distinctive measurement of lexical diversity. *Assessing Writing*, 53, 100640. <https://doi.org/10.1016/j.asw.2022.100640>
- McLean, S., & Kramer, B. (2015). The creation of a new vocabulary levels test. *Shiken*, 19(2), 1–11.
- Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis* (Vol. Eurosla Monographs Series 2, pp. 57–78). Eurosla.
- Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, 28(2), 291–304. [https://doi.org/10.1016/S0346-251X\(00\)00013-0](https://doi.org/10.1016/S0346-251X(00)00013-0)
- Nation, P. (2021). Thoughts on word families. *Studies in Second Language Acquisition*, 43(5), 969–972. <https://doi.org/10.1017/S027226312100067X>
- Nation, P., & Bauer, L. (2022). Morphological awareness. <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/word-parts/morphological-awareness/nation-and-bauer-morphological-awareness-docx>
- Qian, D. D., & Lin, L. H. F. (2019). The relationship between vocabulary knowledge and language proficiency. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies*, 66–80.
- Sasao, Y., & Webb, S. (2017). The word part levels test. *Language Teaching Research*, 21(1), 12–30. <https://doi.org/10.1177/1362168815586083>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word

- associations and verbal suffixes. *Studies in Second Language Acquisition*, 19(1), 17–36. <https://doi.org/10.1017/S0272263197001022>
- Stoeckel, T., Ishii, T., & Bennett, P. (2020). Is the lemma more appropriate than the flemma as a word counting unit? *Applied Linguistics*, 41(4), 601–606. <https://doi.org/10.1093/applin/amy059>
- Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, 37(3), 461–469. <https://doi.org/10.1016/j.system.2009.01.004>
- Webb, S. (2010). A corpus driven study of the potential for vocabulary learning through watching movies. *International Journal of Corpus Linguistics*, 15(4), 497–519. <https://doi.org/10.1075/ijcl.15.4.03web>
- Webb, S. (2021). The lemma dilemma: How should words be operationalized in research and pedagogy? *Studies in Second Language Acquisition*, 43(5), 941–949. <https://doi.org/10.1017/S0272263121000784>

## APPENDIX 1

### Summary of the Bauer and Nation (1993) Word Family Levels

Level 1	A different form is a different word. Capitalization is ignored.
Level 2	Regularly inflected words are part of the same family. The inflectional categories are – plural; third person singular present tense; past tense; past participle; -ing; comparative; superlative; possessive (8 affixes).
Level 3	-able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un-, all with restricted uses (10 affixes).
Level 4	-al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, -ous, in-, all with restricted uses (11 affixes).
Level 5	-age (leakage), -al (arrival), -ally (idiotically), -an (American), -ance (clearance), -ant (consultant), -ary (revolutionary), -atory (confirmatory), -dom (kingdom; officialdom), -eer (black marketeer), -en (wooden), -en (widen), -ence (emergence), -ent (absorbent), -ery (bakery; trickery), -ese (Japanese; officialese), -esque (picturesque), -ette (usherette; roomette), -hood (childhood), -i (Israeli), -ian (phonetician; Johnsonian), -ite (Paisleyite; also chemical meaning), -let (coverlet), -ling (duckling), -ly (leisurely), -most (topmost), -ory (contradictory), -ship (studentship), -ward (homeward), -ways (crossways), -wise (endwise; discussion-wise), anti- (anti-inflation), ante- (anteroom), arch- (archbishop), bi- (biplane), circum- (circumnavigate), counter- (counter-attack), en- (encage; enslave), ex- (ex-president), fore- (forename), hyper- (hyperactive), inter- (inter-African, interweave), mid- (mid-week), mis- (misfit), neo- (neo-colonialism), post- (post-date), pro- (pro-British), semi- (semi-automatic), sub- (subclassify; subterranean), un- (untie; unburden) (50 affixes).
Level 6	-able, -ee, -ic, -ify, -ion, -ist, -ition, -ive, -th, -y, pre-, re- (12 affixes).
Level 7	Classical roots and affixes

Adapted from “Making and using word lists for language learning and testing,” by I. S. P. Nation, 2016, p. 27. (<https://doi.org/10.1075/z.208>). Copyright 2016 by the John Benjamins Publishing Company.

## ABSTRACT

### **A Review of McLean’s Evidence for the Adoption of the Flemma as an Appropriate Word Counting Unit**

Yajie LI

Graduate School of Integrated Arts and Sciences

Hiroshima University

Simon FRASER

Institute for Foreign Language Research and Education

Hiroshima University

Jon CLENTON

Graduate School of Integrated Arts and Sciences

Hiroshima University

This review article examines a paper by McLean (2018) in which he addresses an important gap in the field of second language (L2) vocabulary research: the ability of L2 learners to understand the inflected and derived forms of words. In his study, McLean questions the inferences made in research that uses the word family as an appropriate unit of counting. Japanese EFL learners ( $N=279$ ) were divided into three proficiency groups (beginner, intermediate, and advanced), and their knowledge of inflectional and derivational forms was measured using an English to Japanese translation test. It was found that learners at all levels differed in their ability to understand the base form, inflected forms, and derived forms of 12 words. McLean concludes that the flemma, a word’s base form and associated inflectional forms, is therefore a more appropriate word counting unit than the word family.

Our review begins with an explanation of the key terms and concepts in this field of research, followed by an overview of McLean’s paper. A critique of the paper follows in which we examine the validity of McLean’s argument, and evaluate the strengths and weaknesses of the study with a view to assessing its importance in the L2 vocabulary research field. We close with a consideration of the pedagogical implications of the study, including the treatment of morphology in the L2 classroom.

## 要 約

### フレマを適切な単語カウント単位として採用した McLean の根拠についての一考察

ヤジェ・リー

広島大学大学院総合科学研究科

サイモン・フレイザー

広島大学外国語教育研究センター

ジョン・クレントン

広島大学大学院総合科学研究科

本稿は、第二言語 (L2) 語彙研究の分野における重要なギャップ、すなわち第二言語 (L2) 学習者が単語の屈折形や派生形を理解する能力を扱った McLean (2018) の論文を検証するものである。彼の研究において、McLean は、ワードファミリーを適切なカウント単位として用いるべきとする研究に於ける結論に疑問を投げかけている。日本人 EFL 学習者 (N=279) を 3 つの習熟度グループ (初級, 中級, 上級) に分け、屈折形と派生形の知識を英日翻訳テストにより測定した。その結果、全レベルの学習者において、12 単語について、基本形、屈折形、派生形を理解する能力に差があることがわかった。McLean は、単語の基本形と関連する屈折形であるフレマが、ワードファミリーよりも適切な単語カウント単位であると結論付けている。

本稿では、まず主要な用語と概念について説明し、次に McLean の論文の概要を述べる。その後、McLean の主張の妥当性を検証し、L2 語彙研究分野での重要性を評価するために、この研究の長所と短所を評価した上で、この論文の批評を行う。最後に、L2 教室における形態論の扱いを含む、この研究の教育的示唆について考察する。