

THIS IS AN EARLIER VERSION OF THE MANUSCRIPT. FOR THE FINAL VERSION, PLEASE CHECK THE JOURNAL WEBSITE: The published version of this paper should be considered authoritative, and any citations or page references should be taken from it: *ITL- International Journal of Applied Linguistics* <https://benjamins.com/catalog/itl>

Exploring the Learning Burden and Decay of Foreign Language Vocabulary Knowledge: The Effect of Part of Speech and Word Length

Samuel Barclay^{1,2} and Ana Pellicer-Sánchez¹

¹UCL Institute of Education, ²Nottingham Trent University

Research has shown that several intralexical factors affect the learning burden of foreign language vocabulary (e.g., Laufer, 1997) and that some accrued lexical knowledge is forgotten. It is often assumed that the lexical items most difficult to acquire are those easiest to forget. However, few studies have provided empirical evidence to support this claim. This study examined the effect of two intralexical factors, part of speech and word length, on the learning burden and decay of intentionally learned foreign language lexical knowledge, as well as the role that learning burden played in the decay process. Forty-eight learners of English studied words of various parts of speech and lengths using flashcard software. Knowledge (form recall and recognition) of target items was assessed immediately after learning and four weeks later. Results of mixed-effects models showed that part of speech and word length have differential impacts on learning burden and decay and that increased burden mitigated loss of form recognition knowledge.

Keywords: learning burden; vocabulary acquisition; lexical decay; intralexical factors; flashcard software

Introduction

New words are learned with different levels of ease or difficulty. The learning burden of a word (i.e., the difficulty with which a word is acquired) is affected by various factors, including “regularity of patterning, the learner’s L1, other known languages, opportunity and experience, personal commitment, the quality of teaching, and the quality of course design” (Nation, 2020, p.15). Several characteristics of the words themselves, i.e., intralexical factors, also contribute to a word’s learning burden (Laufer, 1997). Two of the intralexical factors that have been studied with relation to lexical acquisition are part of speech (PoS) and word length. Previous studies have suggested that these factors affect learning burden, with shorter words and nouns generally found to be easier to learn (see Laufer, 1997; Schmitt, 2010). Research to date has measured learning burden by looking at learning gains. That is, learners interact with target items for a fixed duration and then knowledge of those items is assessed. In such a design, items for which many learners cannot demonstrate learning are considered to pose difficulty. However, measuring learning difficulty in this way may not accurately reflect the effort required to learn new vocabulary. Not all items that are successfully learned require the same number of exposures to be learned. For example, if a learner encounters a set of words in an activity five times, some of the items may require those five exposures to be learned whereas others might be learnt earlier. Thus, measuring learning burden by looking at learning gains does not provide nuanced insight into how burdensome the learning process was. Therefore, in this study we measure learning burden as the number of times each learner needs to see each word for it to be learned.

Another robust finding of the vocabulary learning literature is that lexical knowledge generally decays after acquisition (e.g., Ellis & Beaton, 1993). Studies typically report, for

instance, a loss of knowledge between immediate and delayed post-tests, reflecting the degradation of memory traces that naturally occurs. However, a similarly robust finding is that not all lexical knowledge decays. Research, for example, generally finds some retention on delayed post-tests and while productive knowledge appears to be more vulnerable than receptive knowledge (Chen & Truscott, 2010), some productive vocabulary knowledge typically remains after a retention interval (e.g., Peters, 2014). While a large literature exists on language loss in conditions of reduced or no target language contact (e.g., Hansen & Chen, 2001), minimal systematic investigation of loss in contexts of continued target language contact, referred to in the current paper as *decay*, has been undertaken. Indeed, beyond the assumption that lexical knowledge decreases as a function of time, the variables that influence the extent and speed of decay have received less research attention than those influencing acquisition. An interesting assumption in this area is that the lexical items most difficult to acquire are those easiest to forget, pointing towards a positive relationship between learning burden and loss. However, there is currently limited empirical evidence to support this assumption (but see Olshtain, 1989).

A greater understanding of the processes of learning burden and decay, as well as of the potential relationship between them, has important implications for vocabulary teaching, as instruction could be modified to foster more sustainable learning (Schmitt, 2010). However, few vocabulary learning studies have overtly examined these processes. In order to address these gaps, the present study examined the role of two intralexical factors, i.e., word length and PoS, on learning burden and decay. Learning burden was operationalised as the number of times needed by participants to learn each item. The use of flashcard software for the

instructional intervention allowed us to obtain such information about the learning process. The potential relationship between learning burden and decay was also explored.

Background

Vocabulary learning burden

The extent to which a variable influences the learning process is associated with its learning burden, with a heavy-learning burden slowing the learning process compared to a light-learning burden. Numerous factors have been claimed to affect the learning burden of new words, making them easier or more difficult to learn (see Peters, 2020, for a comprehensive review of factors affecting vocabulary learning). Studies have demonstrated that learning gains are influenced by contextual factors such as the relatedness of a group of words (e.g., Tinkham, 1993) and the frequency with which a word occurs in a text (e.g., Horst, Cobb, & Meara, 1998), as well as by individual factors such as language learning aptitude (e.g., Li, 2016) and learner vocabulary size (e.g., Webb & Chang, 2015).

Word-related factors can be divided into interlexical and intralexical factors (Laufer, 1990). The former relates to the relationship between novel L2 items and words (L1 or L2) already known to a learner. These have been shown to affect learning, with effects found for orthographic wordlikeness (e.g., Bartolotti & Marian, 2017), L1 frequency (e.g., de Groot & Keijzer, 2000), phonotactical typicality (Ellis & Beaton, 1993), and cognateness (e.g., de Groot & Keijzer, 2000). Intralexical factors are those connected to the linguistic representation (e.g., orthography, pronunciation), the meaning (e.g., concreteness, imageability), and/or the use (e.g., L2 frequency) of a lexical item (Laufer, 1991; 1997). Such factors have received considerable research attention (see Laufer 1997 or Peters 2020 for overviews). Previous

studies have shown that increased concreteness (de Groot & Keijzer, 2000) and morphological transparency (see Laufer, 1997) are associated with lighter learning burdens while polysemy/homonymy can make learning more challenging (Schmitt, 1998). However, research on the effect of two variables, PoS and word length, has received comparatively less attention (see Table 1 for a summary of the word-related factors on learning burden).

<INSERT TABLE 1 AROUND HERE>

Some studies have shown that PoS affects the learning of L2 vocabulary, with nouns having an advantage for both incidental (e.g., Horst & Meara, 1999) and intentional learning (e.g., Ellis & Beaton, 1993; Rodgers, 1969). Regarding intentional learning contexts (the focus of this study), Rodgers (1969) used a Russian-English paired associate learning design and found that more target nouns were learned than target verbs. More recently, Ellis and Beaton (1993) examined the effect of PoS on the learning gains of novice students of German using high-frequency nouns and verbs, and a variety of learning strategies. They found that a significantly greater number of nouns were learned than verbs. This learning advantage of nouns over verbs has been attributed to a range of causes: the superior imageability of nouns (Gentner, 1982), the greater syntactical complexity of verbs (Tomasello, 2003) and the comparative importance placed on nouns by some cultures (Gopnik & Choi, 1990). Crucially, while these studies suggest that nouns are easier to learn than other parts of speech, the research evidence to date is inconclusive (Peters, 2020). The noun advantage reported in studies to date may have stemmed from confounding intralexical variables such as imageability or concreteness (Peters, 2020). Thus, the investigation presented in this paper manipulated PoS while controlling for concreteness.

The length of L2 lexical items also seems to impact their learning burden. Length has been operationalised as the number of syllables, phonemes, and letters a word contains. Research suggests that word length positively correlates with learning difficulty. For example, Gerganov and Taseva-Rangelova (1982) reported that monosyllabic words were easier to learn than disyllabic words, and Ellis and Beaton (1993) found word length (number of letters) and learning gains to be negatively correlated. Additionally, word length has been found to impact processing difficulty, with shorter words (number of phonemes) associated with faster and easier processing than longer words (Tehan & Tolan, 2007). Psycholinguistic studies also typically find shorter words (number of phonemes) to be better retained over brief (e.g., Baddeley, Thomson, & Buchanan, 1975) and longer retention intervals (Tehan & Tolan, 2007). Finally, investigations of vocabulary knowledge have shown that learners often know more shorter words than longer words (number of letters, phonemes, and syllables) (Willis & Ohashi, 2012). This word-length effect likely occurs because there is more content to encode with longer words, so they are more prone to error on retrieval (Ellis & Beaton, 1993).

However, word length has received less research attention than other word-related factors (Peters, 2020), pointing towards a need for further empirical consideration. Additionally, as word length is related to other factors such as L2 frequency, with frequent words likely to be comparatively short (Peters, 2020), studies that did not control for L2 frequency may have confounded length with frequency (e.g., Willis & Ohashi, 2012). The research presented in this paper examined the role of word length while controlling for potentially confounding variables such as L2 frequency.

Importantly, studies of learning burden have generally used learning gains to determine learning burden (cf. Tinkham, 1993). In many of the studies reviewed above, learning burden was understood as the amount of learning that occurred when the number of exposures was held constant, with lower gains considered a reflection of higher learning burden (e.g., Peters, 2014). The idea here is that items that can be acquired within a fixed number of exposures pose a lighter learning burden than items that cannot be learned given the same amount/quality of input.

Another method to operationalise learning burden, less exploited in vocabulary learning research, is to look at the number of exposures necessary for a word to be acquired (Higa, 1965). The principle behind this approach is that items that require more exposures to be learned possess a heavier learning burden (e.g., Tinkham, 1993). This approach is preferable for two reasons. First, if learning gains are used to determine burden, it is not possible to measure differences in the effort required when items are successfully acquired, as the outcome variable is inherently binomial; an item is either learned or not. Thus, all learned items are considered equally easy, and all unlearned items equally difficult, when this is unlikely to have been the case. Typically, studies circumvent this problem by either calculating mean gains per word across a sample of learners or by summing the gains of a set of target items according to a grouping variable (e.g., Ishii, 2015). With such a metric, the learning burden of an item is assumed to be constant across a group of learners and/or all items in a set of words are thought to pose a similar level of burden. Crucially however, the learning burden of L2 lexis varies by item and by learner (Higa, 1965).

Secondly, studies that measure learning burden via learning gains inherently assume that controlling the amount of exposure to target items results in all items being processed in a similar fashion; however, this is also unlikely to be the case. When learners are presented with a set of words to learn in a specific time, the difficulty of one word may impact the likelihood of another word being learned in the same set, because the increased time spent acquiring one word may reduce the time available for all other words. In such a method, the learning burden of one item is dependent on the items with which it is presented. Moreover, a key distinction can be made between exposures during the process of encoding and exposures that result in the retrieval of already learned knowledge. Controlling the frequency of exposure does not necessarily reflect the speed with which learners actually encode items. Even if research prescribes a fixed number/length of exposure, learners will encode some items before others and thus items will differ in the number of retrievals they have. This is problematic for research designs that consider language loss because retrieval facilitates retention to a greater extent than presentation (Baddeley, 1990). This means that studies to date may have biased the retention data by controlling for frequency of exposure during the learning process, rather than frequency of retrieval. In order to accurately measure learning burden in a manner that does not artificially strengthen knowledge of the easiest items, a methodology is needed that can capture the number of exposures needed by each individual learner to encode a particular item while controlling for frequency of retrieval. The present study employed such a method.

Factors Affecting Lexical Decay

Despite potentially impactful implications for theory and practice (Schmitt, 2010), few studies have investigated the effect of word-related, contextual, and individual factors on the process of lexical decay. Lexical decay relates to the loss of vocabulary knowledge. Language

loss has mostly been considered in terms of language attrition. In the SLA literature, the term *attrition* has been defined as “the (total or partial) forgetting of a language by a healthy speaker (...) in a setting where [the attriting] language is only used rarely” (Schmid, 2011; p. 3). Crucially therefore, *language attrition* in most cases concerns language loss that occurs in contexts of limited or no contact with the attriting language. In an attempt to establish a difference with the attrition process, as described above, the term *decay* is used in this paper. Following Schmitt (2010), lexical decay is used in this paper to refer to the loss of L2 linguistic knowledge by healthy individuals in contexts of continued target language exposure; that is, settings where learners may not be exposed to specific target items despite continued exposure to the target language more generally. An example here might be a learner studying vocabulary that is not subsequently recycled or encountered despite continued exposure to the L2 more broadly. The terms *loss* and *forgetting* are also used in the literature, yet these terms are often criticised as research has suggested that linguistic knowledge, once acquired, is not likely to be truly lost or fully forgotten (de Bot & Weltens, 1995). This paper, while recognising this criticism, employs *loss* and *forgetting* as umbrella terms to refer in general to language loss in healthy individuals (i.e., encompassing attrition and decay) in line with previous use in the field (e.g., de Groot, 2006). Lexical decay is used to refer to the specific process investigated in the present manuscript, the loss of L2 lexical knowledge by healthy individuals in settings of continued exposure to the L2.

Research has considered the impact of interlexical factors on lexical decay, finding that cognateness (e.g., Tonzar, Lotto, & Job, 2009), and high L1 frequency (e.g., Lotto & de Groot, 1998) affect decay. However, only a few studies have targeted the effect of intralexical factors on this process. For example, de Groot (2006) measured the effect of concreteness on the

forgetting of lexical items, finding that more loss occurred with abstract items than concrete items. Similarly, de Groot and Keijzer (2000) reported an effect for concreteness, with abstract items being harder to learn but suffering the most decay.

Research has also considered the specific intralexical variables investigated in this study. Ellis and Beaton (1993) considered the decay of 36 English-German translation pairs over an interval of four weeks. The target items were selected to explore the influence of PoS (nouns vs. verbs). The participants studied the target language in a series of four learning trials. After each trial, tests of meaning and form recall were conducted. This procedure was repeated after a four-week interval. The results indicated that form and meaning recall knowledge suffered considerable decay and that there was a significant effect for PoS with knowledge of nouns suffering less loss than knowledge of verbs.

There has been limited investigation of the role of word length on lexical decay. Studies have found that word length is related to vocabulary knowledge, with learners more likely to know shorter words (e.g., Willis and Ohashi, 2012). However, no previous studies have experimentally examined the effect of word length on decay. Table 2 presents a synthesis of the results discussed in this review.

< TABLE 2 AROUND HERE >

The Relationship between Learning Burden and Vocabulary Loss

It is often assumed that the lexical items most difficult to acquire are those easiest to forget (e.g., Webb & Nation, 2017), pointing towards a positive relationship between learning burden and lexical decay. However, few studies have provided empirical evidence to support this assumption. This is partly the case because studies of language loss have generally not

measured how target items have been acquired. However, it is also the case that many vocabulary learning studies that have measured learning burden have not conducted delayed post-tests (e.g., Tinkham, 1993), precluding the type of comparison needed to understand patterns of decay.

The limited research evidence available has generally found that an increased learning burden is associated with greater loss. For example, Bahrick and Phelps (1987) looked at the attrition of L2 Spanish by L1 English university students over a period of eight years. They correlated initial learning burden (number of exposures needed for learning) with target item retention at the form recall level, finding that items initially easier to learn were better retained. A similar finding was reported by de Groot (2006) who investigated the effect of several variables on the decay of L2 lexical items using Dutch-nonword translation pairs. These items were presented over three learning sessions with interim test scores taken as a metric of learning burden (i.e., items learned in later session were considered more difficult). These results were compared to a one-week delayed meaning recall test. The findings suggested that the words hardest to learn (i.e., were acquired later in the learning procedure) were those most likely to be forgotten.

However, methodological characteristics of de Groot (2006) mean that her findings might have been confounded with retrieval frequency. In the study, target items were encountered the same number of times, but interim assessment scores indicated that some items were encoded earlier than others. These encoded items were not removed from the learning procedure and thus received more retrievals than other target items that took longer to encode, potentially strengthening the knowledge of the easiest items (Baddeley, 1990).

Bahrack and Phelps (1987) circumvented this issue by controlling the frequency of retrieval during the learning phase; however, the extended length of their study means that intersessional target item exposure may have occurred. Thus, more evidence is required to gain a better understanding of the role of learning burden on lexical decay. The current study addressed this gap while controlling for frequency of retrieval and employing a comparatively short retention interval, reducing the chance of intersessional exposure to the target items.

The Study

Overall, with limited empirical evidence available, we lack a detailed understanding of lexical decay and the effect intralexical factors such as word length and PoS have on this process. We also have limited evidence regarding the effect these variables have on learning burden, and the role learning burden plays in the decay of knowledge. The research presented in this paper aimed to address these lacunae.

The following three research questions were examined:

- 1) What is the effect of PoS and word length on the learning burden of foreign language lexis?
- 2) What is the effect of PoS and word length on the decay of foreign language lexis?
- 3) What is the effect of learning burden on the decay process (irrespective of the target intralexical variables)?

To answer these questions, English language learners studied previously unknown items using electronic flashcards. Flashcards were studied *productively* (see Webb & Nation, 2017). Using digital flashcards allowed us both to measure the number of times each participant saw

an item, and to differentiate exposures prior to lexical encoding (i.e., learning burden) from retrievals of encoded knowledge. Participants then completed a test that measured written form recognition and written form recall of the 32 target items. Four weeks later, the participants completed the same test again. The data of the first test were then compared to the second to determine the extent of knowledge of each item for each participant. The effect of intralexical factors on both the learning burden and decay of intentionally learned lexical knowledge, as well as the effect that learning burden had on the decay of foreign language vocabulary knowledge were examined.

Methodology

Participants

Forty-eight English learners of differing L1s participated in the study (30 females, 18 males; mean age = 24 years; age range = 19-36 years). All participants were enrolled on a pre-sessional English for Academic Purposes course at a UK tertiary institution at the time of data collection and had B2 proficiency as shown by their entry examinations. Participation in the study was voluntary and learners did not receive compensation for their participation. Students gave their informed consent and were told they could withdraw at any time.

Target items

Target items (N = 32) were chosen controlling for L2 frequency, concreteness, morphological transparency, conceptual familiarity, and orthographic neighbourhood size, and manipulating the experimental variables PoS and word length. The target items were sampled from the tenth and eleventh one-thousand-word frequency bands of a list compiled from the British National Corpus and the Corpus of Contemporary American English (Nation, 2012). The

procedure for item selection was as follows. Each word was first classified according to its PoS and length (i.e., number of letters). Subsequently, nouns were categorised as either abstract or concrete. An initial judgment by the first author was verified by a group scaling procedure with seven speakers of English as first language (L1). This procedure was based on the methodology of Spreen and Scquziulz (1966). Only items evaluated as concrete were included in the pool of potential target items. Items were then deleted from this pool if the participants' teacher considered it likely that they were known to them, were morphologically transparent/had deceptive morphological transparency (deceptive morphological transparency relates to words that appear to be made up of meaningful morphemes, but are in fact not [see Laufer, 1997]), or had referents likely to be unfamiliar to the participant group (e.g., culture-specific concepts). Furthermore, using Medler and Binder's (2005) procedure, potential target items were analysed to determine the size of the orthographic neighbourhood of each item (i.e., the number of target language words of the same length as a target item that differ from it by one letter), which was taken as a metric of orthographic distinctiveness (see Hunt & Elliot [1980] for discussion of how orthographic distinctiveness can impact form recall). Items were chosen so that target items of similar length were broadly homogenous in terms of orthographic neighbourhood size. A stratified sampling procedure was used to select target words from the resulting pool of potential items. An equal number of verbs and nouns were selected for each word length. This is represented in Table 3. The 32 target items were separated into four blocks. This was done to make the learning task less arduous and enable words to be learned and recycled in a systematic manner. Each block contained eight items, one noun and one verb of each word length (three, six, eight, and eleven letters).

<TABLE 3 AROUND HERE>

Definitions

Because the participants did not share the same L1, the meaning of the target items was disambiguated using English definitions. These were taken from a monolingual learner's dictionary of English. As such they were written in controlled language and, therefore, were at an appropriate level of grammatical and lexical difficulty for the learners; however, to check comprehensibility, the 2,000, 3,000, and AWL sections of the vocabulary levels test (Schmitt, Schmitt, & Clapham, 2001) were administered to all participants. Results ($M = 66.30$, $SD = 13.22$) showed that participants were likely to have good knowledge of the definitional vocabulary. The lexical frequency of the target definitions was measured, and some were modified to ensure that all defining vocabulary came from the 2,000 most frequent word families of the BNC/COCA lists (Nation, 2012). These definitions were piloted with a similar group of learners ($n=15$), who confirmed that all definitions were easy to read and that the definitional vocabulary was known to them. A full list of the target items and definitions is given in Appendix 1.

Learning software

The target items and definitions were loaded onto electronic flashcard software. Such software has been shown to be effective for vocabulary acquisition (Hung, 2015; Nakata, 2017) and popular with learners (Hung, 2015; Stroud, 2014). The software *Anki* was chosen for two reasons. Firstly, unlike the majority of flashcard platforms freely available, *Anki* allows users to control the appearance of each flashcard and, therefore, we were able to modify the structure of the flashcards to suit our research purposes and ensure that all flashcards were presented in exactly the same manner. Secondly, using *Anki*, it was possible to access user data regarding

the number of exposures each participant had to every target item. This was used as the metric for learning burden.

The flashcards were set up productively (Webb & Nation, 2017): participants saw an L2 definition and had to type the form of the matching target word. Upon attempting an answer (or leaving a card blank if the target form was unknown or could not be recalled), the correct written form was presented to a participant below his/her effort and any errors in learner production were indicated by the software. Thus, learners were able to compare their production to the correct written form and their mistakes were highlighted. Following this, learners evaluated the accuracy of their production by selecting one of three rating options. *Again* caused a target item to be presented within a period of one minute. Learners were instructed to select this option if their production was inaccurate or they did not write anything. *Good* resulted in a word being presented within a period of ten minutes. Learners were told to choose this option if their production was accurate. Finally, *Easy* meant a word was not presented again within a learning session. Participants were told to select this option if a target item was known to them prior to study. Once knowledge of an item was twice rated *Good* or once rated *Easy*, it was considered to have reached criterion and not repeated in that session. In practice, we would expect most items to be unknown on the first encounter and learners to select *Again*, but that in subsequent exposures, learners would provide a response and select *Good*.

This methodology allowed items to be seen as often as necessary for encoding to occur but ensured that, once encoded, all items received an equal number of retrievals. Retrieval here refers to retrieving already learned words, with each item retrieved twice before removal

from the pool. This dropout procedure allowed us to examine the number of exposures during the process of encoding (i.e., the number of times an item was rated *Again*) on a by-item basis which was adopted as the metric of learning burden.

Measurement Instrument

As the learning procedure involved self-report, a test of target item knowledge was developed. This test was administered upon completion of the learning task and after a delay of four weeks. The testing battery consisted of two instruments of 40 items; 32 target words and 8 previously unstudied words which were included to allow post hoc analysis of a test effect. The first test measured written form recall knowledge. Participants were presented with a definition (the same as was used during the learning procedure) and typed the equivalent form. Learners were not presented with the first letter of the target items as cue usage would have made distractor selection for the multiple-choice test of written form recognition (the second part of the testing battery) problematic. Thus, this instrument is a stricter measure of form recall than those employed in some other studies (e.g., Laufer & Rozovski-Roitblat, 2015), and may not have tapped into some partial productive lexical knowledge.

Responses were scored in both strict and lenient conditions. At the strict form recall level, correctly spelled responses were awarded a score of one, while incorrect responses were given a score of zero. The lenient form recall level allowed for one-third of an item to be spelled incorrectly. However, as there was little difference in the results of these two scoring systems, only the findings relevant to the strict scoring are presented in this paper.

The second instrument measured written form recognition knowledge of the target items. It employed a five-option multiple choice format with three distractors, the key, and an

additional *I don't know* option. The first distractor was a randomly selected target item matched for PoS with the key. The second distractor was a non-target item matched with the key for frequency and PoS. The final distractor represented either another randomly selected target item matched for PoS or another non-target item. Additionally, an *I don't know* option was included to minimise guessing (see Lucovich, 2014; Zhang, 2013).

The instrument was administered via Microsoft Excel. If a participant provided the correct response on the first test of form recall, that word was automatically omitted from the second test of form recognition. This design is based on the Computer Adaptive Test of Size and Strength (Laufer & Goldstein, 2004) which suggests a hierarchical relationship between recall and recognition: if a test taker is able to recall the form of an item, it can be assumed he/she would correctly identify the key in a form recognition test. Recent studies utilising implicational scaling have confirmed this assumption (González-Fernández & Schmitt, 2020). The full test battery was 80 items in length (32 target items and eight items included to measure a test effect, on a recall and a recognition instrument); however, as the test adapted to the responses of learners, a participant only saw all 80 questions if he/she answered all items on the form recall test incorrectly.

Procedure

Prior to collecting data, all aspects of the study were extensively piloted with a group of fifteen learners similar in proficiency to the experimental group. Feedback from this pilot showed that the target items were unknown, the L2 definitions posed no comprehension difficulty, the instructions regarding the flashcard software were clear, and learners interacted with the flashcard software as expected.

The study began with a thorough induction and practice session using the flashcard software, after which participants studied the vocabulary during class under the supervision of their teacher and the first author using identical computers. The learning procedure lasted for two sessions. In the first session, participants studied the thirty-two items split into four blocks of eight items. The next day, students restudied the same vocabulary, but to control for a sequence effect, the presentation order of the blocks was reversed and the order of items within each block was varied. This one-day between-session interval was chosen as it has been shown to facilitate retention (Cepeda et al., 2009; Dumay & Gaskell, 2007). Immediately after the second learning session, the participants completed the test of form recall and recognition. This marked the last time the participants were exposed to the target items until the same test was administered four weeks later. The first administration of this instrument will be referred to as the immediate test and the second as the delayed test.

Analysis

Prior to analysis, items viewed only twice during the learning procedure were deleted. Two exposures equated to a learner clicking *Easy* in both learning sessions, and thus indicated either disengagement with the learning procedure, or prior knowledge of a target item. In total, 140 data points were deleted. In doing so we controlled for previous knowledge of target items on a by-learner basis. Analysis was conducted with items seen three or more times, and thus only considered items previously unknown to the participants. The number of exposures required for learning were analysed to determine the effect of the target intralexical factors on learning burden (research question 1). Importantly, this analysis only included items for which knowledge was demonstrated on the immediate test. Separate analyses were conducted for the different strengths of knowledge tested. Thus, frequency of exposure data was considered

at two levels: items learned to the level of form recall and items learned to the level of form recognition (as measured by the immediate test).

Mixed-effects models (MEMs) were fitted onto the continuous variable, learning burden. Different models were fitted on the frequency of exposure data at the levels of form recall and form recognition. The fixed effects were word length, PoS, L1, and a vocabulary measure (the first three levels of the VLT [Schmitt, Schmitt, & Clapham, 2001]). Random effects were by-participant random slopes for length and PoS, and by-item random intercepts. MEMs are a form of regression analysis that allows for the investigation of both individual differences and systematicity within a data set (Murakami, 2016). Such models allow for the consideration of random effects that occur from population sampling, and fixed effects that typically relate to the independent variables (Cunnings, 2012). Thus, MEMs were used to determine any systematicity in learning burden/decay resulting from the target variables, while also taking into consideration individual variation.

To determine the effect of the two target variables, in addition to numerous predictors, on the decay of target item knowledge (research question 2), MEMs were fitted to the delayed test data. To ensure that we isolated the decay of learned knowledge, we only included items for which knowledge had been demonstrated on the immediate test. This prevented us from confounding the analysis of decay with learning that had occurred during the retention interval. As the outcome variable, retention, was binomial, logistic mixed-effect models (GLMER) were used. The fixed effects were word length, PoS, L1, number of exposures during the learning phase (learning burden), proficiency (IELTS), and the vocabulary measure. Random effects were by-participant random slopes for length and PoS, and by-item random intercepts. One point to

stress here is that by including number of exposures (learning burden) as a covariate in the analysis of decay, we were able to determine whether items that were harder to learn were more likely to be forgotten. This allowed us to answer research question 3.

For all models computed in this study, individual target items were nested in one of eight conditions according to the fixed effects of length (3, 6, 8, and 11 letters) and class (noun and verb). Also, continuous variables were log-transformed and, for the GLMERs, the BOBYQA algorithm was adopted and iteration number increased to 100,000, as recommended by Singmann (2014). All analyses were conducted using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in *R* (R Core Team, 2016). In all cases, model fitting began with the computation of a core model in which the outcome variable was predicted by word length and PoS, and their interaction. Maximal models were fitted with all potential covariates listed above. Backward elimination was then conducted using likelihood ratio tests.

Results

The eight items included to measure the possibility of gains from the immediate test were analysed. The results showed that less than one word was learned on average ($M = 0.77$, $SD = 1.07$). Thus, the results suggest that the first administration of the test led to minimal learning.

Learning Burden

In response to the first research question, this section discusses the effect of the target variables on learning burden. Table 4 shows the descriptive statistics for learning burden (i.e., the number of exposures needed for learning) by word length and PoS. Additionally, the number of items learned at each length and PoS is presented.

<TABLE 4 AROUND HERE>

MEMs were fitted to determine the effect of the target variables on learning burden. At the level of written form recall, there was a significant effect for length, with longer words requiring more exposures to be learned ($p < .03$). No statistically significant effects were found for PoS ($p = .93$) or any other covariate considered. At the level of written form recognition, the best-fitting MEM indicated that word length significantly impacted the learning burden ($p = .003$), with longer items associated with more exposures. However, in line with the form recall learning burden data, no statistically significant effect was found for PoS ($p = .70$). The MEM results are reported in Appendix 2.

Lexical Decay

With regard to the second research question, the descriptive statistics for the delayed test are presented in Table 5. The table presents the mean items recalled per word length and PoS. Only items for which knowledge was demonstrated on the immediate test were included in the analysis. This trimming was conducted according to the various strengths of knowledge considered; therefore, the analysis of form recall only considered those items shown to have been learned on the form recall instrument. This was also the case with the form recognition analysis.

<TABLE 5 AROUND HERE>

The scores from the form recall test indicated a floor effect, showing that most of the learned knowledge had been forgotten. This was likely caused by the length of the retention interval and/or the strength of knowledge facilitated by the learning procedure. Therefore, it was not possible to compute a model for the retention of form recall knowledge. Nevertheless,

to begin to understand the role of learning burden on the retention of form recall knowledge (research question 3), descriptive statistics are reported (Table 6). These show the number of items learned and retained according to their learning burden during the learning stage. The relative retention indicates the proportion of learned words that were subsequently retained.

<TABLE 6 AROUND HERE>

Although no clear pattern emerged, relative retention figures suggested that there was greater retention of items that received fewest exposures. Thus, it is perhaps the case that items with a lighter learning burden were better recalled than items with a heavier learning burden. However, further research is, of course, needed to confirm this initial observation.

Moving on to the analysis of decay at the level of written form recognition. The results indicated that neither word length ($p = .18$) nor PoS ($p = .55$) had a statistically significant effect on the decay of learned knowledge. There was similarly no significant effect for the interaction of the two target variables ($p = .27$) (see Appendix 3 for the coefficients of the random effects and the fixed effect structures of the best-fitting model). In response to the third research question, learning burden was found to significantly affect decay ($p < .001$), with more exposures (i.e., a higher learning burden) associated with the mitigation of loss of form recognition knowledge. Table 7 presents the number of items learned and retained by the number of exposures needed for learning. The data show that many items needed comparatively few exposures to be learned, but that relative retention was better when learners had more exposures to the target items. This finding differs from the trend reported for form recall; crucially however, that trend was not based on inferential statistics.

<TABLE 7 AROUND HERE>

Discussion

The present study examined the effect of PoS and word length on the learning burden and the decay of foreign language lexical knowledge that was intentionally learned using flashcard software. The effect of learning burden on decay was also examined. In this section, the results are discussed with reference to the three research questions the study set out to answer.

RQ1: What is the effect of PoS and word length on the learning burden of FL lexis?

Considering the effect of the target variables on learning burden, the results showed that length was associated with burden, with the shortest words posing the least burden and the longest words the greatest burden. This finding is in line with previous studies that operationalised burden in terms of learning gains (e.g., Baddeley, Thomson, & Buchanan, 1975; Gerganov & Taseva, 1982). No effect for PoS was found on learning burden. This contradicts the findings of some previous studies that have found nouns to pose a lighter learning burden than verbs (e.g., Ellis & Beaton, 1993). There are several explanations for this result. First, previous research has suggested that the effect of PoS may be moderated by learner proficiency, with the noun effect less prevalent with higher proficiency learners (Phillips, 1981; cited in Laufer, 1997). Studies that have shown an effect for PoS have often involved beginners (e.g., Ellis & Beaton, 1993), whereas the current study considered learners with B2 proficiency. Another possibility relates to the manner in which target item meaning was disambiguated in the present study. Studies that have demonstrated a noun advantage have tended to use L1 equivalents to convey the meaning of target items (e.g., de Groot, 2006). However, due to the multilingual nature of the participant group, it was necessary to employ L2 definitions in this

study. The definitions used did not explicitly state the PoS, communicating it instead via syntactic cues. The definitions of nouns began with an indefinite article while verb definitions began with a to-infinitive. It may be that these cues were not sufficiently salient to alert learners to the PoS of the target items. Therefore, future research that includes measurements of grammatical aspects of word knowledge and/or directly compares learning with L1 and L2 meaning presentation codes may help to clarify the moderating role of meaning presentation code on the PoS effect.

Overall, the finding for PoS speaks to the general inconclusiveness of research in this area (Laufer, 1997; Peters, 2020), with some studies finding a noun advantage (e.g., Ellis & Beaton, 1993), some reporting no effect for PoS (e.g., the present study), and others finding an advantage for verbs (e.g., Pigada & Schmitt, 2006). This difference might be explained by moderating variables such as learner proficiency or the code of meaning presentation. Further research looking at PoS and potentially impactful moderating factors is therefore needed before firm conclusions can be drawn.

RQ2: What is the effect of PoS and word length on the decay of FL lexis?

The results indicate that considerable decay took place over the four-week retention interval. A comparison of the form recall data from the immediate test with the delayed test shows decay figures of 92.5%. A similar comparison with the written-form recognition data showed that knowledge also decayed over the four-week retention interval. This is in line with previous research findings showing that learned knowledge typically decays over a retention interval (e.g., Waring & Takaki, 2004).

A comparison of the written-form recall scores with those of written-form recognition indicate there was considerably more decay of the former. In contrast to the high rates of decay reported for form recall, at the level of form recognition, 44.5% of the acquired knowledge was found to have been forgotten when tested four weeks later. To date, research has found that the acquisition of aspects of vocabulary to the level of form recall is more difficult than to the level of form recognition (e.g., González-Fernández & Schmitt, 2020; Laufer & Goldstein, 2004). This study extends this finding by suggesting that in addition to being harder to develop, form-recall knowledge also decays at a faster rate than form-recognition knowledge.

With regard to the effect of the target variables on the decay of form-recall knowledge, the large amount of decay learners experienced meant that we could not perform statistical analysis. Presently therefore, it is not possible to determine the effect of PoS and word length on the decay of written form recall knowledge. Future studies would need to employ a shorter retention interval to avoid a floor effect and facilitate statistical analysis.

Concerning the effect of the target variables on decay at the form recognition level, no effect for PoS or word length was found on the decay that occurred. This finding differs from previous research. Ellis and Beaton (1993) found a significant effect for word class after a four-week period of reduced input. The difference between their findings and those of the present study likely stems from methodological and/or measurement differences. Regarding the former, as previous studies have required learners to view target language an equal number of times, target words with a heavier learning burden (e.g., verbs) may have been encoded later and thus received fewer opportunities for retrieval than items with a lighter learning burden

(i.e., nouns). As retrieval frequency is associated with retention, this may have biased items with lighter learning burdens to be better retained. In the present study, a different approach to frequency of exposure was adopted in which items were seen as often as needed for them to be encoded but were removed from the learning procedure after two retrievals. Thus, unbalanced retrieval frequency is not likely to have impacted the results of the current study. Therefore, by controlling for retrieval frequency, the present findings may better represent the effect of PoS and word length on the decay process. Importantly, we believe the results from the present study show the need to employ designs that allow for a distinction between opportunities for learning and opportunities for retrieval of learned knowledge. Another explanation is that the difference in findings stems from the measurement instrument. This study measured knowledge at the levels of form recognition and recall, while Ellis and Beaton (1993) measured meaning and form recall.

From a methodological perspective, the results illustrate the challenge of exploring the decay of recall and recognition knowledge in the same study. A lengthy retention interval is likely to produce a floor effect on a recall measurement, while a brief retention interval may well produce a ceiling effect on a recognition instrument. Furthermore, as testing target items multiple times would lead to a test effect and likely impact retention, it may well be necessary for future research to prioritise one level of knowledge per study and adjust the length of the retention interval accordingly.

RQ3: What is the effect of learning burden on the decay process (irrespective of the target intralexical variables)?

The results showed that the number of exposures needed for learning was positively related to retention at the level of form recognition. Items that were viewed more often during

the learning procedure were better retained. Thus, these data suggest that a higher learning burden, reflected in more exposures, leads to less decay. More effort at the encoding stage when learners are developing knowledge of the form-meaning link seems to lead to less decay of form-recognition knowledge.

However, there is seemingly a contradiction here; word length was found to impact learning burden (with longer words requiring more exposures to be learned), and learning burden was found to affect decay (with fewer exposures associated with decay); however, word length did not significantly affect the decay process. Therefore, it may be that other intralexical, interlexical, and/or learner factors influenced decay. Numerous factors are associated with the learning burden of lexis (see Webb & Nation, 2017 for an overview). Perhaps one or more of these was ultimately responsible for the effect of learning burden on the decay of form recognition knowledge. Alternatively, individual differences such as learners' L1, aptitude, or motivation may explain these patterns.

Limitations and Future Directions

Due to the length of the retention interval, insufficient form recall knowledge remained after four weeks to allow statistical analysis. Because of the differential decay rates of lexical recall and lexical recognition knowledge, it may be methodologically challenging to investigate loss at both levels of word knowledge in one study. Future research, therefore, may need to prioritise recall or recognition knowledge and adjust the length of the retention interval accordingly. A further limitation relates to the characteristics of the participant group. Due to the multilingual make-up of the participants, it was not possible to investigate potentially confounding interlingual factors on the decay process. Therefore, future research utilising

participants with a common L1 is needed. An additional limitation is that student production during learning was not automatically evaluated; rather, learners were provided with the correct answer and feedback indicating the errors in their response, and they evaluated their performance by selecting *Good* when their production was correct or *Again* when their production was incorrect or no gave no response. This selection was based on feedback from the learning software. To mitigate the impact of this limitation, a thorough induction was conducted, and learners were monitored while they completed the task. Thus, it is unlikely that learners ignored instructions and evaluated their knowledge in a different manner. Fourthly, it was only possible to assess knowledge at the level of form recognition and form recall. These word-knowledge aspects have been shown to pose greater difficulty than meaning recognition and recall respectively. Future research should consider using measures of both form and meaning (as well as other word-knowledge aspects). Moreover, the findings need to be understood in light of the learning task. It has been shown that acquisition of form recall knowledge (Laufer & Goldstein, 2004) and flashcard learning using form production (Webb, 2005) are most challenging. Therefore, the findings are likely to differ were other learning activities to be employed. Finally, while it was necessary to avoid target item exposure during the retention interval for reasons of experimental validity, in an instructed context it is likely that learners will engage with target items more frequently than the retention interval used in this study. Therefore, more encounters with the target items in between the learning sessions or during the retention interval could affect the decay rates reported in this study. This speaks to the need for future research to employ different retention intervals in classroom settings.

Conclusion

The current study targeted the effect of two intralexical variables, word length and PoS, on the learning burden and subsequent decay of thirty-two English words. Overall, considerable loss of vocabulary knowledge that had been intentionally acquired was demonstrated at the levels of form recall and form recognition, with loss of the former greater than the latter. Word length had a significant effect on the learning burden, but not on the decay, of intentionally learned knowledge over a period of four-weeks. This suggests that the effect of intralexical factors on learning burden might not equate to their impact on decay, and points to the need to examine the effect of factors and learning conditions on both of these indices in vocabulary studies. No effect was found for PoS on either learning burden or decay. Importantly, this study has shown that a higher learning burden during the period of acquisition was associated with less decay at the level of form recognition. Items that were more burdensome for learners seem to be better retained at that level of lexical mastery. Crucially, results of this study suggest that lexical decay seems to be more clearly affected by the amount of effort required during learning, i.e., learning burden, than by intrinsic properties of the words. Overall, this study has added to our limited understanding of learning burden, lexical decay, and the interface between them; as such, it represents an initial foray into this area.

References

- Baddeley, A. D. (1990). *Human Memory and Theory and Practice*. Hove: Lawrence Erlbaum Associates.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575-589.
- Bahrack, H. P., & Phelps, E. (1987). Retention of Spanish Vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(2), 344-349.
- Bartolotti, J., & Marian, V. (2017). Orthographic Knowledge and Lexical Form Influence Vocabulary Learning. *Applied Psycholinguistics*, *38*(2), 427-456.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software; Vol 1, Issue 1 (2015)*.
- Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*(4), 236-246.
- Chen, C., & Truscott, J. (2010). The Effects of Repetition and L1 Lexicalization on Incidental Vocabulary Acquisition. *Applied Linguistics*, *31*(5), 693-713.
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–82.
- de Bot, K., & Weltens, B. (1995). Foreign Language Attrition. *Annual Review of Applied Linguistics*, *15*, 151-164.
- de Groot, A. M. B. (2006). Effects of Stimulus Characteristics and Background Music on Foreign Language Learning and Forgetting. *Language Learning*, *56*(3), 463-506.

- de Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning, 50*(1), 1-56.
- Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35-39.
- Ellis, N. C., & Beaton, A. (1993). Psycholinguistic Determinants of Foreign Language Vocabulary Learning. *Language Learning, 43*(4), 559-617.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In. Technical report no. 257 No. BBN-R-4854.
- Gerganov, E., & Teseva-Rangelova, K. (1982). The impact of association value and number of syllables of English words on memorization in teaching English to Bulgarian learners. *Supstavitelno Exikoznarie, 7*(4), 3-12.
- González-Fernández, B., & Schmitt, N. (2020). Word Knowledge: Exploring the Relationships and Order of Acquisition of Vocabulary Knowledge Components. *Applied Linguistics, 41*(4), 481–505.
- Gopnik, A., & Choi, S. (1990). Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development. *First Language, 10*(30), 199-215.
- Hansen, L., & Chen, Y.-L. (2001). What counts in the acquisition and attrition of numeral classifiers? *23, 1*, 90-110.
- Higa, M. (1965). The Psycholinguistic Concept Of “Difficulty” And The Teaching Of Foreign Language Vocabulary. *Language Learning, 15*(3-4), 167-179.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring Second Language Vocabulary through Reading. *Reading in a Foreign Language, 11*(2), 207-223.

- Horst, M., & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *Canadian Modern Language Review*, 56(2), 308-328.
- Hung, H. T. (2015). Intentional Vocabulary Learning Using Digital Flashcards. *English Language Teaching*, 8(10), 107-112.
- Hunt, R. R., & Elliot, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, 109(1), 49-74.
- Ishii, T. (2015). Semantic connection or visual connection: Investigating the true source of confusion. *Language Teaching Research*, 19(6), 712-722.
- Laufer, B. (1990). Words you know: how they affect the words you learn. In J. Fisiak (Ed), *Further Insights into Contrastive Linguistics*. Benjamins: Holland, 573-593.
- Laufer, B. (1991). *Similar Lexical Forms in Interlanguage*. Tübingen: Narr.
- Laufer, B. (1997). What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition, and Pedagogy* (pp. 140-155). Cambridge: Cambridge University Press.
- Laufer, B., & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language Learning*, 54(3), 399-436.
- 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.
- Li, S. (2016). The Construct Validity Of Language Aptitude: A Meta-Analysis. *Studies in Second Language Acquisition*, 38(4), 801-842.

- Lotto, L., & de Groot, A. M. B. (1998). Effects of Learning Method and Word Type on Acquiring Vocabulary in an Unfamiliar Language. *Language Learning*, 48(1), 31-69.
- Lucovich, D. (2014). Test Taking and DK Use on the Vocabulary Size Test. *Vocabulary Learning and Instruction*, 3(2), 69-77.
- Medler, D. A., & Binder, J. R. (2005). MCWord: An On-Line Orthographic Database of the English Language. Retrieved from <http://www.neuro.mcw.edu/mcword/>
- Murakami, A. (2016). Modeling Systematicity and Individuality in Nonlinear Second Language Development: The Case of English Grammatical Morphemes. *Language Learning*, 66(4), 834-871.
- Nakata, T. (2017). Does Repeated Practice Make Perfect? The Effects of Within-Session Repeated Retrieval on Second Language Vocabulary Learning. *Studies in Second Language Acquisition*, 39(4), 653-679.
- Nation, I. S. P. (2012). The BNC/COCA word family lists [PDF file]. Retrieved from https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf
- Nation, P. (2020). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 15-29). NY: Routledge.
- Olshtain, E. (1989). Is Second Language Attrition the Reversal of Second Language Acquisition? *Studies in Second Language Acquisition*, 11, 151-165.
- Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, 18(1), 75-94.
- Peters, E. (2020). Factors Affecting the Learning of Single-Word Items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies*. (pp. 125-142) London: Routledge.

- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1-28.
- R Core Team (2016). R: A Language and Environment for Statistical Computing. In. Vienna: R Foundation for Statistical Computing.
- Rodgers, T. S. (1969). On Measuring Vocabulary Difficulty. *IRAL, VII*(4), 327-343.
- Schmid, M. S. (2011). *Language Attrition*. Cambridge: Cambridge University Press.
- Schmitt, N. (1998). Tracking the Incremental Acquisition of Second Language Vocabulary: A Longitudinal Study. *Language Learning, 48*(2), 281-317.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363.
- Schmitt, N. (2010). *Researching Vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- 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.
- Singmann, H. (2014, May 15). Re. Warning messages from mixed model (glmer) [Online discussion post]. Retrieved from <https://stats.stackexchange.com>
- Spreen, O., & Scquziulz, R. W. (1966). Parameters of Abstraction, Meaningfulness, and Pronunciability for 329 Nouns. *Journal of verbal learning and verbal behaviour, 5*, 459-468.
- Stroud, R. (2014). *Student engagement in learning vocabulary with CALL*. Paper presented at the CALL Design: Principles and Practice; Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands.

- Tehan, G., & Tolan, G. A. (2007). Word length effects in long-term memory. *Journal of Memory and Language, 56*(1), 35-48.
- Tinkham, T. (1993). The Effect of Semantic Clustering on the Learning of Second Language Vocabulary. *System, 21*(3), 371-380.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Tonzar, C., Lotto, L., & Job, R. (2009). L2 Vocabulary Acquisition in Children: Effects of Learning Method and Cognate Status. *Language Learning, 59*(3), 623-646.
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130-163.
- Webb, S. (2005). Receptive and Productive Vocabulary Learning: The Effects of Reading and Writing on Word Knowledge. *Studies in Second Language Acquisition, 27*(1), 33-52.
- Webb, S., & Chang, A. C. S. (2015). How Does Prior Word Knowledge Affect Vocabulary Learning Progress In An Extensive Reading Program? *Studies in Second Language Acquisition, 37*(4), 651-675.
- Webb, S., & Nation, P. (2017). *How Vocabulary Is Learned*. Oxford University Press: Oxford.
- Willis, M., & Ohashi, Y. (2012). A model of L2 vocabulary learning and retention. *The Language Learning Journal, 40*(1), 125-137.
- Zhang, X. (2013). The I Don't Know Option in the Vocabulary Size Test. *TESOL Quarterly, 47*(4), 790-811.

Table 1

Word-related factors and their effect on the learning burden of foreign language lexical knowledge (adapted from Peters, 2020)

<i>Facilitative factors</i>	<i>Difficulty-inducing factors</i>	<i>Inconclusive</i>
Cognates	False cognates	
Phonotactic regularity	No phonotactic regularity	
More frequent L1 equivalent	Less frequent L1 equivalent	
Orthographic wordlike	Orthographic unwordlike	
		Word length
		Part of speech
Morphological transparency	Deceptive morphological transparency	
Concrete words	Abstract words	
One form, one meaning	Polysemy/homonymy	

Table 2

Lexical factors affecting the decay of L2 vocabulary

Facilitating factors	Difficulty-inducing factors
<ul style="list-style-type: none"> • word class (nouns) • frequency (high) • concreteness • word length (short) • cognate 	<ul style="list-style-type: none"> • word class (verbs) • frequency (low) • abstractness • word length (long) • non cognates

Table 3

A description of the target items

<i>Word length</i>	<i>Number of nouns</i>	<i>Number of verbs</i>	<i>Total</i>
3	4	4	8
6	4	4	8
8	4	4	8
11	4	4	8
<i>Total</i>	16	16	32

Table 4

Immediate test scores (form recall and recognition) by PoS and word length

Length	PoS	Form Recall		Form Recognition	
		Items learned (m)	Number of exposures (m)	Items learned (m)	Number of exposures (m)
3	Noun	2.23 (1.31)	5.23 (0.65)	3.33 (0.78)	4.75 (0.36)
	Verb	2.15 (1.15)	5.05 (0.13)	3.38 (0.87)	4.69 (0.18)
	Total	4.38 (2.14)	5.14 (0.45)	6.73 (1.32)	4.72 (0.27)
6	Noun	1.52 (1.22)	5.50 (0.64)	3.17 (0.97)	4.90 (0.23)
	Verb	1.77 (1.28)	5.78 (0.69)	3.29 (0.90)	5.13 (0.54)
	Total	3.29 (2.22)	5.64 (0.64)	6.46 (1.66)	5.02 (0.41)
8	Noun	1.65 (1.26)	5.91 (0.92)	3.31 (0.78)	5.38 (0.41)
	Verb	1.58 (1.29)	6.63 (1.16)	3.38 (0.89)	5.85 (0.30)
	Total	3.23 (2.29)	6.27 (1.04)	6.69 (1.39)	5.61 (0.42)
11	Noun	1.33 (1.17)	6.40 (0.94)	3.48 (0.77)	5.53 (0.31)
	Verb	1.08 (1.20)	6.89 (0.96)	2.92 (1.01)	5.85 (0.11)
	Total	2.42 (2.18)	6.65 (0.92)	6.40 (1.45)	5.69 (0.28)
Total	Noun	6.73 (4.06)	5.76 (0.85)	13.29 (2.60)	5.14 (0.45)
	Verb	6.58 (3.72)	6.09 (1.05)	12.98 (2.67)	5.38 (0.59)
	Total	13.31 (7.78)	5.92 (0.96)	26.27 (5.27)	5.26 (0.53)

Note: Maximum score per word length and PoS = 4; maximum score per word length = 8; maximum score per PoS = 16; and maximum total score = 32.

Table 5

Mean delayed test scores relative to learning by PoS and word length

Length	PoS	Written form recall	Written form recognition
3	Noun	0.25 (0.48)	2.00 (1.22)
	Verb	0.21 (0.50)	1.96 (1.10)
	Total	0.46 (0.71)	3.96 (2.00)
6	Noun	0.10 (0.37)	1.67 (1.23)
	Verb	0.17 (0.47)	1.63 (1.27)
	Total	0.27 (0.73)	3.29 (2.23)
8	Noun	0.08 (0.28)	1.98 (1.22)
	Verb	0.10 (0.37)	1.73 (1.27)
	Total	0.18 (0.44)	3.71 (2.16)
11	Noun	0.08 (0.28)	2.19 (1.18)
	Verb	0.00 (0.00)	1.44 (1.15)
	Total	0.08 (0.28)	3.63 (2.11)
Total	Noun	0.52 (0.79)	7.83 (3.87)
	Verb	0.48 (0.91)	6.75 (3.83)
	Total	1.00 (1.41)	14.58 (7.25)

Note: Maximum score per word length and PoS = 4; maximum score per word length = 8; maximum score per PoS = 16; and maximum total score = 32.

Table 6

Number of items learned (immediate test) and retained (delayed test) on the measure of form recall by frequency of exposure

Exposure frequency	Items learned	Items retained	Relative retention
3	73	9	0.12
4	187	15	0.08
5	97	9	0.09
6	75	5	0.07
7	48	5	0.10
8	35	1	0.03
9	31	4	0.13
10	14	0	0
11	11	0	0
12	10	0	0
13	6	0	0

Note: Maximum number of words that could be learned = 1396 (48 participants x 32 target items – 140 deleted data points)

Table 7

Number of items learned (immediate test) and retained (delayed test) on the measure of form recognition by frequency of exposure

Exposure frequency	Items learned	Items retained	Relative retention
3	151	86	0.57
4	377	216	0.57
5	178	101	0.57
6	123	80	0.65
7	75	48	0.64
8	64	46	0.72
9	46	34	0.74
10	29	20	0.69
11	24	16	0.67
12	18	15	0.83
13	14	10	0.71
14	5	3	0.60
15	4	3	0.75
16	3	3	1.00

Note: Maximum number of words that could be learned = 1396 (48 participants x 32 target items – 140 deleted data points)

Appendix 1

The Target items and Definitions

Item	PoS	Length	L2 Definition
bib	Noun	3	a piece of cloth or plastic tied under a baby's face
keg	Noun	3	a round wooden container with a flat top and bottom
tic	Noun	3	a sudden movement of a muscle in your face
orb	Noun	3	a bright ball-shaped object such as the sun or the moon
voyeur	Noun	6	a person who enjoys watching other people
cinder	Noun	6	a very small piece of burnt wood
clique	Noun	6	a small group of people who spend their time together
zealot	Noun	6	a person who has very strong feelings
asterisk	Noun	8	an image placed next to a word to make people notice it
spinster	Noun	8	an unmarried woman who is old
cauldron	Noun	8	a large round metal pot for boiling water over a fire

mackerel	Noun	8	a sea fish that is blue and silver, and has a strong taste
contraption	Noun	11	a machine that looks strange and is unlikely to work well
archipelago	Noun	11	a group of small islands
harpsichord	Noun	11	a musical instrument like a piano
condominium	Noun	11	an apartment in a building with several apartments
kip	Verb	3	to sleep somewhere that is not your home
irk	Verb	3	to make someone feel annoyed
nab	Verb	3	to catch or arrest someone who is doing something wrong
bop	Verb	3	to hit someone gently
decant	Verb	6	to pour wine from one bottle into another
prance	Verb	6	to walk with high steps
frolic	Verb	6	to play and move around in a happy way
heckle	Verb	6	to interrupt and try to embarrass someone who is speaking in public
conflate	Verb	8	to combine two or more things to form a single new thing
truncate	Verb	8	to make something shorter
venerate	Verb	8	to respect someone because they are old or important
expedite	Verb	8	to make a process happen more quickly
matriculate	Verb	11	to officially begin studying at a university
scintillate	Verb	11	to turn on and off quickly
pontificate	Verb	11	to give your opinion about something
regurgitate	Verb	11	to bring food that you have already eaten back into your mouth

Appendix 2

Fixed Effects for Selected Recall Model (Learning)

Parameter	Estimate	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	1.41	0.09	16.06	<.001
Word length	0.12	0.05	2.31	.03*
PoS	0.01	0.12	0.09	.93
Length*Class	-0.02	0.07	-0.31	.76

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Random Effects for Selected Recall Model (Learning)

Parameter	Variance	<i>SD</i>
Item	.01	.06
Participant	.01	.09
Word Length	.02	.12
PoS	.01	.08
Length x Class	.01	.07

Fixed Effects for Selected Recognition Model (Learning)

Parameter	Estimate	SE	<i>t</i>	<i>p</i>
Intercept	1.40	0.06	23.67	<0.001
Word length	0.12	0.04	3.12	.003**
PoS	0.03	0.08	0.39	.70
Length*Class	-0.03	0.04	-0.66	.51

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Random Effects for Selected Recognition Model (Learning)

Parameter	Variance	SD
Item	.01	.04
Participant	.01	.12
Word Length	.03	.16
PoS	.01	.08
Length x Class	.01	.08

Appendix 3

Fixed Effects for Selected Recognition Model (Decay)

Parameter	B	SE	z	p
Intercept	-0.34	0.69	-0.49	0.06
Word length	-0.46	0.34	-1.36	0.17
PoS	-0.54	0.89	-0.6	0.55
Exposures	0.69	0.19	3.73	.0001***
L1:				
Mandarin	1.01	0.3	3.43	.001***
Arabic	0.47	0.52	0.91	0.36
Thai	0.57	0.42	1.35	0.18
Malay	1.1	0.7	1.57	0.12
Hungarian	0.58	0.73	0.8	0.43
Length*Class	0.52	0.47	1.1	0.27

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Random Effects for Selected Recognition Model (Decay)

Parameter	Variance	SD
Item	.64	.80
Participant	.21	.45
Word Length	.68	.81
PoS	.12	.35
Length x Class	.27	.52

Author's addresses

Samuel Barclay

Nottingham Language Centre
Nottingham Trent University
50 Shakespeare Street
Nottingham
NG1 4FQ

sam.barclay@ntu.ac.uk

Ana Pellicer-Sánchez

UCL Institute of Education
20 Bedford Way
London
WC1H 0AL

a.pellicer-sanchez@ucl.ac.uk