

Challenges in the Assumptions of Using a Flemma-based Word Counting Unit

Tomoko Ishii^a, Phil Bennett^b, and Tim Stoeckel^b
^aMeiji Gakuin University; ^bUniversity of Niigata Prefecture

Abstract

The choice of word counting units (i.e. word family, flemma, or lemma) is of great importance in vocabulary list and test creation, as there are assumptions underpinning the use of each. Flemma-based counting assumes that if a learner can understand the meaning of a word in one part of speech (POS), they can also understand its meaning when the same word form is used in another POS. A previous quantitative study showed that such an assumption is not always valid, but it did not provide reasons as to why. Therefore, this article presents an interview study probing the challenges learners face when they fail to understand the meaning of a known word form used in a new POS. The data were collected through one-on-one interviews with 16 university students in Japan, where they were asked to demonstrate comprehension of target words embedded in short sentences, as well as to explain how they approached the task. The interviews revealed that factors related to both the words and the learners play a role: some words have more complex meaning relationships across different POS, and some learners did not have advanced enough parsing skills to overcome such complexity. Based on these findings, this article concludes that the flemma might not be a suitable word counting unit for Japanese university students with relatively low levels of proficiency.

Keywords: word counting unit, flemma, lemma

1 Background

Decades have passed since vocabulary became one of the focal points of second language (L2) research, and a number of vocabulary lists and tests have been proposed to help learner's development in this area. A fundamental assumption of word lists, and by extension the tests that are based on them, is that word forms can be grouped into units. The words in each unit are treated as equivalent for the purposes of assessing text difficulty and measuring learner knowledge. That is, if a learner demonstrates knowledge of a single member of a unit, then knowledge of the other members can be reasonably assumed. Bauer and Nation (1993) suggested seven levels of word grouping, or word families, two of which are commonly used in word lists and tests. When the term *word family* is used without annotation, it typically refers to Level 6 of their word family system. This level groups a head word, its inflections and some of its derivations (e.g., *act*, *acts*, *action*, *actor*, *acting*). On the other hand, Level 2 includes headwords and inflected forms within the same grouping. The question of whether knowledge is assumed to cross part of

speech (POS) boundaries is not explicitly addressed, which has led to two subdivisions within this level: lemma and flemma. In the *lemma*, different POS are treated as distinct units (e.g., verbal *act* and nominal *act* and their respective inflections are separate), while in the form-based lemma, or *flemma* (Pinchbeck, 2014), members of the lexical unit may be from different POS (e.g., *act* as a noun and as a verb are both included).

Among these units, the Level 6 word family (hereafter “word family”) has been the unit of choice for the word lists which some widely used vocabulary tests are based on, including the Vocabulary Levels Test (VLT; Schmitt et al., 2001) and the Vocabulary Size Test (VST; Nation & Beglar, 2007). The use of a word family assumes that the knowledge of a head word allows learners to access its inflections and derived forms (e.g., when learners know the word *act*, they can understand the meaning of *actor* as well, or at least they can learn the new word form *actor* with a small amount of effort and time). However, multiple studies have shown that such an assumption might not hold true in all contexts (e.g., Gardner, 2007; Mochizuki & Aizawa 2000). Iwaizumi and Webb (2021) showed how the knowledge of derivations might differ amongst learner groups of different proficiency levels, which again suggests the inadequacy of using word families across all learners. In his study, McLean (2018) demonstrated that Japanese learners ranging from beginner to advanced levels might not possess enough knowledge of derivation. As such, he argued for the use of flemma to avoid the overestimation of vocabulary knowledge. Laufer et al. (2021) also showed different degrees of mastering derived forms between two proficiency level groups; they argued, however, that these differences did not disqualify the use of word families, since the derived forms that are typically not known by less advanced participants do not frequently appear in actual texts.

Although to a lesser extent than word family-based counting, flemma-based counting still assumes that learners successfully use their knowledge of known word forms when decoding the meaning of unknown ones. In the case of flemmas, the assumption is that knowing the meaning of a word form or its inflectional forms in one POS allows learners to understand the meanings of the same word forms when they occur in another POS (e.g., the knowledge of *act* as a verb would lead them to understand what *an act* might mean). Therefore, building on the work of McLean (2018), Stoeckel, et al. (2020) investigated whether such an assumption would hold true among Japanese learners. The study asked participants to demonstrate their understanding of words with two POS (e.g., *twist*, *rise*, *fool*) by translating 24 short sentences containing 12 target words in each POS. In this task, there were no penalties for mistranslations of non-target words and only the comprehension of the target words was assessed. The results suggested that participants sometimes failed to demonstrate knowledge of the meaning of a word even though they knew the same orthographic form in another POS. Namely, even when a participant successfully translated “We *edit* it well,” it did not always mean that the same participant could translate “We made an *edit* quickly.” Amongst the cases where participants knew the word in at least one POS, 44% indicated a lack of understanding of the word in

the other POS. This led to the conclusion that lemma-based counting may be more suitable for Japanese learners.

This phenomenon may seem rather puzzling for native speakers or advanced English users. In theory, all learners have to do is transfer the meaning of a word they already know to make it fit a new context of meaning and grammar. It is also quite common to have a noun become a verb over time (e.g., *to Google*, *to Photoshop*, *to friend/unfriend*) or vice versa. Upon encountering a known word form in a new POS, learners would not seem to require any special skills to cross the boundary of POS. For example, it is hard to imagine how a learner who knows the verb *to drink* would struggle to understand that *a drink* is something to drink.

No matter how simple it might seem, some learners do have difficulty handling such a shift and fail to understand the meaning of known words in an unknown POS. Although Stoeckel et al. (2020) showed this quantitatively, their report on participants' performance as a group does not tell us why some learners miss what seems apparent. We need qualitative examination in order to see the obstacles each learner might face. Therefore, to complement previous findings, this article reports on an interview study to examine what learners know, what they do not know, and what causes difficulty in applying the knowledge of a word form across different word classes.

2 Methods

2.1 Participants

Sixteen students from three universities in Japan participated in this study. They were recruited via e-mail by their teachers and were offered a small monetary incentive for taking part in the interviews, which took around 30–40 minutes each. No attempt was made to control the English level of the participants, and the group was made up of learners ranging from limited proficiency (Eiken pre-2nd grade) to intermediate proficiency (TOEIC 780), which is considered to be equivalent to a range from A2 to B2 on the Common European Framework of References (CEFR) scale.

2.2 Materials

The target items used in this study were exactly the same as those used by Stoeckel et al. (2020). To determine whether the participants understood the meaning of a single word form in a different POS, they were asked to translate short sentences containing each target word. The target words and sentences are listed in Table 1.

All the target words came from the first 3,000 flemmas of the Cambridge English Corpus, and had more than one POS. The 12 words represented the average and range of difficulty of a broader pool of items appearing on the New General Service List test (Stoeckel et al. 2020). The target words in each POS had common, similar meanings, as determined by the definitions of standard dictionaries. As Zipf (1945) showed, high frequency words typically have multiple senses in dictionaries, and therefore a choice of word meaning had to be made. For the selection of meaning, 100 randomly selected concordance lines from the Corpus

Table 1. Target Words and Prompt Sentences

Target item	POS	Prompt sentence
Edit	Noun	We made an edit quickly.
	Verb	We edit it well.
Rise	Verb	They did not rise this year.
	Noun	We did not see the small rise.
Compound	Verb	This compounds the problem.
	Noun	We need those compounds.
Function	Noun	This has several functions.
	Verb	This functions very well.
Fool	Verb	They fool me every time.
	Noun	That fool did it again.
Quote	Verb	Can I quote you?
	Noun	Can you remember this quote?
Result	Verb	Your action results in three things.
	Noun	We don't know the results.
Export	Noun	They buy our most important export.
	Verb	They export many different things.
Variable	Adjective	They have variable prices.
	Noun	This is an important variable.
Twist	Noun	I gave it several twists.
	Verb	He really twists it a lot.
Pause	Verb	We usually pause here.
	Noun	I took a pause here.
Extra	Noun	I have one extra at home.
	Adjective	We did some extra things.

of Contemporary American English (COCA) for each target word in each POS were examined to identify typical meaning senses. A meaning sense that was frequent in both POS was the intended meaning of the target items in each pair of prompt sentences. Note, however, that in some cases, the most frequent meaning sense in the COCA was a sub-entry under a broader, numbered dictionary definition, and in such cases, the broader definition was the intended meaning in the test items. An example is the noun form of *extra*. The most common meaning sense in the COCA for this word pertains to an extra in a movie. In the dictionary, such use of *extra* is listed under a numbered entry with the broader meaning of an item in addition to what is usual or needed. The latter was the intended meaning of *extra* (n) in the test. We must concede that such variation in the selection of intended meaning was a limitation of this study.

Words with substantially different meanings such as *book* (to reserve / a book to read) or *affect* (to have an effect on / emotion) were not included. In addition, words that have become part of the Japanese vocabulary as loanwords (e.g., *copy*, *dry*) were not selected because of the risk that participants could provide responses based on phonological similarity to their native language (L1), which would make it difficult to tell if the participants really grasped the meanings of the words through the experiment.

In order to eliminate the possibility of participants not being able to respond properly due to a lack of grammatical knowledge, Stoeckel et al. (2020) included items to check the participants' grammatical ability by translating sentences of similar structure and complexity as those used in the main study. In the current study, however, such items were skipped, as the sources of difficulty would be naturally revealed during the interviews.

2.3 Procedure

The interviews were conducted in participants' L1 (Japanese) via video chat and using the screen sharing function. The following steps were taken:

1. Prompt sentences were presented on the screen one by one.
2. The participants translated them orally, while the interviewer took notes of their responses.
3. After completing all the items, the interviewer and the participants reviewed the items together, and the participants were invited to explain how they processed the prompt sentences.

Leow and Morgan-Short (2004) presented verbal reports as a way to elucidate the internal processes employed by learners, pointing to introspective-retrospective distinctions in data gathering procedures. The introspective or real-time approach has some advantages, as unlike the retrospective approach, it does not rely on participants' memory. However, for verbalisation to be captured efficiently and to be reliable and complete, established and dependable procedures need to be followed. Additionally, we did not want concurrent interviews to influence the way learners completed the task. Once the issue of POS was broached, for example, participants may have been more alert to it than they would have been otherwise. Therefore, a retrospective approach was adopted in this study. The participants focused on the task first, and then immediately after the task, when their memory was still fresh, they were invited to explain how they completed the task.

In order to help participants recall the process, as well as describe the state of knowledge that they might not be aware of, the interviewer asked if they knew the target word, if they knew the word in the POS as used in the prompt sentence, and if they applied knowledge of the word in another POS when translating the sentence. When the participants missed an item, the interviewer also provided a quick explanation to help them comprehend the item better, and sometimes asked for further ideas after providing some hints regarding the meaning of the word or the grammatical structure of the prompt sentence.

During the second step of the interview process, the responses were coded following the criteria established in Stoeckel et al. (2020). That is, each response was marked dichotomously as correct or incorrect. Although each target word form had a pair of items with a shared, common meaning sense, any correct meaning of the tested word that fit the prompt sentence was judged as correct. This was done by just one coder, which is a limitation of the present study. However, this person achieved a high level of inter-rater reliability ($k = 0.787$)

Table 2. Summary of Task Performance

Participant ID	Known meanings (max 24)	Word form known in at least one POS (max 12)	Word form known in both POS (max 12)	Word form known in only one POS (max 12)	Target POS known (max 24)	Another POS knowledge applied (max 12)	Other ^a (max 24)
HS	21	11	10	1	15	4	2
TS	20	11	9	2	13	5	2
NK	20	10	10	0	11	8	1
RS	18	10	8	2	14	4	0
SW	18	10	8	2	16	2	0
NW	16	8	8	0	14	2	0
MK1	12	8	4	4	9	3	0
AU	11	7	4	3	9	2	0
TK	10	7	3	4	8	0	2
SY	10	6	4	2	7	2	1
AS2	9	6	3	3	6	3	0
MK2	7	6	1	5	6	1	0
SM	5	4	1	3	4	1	0
HI	5	4	1	3	3	2	0
YK	4	3	1	2	3	1	0
AS1	3	3	0	3	0	0	3
Mean	11.81	7.13	4.69	2.44	8.63	2.50	0.69
SD	6.23	2.75	3.57	1.36	4.86	2.03	1.01

^a“Other” refers to the cases where participants applied their knowledge of other words (e.g., they guessed the meaning of *export* from the word *import*).

with a second coder on the same item set in Stoeckel et al. (2020). Additionally, when word comprehension was unclear, the coder had the opportunity to probe learners’ understanding of each target word during the subsequent interview. We believe this approach gives reasonable credibility to the consistency of judgements in the present study.

3 Results

Table 2 provides a summary of the overall performance of the 16 participants. As the focus of this study was on the understanding of the target words, translation errors in other parts of the prompt sentences were disregarded.

As is shown Table 2, participants exhibited a great deal of variety in both the number of items they could understand and the extent to which they understood target word forms in both POS. Participant HS translated 21 items out of 24 correctly, while AS1 translated only 3. Participants NW and MK1 both knew 8 words out of 12, but NW provided correct translations for both POS in all cases, whereas MK1 gave correct translations in both POS for just 4 of the 8 words. Overall, among the cases where at least one POS was known, the other POS was also correctly translated 65.78% of the time (4.69/7.13).

3.1 Interview Data

Seeking explanations for why some participants fail to apply their partial knowledge of words when trying to understand a sentence, one-on-one interviews were held with the participants immediately after they completed the translation task. These interviews showed how learners of different levels approached the task and revealed complex issues for both words and participants.

First, not all words seemed to pose the same degree of difficulty across the two POS. L1 speakers or advanced learners of English might recognise the meaning connection between two POS of the same word form as something straightforward, but for the less proficient learners in this study not all connections were equally transparent.

Even when the concepts are clearly connected, there are sometimes multiple possible meanings when a word is used in another POS. For instance, *Oxford Learner's Dictionaries* (<https://www.oxfordlearnersdictionaries.com>) defines the noun form of *export* as both: 1) *the selling and transportation of goods to another country*, and 2) *a product that is sold to another country*. Simply recognising *export* as the noun form of the verb *to export* does not indicate which of these definitions is appropriate in the given context. In other words, when a verb is used as a noun, the aspects of meaning transferred are not always systematic. For example, one meaning of *twist* as a noun is the act of twisting something, while *fool* as a noun is meant to be interpreted as a person with certain characteristics. Why could it not describe a foolish act? When a similar word like *tease* can refer to an act of teasing someone, there seems to be no logical reason why *fool* as a noun cannot. As shown in the extract below, some participants had problems with this. All the interview extracts presented in this article were originally in Japanese and were translated into English by the interviewer.

- Interviewer: You seemed to have no problem translating “They fool me every time.” Did you know all the words in this sentence?
- MK1: Yes.
- Interviewer: Okay, the next one “That fool did it again,” you translated this sentence as “It was a lie again.” Did you know the word *fool* as a noun?
- MK1: Honestly, I had no idea what this sentence might mean. As “fool” [verb] is like *damasu* (to trick someone by telling an untruth), I thought it might be ‘it was a lie.’”

Another participant saw a gap in meaning transfer between the adjective and noun forms of *variable*.

- Interviewer: “They have variable prices...”
- HS: I knew *variable* as an adjective, but didn’t know it as a noun.
- Interviewer: I see. You have translated the noun form as *henka* (changes).
- HS: As *variable* [adjective] means something is changing, I thought its noun might mean “changes.”

- Interviewer: I see. One typical translation of this word is *hensu*, a factor that changes.
HS: Oh, hmm, I couldn't guess that.

This participant had no problem seeing *variable* as something that has to do with changes, but could not identify what part of the changes were being referred to. Logically, “the state of being changed” and “something that causes a change” both seem plausible, and it is no surprise that “a factor that changes” was not always their first choice when learners have only learned this word as an adjective.

As shown above, the meaning relationship across different POS is not always as straightforward as it might appear, and learners must explore different possibilities that are logically plausible for them to reach the correct deduction. Such differences in the meaning relationship are sometimes evident in the translation of target items. Some POS pairs have very similar translations in Japanese with the simple addition of a grammatical marker (e.g., *kinou* and *kinou-suru* for the noun and verb forms of *function*, where *suru* indicates a verb), while other pairs are translated quite differently (e.g., *kawariyasui* and *hensu* for the adjective and noun forms of *variable*). For some word pairs that have similar translations, students who did not even know the word at all were able to provide the correct meaning in one POS when its definition was explained in the other.

- SY: I don't know “function.”
Interviewer: It means *kinou* (function) and is used as a noun. Can you tell what the other sentence might mean now?
SY: Ah, *kore-wa yoku kinou-suru* (It functions well), then.
Interviewer: That's right.

Such differing degrees of closeness in the translation pairs, combined with the complication of having multiple logically plausible meanings, seemed to influence students' performance. The more similar the translations, the easier it seemed for even the lowest-level learners to answer correctly.

In addition to the varying degrees of challenge that different word pairs posed, learners also had a great deal of variation in their ability to process meanings in different POS. The first thing we should note is that learners varied in their declarative knowledge of POS. Some participants actively used grammar knowledge and derived the meaning of words based on their awareness that the word was being used in another POS. NK was one such participant, and the case below is typical of the responses he provided throughout the interview.

- Interviewer: Did you know this word (*edit*)?
NK: I only knew this word as a verb, but seeing it being used as a noun, I thought it would be a noun, so I translated it that way.
Interviewer: How could you tell it was a noun?
NK: It's following “an.”
Interviewer: I see. How about this word (*quote*)?

NK: I didn't know this word could also be a noun, but it is used as a noun here, right?

On the other hand, some participants seemed to have very limited awareness of POS itself (even for words they knew) and expressed difficulties when asked if they knew a word as a noun or a verb. The coding presented in columns six and seven of Table 2 is based on what the participants reported, although not all of them were sure about how they knew the POS of certain words. One participant (AS1) completely gave up answering such questions, while others hesitated and used a substantial amount of hedging such as “probably” and “I think” when describing their knowledge. Another participant (SW), who translated 75% of the sentences correctly and exhibited a good command of grammar in the follow-up interview, said she was not quite sure how she knew each word in terms of POS. She explained that she had some understanding of the core meaning of words, without specification of POS, and tried to find the best translation to fit the sentence both in terms of meaning and grammar. For instance, she associated the concept *henshu* with the word *edit*, which in her mind could both mean *to edit something* or *an action of editing something*; this is similar to us not being able to tell if the word *edit* is a noun or a verb simply by looking at it on its own. She then saw the word in a sentence “*We edit it well,*” from which she gathered that *edit* was being used as a verb, and correctly translated it into the verb form *henshu-suru*, with the verb marker *suru*. She added that although she claimed her previous knowledge of some words “as a noun” when she thought she had seen them being used as a noun, she did not think they were a part of her productive vocabulary and she was not sure whether she really knew their POS.

Not being able to recognise the grammatical features of a word on its own might not be an issue as long as we discuss learners' receptive knowledge. Learners like participant SW should be able to identify the words' grammatical features from the way they are used in a sentence. However, it can become a problem when a lack of awareness about POS is combined with weak parsing skills. Namely, some learners could not tell if a word was being used as a noun or a verb in a sentence, and therefore they stuck to the POS they knew.

Interviewer: You knew the word *function* as you translated “This has several functions” smoothly.
 TK: Yes, it's *kinou* (function [noun]), right?
 Interviewer: Alright, then what about “This functions very well”?
 TK: *Sono kinou wa totemo yoi.* (This function is very good.)
 Interviewer: Well, ... what is the verb in this sentence?
 TK: ... Oh, there is no verb. So, is *function* a verb here?
 Interviewer: That's right.
 TK: Then, *kinou-suru* (function [verb])?
 Interviewer: That's right.

This participant failed to see that the word *function* was being used as a verb in her first attempt and applied the noun-like translation *kinou* because that was how she remembered the word. When the interviewer drew her attention to

the grammatical structure, she could parse the sentence correctly, but until then, it did not occur to her that a word that she knew as a noun could also be used as a verb. What is interesting here is the fact that when it was pointed out to her that her translation gave no indication of a verb in the English sentence, she was able to immediately correct her answer. Therefore, the problem was not that she had no parsing skills. She would have been able to parse other sentences with the same structure, but it appears that she jumped to “this function” (*this* + noun) combination as she knew the word as a noun. It seems her parsing skills were not strong enough to alert her to the fact that such a reading would leave the sentence without a verb.

Here is another example of a participant who translated the prompt “I have one extra at home” as “I have an extra home.”

- Interviewer: The next item ... What’s your understanding of this word?
SM: *Extra* ... it is like “there is a lot,” like in “extra paper.” That is how I know this word.
So, I translated this sentence as “another home.”
Interviewer: Yes, the word *extra* means *yobun-na* (extra, additional).
SM: Yes, yes, *yobun-na*.
Interviewer: That’s right. The meaning is right. In the second sentence, “We did some extra things,” it fits well. Your translation is good.

But this one, “I have one extra at home,” this is not about a rich person owning a second house.

- SM: Really? (Laughs)
Interviewer: It says “one extra,” which means *extra* is used as a noun.
SM: Huh?
Interviewer: What do you think this might mean?
SM: Does that mean there is one more person?
Interviewer: Well, probably there is one more of something. For example, when you suddenly need an umbrella, “Ah, I have another one at home” or something like that. It can be about anything, you have one that you don’t need and you can lend it to someone, or something like that.
SM: Oh, I see. Yeah, “at home,” not “extra home.” Now I see it.

As we can see from these examples, learners need to be able to detect the correct POS before they can consider what a word might mean in another POS. In some cases, it is possible that by remembering a translation with grammatical markers (e.g., a verb marker *-suru* or an adjective marker *-na*), learners with weak parsing skills try to put the words together based on the grammatical category they associate the word with, making it even harder to parse a sentence. Namely, unless a learner has sufficient parsing skills to recognise that an adjectival form does not fit the context, a learner that understands *extra* as *yobun-na* (adjective)

might be more likely to interpret “extra at home” as “extra home” (*yobun-na ie*). In other words, they may disregard the “at” before “home.”

In addition to the variation in parsing skills, there were also differences in the level of flexibility when exploring meanings. As explained earlier, some words have a straightforward connection between the meanings in the two POS, while others have a slightly less obvious connection in the eyes of learners. Some participants seemed more capable than others to handle such connections by exploring various possible interpretations in a flexible manner. For example, for *export*, some participants could deduce the meaning correctly even if they had not seen the word being used to mean “something to be exported.”

Interviewer: You translated “They buy our most important export.” without difficulty...

HS: I initially did not know what the noun form of *export* meant, but as it follows *buy*, it must be something people buy, so “things to be exported”? Was I right?

Not all participants had such good guessing skills, and some struggled more. Participant SY associated the word *export* with a translation word *yushutsu* (the act of exporting something), which she transformed to a verb form *yushutsu-suru*, when seeing it used as a verb. However, she could not translate the same word when it was used as a noun, as she did not see how the translation word she knew (*yushutsu*) would fit the sentence.

Interviewer: You translated the sentence “They export many different things” correctly. Did you know about the word *export*?

SY: Yes.

Interviewer: And you did not provide any translation for “They buy our most important export.”

SY: I could tell that *export* was used as a noun, but I didn’t know what “to buy *yushutsu*” could mean.

Had the prompt been “Export is important to the economy of our country,” this participant might have been able to come up with a good translation, as she could tell that the word *export* was being used as a noun with a familiar meaning sense in this sentence. However, the shift to “things to be exported” from “the act of exporting something” seemed to be difficult for her to handle. It seems that the first translation she came up with, *yushutsu*, further hindered her processing and blocked her ability to comprehend the sentence. With more flexibility to explore different possibilities, thinking about what is involved in transactions described as *exports* and what could be bought, she might have reached the right answer.

4 Discussion and Conclusions

This study was a follow-up of Stoeckel et al.’s (2020) work, which showed that some learners fail to recognise the meaning of a word form even when they can recognise the same word in another POS. As was the case in their study, some

participants missed an item in one POS even if they knew the word itself in the other POS. Interviews with participants revealed the factors that are involved in applying the knowledge in one POS to another, and what problems learners might face in decoding the meaning of words.

To apply previous knowledge of a word in one POS to another, learners need to take the following steps:

1. Recognise the word form.
2. Note that the word is used in a different POS than the one they are familiar with.
3. Identify the new POS of the word.
4. Deduce the meaning based on context and the meaning of the word in the known POS.

L1 or advanced L2 speakers of English naturally go through these steps and do not recognise many obstacles. However, as has been shown in the examples from the interviews, some learners lack the parsing skills to notice or figure out the POS (Steps 2 and 3). In some cases, the translation they are familiar with in one POS contains an L1 grammatical marker (i.e., verb marker *-suru* or adjective marker *-na*), which could possibly hinder accurate parsing. It seems that reliable parsing skills are required to make adjustments to the previously acquired knowledge that contradicts the sentence being presented to them.

When comprehension broke down at steps 2 or 3, it could be argued that a lack of grammatical knowledge, not lexical understanding, was the primary reason learners had difficulty demonstrating an understanding of target words that were known in another POS. However, this belies the fact that in such cases, the correct translation of the sentence containing the target word in the known POS was a clear demonstration of the grammatical knowledge needed: the two sentences were grammatically similar. Moreover, it is reasonable to conclude from the interview results that, in many cases, if learners had been familiar with the target word in each POS, they could have correctly parsed both sentences with no additional global grammatical knowledge. Some participants could also correct their parsing once subtle hints were provided by the interviewer. This suggests that the problem is a poor understanding of some grammatical aspects of the tested words, rather than weak global grammatical knowledge. As Nation (2001) suggested in his list of components of vocabulary knowledge, understanding how a word behaves grammatically is a part of lexical knowledge.

There were also cases where participants accurately captured the grammatical structure of a sentence, but still could not correctly interpret the target word. This is because the relationships of meaning across different POS might not be as straightforward as they seem. For some words, determining the meaning (Step 4) is almost automatic after detecting the POS of the word. However, there are also cases where learners need to choose between multiple logically possible meanings (e.g., the noun form of *export*), and discrepancies between the meanings in different POS and distance in the translation words can also pose challenges to differing degrees (e.g., *kinou/kinou-suru* vs. *kawariyasui/hensu*).

Despite attempts made to select words with similar meanings, there was inevitably some variation in the semantic distance between the two POS tested. In some cases, the meaning of the target words in the test materials reflected the literal, core senses of the words, rather than a more frequent but specialised sense (e.g., *extra* as a noun is primarily used to refer to extras in movies, while the prompt in this study suggested an extra thing). Controlling for such variation may not be entirely possible, owing to the unique characteristics of each word pair, but as we state above, differences in relationships of meaning seem to contribute to the difficulties that some of our participants faced. To some extent, this also demonstrates the challenge of using flemma-based groupings in vocabulary tests to determine learners' knowledge.

Due to the limited sampling of the target items, we must concede that this study might not have captured the entirety of the issue. First, this study limited target words to those that supposedly had similar meanings in the two POS. It was rather surprising how much difficulty some, although not all, participants faced in identifying the different meanings. The interviews revealed that detecting meaning might require an active and flexible search of multiple possibilities, even when the meanings seem to be closely related to each other across different POS. In the broader English lexicon, there are words that have multiple POS with considerably different meanings (e.g., *rate*, *project*). The risk of flemma-based word counts leading to overestimations of lexical knowledge may therefore be even more substantial than what the current study suggests.

Second, the decision to avoid loanwords inevitably distorted sampling. In selecting the items, 900 out of the most frequent 3,000 flemmas were identified as having more than one POS. After eliminating the words that were considered to be part of the Japanese vocabulary (i.e., loanwords), nearly 40% of those words were discarded. Although these items did not qualify for the research purpose in this study, they are also an important part of English vocabulary that Japanese learners know in some way. We do not know how excluding them could have affected the results. If such items had been included, participants might have taken advantage of their knowledge in their first language, or the differences in the usage of the loanwords from the original English might have created an extra challenge. How the elimination of loanwords has influenced the results is unknown, but we need to acknowledge the fact that there is a large missing piece in this context.

There is also a limitation in the presentation of the target words. All the items were presented either in their base forms or base forms inflected with plural *-s* or third person singular *-s*. To obtain a better understanding of the utility of the flemma versus lemma, further research might also need to include other inflectional forms (e.g., past tense *-ed* or present participle *-ing*). In addition, in the current study, the contexts were kept to a minimum to measure the knowledge of target words rather than participants' skills at guessing the meaning of unknown words based on the context. However, there is a chance that this decision added to the difficulty in deducing the meaning of words known in one POS. If the context were richer, some participants might have performed better.

We should also concede some limitations in the coding procedure. The responses from the participants were coded by a person with previous experience

in a study using the same experiment materials as the current study. Additionally, the interview data served to clarify the presence or lack of understanding of the target items. Even so, it would have been better if we had had another coder so that we could assess the reliability of the judgements.

Despite the small number of participants and potential issues with target word sampling, the current interview study provides some warnings regarding the use of flemmas in vocabulary testing. Whichever word counting units we apply, there are two assumptions that underlie vocabulary test construction and use. First, when a learner correctly answers a test item assessing knowledge of a word form, it is assumed that they also know most or all the other word forms in the lexical unit. Second, it is assumed that this knowledge is employable in specific, typically real-world, uses of the language. Accordingly, with a flemma-based test of written receptive vocabulary knowledge like the one used in this study, it is assumed that a correct response corresponds with the ability to receptively understand all flemma constituents when reading.

Stoeckel et al. (2020) have already shown that this assumption may not be viable in all cases, and the interviews in this study provide further insights into why this is the case. There is more complexity than we often assume in the relationship between the meanings of a single word form in two different POS. Examples of this complexity include differences in how the two POS are translated into the L1 as well as the presence of multiple possible meanings. Such complexity might be overcome if the learners have advanced parsing skills as well as the ability to explore different meanings flexibly; however, some participants in this study were not yet at that level.

As the proficiency levels of the participants in this study were within the range typically found in Japanese universities, the flemma appears to be too encompassing a word counting unit for written receptive vocabulary tests with this population. Regarding pedagogical word lists, those based on the flemma or word family may help Japanese learners to see the relationship among morphologically related word forms, but such resources need to overtly list the various lemmas that make up the larger lexical units in the list.

References

- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6, 253–279. <https://doi.org/10.1093/ijl/6.4.253>
- Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28, 241–265. <https://doi.org/10.1093/applin/amm010>
- Iwaizumi, E., & Webb, S. (2021). 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>
- 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*. Advance online publication. <https://doi.org/10.1075/itl.20020.lau>
- Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition*, 26, 35–57. <https://doi.org/10.1017/S0272263104261022>
- McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, 39, 823–845. <https://doi.org/10.1093/applin/amw050>
- Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, 28, 291–304. [https://doi.org/10.1016/S0346-251X\(00\)00013-0](https://doi.org/10.1016/S0346-251X(00)00013-0)
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31, 9–13. https://jalt-publications.org/tlt/issues/2007-07_31.7
- Pinchbeck, G. G. (2014, March 22–25). *Lexical frequency profiling of a large sample of Canadian high school diploma exam expository writing: L1 and L2 academic English* [Roundtable presentation]. American Association for Applied Linguistics Annual Conference.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88. <https://doi.org/10.1177/026553220101800103>
- 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>
- Zipf, G. K. (1945). The meaning-frequency relationship of words. *The Journal of General Psychology*, 33, 251–256. <https://doi.org/10.1080/00221309.1945.10544509>