Examining the Learning Burden and Decay of Second Language Vocabulary Knowledge

by

Samuel Christopher Barclay

A thesis submitted to UCL Institute of Education
in fulfilment of the requirements for the degree of
Doctor of Philosophy

UCL Institute of Education

2021

I, Samuel Barclay, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ABSTRACT

Research in second language (L2) vocabulary learning has shown that not all words are equally easy to learn, and that several factors affect the difficulty with which words are acquired, i.e., their learning burden. However, research to date has explored only a few of the many factors affecting learning burden and existing findings are inconclusive. Another important finding in the L2 vocabulary learning literature is that L2 lexical knowledge is forgotten after learning but, to date, there has been minimal investigation of the variables that influence lexical decay. It has also been assumed that the lexical items most difficult to acquire are those easiest to forget, pointing towards a positive relationship between learning burden and decay (Webb & Nation, 2017). However, there is currently limited empirical evidence to support this assumption.

This thesis reports research undertaken to explore the effect of different variables on learning burden and lexical decay, and the relationship between burden and decay. It consists of three empirical studies that investigated the effect of intralexical (i.e., part of speech, word length), contextual (i.e., meaning presentation code, form presentation mode), and individual (i.e., perceived target item usefulness, language learning aptitude) factors on the learning burden and decay of vocabulary knowledge that was intentionally learned with flashcard software. Each study also considered the effect of learning burden on lexical decay. Additionally, a cross-study analysis was conducted to explore the effect of the retention interval length on decay. The empirical studies showed that word length, aspects of language learning aptitude, and form presentation mode impacted learning burden but not decay, with shorter words, higher associative memory capacity, and bimodal form presentation related to less burden. Perceived target item usefulness was found to have no effect on burden or decay. Meaning presentation code and PoS were found to affect both burden and decay. Lexical items presented with an L2 definition and verbs were more burdensome and more likely to decay than items presented with an L1 equivalent and nouns. The findings also indicated that more learning burden was associated with a higher likelihood of decay. The cross-study analysis showed that decay was not directly proportional to the retention interval length and that form recall knowledge was more susceptible to decay than form recognition. Additionally, this thesis explores implications for vocabulary research and L2 pedagogy.

IMPACT STATEMENT

This thesis sought to better understand why some second language words are harder to learn, why some words seem to be forgotten more quickly, and how learning difficulty and forgetting are related. Over three studies, various factors that we might expect to impact the learning burden and decay of lexical items were investigated. Study 1 looked at the role of factors relevant to the form and meaning of words themselves. Study 2 focused on how the meaning of target words is communicated to learners and the degree to which a word is perceived to be useful. Study 3 considered the role of language learning aptitude and also investigated how the form of a target word is presented to a learner. Finally, the data of all studies were compared to determine how the length of the retention interval (the period in between learning sessions when a learner is not exposed to a target item) impacted the amount of decay that took place. Understanding these areas has meaningful applications for second language learning, teaching, and research.

In relation to learning and teaching, a better understanding of lexical decay can improve the efficacy of language pedagogy. Clarifying the manner and speed of foreign language lexical loss, as well as the variables associated with it, will lead to more robust learning procedures that offset the natural process of forgetting. This thesis found that knowledge of verbs decays faster than nouns. This suggests that nouns and verbs may need different pedagogical treatments, with verbs receiving richer instruction and more frequent recycling to ensure they are retained. Furthermore, this thesis found that presenting the meaning of foreign language words via a first language equivalent led to faster learning and less forgetting, while presenting the spoken and written forms of words at the same time led to faster learning without increasing decay. Thus, teachers should be advised to use the L1 and present novel lexical items using the spoken and written forms. Additionally, more burden while learning a word increased the probability of forgetting. This finding has implications for developers of EdTech programmes that employ algorithms to determine recycling patterns. Such algorithms should take initial burden into consideration to ensure that target items are recycled in a timely manner best suited to the individual learner and item being learned. The findings of the present study directly inform the design of such algorithms.

With regard to research, the thesis employed an innovative methodology. Flashcard software was used to develop lexical knowledge. This allowed items to be removed from the learning treatment after they had been learned. It also allowed exposures during the learning process to be distinguished from retrievals of learned knowledge. This is an improvement on previous methodologies. In Study 3, this software was paired with key-stroke logging software, which, to the best of my knowledge, is the first time these tools have been used in partnership. The data provide considerable detail regarding the behaviour of learners while they are engaged in a vocabulary learning activity. Additionally, using flashcard software allowed learning burden to vary by learner and by target item, which better reflects the burden construct. Such an operationalisation differs from much of the literature to date and represents a methodological improvement. Thus, the studies presented in this thesis represent considerable methodological innovation that have allowed learning burden and decay to be explored in a level of detail rarely seen in classroom-based studies of vocabulary learning and they will inform future research in this area.

ACKNOWLEDGEMENTS

Reflecting on the past six-or-so years, I realise that I could not have completed this thesis, not even come close, without the help and guidance of several individuals. Firstly, my supervisor, Dr Ana Pellicer-Sánchez, has been truly incredible. Not only has she generously dedicated her time and expertise to my development, but has done so with passion, kindness, and patience. I looked forward to each and every supervision, as we threw about ideas and discussed nuances of theory, design, and application. I left those sessions motivated, energised, and with a renewed sense of self-belief. Coming into this project, I thought that the role of a supervisor was just to foster the academic development of students; to advise on research design and provide comments on writing. Ana has certainly done all of that, but she has also done so much more. Chiefly, she has given me the belief that I can contribute to this field and the motivation to pursue the answers to difficult questions and embrace nuanced, complex solutions. My thinking about second language acquisition has been shaped and will forever be affected by my interactions with Ana. I would not have started, let alone finished, a PhD if I had not been assigned to her for my MA dissertation. It was a lucky coincidence for which I will be eternally grateful. For the last eight years, she has guided me, pushed me, and given me agency. I feel so fortunate and proud to have been mentored by her all these years – thank you so much Ana.

I also want to thank my wider academic family. Undertaking a project like this, part-time, over a number of years, could easily lead to feelings of isolation. I am incredibly fortunate that this was not my experience. Before starting the PhD, I met Prof. Norbert Schmitt at a conference in the USA and told him that I had just been accepted to study with his former student, Ana. His response was short but life changing: “That’s great, congratulations. Come to Nottingham”. He was right; the PhD, while beastly at times, would have been a whole different beast had I been 12,000 miles from my supervisor and the University of Nottingham’s Vocabulary Research Group (VRG). It is difficult to overstate the important role that this wonderful community of vocabulary researchers have played in supporting me through this thesis. It was in a VRG meeting, during a discussion with Dr Michael Rodgers after a particularly terrible presentation I gave, that I stumbled upon the idea of using flashcards for my research. And it was while reviewing a paper that I turned to Dr Beatriz González-Fernández and first considered using a dropout methodology. It was later that year that Dr Laura Vilkaitė-Lozdienė took time away from her own work to introduce R and mixed-effects modelling to me. And it was over dinner with Dr Benjamin Kremmel that I thought about comparing the length of retention intervals in the final chapter. There were numerous other interactions that shaped the research in this thesis, and me as a researcher, with these people as well as with Dr Pawel Szudarski, Dr Marijana Macis, and Dr Hana Al-Mutairi. I feel incredibly lucky to have such a generous academic family, with Norbert and Diane Schmitt at the head of the table. Norbert and Diane have believed in me, trusted me, and generously supported me with their time and expertise. I only hope I can help others in the manner that they have helped me. Thank you, and you all, for your friendship, guidance, and encouragement.

I would also like to thank members of the BAAL Vocabulary Special Interest Group and the London Second Language Acquisition Research Forum. Over the years, I have been lucky enough to present my work at the conferences of these organisations. The feedback I received has made a demonstrable impact to the work presented in this thesis, introducing me to theoretical positions I would not have come to alone and helping to shape the design of my research. I am thankful and humbled by their interest in my work and their kindness.

I would also like to thank my family and friends. My family have supported me in myriad ways over this long journey. From continuing to love me even when I talked about my research, giving me space to explore knotty issues and asking probing questions to help me get untangled, and putting me up (and putting up with me) when I visited London for supervisions or Bath for a rural getaway. Thanks go to mum, dad, Alice, Bill, Luke, Diana, Anna, Matt, Gabby, Liz, Ted, Tessa, and my wonderful niblings. Thanks to my friends for their unwavering support – Tom, Ben, Gianluca, Giuliana, Vivien, Tomoki, Rina and many more. Thanks also to my colleagues and friends at Nottingham Trent University. Particular thanks go out to Walter, Martin, Sarah, and Amy. I want to make special mention and give special thanks to Miho Ogura. Her kindness and energy have been constant sources of support and continue to inspire me. She sees the world differently from others, is incredibly compassionate, and offers wonderful advice. Quite simply, she is the best person I know. もし美穂に出会ってなかったら、ここにいないだろう. 本当に感謝しています.

Most importantly, I would like to thank my students; those who took part in my research studies and those who have inspired me and continue to do so. Without them there would be no thesis. Without them I would not have asked the questions this thesis sought to answer. Many years ago, in a fifth-period class on a rainy Tuesday, a student at Ehime University, Japan, said in a *sairishu* class that “there is value in finishing anything”. He was right; over the past few years there have been many highs and lows as I struggled to complete this thesis. I considered throwing in the towel several times, but there is value in finishing and there is a pride that comes from seeing something, anything, through to its conclusion. It is only thanks to all those mentioned here, and numerous others, that I have been able to make it this far, to see this doctorate through to its conclusion. Thank you all; I am forever grateful.

Parts of this thesis have been published in peer-reviewed publications:

- Fragments of the literature review were published in co-authorship with Prof. Norbert Schmitt in the *Second Handbook of English Language Teaching* published by Springer in 2019 (Barclay & Schmitt, 2019).

    Barclay S., & Schmitt, N. (2019) Current Perspectives on Vocabulary Teaching and Learning. In Gao X. (Eds), *Second Handbook of English Language* Teaching (pp. 799-819). Springer International Handbooks of Education.

- Study 1, presented in Chapter 4, was published in co-authorship with Dr Ana Pellicer-Sánchez in *ITL – International Journal of Applied Linguistics* (Barclay & Pellicer-Sánchez, 2021).

    Barclay, S. & Pellicer-Sánchez, A. (2021). Exploring the learning burden and decay of foreign language vocabulary knowledge: The effect of part of speech and word length. *ITL - International Journal of Applied Linguistics.* Advanced online publication.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CS	Codeswitching
EAP	English for Academic Purposes
EFL	English and a foreign language
ELT	English language teaching
ESL	English as a second language
GLMER	Generalised logistic mixed-effects model
HE	Higher education
ID	Individual differences
ILH	Involvement load hypothesis
L1	First language
L2	Second language
LLA	Language-language aptitude
MEM	Mixed-effects model
PAL	Paired-associate learning
PoS	Part of speech
SLA	Second Language Acquisition
TESOL	Teaching English to speakers of other languages
TNE	Transnational Education
TOPRA	Types of Processing – Resource Allocation
VKS	Vocabulary Knowledge Scale
VLT	Vocabulary Levels Test
VST	Vocabulary Size Test

Chapter 1: Introduction

1.1 Background and Rationale

Vocabulary learning is a critical part of developing second language (L2) competence. Vocabulary knowledge has been shown to play a significant role in L2 comprehension (Laufer & Ravenhorst-Kalovski, 2010; van Zeeland & Schmitt, 2013a; Webb & Rodgers, 2009) and production (de Jong et al., 2012; Leki & Carson, 1994), and to affect learner performance on general proficiency measures (Kaneko, 2017; Laufer & Goldstein, 2004). Unsurprisingly therefore, vocabulary development is considered by researchers, teachers, and students alike to be among the most important aspects of learning a language (Webb & Nation, 2017). Reflecting this importance, there has been a wealth of research in this area in recent years (Meara, 2018; Nation, 2013), with hundreds of research papers (Meara, 2018), research manuals (Nation & Webb, 2011; Schmitt, 2010), a handbook (Webb, 2020), and several special issues (e.g., Laufer, 2017) dedicated to vocabulary studies published within the last decade.

One key strand of this research relates to the efficacy of different approaches to learning and teaching vocabulary. Studies have shown that vocabulary can be intentionally learned from flashcards (Nakata, 2016), wordlists (Ishii, 2015), and dictionaries (Laufer, 1994), as well as through incidental exposure from reading (Pellicer-Sánchez, 2016, 2017; Waring & Takaki, 2003), listening (van Zeeland & Schmitt, 2013b), and viewing TV and films (Peters & Webb, 2018). One particularly robust finding from this research is that accrued lexical knowledge decays after it has been acquired (Ellis & Beaton, 1993). Experimental studies typically report, for instance, a loss of knowledge between immediate and delayed post-tests. However, the majority of studies conducted in this area have not explicitly considered the amount of forgetting that takes place nor the factors that affect the forgetting process. Thus, there is currently minimal research to inform our understanding of L2 lexical decay (Schmitt, 2010).

Another important finding from the vocabulary learning literature is that lexical knowledge does not seem to decay in a uniform manner. Studies, for example, generally report some loss and some retention of learned knowledge on delayed post-tests, suggesting that some lexical items are more susceptible to decay than others (Berman & Olshtain, 1983). While a large literature exists on language loss in conditions of reduced or no target language contact (Hansen & Chen, 2001), minimal systematic investigation of loss in contexts of continued target language contact, referred to in this thesis as *decay*, has been undertaken. Indeed, beyond the assumption that lexical knowledge decays as a function of time, the variables that influence the extent and speed of this loss have received less research attention than those influencing acquisition (Schmitt, 2010). Thus, we currently know very little about how various factors impact the decay of L2 vocabulary knowledge. The research presented in this thesis was conducted, in part, to better understand the phenomenon of lexical decay and the variables that impact this process.

Another common concern of vocabulary learning research has been to examine the factors that make certain words harder to learn. The extent to which a word is easy or hard to learn is associated with its learning burden. The learning burden of an L2 lexical item is the amount of effort a learner needs to exert in order to learn it (Webb & Nation, 2017). Words with a heavy learning burden are harder to learn, while words with a light learning burden are easier to learn. The learning burden of a word 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). Additionally, characteristics of words themselves, i.e., intralexical factors, also contribute to a word's learning burden (Laufer, 1990).

Research has delineated the effect of several factors on the learning burden of L2 vocabulary, finding that some variables impact learning difficulty (e.g., cognateness, concreteness, and phonotactic typicality). There are also numerous factors that have received minimal research attention; for instance, we currently know little about how the manner of form and meaning presentation, and the length of the interval between learning sessions, impacts burden. Similarly, research has yet to satisfactorily explore the effect of part of speech (PoS) and word length, on learning burden (Peters, 2020), and there has been almost no consideration of how burden is impacted by learner-related factors such as the perceived usefulness of a lexical item or language learning aptitude. The studies presented in this thesis investigated these areas and thus contribute to our understanding of burden.

Several factors can impact the learning burden of L2 vocabulary and research has considered the effect of many of these variables on the learning process (Laufer, 1997). However, we currently know very little about their comparative effect on lexical decay. One common assumption in vocabulary research is that the rate of decay relates to the difficulty of learning, with words that were initially harder to learn most likely to be forgotten (Webb & Nation, 2017). This assumption suggests a positive relationship between learning burden and lexical decay; however, there is little research evidence to support this claim. One goal of this thesis was to understand this relationship between burden and decay.

Another common assumption in vocabulary research is that the amount of decay that occurs is proportionate to the length of the retention interval. The retention interval is the period of time between a learner's exposure to lexical items and the administration of tests measuring retention of those items. This is often realised as the length of time between an immediate test and a delayed test. In pedagogy, it might equate to the period between language classes. Studies have used an array of retention intervals ranging from a few minutes (Ishii, 2015) to many years (Bahrick, 1984). We typically expect there to be more decay over a longer retention interval, yet this area has received minimal research attention within the field of vocabulary studies. One aim of this thesis was to delineate the effect of the retention interval length on the decay process.

The rationale for attempting to better understand learning burden, lexical decay, and the relationship between learning burden and decay stems from my experience as a teacher. I have taught English as a Foreign Language (EFL) for over fifteen years. In class, I have personally observed that students are able to learn some words more quickly than others, and also that some learners can acquire vocabulary faster than their peers. During this time, I have also noticed that learners struggle to retain some lexical items while other words seem particularly resilient. Furthermore, I have observed that some students suffer considerable loss in the interval between classes while others seemingly forget little. Initially, I considered this loss to be merely a function of time, ignoring observable by-item and by-learner variation. Moreover, discussions with colleagues typically invoked variations on clichés such as "students need to use their vocabulary knowledge or they'll lose it". However, such advice also contradicted the differential patterns of loss I observed in my classroom: it seemed that some learners did not need to use the target items, nor did some items need to be used, to avoid being forgotten. The research presented in this thesis was conducted to understand these observations, with the belief that a better understanding of lexical decay and learning burden, as well as the various factors affecting these complex phenomena, would lead to meaningful pedagogical interventions to mitigate such loss.

1.2 Defining Vocabulary Knowledge

Vocabulary knowledge is commonly conceptualised as a multidimensional construct (Henriksen, 1999; Richards, 1976). The most established framework of vocabulary knowledge is Nation (2013). This framework includes nine aspects of word knowledge organised into three categories. Each aspect can be known receptively and productively. Receptive knowledge is associated with decoding language input through listening or reading, while productive knowledge is related to producing vocabulary through speaking or writing (Nation, 2013). Nation's (2013) framework is presented in Table 1.1. Table 1.1 shows that vocabulary knowledge is complex, with full mastery of a word involving the attainment of multiple aspects of knowledge at both the receptive and productive level.

Table 1.1

Nation's Aspects of Word Knowledge Framework (Nation, 2013)

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concepts and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does this word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Note. R = receptive knowledge, P = productive knowledge

Studies investigating the multidimensionality of lexical knowledge have found some key differences between the various components of word knowledge. Research has shown that some components are easier to learn than others (Laufer & Goldstein, 2004; González-Fernández & Schmitt, 2019), with receptive knowledge easier to acquire than productive knowledge (Laufer and Goldstein, 2004) and the form-meaning link easier than other aspects of word knowledge such as word parts, concepts and referents, and collocation (González-Fernández & Schmitt, 2019). The various aspects also differ in terms of significance. At the earliest stages of learning a word, the form-meaning link is perhaps the most essential aspect of word-knowledge (Schmitt, 2010). This is the case because knowledge of other aspects (e.g., collocation, register) presumes knowledge of the form-meaning link. Thus, the form-meaning link is likely to be one of the first and among the most critical components to be learned. Reflecting this importance, the research presented in this thesis investigated the learning and decay of the form-meaning link.

1.3 Approaches to Vocabulary Teaching and Learning

Vocabulary can be learned incidentally or intentionally (Schmitt, 2000). Intentional learning relates to learning which occurs when the primary intention of the learner/task is the development of linguistic knowledge (Webb & Nation, 2017). Incidental learning occurs as an unintended consequence of language use or exposure (Schmitt, 2010). Studies have shown that vocabulary can be incidentally learned while learners are reading (Pellicer-Sánchez, 2016, 2017; Waring & Takaki, 2003), listening (Van Zeeland & Schmitt, 2013b), viewing TV or films (Peters & Webb, 2018), and gaming (Sundqvist & Wikström, 2015). Research shows that some of these sources of input foster more learning gains than others; however, it is generally the case that the learning gains from incidental conditions are small

(Webb, 2020). This is particularly the case when compared to intentional learning conditions, which have been found to foster considerable learning gains (Laufer & Shmueli, 1997; Nakata, 2016; Rogers, 1969). An intentional learning approach may simply involve discussion of specific target items, or activities that require learners to focus on a group of words. Additionally, however, several pedagogical tools are associated with intentional learning, including word lists, dictionaries, glosses, and flashcards. Each has received considerable research attention and results show that some are more effective than others (Laufer, 2006). In particular, flashcards have been argued to be an efficient means of developing knowledge of the form-meaning link (Webb & Nation, 2017). While acknowledging that both approaches are important for vocabulary development (Webb & Nation, 2017), the research presented in this thesis focused on intentional learning of the form-meaning link through flashcards.

Traditionally, a flashcard is a small card with a word form written on one side and the relevant meaning(s) given on the reverse. Learners can interact with flashcards by looking at the form and recalling the meaning (receptive flashcard use) or by looking at the meaning and recalling the form (productive flashcard use). Advancements in computer and mobile technology have led to the development of numerous electronic flashcard programmes (see Nakata, 2011, for a review). Electronic flashcards have several benefits for learning and research that are discussed in detail in Chapter 3. Because they are useful for creating the initial form-meaning links of words and they are popular with learners, the research presented in this thesis employed electronic flashcards to teach the form-meaning connection of the target lexical items.

1.4 Defining and Conceptualising Learning Burden and Lexical Decay in this Thesis

Learning burden involves the difficulty with which L2 lexical items are acquired. It is associated with the speed of learning, with a heavy learning burden slowing the learning process and a lighter learning burden leading to comparatively faster acquisition. In contrast to learning burden, which has a relatively uncontentious definition, there are important terminological issues around the definition and conceptualisation of decay. Lexical decay relates to the loss of vocabulary knowledge. Language loss has mostly been considered in terms of language attrition. The term *attrition* has been used to refer to diverse phenomena such as language death, language shift, and language loss stemming from neurological impairment (Freed, 1982; Van Els, 1986). However, in SLA, it has been understood 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* concerns language loss that occurs in contexts of limited or no contact with the attriting language.

The term *decay* is typically used synonymously with *attrition*. It has been used to refer to language shift and language death (Schmidt, 1991), the effects of brain trauma on linguistic knowledge (Watkins et al., 2012), and language loss in contexts of reduced or no target language contact (Hutz, 2004). Thus, *attrition* and *decay* have been used to discuss various types of language loss and are often used interchangeably. However, this thesis draws a distinction between the two terms. While *attrition* considers language loss in environments with reduced or no target language exposure (Freed, 1982; Hansen & Reetz-Kurashige, 1999), *decay* is used in this thesis to describe the forgetting of L2 linguistic knowledge in contexts of continued exposure to the target language (see Schmitt, 2008).

Adapting Schmid's (2011) definition given earlier in this section, the term *lexical decay* is therefore defined in this thesis as the loss of lexical knowledge by healthy individuals in contexts of continued language exposure. That is, contexts 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 despite continued exposure to the L2 more broadly, or a learner who, while residing in the L2 environment, is exposed to certain words that are not encountered again.

Other terms, such as *loss* and *forgetting,* have been used in the literature. However, the use of these terms is 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; Paradis, 2007). This thesis, while recognising this criticism, employs *loss* and *forgetting* as umbrella terms to refer to language loss in healthy individuals (i.e., encompassing attrition and decay), in line with previous use in the field (de Groot, 2006; de Groot & Keijzer, 2000).

1.5 Operationalising Learning Burden and Lexical Decay in this Thesis

Research to date has generally 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 knowledge are considered to pose difficulty. However, measuring learning burden 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 encountered 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. In this thesis, learning burden is operationalised as the number of times each learner needs to see each word for it to be learned, which is likely to provide a clearer picture of the effect different variables have on the learning burden of L2 lexis.

Much of the research into L2 vocabulary learning and teaching employs experimental research designs (e.g., Hulstijn & Laufer, 2001). Nation and Webb (2011) explain that an experimental study should have an immediate and a delayed test and may also have a pre-test. Such studies typically compare a pre-test to the immediate test to measure learning, and a pre-test to the delayed test to show retention (Nation & Webb, 2011). It is less common, however, for such studies to directly compare the immediate and delayed tests in order to measure lexical decay and to explicitly explore factors affecting the extent of decay that occurs (but see de Groot, 2006; Ellis & Beaton, 1993).

The research presented in this thesis explored lexical decay by investigating the maintenance of individual lexical items between an immediate and a delayed test. Comparing knowledge in this manner leads to two potential outcomes; word knowledge can either be maintained or it can decay between the test administrations. These two results are essentially two sides of the same coin and look at the same phenomenon from different perspectives. Following de Groot (2006), the research presented in this thesis considered decay by analysing items that had been learned on the immediate test and maintained to the delayed test.

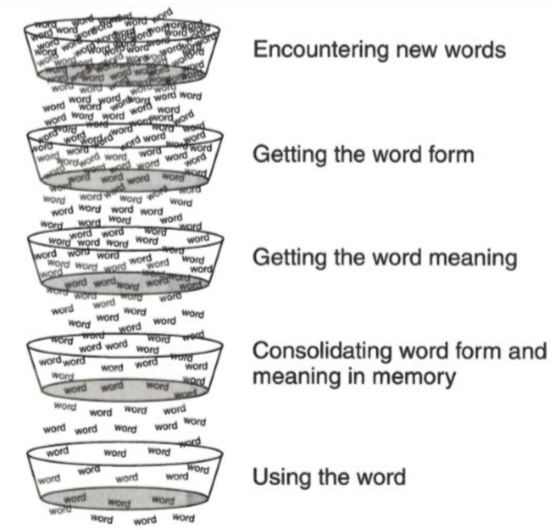
1.6 Potential Significance of Research on Learning Burden and Lexical Decay

Research into lexical decay may have important implications for both theory and practice. With regard to theoretical implications, the field of vocabulary studies currently lacks a unified model of vocabulary learning (Schmitt, 2019). Crucially, the theories that have been proposed to date typically envisage vocabulary learning as a linear phenomenon, making little allowance for the loss of accrued knowledge. For instance, Brown and Payne (cited in Hatch & Brown, 1995) conceptualised vocabulary acquisition as a series of sieves through which words pass as they are acquired, with some words progressing through all five sieves and others impeded along the way (see Figure 1.1). This model does not overtly refer to or account for lexical decay. Instead, it seems to conceptualise decay as failure to progress from one sieve to the next, with no consideration given to how accrued knowledge may backslide. The Five-Step Model implies, therefore, that the vocabulary learning process is unidirectional and ignores lexical decay. Unlike such models, it is crucial for theories of vocabulary acquisition to account for lexical decay. In order to develop such a comprehensive model of vocabulary development, it is first necessary to extend our current limited understanding of lexical decay. The research presented in this thesis is an attempt to do that.

Brown and Payne's (cited in Hatch & Brown, 1995) model also fails to explicitly account for the role of learning burden in the acquisition process. Presumably, items that have a heavy learning burden are unable to pass from one sieve to the next, suggesting that words with a heavy learning burden often do not progress to full mastery. The model also implies that lexical items have multiple discrete learning burdens relevant to the form, the meaning, the ease of consolidating the form-meaning link, and the ease of use. While there certainly are numerous factors that impact the learning burden of L2 lexical items (see Laufer, 1997 and Peters, 2020) that relate to these categories (e.g., form - word length; meaning - concreteness), it is currently unclear whether they impact learnability in a sequential manner as suggested by the model. Furthermore, the model does not speak to the relationship between learning burden and forgetting; for instance, it does not predict what would happen to words that are learned despite possessing heavy learning burdens. A comprehensive model of lexical acquisition needs to overtly account for the impact of various factors on learning burden, and the impact of burden on decay.

Figure 1.1

Graphic Representation of the Five-Step Model of Lexical Acquisition (Taken from Hatch and Brown, 1995)



Regarding language learning and teaching practice, a clearer understanding of burden and decay may have considerable implications for second language pedagogy. By better understanding lexical decay and the factors that contribute to more or less loss, it may be possible to engineer the learning procedure to offset decay, leading to more sustainable learning (Hayashi, 2011; Weltens & Cohen, 1989; Weltens & Grendel, 1993). In fact, Schmitt (2010) argued that the mitigation of decay “should drive most of pedagogy" (p. 257). Furthermore, a better understanding of learning burden and the relationship between burden and decay may help materials designers develop effective pedagogical tools. For example, by understanding the relationship between burden and decay, materials writers can ensure that items most likely to be forgotten are recycled on a just-in-time basis. This is particularly the case with electronic materials and language learning smartphone applications that can adapt to individual learners. However, such pedagogical benefits are contingent on a better understanding of learning burden and decay; therefore, there is a need to increase our limited understanding of this area.

1.7 Aims of the Thesis

As outlined above, a better understanding of learning burden may impact language learning theory and practice and there are still several variables about which we know little. Furthermore, there has been limited systematic investigation of lexical decay, the factors that affect decay, as well as the relationship between the learning burden of L2 vocabulary items and decay. This is despite the potentially impactful contribution such research could make to an overall theory of lexical development and EFL/ESL pedagogy. The research presented in this thesis was conducted to fill these gaps. Specifically, it aimed to achieve the following:

(1) To determine the factors that impact learning burden.

(2) To determine the factors that impact lexical decay.

(3) To determine the relationship between the learning burden of L2 vocabulary and lexical decay.

In order to address these areas, three experimental studies were conducted in which the role of several intralexical, contextual, and learner-related factors were examined. More specifically, the following factors were investigated:

- PoS (Study 1)

- Word length (Study 1)
- Code of meaning presentation (Study 2)
- Perceived target item usefulness (Study 2)
- Mode of form presentation (Study 3)
- Language learning aptitude (Study 3)

Additionally, all three studies considered the role of learning burden in the decay process. The thesis also compared the data from the three studies to determine how the length of the interval between learning sessions affected the extent of decay that took place.

1.8 Organisation of the Thesis

This thesis consists of seven chapters that are centred around three experimental studies. Chapter 2 considers the constructs and measurement of lexical decay and learning burden and reviews important literature relevant to these areas. As indicated above, the experimental studies presented in this thesis employed electronic flashcard software to present the target items to the learners. The selection and initial piloting of the chosen flashcard platform are explored in Chapter 3. Chapters 4, 5, and 6 outline the three empirical studies at the centre of this thesis. These studies investigated different factors that potentially affect lexical decay. They also examined the effect of the target factors on learning burden and determined the relationship between burden and decay. Study 1 (Chapter 4) considered the effect of two intralexical factors, word length and part of speech, employing a retention interval of four weeks. Study 2 (Chapter 5) investigated the effect of a contextual factor, code of meaning presentation (i.e., L1 or L2 meaning presentation), and a learner-related factor, perceived usefulness of the target items, over a one-week retention interval. Finally, Study 3 (Chapter 6) looked at further contextual and learner-related variables, the mode of form presentation (i.e., written form only or written and spoken form) and language learning aptitude, over a two-week retention interval. All three chapters include a literature review relevant to the specific variables targeted. The final chapter (Chapter 7) synthesises and discusses the main findings from these studies. Additionally, this chapter compares data from the three investigations to explore the effect of an additional contextual variable, the retention interval length, on the extent of decay recorded. Chapter 7 also draws final conclusions, reflects on the impact of the research explored in this thesis, and calls for greater empirical investigation of this area.

Chapter 2: General Literature Review

This chapter considers the literature on learning burden and factors that have been found to increase or alleviate burden. It also discusses the literature on the forgetting of foreign language lexical knowledge and the variables that impact such loss. An additional area of focus is the relationship between learning burden and lexical decay. However, it does not discuss in detail the various foci of the experimental studies presented in this thesis. Comprehensive literature reviews of the particular factors explored in each of the studies are presented in the subsequent chapters. The chapter begins by exploring learning burden and outlines two ways in which it has been measured. After this, it considers the history of research on the loss of accrued knowledge from the fields of education and SLA. It then discusses the literature on vocabulary loss, the various factors affecting this phenomenon, and issues relating to the measurement of the lexical decay construct. The relationship between learning burden and lexical decay is then discussed. Finally, based on the research gaps identified, the research questions this thesis sought to answer are presented.

2.1 *Learning Burden*

New words are learned with different levels of ease or difficulty (Laufer, 1997; Peters, 2020). The extent to which a word is harder or easier to learn is associated with its learning burden, with a heavy learning burden slowing the learning process compared to a light learning burden. Understanding the construct of learning burden is important because the learning burden of a word may affect the probability of it being forgotten (Nation & Webb, 2011). The learning burden of a word is affected by various factors (Nation, 2020), which can be organised into different categories. Several taxonomies of factors have been proposed to date, with two particularly prominent taxonomies being suggested by Higa (1965) and Peters (2020). These are discussed below.

2.1.1 Models of Learning Burden

Higa (1965)

A seminal model of learning burden was proposed by Japanese psychologist Masanori Higa (1965). This highly influential model (Webb & Nation, 2017) suggested that learning burden consists of five areas:

1. The intrinsic difficulty of a word to be learned

   Higa (1965) argued that learning burden is partly derived from the intrinsic difficulty associated with a word form or its meaning. In particular, he discussed two dimensions relevant to the semantic value of lexical items, concreteness and semantic complexity, although research has since shown other factors associated with form and meaning to impact the extent of burden (Laufer, 1997; Peters, 2020).

2. The interaction between previously learned words and a new word to be learned

   Higa (1965) stated that the relationship between novel L2 lexical items and existing L1 and L2 lexical knowledge can also impact learning burden, specifically referring to two variables, meaningfulness and familiarity. He interpreted meaningfulness as relating to association value, phonotactical typicality, and referent familiarity. Familiarity, meanwhile, was associated with L2 frequency.

3. The interaction within a group of words to be learned at the same time

   The third category is related to the relationship between a novel item and the words with which it is presented. Higa (1965) identified three sources of burden in this regard: position, similarity, and volume. Position relates to the sequential position in which a target item is placed. Similarity refers to the semantic relatedness of the target items. Volume relates to the number of target items expected to be learned at the same time.

4. The interaction between groups of words to be learned in sequence

   Higa (1965) also argued that the learning burden of lexis is influenced by the order in which words are presented. He suggested that moving from easier to more difficult items leads to a lighter learning burden.

5. The effect of repeated presentation of words to be learned

   The final component of Higa's (1965) classification related to the frequency with which a target lexical item is presented. He suggested that providing repeated exposure to target lexical items makes them easier to learn.

Thus, Higa (1965) argued that an item’s learning burden stems from numerous sources and is, therefore, a complex amalgam of disparate variables that interact in complicated ways. Importantly however, Higa (1965) recognised that learning burden does not exist in an abstract sense, in that it is not uniformly experienced by all learners. Rather, burden is a situated construct and the extent to which it is felt varies from one learner to the next. That is, although we may derive patterns of burden by looking at groups of participants, consideration at the level of the individual language learner may show that what is easy for one is burdensome for another. This is a crucial distinction because, although it is pedagogically convenient to think of learning burden as a standardised phenomenon (Laufer, 1997; Schmitt, 2010; Webb & Nation, 2017), for research purposes, considering it as a by-item by-learner variable more accurately embodies the burden construct and facilitates more robust conclusions.

Issues with Higa (1965)

While Higa's (1965) model recognised many useful distinctions, it suffers from several limitations. First, in his model, Higa included the frequency with which target items are repeated, arguing that increased exposure frequency leads to a lighter burden. While research has shown exposure frequency to be positively associated with learning vocabulary (Brown et al., 2008; Saragi et al., 1978; Webb, 2007), including repetition in the burden construct is problematic. This is the case because Higa also suggested that repetition can be used as a metric for burden. He asserted that burden is reflected in the frequency or time of exposure necessary for a person to learn an item, with words that require more exposures deemed to pose a heavier burden. Thus, he suggested that increased repetition is simultaneously a mechanism to alleviate learning burden and an indication of a heavy learning burden, which is incongruous.

Second, he implies that the identified factors impact learning burden in the same manner; however, it is likely the case that the numerous sources of burden operate at different levels of influence. Some factors affect burden at the item level, meaning that some words are more burdensome than others, while other factors operate at the task level, influencing the learning burden of multiple items simultaneously. Importantly, burden at the word-level is static while task-induced learning burden is dynamic and can be manipulated by materials writers and teachers. This conceptualisation of burden aligns with Suzuki, Nakata, and DeKeyser's (2019, based on Housen & Simoens, 2016) distinction between linguistic difficulty and condition of practice. The former relates to linguistic features such as formal complexity; the latter to characteristics of the instructional context. Seen in this light, learning burden is a multidimensional construct with item-level burden having a static influence and the extent of task-related burden dependent on several factors associated with the instructional context.

An important notion that was not considered by Higa (1965) is that learning burden can be moderated by individual differences. Examples include individual factors such as motivation, L2 proficiency, and language learning aptitude. For instance, all things being equal, a learner with comparatively low language learning aptitude would require more exposures to learn an item than a learner with higher levels of aptitude (Carroll & Sapon, 1959). Similarly, more motivated learners are likely to acquire content more quickly than less motivated peers (Dornyei, 2005). Peters (2020) considered such factors as part of the learning burden of foreign language vocabulary, but in this thesis learner-related variables are not included in the learning burden construct. Such individual factors do interact with contextual factors (e.g., Robinson [2002] suggests that aptitude profiles should be aligned to instructional conditions to foster effective learning) as well as intra and interlexical factors (e.g., the impact of phonotactical typicality on learning likely varies with L2 proficiency, with more proficient learners experiencing less burden from phonotactically atypical items). Thus, while not included in the burden construct, individual factors can be seen to have a regulatory effect on learning burden, impacting the extent to which learning burden is experienced by a learner.

Peters (2020)

In a later model, Peters (2020) organised the variables that impact the learning burden of a word into word-related, context-related, and learner-related factors.

- Word-related factors are associated with the formal and semantic properties of a word. Cognateness, word length, and concreteness are examples of word-related factors.
- Contextual factors relate to the context in which a word is used/presented. Examples are the frequency with which a word occurs in a specific text and the instructional context in which a target item is presented.
- Learner-related factors are associated with the individual language learner. The inferencing skill of a learner as well as his/her language learning aptitude and vocabulary size are examples of learner-related factors.

Thus, Peters (2020) also suggests that learning burden is influenced by several variables which relate to different aspects of the learning environment and can interact in complex ways.

Issues with Peters (2020)

However, Peters' (2020) taxonomy also suffers from some limitations and requires increased granularity. Most importantly, the category of word-related factors includes both intralexical and interlexical factors (see Laufer, 1990). Intralexical describes variables associated with the formal or semantic properties of a word while interlexical denotes the interaction between the formal and semantic properties of a word and the existing language knowledge of the learner. This knowledge could be relevant to the L1, L2 or any other known language. Word length and concreteness are examples of intralexical factors, while cognateness is an example of an interlexical factor. Importantly, intra and interlexical factors may impact learning in different ways. Intralexical factors are likely to be relatively consistent across a multilingual learner group, while interlexical factors necessarily depend on the individual learner. Word length and part of speech do not vary by learner L1, but cognateness does, for example. This difference is not reflected in Peters' (2020) model.

2.1.2 The Learning Burden Construct Adopted in this Thesis

Analysing the models of Higa (1965) and Peters (2020), several axioms can be derived from which it is possible to construct an improved model of learning burden. The first is that there are different factors that influence the learning burden of L2 lexical items. These factors relate to the target items (i.e., intralexical and interlexical factors) and the context in which they are learned (i.e., contextual factors). Such variables operate at different levels and interact in complex ways. Therefore, a model of learning burden needs to account for the different factors and the nature of their influence on the learning process. Additionally, although learning burden is often conceptualised as existing in an absolute sense (Laufer, 1997), this is not the case; learning burden is not external to the learner. Rather, as Higa (1965) argued, it is situated in the individual, meaning that the extent of learning burden experienced varies from one learner to another. Thus, it should be considered a by-learner by-item construct. Moreover, as it is a situated phenomenon, individual differences (e.g., aptitude, motivation, proficiency) influence the amount of burden experienced, although these variables are not included in the burden construct.

A comparison of Higa's (1965) and Peters' (2020) frameworks is presented in Table 2.1, along with the terminology employed in this thesis. As can be seen, this thesis considers learning burden as a by-item by-learner multidimensional construct consisting of intralexical, interlexical, and contextual factors, and moderated by learner-related variables. These four categories are defined as follows:

- intralexical factors: these are internal to a lexical item (Laufer-Dvorkin, 1991) and relate to the form (e.g., orthography, pronunciation), meaning (e.g., concreteness, imageability) or use (e.g., L2 frequency) of a word (Laufer-Dvorkin, 1991).
- interlexical factors: these relate to the interaction between a learner and a lexical item. Examples include cognateness, L1 frequency, and the relationship between known L2 items and the novel vocabulary to be learned (Laufer-Dvorkin, 1991).
- contextual factors: these relate to the environment in which a lexical item is learned. Examples here include the instructional method, the relatedness of the target items, and the manner in which word meaning is presented.
- learner-related factors: these relate to the language learner and include predictor variables such as age, L2 proficiency, and length of experience with the L2 (Olshtain, 1989), in addition to affective (e.g., anxiety, motivation), cognitive (e.g., language learning aptitude), and strategic (e.g., preferred learning strategies) factors.

Table 2.1

The Different Conceptualisations of Learning Burden

Classification of Higa (1965)	Classification of Peters (2020)	Terminology used in this thesis	Type of learning burden
1. The intrinsic difficulty of a word to be learned	Word-related factors	Intralexical factors	Item-level learning burden
2. The interaction between previously learned words and a new word to be learned	Word-related factors	Interlexical factors	Item-participant level learning burden
3. The interaction within a group of words to be learned at the same time	Contextual factors	Contextual factors	Task-induced learning burden
4. The interaction between groups of words to be learned in sequence	Contextual factors	Contextual factors	Task-induced learning burden
5. The effect of repeated presentation of words to be learned	Contextual factors	Contextual factors	Task-induced learning burden
	Learner-related factors	Learner-related factors	Not part of the burden construct, but can influence learning burden

2.1.3 Empirical Research on Learning Burden

The three experimental chapters of this thesis contain targeted literature reviews pertinent to the specific variables under investigation in each chapter. The current section provides a summary of the main findings from studies looking at learning burden. Overall, research shows that each of the four categories outlined above can affect learning burden.

For instance, studies have found that intralexical factors such as imageability, bigram frequency, orthographical neighbourhood size, and structural complexity influence the learning process (see Laufer, 1997; Schmitt, 2010). Previous studies have also 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). Findings are less clear with regard to the effect of PoS and word length (Peters, 2020), with some studies reporting a learning advantage for nouns (Ellis & Beaton, 1993; Horst & Meara, 1999; van Zeeland & Schmitt, 2013b) and shorter words (Ellis & Beaton, 1993; Barcroft & Rott, 2010; Willis & Ohashi, 2012), while other studies have reported conflicting results for PoS (Pigada & Schmitt, 2006) and length (Puimége & Peters, 2019). Furthermore, studies may have conflated different factors; for instance, Peters (2020) suggested that PoS has been conflated with concreteness and word length with L2 frequency in studies to date. Taken together, the inconsistency of research findings and presence of potential confounds means that we currently do not have a clear picture of how PoS and length impact burden (Peters, 2020). In light of this, Study 1 of this thesis focused on the effect of PoS and word length on the learning burden of L2 vocabulary (a detailed review of the literature on PoS and word length is provided in Chapter 4).

Research has also demonstrated that interlexical factors affect burden, with effects shown for variables such as orthographic wordlikeness (Bartolotti & Marian, 2017), phonotactical typicality (Ellis & Beaton, 1993), L1 frequency (de Groot, 2006; de Groot & Keijzer, 2000), and cognateness (de Bot & Stoessel, 2000; de Groot & Keijzer, 2000).

Studies have also considered contextual factors such as the task (Barcroft, 2002; Lado et al., 1967; Laufer & Shmueli, 1997), the informativeness of the context in which an item is presented (Webb, 2008), and the relatedness of target words (Ishii, 2015; Tinkham, 1993; Tinkham, 1997; Waring, 1997). These studies have shown that such factors can influence both the efficacy (i.e., the extent of learning) and efficiency (i.e., the speed of learning) of vocabulary-focused instruction. However, there are also several contextual factors that we currently know little about. For instance, minimal consideration has been given to the manner in which word meaning is conveyed to learners (i.e., use of L1 equivalents or L2 definitions) and the mode of form presentation (i.e., written form only or co-presentation of the spoken and written form). Studies that have investigated these variables (e.g., Lado, Lobo, & Baldwin, 1967, Laufer & Shmueli, 1997, Mishima, 1967) have generally found that using L1 equivalents and co-presentation of the written and spoken forms leads to more learning; however, no research has looked at the effect of these variables on the amount of burden posed by L2 lexical items. Study 2 of this thesis considered the effect of meaning presentation code and Study 3 looked at the impact of form presentation mode (detailed reviews of the literature on meaning presentation code and form presentation mode are given in Chapters 5 and 6 respectively).

Finally, research has found that learner-related variables such as age of onset (Granena & Long, 2012), motivation (Gardner & MacIntyre, 1991; Tseng & Schmitt, 2008), and aptitude (Dahlen & Caldwell-Harris, 2013; Granena & Long, 2012) impact vocabulary learning. However, in comparison to other aspects of language knowledge such as pronunciation (see Saito, 2016), few studies have considered the role of language learning aptitude in vocabulary acquisition. Furthermore, studies that have looked at this factor, have generally considered its impact on attainment rather than the burden posed by lexical items.

Moreover, there are other learner-related factors, such as the perceived usefulness of a lexical item, that may moderate the amount of learning burden experienced. Research to date has considered usefulness to be a function of the learning task (Hulstijn & Laufer, 2001) or to be consistent across a learner cohort (Webb & Peters, 2018), yet the extent to which a word is perceived to be useful is likely to vary with the individual learner. Study 2 looked at the effect of perceived usefulness on burden, conceptualising it as a by-item by-learner factor. Study 3 considered the effect of language learning aptitude on burden (literature reviews relevant to these variables are provided in Chapters 5 and 6 respectively).

A summary of these findings is presented in Table 2.2. As can be seen, while we know quite a lot about the effect of several factors on the learning burden of L2 vocabulary, there are also numerous variables that have either received minimal research attention or which have produced inconclusive findings. For instance, with regard to intralexical factors, the effect of word length and part of speech is currently unclear (Peters, 2020), justifying further research on these two variables. The effect of several contextual variables also requires delineation. For instance, factors associated with the manner in which the meaning and the form are presented to learners have received limited research consideration. Furthermore, some studies that have been conducted on the effect of these variables have employed a research methodology that may have biased the findings (see Section 2.1.4). Additionally, almost no research has considered the moderating effect of learner-related variables on the effect of intralexical, interlexical, or contextual factors. The three experimental studies presented in this thesis sought to meet these lacunae.

Table 2.2

Factors Affecting the Learning Burden of L2 Vocabulary (Partially Adapted from Laufer [1997] and Peters [2020])

Factor type	Facilitating factors	Difficulty-inducing factors	Inconclusive
Intralexical	• concreteness • more imageable • less structurally complex • congruent sound-script relationship • inflexional regularity • derivational regularity • morphological regularity • one meaning, one form	• abstractness • less imageable • more structurally complex • incongruent sound-script relationship • inflexional complexity • derivational complexity • deceptive morphological transparency • polysemy	• word length • PoS
Interlexical	• cognateness • orthographic wordlikeness • phonotactic typicality • high L1 frequency	• non cognateness • orthographic non-wordlikeness • phonotactic atypicality • low L1 frequency	
Contextual	• high frequency of occurrence • informative context • presentation with semantically dissimilar items	• low frequency of occurrence • uninformative context • presentation of items in semantic sets	• mode of form presentation • meaning presentation code

2.1.4 Measuring Learning Burden

Two Approaches to the Measurement of Learning Burden

Research has adopted two approaches to the measurement of learning burden: consideration of learning gains and trials to criterion. The measurement of learning gains typically uses a longitudinal design involving a pre-test, treatment, and post-tests (Webb & Nation, 2017). Gains are determined by comparing scores on the pre-test and post-test, and such designs often control for confounding variables such as time on task by fixing the quantity of exposure participants have with the target lexical items. In this way, gains between the two tests can be attributed to the treatment. This is a measurement of learning burden because items that can be acquired within a fixed number of exposures are thought to pose a lighter learning burden than items that cannot be learned given the same amount/quality of input. For example, Ishii (2015) employed this methodology to determine which of three types of word sets (semantically unrelated, semantically related, and visually related) bore the heaviest learning burden. Participants were presented with sets of six L1-pseudoword word pairs and were tasked with memorizing the target forms within 45 seconds. A test of meaning recall was administered immediately after the treatment and again twenty minutes later. The results showed that less learning occurred in sets that were visually related than sets that were either semantically related or unrelated. Ishii therefore concluded that the co-presentation of visually-related items increased the learning burden. This study illustrates that learning gains have been used to measure learning burden. Other studies employing this metric include Hashemi and Gowdasiaei (2005) and Hoshino (2010). This operationalisation of burden is referred to in this thesis as the gains-based criterion.

Although not as commonly employed, learning burden can also be measured by the number of exposures necessary for a word to be acquired (Higa, 1965). The principle is that items requiring more exposures to be learned possess a heavier learning burden than items that require fewer exposures. An example of this approach is Tinkham (1993), who like Ishii (2015), investigated the relative difficulty of semantically related and unrelated sets. Learners were presented with oral L1 equivalents and were required to produce target nonwords. After each trial (i.e., exposure), learners heard the L1-nonword word pair. Tinkham measured the number of repetitions needed for the nonword form to be produced correctly. He found that semantically related words required more trials to be successfully produced than semantically unrelated words and thus concluded that organising lexical items by semantic category increased the learning burden of foreign language lexis. This approach to the measurement of learning burden is referred to in this thesis as the frequency-based criterion.

The manner in which learning burden is operationalised typically depends on the research context and the research instruments employed. Studies that adopt a frequency-based criterion require a means of measuring the learning process of each individual learner. For example, Tinkham (1993) met and recorded the spoken output of each learner in order to calculate the number of trials needed to reach criterion. Such instruments can be difficult to use if researchers wish to retain the environmental validity of pedagogical studies and, particularly, if an investigation is conducted in a classroom setting. However, the choice of metric should not entirely be governed by the research environment because the manner in which learning burden is measured can have a considerable impact on research findings. As shown above, Ishii (2015), using a gains-based criterion, found semantic relatedness not to affect learning burden, while Tinkham (1993), using a frequency-based criterion, found that it did. In fact, the research on semantic relatedness is largely split along these methodological lines, with studies employing a frequency-based approach finding that relatedness leads to more burden (Tinkham, 1993; Tinkham, 1997; Waring, 1997), and studies using a gains-based metric reporting that semantic relatedness does not affect burden or actually results in a lighter learning burden (Hashemi & Gowdasiaei, 2005;

Hoshino, 2010). This suggests that the conceptualisation of learning burden may have considerable bearing on the results that are obtained.

Advantages of a Frequency-Based Approach

Of the two approaches, this thesis maintains that a frequency-based measure is preferable. This is because the outcome variable of a gains-based approach is inherently binomial, with an item either learned or not. Such an approach, therefore, assumes the burden of all learned items to be equal, which is clearly not the case. Typically, studies circumnavigate this problem by either calculating mean gains per word across a sample of learners (e.g., Laufer & Shmueli, 1997) or by summing the gains of a set of target items according to a grouping variable (e.g., Ishii, 2015). This approach, therefore, considers the burden posed by an item to be constant across a group of learners and/or holds that all items in a set of words are thought to pose a similar level of difficulty. Crucially however, as explained in Section 2.1.2, learning burden is a by-item by-learner variable, meaning that a gains-based approach does not align with the learning burden construct adopted in this thesis.

Furthermore, and of particular importance to this thesis, the manner in which learning burden is operationalised may also have considerable implications for the amount of decay that occurs. To understand these implications, it is first necessary to consider the learning process during the gains-based criterion in more detail. In this tradition, there is an assumption that controlling the amount of exposure to target items results in all items being processed in a similar fashion; however, this is unlikely to be the case. A key distinction can be made here between exposures during the process of word encoding and exposures that result in the retrieval of previously encoded knowledge. It is important to recognise that controlling the frequency of exposure does not necessarily reflect the speed with which learners actually encode items. This means that even if research prescribes a fixed frequency/length of exposure, learners will encode some items before others and thus items will differ in the number of retrieval opportunities they have. This distinction between exposures during the encoding phase and retrievals is critical for studies that consider decay. As retrieval facilitates retention to a greater extent than presentation (Baddeley, 1990; Karpicke & Roediger III, 2008; Royer, 1973), words that are retrieved more often during a treatment are more likely to be retained. This means that, if research prescribes a fixed number of exposures for all items, the words that are encoded first will have more opportunities to be retrieved. In turn, by virtue of this increased retrieval, these items are more likely to be retained. Therefore, studies adopting a gains-based criterion risk biasing the retention data by controlling the frequency of exposure, rather than the frequency of retrieval. Ideally, studies should employ a dynamic approach to repetition in which lexical items are encountered as often as needed for encoding to take place, but once encoded, the frequency of retrieval is controlled. Critically, such a design is only possible if learning burden is operationalised using the frequency-based criterion and a methodology is used that distinguishes exposures prior to learning from retrievals. The research presented in this thesis used such a methodology.

The need to control the frequency of retrieval becomes more important if the interaction of frequency of exposure and individual variables is considered. Learner-related factors like language learning aptitude can impact the speed with which lexical items are encoded (Carroll & Sapon, 1959). This means that learners with high aptitude require fewer exposures to encode target items than learners with low levels of aptitude. Importantly, in gains-based methodologies, exposure is held constant across items and learners, potentially leading to learners with comparatively high aptitude having more opportunities for retrieval, which would increase the probability of retention. In a study concerned with lexical decay, it is therefore important that the input can vary both by item and by learner as this will mitigate the confounding effect of item-level and participant-level variation. The studies described in this thesis employed such a method and thus are likely to provide a clearer picture of learning burden and its impact on decay than studies which have prescribed a fixed number/length of exposure for all items and learners.

2.2 *Lexical Decay*

This literature review now moves on to consider research on language and lexical loss, the various factors that have been found to impact the forgetting of L2 vocabulary, and the manner in which loss has been measured. As explained in Section 1.4, terminology is an issue in the area of language loss, with several terms used interchangeably to refer to diverse and discrete types of loss. The term decay is used in this thesis to refer to the loss of linguistic knowledge in contexts of continued exposure to the target language. Attrition is used for language loss in contexts of reduced exposure to the target language. The terms loss and forgetting are used as umbrella terms encompassing contexts of language decay and language attrition. Due to the paucity of research on language decay, the following review covers studies of language loss, including both contexts of no exposure (attrition) and contexts of continued exposure (decay).

2.2.1 Background

In the field of education, there is a long tradition of research on the forgetting of accrued knowledge. One of the earliest formal investigations in this area was conducted by White (1908), who considered the effect of a summer vacation on two aspects of mathematical ability: accuracy and speed of computation. He found there to be little change in terms of accuracy but a general decline in computational speed over a summer interval. In the years since this early investigation, other studies have similarly considered the forgetting/maintenance of knowledge over summer vacations. Investigations have looked at mathematics (Paechter et al., 2015; Saelinger, 1928), L1 reading (Paechter et al., 2015; Saelinger, 1928), and L1 vocabulary knowledge (Botwin, 1965; cited in Cooper et al., 1996). Reflecting the findings of White (1908), these studies have generally shown that some forgetting occurs over a summer vacation, but that knowledge in some areas is more likely to be forgotten than in others. This topic still engenders research interest; for instance, a meta-analysis of summer vacation studies found that loss of computation and spelling proficiency was greater than reading proficiency (Cooper et al., 1996). Thus, over the last hundred years, there has been continued investigation within the field of education into the loss of accrued knowledge.

By comparison, the investigation of language loss in the field of applied linguistics has more modern origins. Although interest in language loss dates back to the sixteenth century (Berko-Gleasen, 1982) and psychologists have considered the forgetting curves of language knowledge for over a hundred years (Ebbinghaus, 1885), it was not until the 1980s that patterns of L2 loss were first studied by modern linguists (Lambert & Freed, 1982). Prior to this period, studies of language loss largely focused on the intergenerational loss of minority languages and the loss resulting from neurological impairment (Park, 2018), while studies of non-pathological intragenerational language loss were somewhat sporadic, consisting of exploratory case studies involving a few learners (e.g., Cohen, 1975) or idiosyncratic contexts (e.g., Cohen, 1974). The current conceptualisation of the field was established in 1980, when the first conference on language loss was hosted at the University of Pennsylvania (Weltens & Cohen, 1989). Motivated by calls for systematic investigations of intragenerational language loss at the conference, scientific studies on this theme steadily increased in the years following (Hansen & Reetz-Kurashige, 1999). The study of language loss is now an important subfield of applied linguistics, as evidenced by the numerous publications dedicated to this phenomenon. For example, the subfield has produced several proceedings from conferences (Lambert & Freed, 1982; Weltens, de Bot, & van Els, 1986), monographs (Schmid, 2011), edited volumes (Hansen, 1999; Kopke, Schmid, Keijzer, & Dostert, 2007), as well as special issues in *ITL-Review of Applied Linguistics* (de Bot, Clyne, & van Els, 1989), *Studies in Second Language Acquisition* (Cohen & Weltens, 1989), and the *International Journal of Bilingualism* (de Bot, 2004). Importantly however, the majority of this research has considered contexts of reduced or non-exposure to the target language (i.e., language attrition settings). In fact, to date there has been minimal investigation of decay settings (Schmitt, 2010). Examples of decay settings include classrooms in which learners study but do not later recycle, specific L2 lexical items despite continued exposure to the L2 more generally, students who learn specific items for one learning activity but do not encounter those items in subsequent activities, and students who move from one class to another with limited continuity of teaching material.

2.2.2 Investigations of L2 Loss

Since the inception of this subfield of applied linguistics, numerous contexts of language loss have been considered. For instance, some studies have explored the loss of L1 as a result of moving to an L2 environment (Schmid, 2012), while others have looked at the loss of L2 as a result of reduced exposure to the target language (Hansen & Chen, 2001). With regard to L2 loss situations, research has mainly focused on three populations: school and university students residing in the L1 environment who cease contact with the L2, returning missionaries, and children repatriating after living in an L2 environment (Bardovi-Harlig & Stringer, 2010). Crucially, these populations involve L2 loss in L1 environments and consider the partial or complete reduction of target language contact. That is, they look at contexts of language attrition.

In addition to multiple loss settings, research to date has considered the loss of several aspects of linguistic knowledge and competence. Studies have looked at morphosyntax (Bahrick, 1984; Berman & Olshtain, 1983; Fukazawa, 1984; Moorcroft & Gardner, 1987; Olshtain, 1989; Scherer, 1957; Smythe, Jutras, Bramwell, & Gardner, 1973; Tomiyama, 1999), oral fluency (Cohen, 1974; Moorcroft & Gardner, 1987; Taura, 2001), listening comprehension (Fukazawa, 1984; Gardner, Lalonde, Moorcroft, & Evers, 1987; Gardner & Lysynchuk, 1990; Murtagh & Slik, 2004; Smythe, Jutras, Bramwell, & Gardner, 1973; Snow, Padilla, & Campbell, 1988), reading comprehension (Bahrick, 1984; Fukazawa, 1984; Gardner & Lysynchuk, 1990; Smythe, Jutras, Bramwell, & Gardner, 1973; Snow, Padilla, & Campbell, 1988), pronunciation (Cole, 1929; Smythe, Jutras, Bramwell, & Gardner, 1973; Tomiyama, 1999), and vocabulary. Such studies have shown that aspects of language knowledge are forgotten at differing speeds (Smythe et al., 1973). Results in this area are somewhat mixed due to differences in how studies have fostered and measured language knowledge (Schmitt, 2010), meaning that there is considerable variance in the extent of loss reported in individual studies. However, on the whole, results suggest that vocabulary knowledge seems to be more prone to forgetting than other language systems such as grammar and phonetics (Bardovi-Harlig & Stringer, 2010; Schmitt, 2000). Studies have shown, for example, that while some forgetting can occur with aspects of language such as morphosyntax (Tomiyama, 1999), knowledge of such aspects is generally better retained than lexical knowledge. In fact, the loss of lexical knowledge appears to be faster and more dramatic than other aspects of language knowledge (Bahrick, 1984; Tomiyama, 1999). This is likely the case because lexical knowledge tends to be less rule-based than other language systems and there are generally more words in a foreign language than there are phonemes and grammatical patterns, meaning there are more elements that could be forgotten (Bardovi-Harlig & Stringer, 2010; Park, 2018; Schmitt, 2010).

However, it would be overly simplistic to suggest that lexical knowledge is universally vulnerable while knowledge of grammar and pronunciation is stoically resolute (Bardovi-Harlig & Stringer, 2013). This is partly the case because lexical items seem to differ in their susceptibility to loss (Berman & Olshtain, 1983), meaning that foreign language words are not forgotten in a uniform manner. Additionally, receptive vocabulary knowledge is generally found to be more robust than productive vocabulary knowledge (Bahrick, 1984) meaning that forgetting patterns might be moderated, at least in part, by the type of knowledge targeted. Also, some lexical items such as idioms and multi-word units appear to be more resilient than single-word items (Bahrick, 1984; Berman & Olshtain, 1983; cf. Bell, 2009). Thus, while vocabulary knowledge does appear to be more susceptible to loss than other aspects of language knowledge, it is not the case that all lexical items and aspects of word knowledge are forgotten in the same manner. These themes are further developed in the next section.

2.2.3 The Loss of L2 Lexical Knowledge

The three experimental chapters contain a detailed literature review pertinent to the specific variables investigated in each chapter, while the current section summarises the main findings of this area. Of the different language systems considered in studies of language loss to date, the most frequently investigated has been vocabulary knowledge (Park, 2018). Many studies have explicitly targeted the forgetting of lexical knowledge, finding that some loss occurs over a period of non-use (Bahrick, 1984). This relative vulnerability is also evident in longitudinal vocabulary learning studies which have shown that knowledge demonstrated on an immediate posttest can be forgotten before a delayed posttest (Laufer & Shmueli, 1997). However, as was argued in the previous section, vocabulary knowledge does not seem to degrade in a uniform manner, with some lexical items, learners, and learning activities associated with more loss, while others appear to be comparatively resilient. Such differential patterns point to the influence of several variables on the process of lexical loss. This section discusses some of these key variables.

Section 2.1 showed that some research consideration has been given to the effect of intralexical, interlexical, contextual, and individual variables on the learning burden of L2 lexis. In comparison, the relative effect of such factors on the loss of lexical knowledge has received little attention. This is despite the fact that vocabulary loss is a highly complex phenomenon that is likely impacted by numerous factors. Research has shown that there is considerable variation in the extent of loss that occurs within and across different cohorts (Bahrick, 1984). Some studies have attempted to explain this asymmetrical loss by exploring the effect of certain variables (Bardovi-Harlig & Stringer, 2010). The idea is that by better understanding the comparative vulnerability of linguistic knowledge and by isolating the variables that contribute to more or less loss, pedagogic strategies can be developed to mitigate the forgetting of language content. However, as will be shown in this section, more research is needed to delineate the role played by some key factors in the process of lexical forgetting.

Canonically, research on language loss draws a distinction between predictor and criterion variables (Lambert & Freed, 1982), or extralinguistic and linguistic variables as they have been recently reimagined (Bardovi-Harlig & Stringer, 2010). The former refers to “the characteristics of the individual and the situation that influence…[the] degree of overall loss” (Lambert & Moore, 1986; p. 184). Key variables here include sociopsychological factors such as attitude to the target language and motivation for learning/retention (Weltens & Cohen, 1989), age, and instructional method (Bardovi-Harlig & Stringer, 2010). Linguistic or language-related variables are those that relate to lexical and morphosyntactic properties of the L1 and L2 (Bardovi-Harlig & Stringer, 2010), as well as characteristics of the language content that is forgotten (Lambert & Moore, 1986). While recognising this convention, in order to be consistent with the discussion of learning burden, this thesis employs the terms intralexical, interlexical, contextual, and individual to refer to the different categories of variables. Predictor/extralinguistic factors are associated with contextual and learner-related factors, while criterion/linguistic factors relate to interlexical and intralexical variables.

The Effect of Learner-Related Factors on the Loss of Lexical Knowledge

Research shows that learner-related factors can impact the process of lexical loss. For instance, studies have found that proficiency is associated with the amount of vocabulary loss that occurs, with more proficient learners suffering less loss (Bahrick, 1984; Hansen, 1999; Mehotcheva, 2010; Murtagh & Slik, 2004; Weltens, 1988; Xu, 2010; cf. Smythe, Jutras, Bramwell, & Gardner, 1973; Morshedian, 2008). Age also appears to be an important factor, with young learners particularly susceptible to forgetting L2 lexical knowledge (Berman & Olshtain, 1983; Cohen, 1974; Howe & Ceci, 1978; Olshtain, 1989;

Tomiyama, 1999). L2 literacy has also been found to impede vocabulary loss (Hansen & Chantrill, 1999; Olshtain, 1989). Finally, the motivation of the learner during both the period of acquisition (Gardner, Lalonde, Moorcroft, & Evers, 1987; Hansen, Umeda, & McKinney, 2002) and the period of reduced input plays an important role in the process of forgetting L2 lexis (Gardner & Lysynchuk, 1990; Russell, 1999). Higher levels of motivation to learn and retain L2 vocabulary knowledge are associated with less lexical attrition (Gardner, Lalonde, Moorcroft, & Evers, 1987; Gardner & Lysynchuk, 1990; Hansen, Umeda, & McKinney, 2002; Russell, 1999; cf. Hansen, 2012).

However, limited research consideration has been given to the effect of other learner-related factors on lexical decay. For instance, although one study has considered the moderating effect of language learning aptitude on the attrition of L1 morphosyntactic knowledge (Bylund, Abrahamsson, & Hyltenstam, 2009), to the best of my knowledge, no study to date has considered the role of aptitude in the decay of L2 lexical knowledge. The role of the perceived usefulness of a lexical item in the decay process has similarly received little research consideration. Like other learner-related factors presented in the previous paragraph, both factors may impact the rate of decay, justifying further research consideration. The effect of perceived usefulness on decay is investigated in Chapter 5 and the impact of aptitude on decay is the focus of Chapter 6. Detailed literature reviews of these variables are provided in the respective chapters.

The Effect of Contextual Factors on the Loss of Lexical Knowledge

Additionally, contextual factors have been found to impact patterns of loss. For instance, the manner in which the target language is learned has been shown to influence forgetting patterns. Research has found that more loss occurs when items are learned incidentally (Pellicer-Sánchez, 2016, 2017; Waring & Takaki, 2003) than under an intentional learning paradigm (Kramer et al., 2019). Vocabulary learning studies have also shown that instructional activities differ in the amount of forgetting that occurs between immediate and delayed posttests (Laufer & Shmueli, 1997). Additionally, the use of the keyword technique (an activity that uses an image linking the meaning of the L2 item to an L1 item with a similar spoken form; Webb & Nation, 2017) has been found to impede loss (Ellis & Beaton, 1993). Another contextual factor particularly associated with intentional learning methods such as flashcards relates to retrieval sequencing. Retrieval opportunities can be massed (i.e., occur together) or spaced (i.e., separated by an interval). Studies have found that spaced retrieval fosters more retention than massed retrieval (Ebbinghaus, 1885; Nakata, 2015; Pimsleur, 1967). Additionally, the length of interval between retrieval opportunities has also been found to impact attrition patterns. For instance, Bahrick, Bahrick, Bahrick, and Bahrick (1993) found that increased spacing between retrievals led to greater learning burden initially, but ultimately facilitated more robust knowledge.

However, there remain several factors that have received limited research consideration. Two examples here are the code used to present the meaning of a word and the manner in which a form is presented. Some research has indirectly looked at the former variable (Ellis & Beaton 1993; Laufer & Shmueli, 1997), reporting that using L1 equivalents positively impacted retention, but characteristics of the designs of these studies mean that they may have artificially strengthened knowledge of the items learned earliest. Thus, further research is needed to delineate the effect of meaning presentation code on the decay of L2 lexical knowledge. Furthermore, although some studies have looked at the effect of form presentation mode on learning, those studies did not conduct delayed tests which would have allowed for an analysis of decay. Thus, we currently have a limited understanding of the effect of form presentation mode on L2 lexical decay. Studies 2 and 3 investigated these variables, and comprehensive literature reviews relevant to these factors are given in Chapters 5 and 6 respectively.

The Effect of Interlexical Factors on the Loss of Lexical Knowledge

Studies have also shown interlexical variables to impact language loss. For instance, research has demonstrated that the formal proximity of L2 items to their L1 equivalents affects the likelihood of loss. Hansen (2011) found that language distance was a significant predictor of attrition, with higher levels of loss associated with greater distance between the L1 and L2. Similarly, Xu (2010) found that Chinese L1 learners of English were more likely to suffer attrition of accrued lexis than Dutch L1 learners. A finding Xu explained by the different linguistic distance between the two L1s and the common L2. Phonotactical typicality has also been found to impact the rate of loss, with knowledge of phonotactically typical items more durable than atypical items (de Groot, 2006). Additionally, de Bot and Stoessel (2000) found that cognateness affected the attrition process. They report that cognates were better recalled after a 30-year interval while items with formally unrelated L1 equivalents were less well remembered. This cognate advantage was also found by Weltens (1988) and de Groot and Keijzer (2000). Furthermore, frequency has been found to impact patterns of loss. Mehotcheva (2010) found that study abroad students of L2 Spanish better retained high-frequency L2 items than low-frequency items. Hansen (2011) also reported that the frequency of the L1 equivalent of L2 target items significantly predicted loss, with more frequent L1 equivalents less likely to be forgotten. De Groot and Keijzer (2000) and de Groot (2006) also found an advantage for items with more frequent L1 equivalents.

The Effect of Intralexical Factors on the Loss of Lexical Knowledge

Finally, the role of several intralexical variables on the process of forgetting have also been considered. In particular, studies have considered concreteness, imageability, part of speech, generality of meaning, and word length. Regarding concreteness, research suggests that there is a difference between the loss of concrete and abstract nouns (Marefat & Rouhshad, 2007). Studies also suggest that more imageable target items are better retained (de Groot & Keijzer, 2000). Part of speech has also been found to affect attrition, with nouns found to be more robust than either verbs or adjectives (Bagherabadi, 2005, cited in Marefat & Rouhshad, 2007). With regard to generality of meaning, research findings are conflicting. Some studies have shown an advantage for general nouns over specific nouns (Olshtain & Barzilay, 1991), other studies have found that technical nouns are less likely to suffer decay than general nouns (Abbasian & Khajavi, 2010), while still other studies report no difference (Jahangard, 2007). Thus, the comparative resilience of general and technical nouns is currently unclear. Studies have also considered word length, finding that shorter words are more likely to be retained than longer words (Hansen, Kim, Lee, & Lo, 2010, cited in Hansen, 2012). Although some research has considered the effect of PoS and length on language loss, these studies have been conducted in contexts of language attrition. It is not currently known whether findings from the attrition literature will generalise to decay settings. Therefore, the extent to which PoS and length impact lexical decay is an open question. Study 1 of this thesis sought to provide an answer by investigating the role of PoS and length in the decay of accrued L2 lexical knowledge (see Chapter 4 for a detailed review of the literature on PoS and word length).

Other Factors

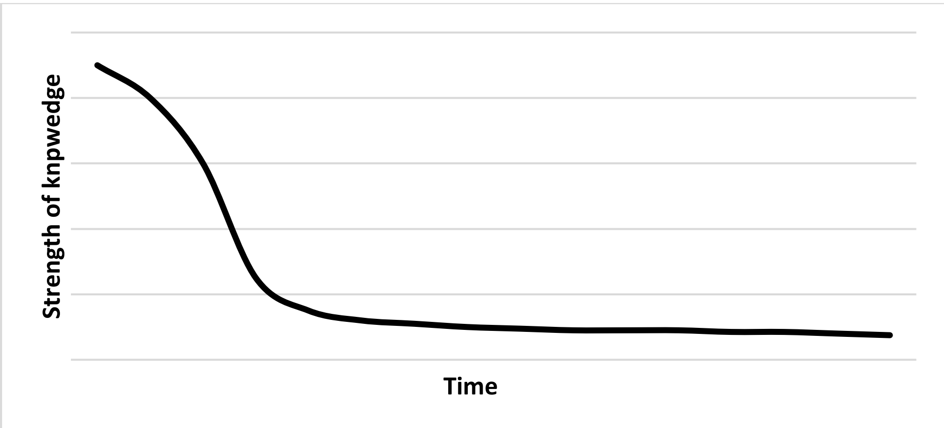
There are two other notable factors that impact the extent of decay observed on empirical studies: the aspect of word knowledge measured and the length of the retention interval. As outlined in Section 1.2, vocabulary knowledge is a multidimensional construct consisting of several aspects of word knowledge (e.g., form, meaning, and use), and receptive and productive mastery. Research shows that receptive vocabulary knowledge can be largely unaffected by a retention interval (Scherer, 1957), with some studies even reporting gains over the period between a learner's exposure to lexical items and the administration of tests measuring retention of those items (Fukazawa, 1984; Kramer, Matsuo, McLean, & Cornwell, 2019; Morshedian, 2008). Such findings are particularly common in studies that have investigated comparatively brief retention intervals. In contrast, productive vocabulary knowledge seems to be more vulnerable to loss over periods of non-use (cf. Marefat & Rouhshad, 2007). This finding has been shown for relatively short (Cohen, 1974; Morshedian, 2008; Waring and Takaki, 2003) and longer (Berman & Olshtain, 1983; Cohen, 1989; Kuhberg, 1992; Tomiyama, 1999) retention intervals. Thus, it seems that the aspect of lexical knowledge targeted is a key variable in studies of language loss (Bardovi-Harlig & Stringer, 2010), with productive vocabulary knowledge suffering more loss than receptive knowledge.

The length of the retention interval may also impact the extent of decay observed. It seems logical to expect language knowledge to be forgotten as a function of time and some studies of language attrition have shown that the longer the retention interval, the more loss occurs (Smythe et al., 1973). Vocabulary learning studies that have used multiple delayed posttests have also generally found more loss on the later delayed tests than the earlier ones, suggesting that the extent of loss increases with the progression of the retention interval. An example of this is Waring and Takaki (2003), who reported that fewer target items were known three months into a retention interval than after a one-week period of non-use. Importantly, however, it seems that the relationship between time and language loss is nonlinear. For instance, Waring and Takaki (2003) found that more loss occurred during the one-week interval between the immediate and first delayed post-test than during the three-month interval between the two delayed post-tests. This suggests that forgetting occurred rapidly and then slowed as the retention interval progressed. Such a finding is broadly in line with the classic forgetting curves reported by Ebbinghaus (1885), who investigated his own learning and subsequent retention of nonsense syllables, finding that many target items were forgotten quickly, while a few were retained for a longer period (see Figure 2.1). This finding has been replicated in the language attrition literature (Bahrick, 1984; Weltens, Els, & Schils, 1989), with extensive loss occurring relatively quickly and knowledge retained beyond this initial period remaining relatively intact for an extended period. This crystallised knowledge is referred to in the attrition literature as being retained in permastore, in reference to the permafrost which can preserve the appearance of objects for a remarkable period of time. For instance, studies have found evidence of lexical items retained in permastore three decades after initial learning (Bahrick, 1984; Hansen & Chen, 2001).

Thus, research suggests that decay sets in soon after attainment, with relatively severe initial loss followed by a period of more sustained retention. However, vocabulary learning studies have not explicitly considered the role of the retention interval length on the extent of knowledge demonstrated on delayed posttests. In fact, the retention interval used in studies typically depends on logistical constraints rather than issues of effective research design. Furthermore, very few studies have compared decay across different intervals of retention (cf. Waring & Takaki, 2003), so we currently know very little about how knowledge decays as a function of time. One goal of this thesis was to explore this area, comparing decay across the three experimental studies, which employed different lengths of retention interval (a review of the literature on the effect of retention interval length on decay is given in Chapter 7).

Figure 2.1

An Example of a Forgetting Curve Based on Ebbinghaus (1885)



2.2.4 Methodological Issues with Studies of Loss to Date

The previous section showed that there are numerous variables associated with vocabulary loss. However, much of our understanding of the effect of different variables on loss is derived from studies of language attrition. Such studies consider language loss in contexts of heavily reduced or no exposure to the target language. In fact, very few studies reviewed in the previous section examined lexical loss in contexts of continued target language exposure (i.e., decay). As reviewed in Section 2.2.3, studies that have considered such loss (e.g., Beaton & Ellis, 1993; de Groot, 2006; de Groot & Keijzer, 2000) have identified some variables that impact the decay process (PoS, code of form presentation, instructional method [Ellis & Beaton, 1993]; cognateness, concreteness [de Groot & Keijzer, 2000]; and phonotactical typicality, L1 frequency, and concreteness [de Groot, 2006]). However, we currently know very little about the role that several other variables play in the decay process, suggesting the need for further research in this area.

While it is perhaps unavoidable to lean on the attrition literature given the limited number of studies that have looked at lexical decay, it is currently unclear how the findings of attrition studies relate to decay contexts. Furthermore, many attrition studies have not employed sufficient methodological control; in particular, several studies have failed to control or measure the manner in which the target items were learned. This means that interpretations about any one variable based on attrition data can be difficult. For instance, take a situation in which, prior to a retention interval, a group of learners studied target items of various parts of speech. Some learners may study certain items (e.g., nouns) in a manner that leads to more engagement than they do other items (e.g., verbs and adjectives). As increased engagement is associated with retention (Schmitt, 2008), subsequent testing would likely indicate that more nouns were retained than either verbs or adjectives. Importantly, although such data may suggest different patterns of loss based on one factor (i.e., part of speech), this effect may actually stem from a confounding variable (i.e., learning behaviour). Thus, as Bardovi-Harlig and Stringer (2010) recognised "the most revealing variables for understanding the nature of language attrition are the duration and nature of input, which must be described and quantified for the time periods applicable to the study" (p. 35). However, this is rarely the case in studies of lexical attrition and so there is currently a need for more tightly controlled investigations in which target variables are sufficiently isolated to delineate the effect of various factors on the loss of L2 lexical knowledge.

2.2.5 Summary of Research on Lexical Loss

It seems that lexical loss can be impacted by numerous factors. These findings are collated in Table 2.3. Importantly, almost none of the findings presented in this table stem from studies that considered decay settings. Studies of lexical decay have shown that intralexical (e.g., concreteness, PoS), interlexical (e.g., cognateness, L1 frequency), and contextual (e.g., learning strategy) factors can impact lexical decay. Thus, although we know a little about the role of some factors, we currently do not know how a large number of factors impact the forgetting patterns of L2 lexical knowledge in conditions of continued language exposure. One goal of the research conducted in this thesis was to determine the effect of several factors on decay processes to meet these gaps.

Table 2.3

Factors Affecting Lexical Loss

Variable type	Factors associated with less loss	Factors associated with more loss
Intralexical	• word class (nouns) • concreteness • imageability • shorter words	• word class (verbs) • abstractness • non-imageability • longer words
Interlexical	• phonotactical typicality • high L1 frequency • high L2 frequency • cognateness • formal similarity to L1	• phonotactical atypicality • low L1 frequency • low L2 frequency • non-cognate • formal dissimilarity to L1
Contextual	• intentional learning activity • use of keyword method • spaced repetition • extended spacing between repetitions	• incidental learning activity • using rote learning • massed repetition • brief intervals between repetitions
Learner-related	• high L2 proficiency • not young learner • L2 literate • high motivated to learn • high motivated to retain	• low L2 proficiency • young learner • L2 illiterate • low motivated to learn • low motivated to retain

2.2.6 Measuring Lexical Loss

Studies of vocabulary loss can be organised into two categories: those that have employed experimental designs and those that have used computational modelling.

Experimental Studies

Studies of language loss typically have four stages: a period of acquisition, a point of peak attainment, a retention interval involving reduced or no contact with the target items (also referred to as an incubation period [Gardner, 1982] or a period of reduced input [Hansen & Reetz-Kurashige, 1999]), and measurement of the knowledge retained after the retention interval. Language loss is calculated by comparing knowledge at the point of peak attainment with the delayed test. Thus, at first sight measuring forgetting seems like a simple proposition involving little more than a comparison of knowledge at two points. However, as the nature and timing of each of the four stages influences the extent of loss that occurs, it is, in fact, a complex endeavour involving the navigation of numerous methodological challenges.

The first challenge relates to the length of the retention interval and its influence on experimental design. This challenge is particularly pertinent to studies of language attrition that can investigate extraordinarily long retention intervals (e.g., Bahrick, 1984). Due to the extended length of the retention interval, it can be difficult to employ longitudinal research designs in such studies. Longitudinal studies trace the development of knowledge within the same participant(s) over a period of time, from initial learning through to a delayed test, with some studies also including periodic tests of retention to track the forgetting of vocabulary knowledge throughout the retention interval (Moorcroft & Gardner, 1987). However, with potentially long retention intervals, sometimes lasting many years (e.g., Bahrick and Phelps, 1987), such studies are liable to suffer from participant attrition and are logistically challenging to conduct. To overcome these problems, cross-sectional designs have often been employed. Such designs consider acquisition and loss using different samples, with the peak attainment of one group of learners compared to the retention of another (Hansen & Chen, 2001). Cross-sectional designs are appealing because they reduce the length of the research study, preventing participant attrition and making participation less onerous for learners. However, because cross-sectional designs consider learning and decay using different learners, the results may be impacted by learner-related factors (e.g., language learning aptitude). Although such designs attempt to control for variables such as

L1 and the learning context, successfully matching participants for all potentially confounding learner-related variables is often not possible. Furthermore, as language loss may be impacted by learning burden (Bardovi-Harlig & Stringer, 2010; Webb & Nation, 2017), which varies with individual learners (see Section 2.1.2), comparing the acquisition of one learner with the retention of another is problematic. Therefore, there have been calls for research on language loss to employ within-subjects designs where feasible (Bardovi-Harlig & Stringer, 2010; Meara, 2004). Importantly, the studies reported in this thesis employed such a design, which is explored in Chapter 3.

The second challenge relates to the measurement of knowledge at peak attainment. Because research on language loss can involve long retention intervals, it can be difficult to measure target item knowledge prior to the start of the retention interval. Three approaches to this measurement have been employed by research to date. Some studies have used self-report instruments, employing retrospection rather than a demonstration of knowledge to determine word knowledge at peak attainment (e.g., Beaton, Gruneberg, & Ellis, 1995). Other studies (e.g., Al-Hazemi, 2000) have simply assumed that learners successfully acquired the target items prior to a retention interval. This assumption is sometimes supported by teacher testimony or through reference to learning material. These two approaches are unreliable measures of lexical knowledge: learners may inaccurately recall the extent of target item knowledge and assuming learner knowledge does not in any way demonstrate actual knowledge of target items. Likewise, the use of teacher testimony and/or learner textbooks merely shows that certain lexical items may have been taught, it does not establish learner knowledge; indeed, there is a critical difference between teaching and learning (Willis, 2003). Furthermore, these two approaches do not delineate the relative difficulty with which target items were acquired. This is important as learning burden may impact the maintenance of lexical knowledge (Nation & Webb, 2011). A more satisfactory metric involves the direct measurement of lexical knowledge at the point of peak attainment. Some longitudinal studies have employed vocabulary tests prior to and following the retention interval. This approach is preferable as it provides a more robust picture of target item knowledge. Critically, the studies presented in this thesis directly measured target item knowledge and thus represent a more reliable treatment of lexical loss than studies that have employed self-report instruments or assumed target item knowledge.

A final challenge faced by studies of language loss is the possibility of intersessional target item exposure. While minimising exposure to target items during a retention interval is important for all vocabulary learning studies, it is especially important for studies of lexical loss. This is because such exposure would likely alter the forgetting process, impacting the reliability of the findings. Studies that have not controlled the learning process are inherently likely to suffer from this limitation, as they are unable to manipulate the target items to ensure both that they have been acquired prior to, and that they are unlikely to be encountered during, the retention interval. This further supports the use of longitudinal designs in studies of language loss. Investigations that have controlled the learning phase, have looked to mitigate intersessional exposure through several methods: the use of non-words (de Groot, 2006), the use of a foreign language which participants would not be exposed to during the retention interval (i.e., a language not studied by a participant group; Ellis & Beaton, 1993), and/or the use of brief retention intervals (de Groot & Keijzer, 2000). The research presented in this thesis mitigated such exposure by using low-frequency target items, brief retention intervals, and by removing from data analysis any items participants reported to have seen during the retention interval. This is likely to have provided a more reliable understanding of lexical decay than studies which did not adopt any mitigation strategies.

The manner in which target item knowledge is measured also impacts the extent of loss observed. Different tasks have been employed to measure vocabulary knowledge in studies of vocabulary loss to date. In fact, this is one cause of the inconsistent results of the research on lexical loss (Bardovi-Harlig & Stringer, 2010; Schmitt, 2010). The item type used to elicit target item knowledge and the aspect of word knowledge targeted are key variables in studies of vocabulary loss. Task types vary in terms of difficulty (Schmitt, 2010), as do different aspects of word knowledge (González-Fernández & Schmitt, 2019). For example, Laufer and Goldstein (2004) showed that tests of the form-meaning connection can vary in the level of challenge they pose, with tasks focusing on meaning easier than those of form, and recognition easier than recall tasks. Recognition tasks involve selecting an answer in response to a stimulus (e.g., a multiple-choice activity), whereas a recall task involves suppling an answer (Laufer & Goldstein, 2004). Task difficulty is associated with measurement strictness, which means that recognition instruments are less strict than recall instruments and are more sensitive to partial knowledge. When measuring language loss, target item knowledge may be relatively weak and so it is important to employ instruments that can detect partial knowledge. Additionally, to better understand language loss, multiple strengths of knowledge should be measured to provide a more nuanced picture of loss. This means that studies should include more than one measurement of vocabulary knowledge, allowing partial and stronger knowledge to be observed. The research reported in this thesis employed tests of both form-meaning recall and recognition.

Computational Modelling

In addition to experimental studies, language loss has been investigated through computational modelling. Meara (2004) constructed a series of models in which each word in a lexicon of 2,500 words was connected to two others. Each *word* in this network could be *activated* or *deactivated* depending on the type of connection it had to other items. Two types of connections were used, *AND* and *OR*. A word with an AND type connection was activated if both of the words to which it was connected were also activated, while a word with an *OR* type connection would activate if one of the words to which it was connected was also activated. Thus, it was easier for a word with an *OR* type connection to become activated than a word with an *AND* type connection. Meara modelled vocabulary loss by sequentially changing 225 *OR* connections to *AND* connections. Importantly therefore, he conceived language loss as a function of changes to the structure of a mental lexicon rather than the loss of knowledge of individual items within that network. He found that making these structural changes led to vocabulary loss, but that the manner in which loss took place was unpredictable, with considerable variance between different models (a sample of models is presented in Figure 2.2).

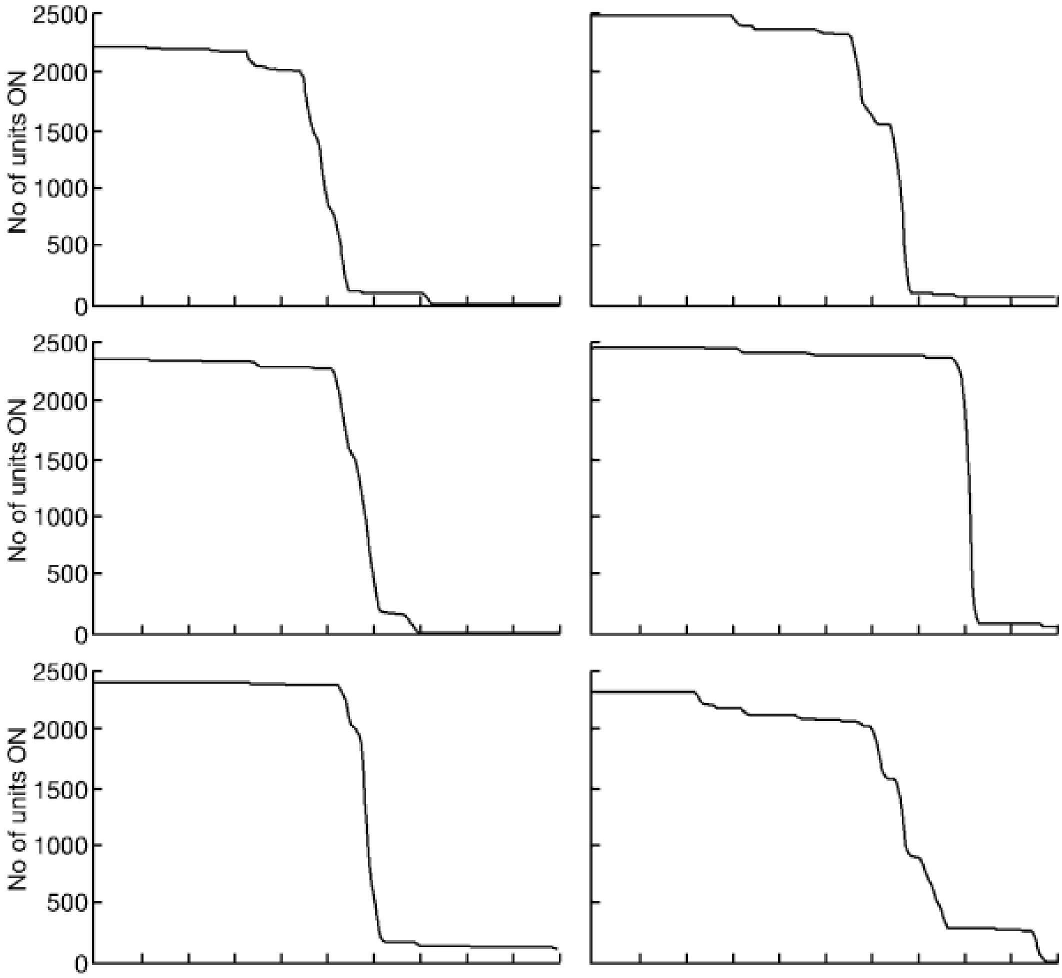
This innovative approach to understanding lexical loss has some advantages. Using computer modelling allows for the effect of several attrition events to be observed; for instance, Meara (2004) conducted 225 such events. In experimental research involving human participants, it would be unfeasible to conduct this number of observations and would likely lead to a testing effect even if it were possible. Additionally, the knowledge of many target items can be considered. Most experimental studies are unable to track learner knowledge of more than a small sample of items known to a learner, while the models

Meara (2004) computed considered all 2,500 items in a mental lexicon. Another advantage relates to the control the modeller has when writing the model. For instance, Meara adopted a k = 2 model in which each word was connected with two others, but this could easily have been altered to k = 3 or 4. While such control is no doubt an oversimplification of the mental lexicon (a fact that Meara acknowledges), it can help to clarify patterns that may be difficult to observe in research using human participants. For instance, the models Meara produced suggested that a mental lexicon could withstand several changes in structure before rapid loss occurred.

It is currently unclear, however, how such computer modelling could be applied to lexical decay, which tends to look at the loss of individual items rather than a lexical network as a whole. To effectively model decay of individual lexical items, we would need to know considerably more about the various factors that impact the loss of individual items. This is because modelling requires several important decisions to be made prior to model computation (Meara, 2004). For these reasons, this thesis has not employed computational models to measure decay; however, this is an interesting methodology to explore in future research.

Figure 2.2

Six Examples of Attrition of Words in a Network. Taken from Meara (2004)



2.3 *The Relationship between Burden and Decay*

It is often assumed that the lexical items most difficult to acquire are those easiest to forget (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 the manner in which 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 (Ishii, 2015; Tinkham, 1993, 1997; Waring, 1997), precluding the type of comparison needed to understand this relationship.

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. A paired associate learning design was employed that removed target items from the item pool once they had been learned to criterion (form recall). The number of trials needed to reach criterion was taken as a measure of burden. This measurement was correlated to target item retention eight years later. The results showed that items initially easier to learn were better retained, pointing to a positive relationship between learning burden and loss.

A similar finding was reported by de Groot (2006) who investigated the effect of several intralexical and interlexical variables on the decay of L2 lexical items using Dutch-nonword translation pairs. Participants studied the target items over three learning sessions, with interim test scores taken as a metric of learning burden (i.e., items learned in later sessions were considered more difficult than those learned in earlier sessions). These results were compared to a delayed meaning recall test administered after a one-week interval. The findings suggested that the words hardest to learn (i.e., were acquired later in the learning procedure) were most likely to be forgotten.

However, methodological characteristics of de Groot (2006) mean that her findings may have been confounded by the number of retrievals each target item received. In de Groot's (2006) study, target items were encountered the same number of times (three learning trials of two exposures, each followed by a test), but interim test scores indicated that some items were encoded earlier than others. These encoded items were not removed from the learning procedure and thus underwent more retrievals than target items that took longer to encode. Research shows that increased retrieval frequency leads to greater retention, meaning that this design may have strengthened the knowledge of the easiest items, making them less likely to decay. Bahrick and Phelps (1987) circumvented this issue by adopting a dropout design in which words were removed from the item pool once they had been learned; however, the extended length of their study means that participants may have been exposed to the target items during the retention interval. Thus, although some research has suggested that the learning burden of a word might be related to its probability of decay, more evidence is required and so the nature of this relationship remains an open question. The studies reported in this thesis addressed this gap by controlling for frequency of retrieval during learning and employing a comparatively short retention interval, reducing the chances of intersessional exposure to the target items.

2.4 *Chapter Summary*

The literature reviewed in this chapter has shown some notable gaps in our understanding of learning burden and decay. First, the literature on the role of several factors on learning burden has yielded conflicting findings (i.e., PoS and word length) and the role of other factors such as the code of meaning presentation, the mode of form presentation, perceived target item usefulness, and language learning aptitude remain unexplored. Secondly, very few studies have examined factors affecting lexical decay and most investigations have examined forgetting in context of reduced or no exposure. Finally, the potential relationship between burden and decay has yet to be empirically demonstrated.

In light of the research gaps identified, three empirical studies were conducted to examine the effect of intralexical (i.e., PoS and word length), contextual (i.e., code of meaning presentation and mode of form presentation), and learner-related (i.e., perceived usefulness and language learning aptitude) factors on learning burden and the decay of accrued knowledge. Additionally, each study considered the effect of the learning burden (irrespective of the specific target factors manipulated) on decay. The studies employed different lengths of retention interval (Study 1 – four weeks; Study 2 – one week; Study 3 – two weeks). The data from the three studies were compared to understand the impact of the retention interval length on the extent of decay. The following research questions were investigated:

Study 1 (Chapter 4):

1. To what extent do PoS and word length affect the learning burden of L2 lexis?
2. To what extent do PoS and word length affect the decay of L2 lexis?
3. To what extent does learning burden affect the decay process (irrespective of the target intralexical variables)?

Study 2 (Chapter 5):

1. To what extent do meaning presentation code and perceived target item usefulness affect the learning burden of L2 vocabulary knowledge?
2. To what extent do meaning presentation code and perceived target item usefulness affect the decay of L2 vocabulary knowledge?
3. To what extent does learning burden affect the decay process (irrespective of the target variables)?

Study 3 (Chapter 6):

1. To what extent does form presentation mode (i.e., unimodal or bimodal) affect the
    a. learning burden of L2 vocabulary?
    b. decay of L2 lexical knowledge?
2. To what extent do aspects of the aptitude complex (i.e., associative memory capacity, PSTM, and phonetic coding ability) influence
    a. the learning process?
    b. the decay process?
3. Irrespective of form presentation mode, to what extent does learning burden affect the decay process?

Cross-sectional analysis (Chapter 7):

1. To what extent is the decay of L2 lexical knowledge affected by the length of the retention interval?

The research presented in this thesis sought to clarify the role of various factors on the learning burden of L2 vocabulary and the subsequent decay of accrued knowledge. This comparison may facilitate a better understanding of the relative effect of certain variables on these two processes, allowing us to determine the extent to which variables act in the same or different manner on the acquisition and decay of lexical knowledge. Additionally, the role of learning burden on decay, irrespective of specific target variables considered, is investigated in each experimental study. This allows further clarification of the role difficulty plays in the process of forgetting.

Chapter 3: The Selection of the Flashcard Platform and Initial Piloting

The studies presented in this thesis employed flashcard software to foster target item knowledge. This chapter discusses the benefits that electronic flashcard software bring to studies of lexical acquisition and loss, as well as the selection criteria that were adopted to ensure the chosen software was maximally useful for research purposes. The chapter evaluates five flashcard platforms against these criteria and presents a pilot study conducted to assess the suitability of the chosen platform for the research studies presented in this thesis.

3.1 Introduction

Chapter 2 argued that the learning burden of L2 lexical items might impact decay and that learning burden should preferably be measured using a frequency-based metric. It also suggested that studies of language loss should attempt to differentiate exposures during the encoding process from retrievals of encoded knowledge, allowing a learner to see an item as many times or for as long as is necessary for them to learn it, but thereafter controlling the number of retrieval opportunities given. Electronic flashcard software can facilitate such a design. Flashcard use is a key vocabulary learning strategy (Nation, 2008) that is commonly recommended by vocabulary scholars (Nation, 1990; 2001; 2013; Webb & Nation, 2017) because it facilitates efficient (Thorndike, 1908) and effective (Crothers & Suppes, 1967) learning of L2 lexis. This intentional, decontextualized learning activity is particularly suited to foundational aspects of word knowledge (Schmitt, 2014) and is therefore often used to develop knowledge of the form-meaning link (Webb & Nation, 2017).

Using flashcards is, of course, not novel to either language learning practice or research. Students have likely used a form of flashcards to study L2 vocabulary for as long as they have been studying languages at all. Today, electronic flashcard platforms such as Anki, Quizlet, and Memrise have a vast userbase; for example, Quizlet (2018) reports having 30 million learners access its site every month. Moreover, research has employed word cards either as the focus of investigation (Burgess & Murray, 2014) or as a research tool to investigate factors that may influence vocabulary learning (Ellis & Beaton, 1993). However, much of the research that has employed flashcards has used physical cards rather than electronic software. Furthermore, while there have been review articles comparing software from an educational perspective (Nakata, 2011), there has been little consideration of which platform is best suited to vocabulary research. As the research presented in this thesis employed electronic flashcards, it was necessary to evaluate existing software to determine which platform was best suited to the research needs of this thesis, as well as vocabulary research more generally.

3.2 The Benefits of Electronic Flashcards for Vocabulary Research

Electronic flashcards have several advantages for vocabulary research. First, they allow the researcher to control item presentation, ensuring all target items are presented in a standardised manner. This is important as variables relevant to the presentation of an item have been found to impact the extent of learning that takes place (Lado et al., 1967; Mishima, 1967). Second, electronic flashcards facilitate a frequency-based operationalisation of burden in an environmentally valid learning context. To date, studies that have adopted a frequency-based metric have tended to use inauthentic learning situations; for example, Tinkham (1993) recorded his participants learning a list of words in one-to-one interviews. Such a design does not reflect typical learner behaviour. In contrast, while electronic flashcards can provide data on the number of exposures needed for learning, they are also used in natural learning contexts (Catalán, 2003). Finally, electronic flashcards allow exposures during the encoding process to be distinguished from retrievals of learned knowledge. Distinguishing encoding from retrieval is important because increased retrieval may counteract the natural decay process (Baddeley, 1990). Controlling frequency of retrieval is thus vital for a study of lexical decay. Learning studies that have used a frequency-based operationalisation of burden (e.g., de Groot, 2006) have generally not differentiated the encoding from the retrieval process. Thus, utilising flashcard software allows for more nuanced consideration of the decay process than research to date.

However, as the primary purpose of flashcard software is developing knowledge rather than researching knowledge development, many platforms are not suited for research purposes. This is because they lack or prevent access to certain features that are crucial to a study of lexical acquisition and/or loss.

3.3 Required Functionality

This section outlines the flashcard functionality required by the studies presented in this thesis. These criteria were used to select a platform from the numerous tools available.

The Capacity to Develop Original Flashcards

Platforms typically use decks of cards that have been developed by third parties, but some software additionally allow users to create their own flashcards. For the learner or teacher, creating flashcards is time consuming and requires knowledge of best practice in card design. For research purposes however, the capacity to develop flashcards is essential because it is unlikely that any pre-existing decks contain the specific target items of a study or are constructed in a manner suitable to the research design of a project. At their simplest, flashcards include an L2 form and the associated meaning(s), but additional features can be incorporated; for example, the spoken form, pictures, and video. The presence of any of these features is likely to affect learner engagement with the target items, impacting the learning process (Schmitt, 2008). It is important, therefore, that the design of the cards be consistent across the target items and match the objectives of a study. Additionally, if word meaning is presented via L2 definitions on pre-existing cards, the frequency profile of the definitions would need to align to the vocabulary level of the target population. This may not be the case, further speaking to the necessity of being able to develop bespoke flashcards.

The Capacity to Use Multiple Languages

Presenting the meaning of target items via L1 equivalents is a common feature of electronic flashcard software. Such a feature affords research to be conducted with participants who lack sufficient lexical coverage to decode L2 definitions and facilitates the investigation of the comparative efficacy of L1 and L2 meaning presentation, which is an important issue is SLA (R. Ellis & Shintani, 2013). The latter was a goal of Study 2 of this thesis. Thus, the capacity to produce flashcards that included learner L1 was adopted as a selection criterion.

The Capacity to Present Both the Spoken and the Written Word Form

Vocabulary studies have investigated learning with spoken and written form presentation. Additionally, studies have investigated how the manner in which word form is presented impacts its learning burden (Lado, Baldwin, & Lobo, 1967). It is important, therefore, that a flashcard platform allows the individual presentation of the spoken and written forms, as well as the concurrent presentation of both modes. This was the focus of the third study presented in this thesis. Thus, for research purposes, it is important that a flashcard platform be able to incorporate both spoken and written forms and allow for their simultaneous presentation.

The Capacity to Control the Formatting and Sequencing of Cards in a Deck

Flashcards can be active, encouraging form recall, or passive, requiring meaning recall (Nation & Webb, 2017). Studies have found that active presentation leads to more learning but is initially more burdensome (see Nation & Webb, 2017). Additionally, flashcards can employ multiple-choice items, targeting form or meaning recognition. Considering the hierarchical relationship between recall and recognition (Laufer & Goldstein, 2004), such cards would be less challenging to learners than either form or meaning recall. Many flashcard platforms manipulate the learning tasks to move from easier to more demanding presentation as knowledge of an item develops; however, research studies are likely to require a consistent learning task. Additionally, flashcard programmes differ in the response required of a learner. Some platforms require users to display knowledge of each target item, for example by typing the target form or selecting the correct meaning, while others do not, relying on self-reported evaluation of knowledge. In general, demonstration of knowledge is a more valid measure of learner knowledge (Schmitt, 2010). Thus, for research purposes it is preferable if a platform requires learners to demonstrate their knowledge, or, at least, allow such provision to be incorporated onto a flashcard.

The Capacity to Manipulate the Size of a Set of Flashcards

There are currently few principles for determining how many items should be included in a deck of flashcards. Studies using word pairs have varied from 6 (Ishii, 2015) to

100 items (Thorndike, 1908) in one learning session. The number of items in a flashcard deck impacts the difficulty with which some words within a deck are learned (Higa, 1965). Therefore, it is important to control this variable. Flashcard platforms differ in this regard, with some allowing users to specify the number of items in a deck and others generating deck size randomly. For research purposes, the former is preferable.

The Capacity of the Software to Evaluate User Knowledge and Provide Feedback

Electronic flashcard programmes often provide feedback on learner production. Based on this feedback, the extent to which an item is known is evaluated. This evaluation can be carried out automatically by software or sometimes manually by a learner. This evaluation is largely conducted to determine the timing of subsequent exposures based on spacing algorithms. For research purposes, this evaluation offers a simple method for differentiating correct and incorrect production. This is one method of distinguishing the processes of encoding (indicated by incorrect production) and retrieval (indicated by correct production) that is necessary to delineate the role of learning burden in the process of lexical decay.

However, self-evaluation has several limitations. Importantly, learners may incorrectly evaluate their knowledge, leading to false positives (claiming incorrect production to be correct) or false negatives (claiming correct production to be incorrect). This may result in an exposure prior to learning being considered a successful retrieval or a successful retrieval being counted as an exposure prior to learning. Therefore, research should use automatic scoring of production where possible. Where this is not possible, as a minimum, flashcard software should require demonstration of knowledge and provide feedback on that demonstration to inform a learner's evaluation. This is likely to reduce the number of false positives and false negatives that may otherwise occur.

The Accessibility of Learner Data

A final consideration is the extent to which individual learner data is accessible. In order to fully exploit the affordances of flashcards for research purposes, it is crucial that usage data be available and that they delineate the learning behaviour at the learner and item level. Previous research has needed to amend the learning task to ensure the availability of this data (Tinkham, 1993; Waring, 1997); ideally however, data should be collected without violating the authenticity and content relevance of the task.

In sum, the effective use of electronic flashcard software in research is contingent on the ability to manipulate the manner of item presentation, the provision of knowledge evaluation, and the accessibility of learner data.

3.4 Software Selection

Five flashcard platforms (Quizlet, Memrise, Anki, VTrain, and iKnow) were analysed against the capabilities presented in the previous section. These software were chosen because they are freely available, have generally been included in discussions of the comparative pedagogical value of flashcard software (Nakata, 2011), and are popular with learners and teachers.

Table 3.1 summarises the analysis of the five platforms. It shows that no platform met all criteria. *Anki* met more criteria than other platforms and it was determined that those criteria it did not satisfy could be met by adding certain procedures. As the software was written with language learning in mind, there are several round research pegs that do not fit into square pedagogical holes. Yet, with the application of a series of workarounds, most notably the use of USB flash disk memory drives to allow access to learner data, this learning tool was able to be adapted for research purposes. The adaptations made are discussed in the next section.

Table 3.1

Summary of the Analysis of Five Flashcard Programmes for Research Purposes

Feature	Quizlet	Memrise	Anki	VTrain	iKnow
Flashcard development	✓	✓	✓	✓	✓
Multilingual input	✓	✓	✓	✓	✓
Multimodal support	✓	✓	✓	✓	✓
Control of presentation	?	?	✓	✓	?
Set size choice	✓	X	✓	?	✓
Presence of feedback	?	?	✓	?	✓
Evaluation of knowledge	?	?	✓	✓	✓
Accessibility of individual learner data	X	?	?	?	X

Note. ✓ equates to presence of a feature, ? indicates inconsistent presence of a feature, and X relates to the absence of a feature.

Anki allows for the development of cards, the inclusion of the spoken form and pictorial support, the control of set size, and meaning presentation via both L1 and L2. It affords meaning and form recognition and, with the addition of a small piece of code, form and meaning recall. Interaction with a card is followed by user evaluation of knowledge, based on feedback from the software. For example, in a productive set up, a learner would see the meaning and be tasked with producing the form (see Figure 3.1). If the learner misspelled the target item, *Anki* would present the correct form beneath the learner's attempt, with the misspelled letters highlighted to indicate the mistake. The learner would then evaluate their knowledge by referring to this feedback and rate their knowledge using one of three options: *Easy, Good,* or *Again*. In the example presented in Figure 3.2, the learner would choose *Again* as he/she has incorrectly produced the target form. It is thus possible to differentiate successful from unsuccessful retrievals. The programme stores learner data on the hard drive, logging the item evaluation and the time taken between presentation and evaluation. The frequency and time of exposure is not automatically available but can be calculated by performing a database join.

Figure 3.1

An Example of the Form Recall Feature from Anki for the Item Conflate

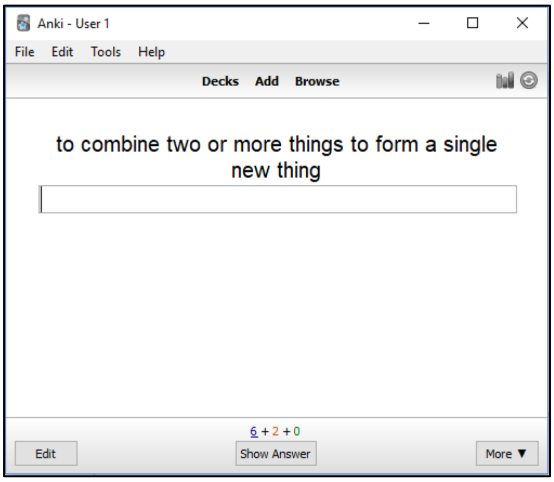
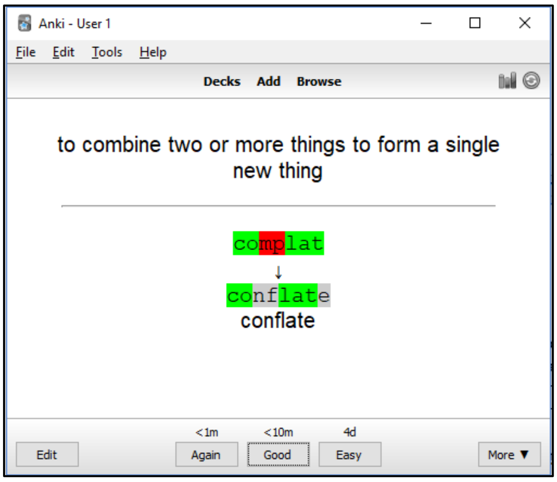


Figure 3.2

Example Feedback (Mistakes Highlighted in Red) and Evaluation Provision from Anki



3.5 Using Anki for L2 Vocabulary Research

This section describes how Anki was used in the three studies in this thesis and examples adaptations made with data from a pilot study.

Flashcard preparation

The flashcards were initially prepared in Microsoft Excel, with the target written form, meaning, and a numerical tag included. The numerical tag was an important addition; although not seen by participants, it allowed the same flashcard to be included in the decks of different days. This facilitated a uniform recycling pattern for the target items and bypassed *Anki*'s in-house spacing algorithm. The spreadsheet was saved as a UTF-8 file and imported to Anki. Once imported, the item presentation order was randomised to avoid a sequence effect. The decks were renamed so learners could navigate without difficulty; for example, the decks in the pilot study were named Day 1 (A), Day 1 (B) and so on. The flashcards were set up to facilitate form recall, requiring demonstration of written form knowledge. To accomplish this, word meaning was assigned to the front of the card and the word form to the reverse, using the following piece of code:

```
<<Front>>
<br>
<<type:Back>>
```

A bat file was written so *Anki* automatically opened when selected by a learner. Crucially, as *Anki* saves learner data to the computer hard drive, the software and the bat file were moved onto USB flash drives.

During the Experiment

During the research, each learner was given a flash drive from which to study. This enabled the calculation of the number of exposures needed by each participant to learn each item and, in Studies 2 and 3, to calculate the time of exposure.

Data Extraction

Data was extracted by accessing the Anki2 file stored on each USB flash drive. This database contained all data associated with the learning procedure in six tables: *cards, col, graves, notes, revlog, sqlite_stat1*. Learning data was stored in the *revlog* table. The raw data in this table does not include the item tags which allow for by-item analysis. Thus, it was necessary to manipulate these tables in Microsoft Excel with a macro (Study 1) and a database join (Studies 2 & 3) using the syntax presented in Figure 3.3. The output was copied into a premade Microsoft Excel file that automatically calculated the number of exposures per item (Studies 1, 2, 3) and total time per item (Studies 2 & 3).

Figure 3.3

Syntax for Extracting Learning Data from Anki2 Database - Studies 2 and 3

```
SELECT revlog.id, revlog.ease, revlog.time, notes.tags
FROM revlog
LEFT OUTER JOIN cards ON revlog.cid = cards.id
LEFT OUTER JOIN notes ON cards.nid = notes.id
```

3.6 Example Output

Tables 3.2 and 3.3 show data sampled from the pilot study, described later in this chapter. The data for two participants and ten items are included in this section. As can be seen, *Anki* provides the number of exposures and the time of exposure on a by-item basis, which can be used to determine learning burden. For instance, Table 3.2 shows that the noun *placard* was comparatively easy while the verb *traumatise* was harder for Learner A. This is indicated by both the greater number of exposures and the longer total length of exposure. Table 3.3 shows that the verb *dote* was comparatively easy, while the verb *traumatise* was more difficult for Learner B. Comparing the two tables, Learner A (Table 3.2) experienced more burden than Learner B (Table 3.3). Importantly, the figures show that the number and time of exposure varied between participants and between target items, illustrating that using electronic flashcards allows for by-item by-learner consideration of learning burden. Thus, using the software *Anki* allowed for nuanced consideration of learning burden using a frequency-based metric without negatively impacting the environmental validity of the study.

Table 3.2

A Selection of Learning Data from the Pilot Study by Item for Learner A

Item	PoS	Length	Number of exposures	Time of exposure (secs)
harpsichord	noun	11	16	290.28
cadaver	noun	7	11	157.58
dote	verb	4	10	76.30
salamander	noun	10	10	139.30
traumatise	verb	10	28	556.44
placard	noun	7	8	92.334
archipelago	noun	11	12	211.15
truncate	verb	8	28	368.68
hibernate	verb	9	8	179.58
scab	noun	4	19	265.94

Table 3.3

A Selection of Learning Data from the Pilot Study by Item for Learner B

Item	PoS	Length	Number of exposures	Time of exposure (secs)
harpsichord	noun	11	10	51.192
cadaver	noun	7	7	139.19
dote	verb	4	6	78.137
salamander	noun	10	9	94.369
traumatise	verb	10	10	150.162
placard	noun	7	9	87.457
archipelago	noun	11	9	43.798
truncate	verb	8	12	74.529
hibernate	verb	9	9	76.324
scab	noun	4	9	64.915

3.7 Pilot Study

This section provides a description of the pilot study. This pilot study was conducted to investigate several aspects of the research methodology. These included the number of target items, the number of target items per learning block, and the number of learning sessions. A secondary aim of the pilot study was to determine if the target intralexical variables chosen for Study 1 (word length and PoS) were likely to impact learning burden. In addition, an innovative assessment instrument was also trialled to determine its functionality. The pilot, as with the other studies investigated in this thesis, adopted a quantitative approach. The methodology employed in this thesis was heavily influenced by the results of the pilot study, so the pilot is described in detail.

3.7.1 Methodology

Participants

Initially, forty-two learners were recruited to participate in the pilot study; however, there was considerable participant attrition leading to a final sample of nineteen. This sample consisted of international students from two tertiary institutions in the UK. Participants varied in terms of gender (male = 9, female = 10), language proficiency (pre-intermediate to advanced), level of study (Pre-sessional English for Academic Purposes, Masters, PhD), L1 (Arabic, Chinese, French, Italian, Japanese, Spanish, Thai, Urdu), and length of residency in the L2 environment (0.25 years to 8.0 years). This diversity allowed for consideration of interactions between these participant characteristics and the research instruments. For example, the inclusion of high-proficiency participants confirmed the likelihood of the target language being largely unknown to even high-level learners, while inclusion of pre-sessional students unaccustomed to educational technology allowed for the usability of the electronic learning platform to be verified. A summary of the participants is given in Appendix 1. The study was run in accordance with the ethical approval (Appendix 2) and participants completed the approved consent form (see Appendix 3).

Target Items

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). Target items were selected to manipulate two intralexical variables: PoS and word length. As target item selection resembled that of Study 1 (see Chapter 4), it is not described in detail here. Table 3.4 presents the target items according to the two target intralexical variables. As can be seen, there were 72 items in total, with 36 nouns and verbs, and 8 items at each target word length.

Table 3.4

A Description of the Target Items from the Pilot Study

Length	Nouns	Verbs	Total
3	4	4	8
4	4	4	8
5	4	4	8
6	4	4	8
7	4	4	8
8	4	4	8
9	4	4	8
10	4	4	8
11	4	4	8
Total	36	36	72

The target items were organised into four decks by part of speech and word length. This was done to make the learning task less arduous for the participants. Each deck contained one noun and one verb of each word length, totalling 18 items per deck.

Flashcard Software

As discussed in Section 3.4, the software *Anki* was used. The target items were installed prior to the participants receiving their USB flash drive and the presentation format was predetermined. Flashcards were set up in the following manner. Participants saw an L2 definition and were tasked with learning the equivalent L2 form (Figure 3.1). After typing an answer, a participant was able to compare his/her answer to the correct form (Figure 3.2). Following this, learners indicated the accuracy of their production by selecting one of three rating options (see Figure 3.2): *Again* caused a target item to be presented again within a period of one minute; *Good* resulted in a word being presented within a period of ten minutes; and *Easy* meant a word was not presented again within a learning session. Learners were instructed to use *Again* for an incorrect response, *Good* for an accurate response, and *Easy* if they had previous knowledge of a target item. Once knowledge of an item was twice rated *Good*, it was not repeated in that session. In practical terms, learners were likely to select *Again* initially when they were encoding a word but then select *Good* when they could successfully retrieve an encoded item.

Instrument

The testing battery contained two instruments of 72 target items. The first instrument measured written form recall. Participants were presented with the definition used during the learning procedure and had to type the equivalent form. No initial-letter cue was provided on the instrument. Answers were scored dichotomously: one point was awarded for a correct response, zero for an incorrect or no response. The second test measured written form recognition. It employed a four-choice multiple choice format with three distractors, the key, and an additional *I don't know* option. Both tests were administered via Microsoft Excel (screen shots are provided in Appendix 4). The test was adaptive in that a correct response on the form recall test resulted in the automatic omission of the equivalent item from the form recognition test. An incorrect response on the form recall test resulted in the automatic inclusion of the relevant multiple-choice question on the form recognition test. The full instrument was 144 items in length (72 target items questioned twice); however, participants only saw all questions if they answered every item on the form-recall test incorrectly.

Procedure

All participants attended an initial session at which they were reminded of the purpose of the study, received instruction in the use of the learning software, and completed a practice to familiarise themselves with the mechanics of the software and the learning procedure. At that point, a USB flash drive was given to each participant. Contained on this were the flashcard software and task instructions. The flashcard software was programmed with daily tasks for the participants to complete. The participants studied the target language independently in their own time. The learning stage lasted for six days. On the first day, participants studied 36 items split into two blocks of 18 items. On the second day, participants studied the remaining 36 items, also divided into two blocks. The items studied on day one were recycled on day three. To control for a sequence effect, the order of block presentation was reversed and the order of items within each block was randomised. On day four, participants studied the target items from day two. On days five and six, participants studied all 72 target items; again, the order in which the items were presented was randomised. The learning procedure is illustrated in Table 3.5.

Table 3.5

The Deck Sequence Employed for the Pilot Study

Day	Target item block
1	1 & 2
2	3 & 4
3	2 & 1
4	4 & 3
5	1, 2, 3, & 4
6	3, 1, 4, & 2

Although the participants had been instructed to complete the learning procedure on consecutive days, for logistical reasons this was not always possible. Some participants took longer than the intended 6 days although all participants completed the six learning tasks in the correct order prior to submitting their USB flash drives. The day after submission, participants attended a testing session in which the testing instrument was administered. Following completion of both the form recall and recognition instruments, learners were asked to provide feedback on the experimental procedure. This feedback addressed the learning instrument, the task, the length of the learning process, and the number of target items. These responses, in addition to the data provided by the learning and measurement instruments, were used to evaluate the design and instruments. The specific areas that the pilot sought to validate were as follows:

1) the choice of the target intralexical factors
2) the appropriacy of the learning procedure
3) the usability of the learning software, *Anki*
4) the effectiveness of the measurement instrument

3.7.2 Analysis

Learning data were analysed to determine the number of exposures necessary for learning. Data was extracted by accessing the *Revlog*, C*ards*, and *Notes* tables in the Anki2 file, and running a macro to join the relevant files. The frequency of exposure data set was used to understand the learning burden of each item and determine the effect of PoS and word length on the learning process. Due to the small number of participants, analysis was conducted with reference to descriptive statistics.

3.7.3 Results

Descriptive statistics relevant to the target intralexical factors are presented in Table 3.6. Scores for the individual items are collapsed according to the two independent variables (length and PoS). The results suggest that both word length and PoS impacted burden. Table 3.6 indicates that longer words required more exposures to be learned and suggests that verbs needed more exposures to be learned than nouns. Additionally, it shows that the noun advantage was consistent at the various word lengths considered. Thus, these findings suggest that longer words and verbs were associated with higher burden.

Table 3.6

Descriptive Statistics for Learning Burden (Freq. of exposure) and Immediate Test Scores (Scores) by Word Length and PoS - Pilot Study (SD in Brackets)

Length	PoS	Freq. of exposure	Scores – form recall	Scores – form recog.
3	noun	8.89 (3.28)	3.47 (1.04)	3.95 (0.22)
	verb	9.96 (3.24)	3.00 (0.92)	3.68 (0.73)
	total	9.43 (3.29)	6.47 (1.67)	7.63 (0.81)
4	noun	9.64 (4.71)	3.21 (1.1)	3.89 (0.45)
	verb	10.04 (4.65)	3.21 (1.15)	3.68 (0.46)
	total	9.84 (4.67)	6.42 (2.18)	7.58 (0.75)
5	noun	9.92 (4.37)	3.16 (1.5)	3.84 (0.36)
	verb	10.97 (5.06)	3.11 (1.21)	3.68 (0.46)
	total	9.84 (4.67)	6.26 (2.65)	7.53 (0.75)
6	noun	10.38 (4.02)	3.05 (1.19)	4.00 (0.00)
	verb	11.01 (4.86)	3.00 (1.30)	3.79 (0.69)
	total	10.7 (4.45)	6.05 (2.30)	7.79 (0.69)
7	noun	9.54 (3.63)	3.32 (0.86)	3.89 (0.31)
	verb	12.72 (5.57)	2.42 (1.43)	3.79 (0.52)
	total	10.7 (4.45)	5.74 (2.10)	7.68 (0.73)

Length	PoS	Freq. of Exposure	Scores – Form Recall	Scores – Form Recog.
8	noun	10.74 (4.52)	2.79 (1.32)	3.68 (0.57)
	verb	11.57 (4.99)	2.63 (1.46)	3.79 (0.52)
	total	10.7 (4.45)	5.42 (2.66)	7.47 (0.94)
9	noun	9.96 (5.05)	2.63 (1.38)	3.79 (0.41)
	verb	11.28 (4.92)	2.58 (1.35)	3.42 (0.94)
	total	10.7 (4.45)	5.21 (2.57)	7.21 (1.20)
10	noun	10.42 (4.51)	2.84 (1.31)	3.58 (0.88)
	verb	11.13 (4.89)	2.53 (1.39)	3.58 (0.67)
	total	10.7 (4.45)	5.37 (2.43)	7.16 (1.42)
11	noun	10.97 (4.02)	2.63 (1.42)	3.79 (0.61)
	verb	11.28 (4.46)	2.79 (1.28)	3.79 (0.61)
	total	11.13 (4.23)	5.42 (2.62)	7.58 (1.23)
total	noun	10.05 (4.28)	27.11 (9.31)	34.42 (2.82)
total	verb	11.11 (4.81)	25.26 (9.93)	33.21 (4.47)
total		10.58 (4.55)	52.37 (9.62)	67.63 (3.65)

To evaluate the effectiveness of the flashcard software, learning gains were analysed. This analysis was organised by length (max = 8) and PoS (max = 36) at the levels of written-form recall and recognition. The descriptive statistics for learning gains are presented in Table 3.6. Taken together, the results suggest that the flashcard software led to considerable learning. The tables show that an average of 52.37 (*SD* = 9.62) items were learned to the level of form recall and 67.63 (*SD* = 3.65) items at the level of form recognition. These gains represent 72.74% and 93.93% of the target items respectfully. Greater learning was demonstrated when knowledge was considered at the level of written-form recognition than recall, but this was expected given that it is harder to develop productive mastery of target items (Laufer & Goldstein, 2004). Learning gains were fairly consistent across word length and PoS when considered at the level of form recognition, but there was a tendency for fewer comparatively long words and verbs to be learned to the level of form recall. Furthermore, there did not appear to be an interaction between word length and PoS, with similar gains for both nouns and verbs at the various word lengths.

3.7.4 Discussion

The results of the pilot and reflections on the procedure will be discussed with respect to the four aims of this investigation.

The Choice of the Target Intralexical Factors

Descriptive statistics from the learning procedure suggested that the proposed target variables for Study 1, length and PoS, impacted learning burden, with verbs and longer items found to pose more burden than nouns and shorter items. Therefore, word length and PoS were deemed to be appropriate intralexical factors to manipulate in Study 1.

The Appropriacy of the Learning Procedure

The data showed that the participants engaged with the learning process, developing knowledge of the majority of the target items. Therefore, it might be argued that the study employed an appropriate number of target items; however, there was also considerable participant attrition. Despite initial recruitment of forty-two learners, only nineteen participants completed the learning study. Furthermore, this participant attrition occurred prior to the onset of the retention interval, when further participant attrition may have also occurred.

This severe participant attrition likely stemmed from the length of the learning procedure. The study consisted of six consecutive days of learning, which was perhaps excessively demanding, practically and/or motivationally, for many learners. For instance, the data of participants who completed the learning procedure showed that many did not engage with the learning programme on consecutive days as intended. Additionally, the number of items may have negatively impacted motivation. The 72 target items and 18 items per deck may have placed excessive burden on the learners. In fact, after completing the learning study, participants commented that the learning procedure was overly challenging, with some learners who completed the learning procedure reporting that they had considered dropping out. The focus on independent learning may also have contributed to the participant attrition. The pilot study required learners to engage independently with the target items; however, some learners may have lacked the intrinsic motivation necessary to complete the learning procedure. In light of the participant attrition and the probable causal factors, the number of items, learning sessions, and items per learning session were reviewed. In addition, the context in which the main studies were conducted was reconsidered.

Regarding the difficulty of the target items and the probability of prior knowledge, the two most proficient students (evidenced by a self-reported proficiency survey and a vocabulary measurement) demonstrated prior knowledge of one item and ten items respectively. This was determined by analysing learner use of the *Easy* option, which learners were told to select if they had prior knowledge of a target item. In general, it appeared that the majority of the target items were unknown to students of advanced proficiency and thus were suitable for inclusion in the main studies of this thesis, which involved intermediate learners. In addition, the pilot study demonstrated that consideration of the *Easy* response data facilitated item deletion on a by-learner basis, which led to less item attrition and greater methodological rigour, as learning data was not confounded by prior knowledge.

The Usability of the Learning Software

Participants did not report any difficulty navigating the flashcard software. This was true even for participants who identified themselves as having poor computer literacy. In fact, participants commented on the ease of use and the effectiveness of the learning process. The learner data contained on each learner's USB flash drive was easily accessed and provided rich data regarding the learning process. Consequently, the pilot showed that *Anki* was an appropriate tool to employ in the three studies of this thesis.

The Effectiveness of the Measurement Instrument

The measurement instrument was adaptive in that a successful response at the recall level led to the omission of the equivalent item at the recognition level. The test utilised conditional formatting in excel to achieve this adaptability. The pilot showed that learners were able to work though the various sections independently and were not fazed by the exclusion of certain items from the recognition instrument. Thus, the same measurement principles were adopted in the main studies of this thesis. Furthermore, the pilot was automatically scored, with items marked dichotomously as correct or incorrect. For a response to be graded correct it needed to have been spelled accurately. However, manual inspection of the responses showed that numerous items included only one misplaced or omitted letter. In the scoring procedure adopted in the pilot, these answers were graded as incorrect. Importantly, there was an interaction between word length and spelling accuracy, explaining the comparatively low learning gains for longer items at the level of form recall, with longer items more likely to contain spelling errors than shorter items. Consequently, the rating procedure for the main studies was amended so that written form recall knowledge was scored at two levels of accuracy, strict and lenient.

3.7.5 Conclusion

The pilot indicated that, in general, the methodological design was appropriate for the studies presented in this thesis. The two target variables seemed to affect burden, the learning software performed as expected, facilitating nuanced consideration of the learning process and allowing all learners to engage with the target items as often as necessary for them to be acquired. However, the pilot also revealed some areas that needed reconsideration. In particular, the number of target items, the research setting, and the leniency with which responses were scored were amended for the main studies presented in this thesis.

Chapter 4: Study 1

The Effect of Part of Speech and Word Length on the Learning Burden and Decay of L2 Vocabulary Knowledge

4.1 Introduction

As argued in Section 2.1.2, several characteristics of words themselves, i.e., intralexical factors, 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); however, studies in this area have tended to use a gains-based metric of learning burden (see Section 2.1.4). The study presented in this chapter adopted a frequency-based metric of burden which, as discussed in Section 2.1.4, offers a more nuanced picture of burden. Additionally, we currently have a limited understanding of how PoS and word length impact the decay of L2 lexical knowledge (see Section 2.2.3) and studies in this area may have confounded the retention data by controlling the frequency of exposure during the learning phase rather than the frequency of retrieval (see Section 2.3). The study reported in this chapter employed a dropout learning procedure that avoided such a confound, and thus better isolated the role of PoS and word length in the decay of L2 lexical knowledge than some studies to date. Furthermore, as argued in Section 2.3, there is a general assumption within vocabulary studies that the lexical items most difficult to acquire are those easiest to forget, pointing towards a positive relationship between learning burden and loss (Webb & Nation, 2017). However, there is currently limited empirical evidence to support this assumption (but see Olshtain, 1989).

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 participants needed to see an item to learn it. The use of flashcard software for the instructional intervention provided access to such information about the learning process. The potential relationship between learning burden and decay was also explored.

4.2 Background

As explained in Section 2.1.3, among the different intralexical factors, PoS and word length have received little attention and the studies we do have, may have confounded these factors with other impactful variables such as concreteness and L2 frequency. The effect of these two factors on burden and decay is discussed in turn.

4.2.1 The Effect of PoS on Learning Burden

Some studies have shown that PoS affects the learning of L2 vocabulary, with nouns having an advantage in both incidental and intentional learning contexts. Within the incidental learning paradigm, Horst and Meara (1999) found that nouns were reported to be easier to learn than verbs, adjectives, and adverbs from exposure to a comic book. The authors argued that the pictorial support provided by the illustrations better supported learning of nouns than other parts of speech. Studies of incidental learning through listening have also demonstrated a noun advantage. For instance, Van Zeeland and Schmitt (2013a) found that more form, meaning, and grammatical knowledge was learned for the target nouns than was learned for the target verbs or adjectives.

Studies of intentional learning contexts (the focus of this study) have also found nouns to be easier to learn than other parts of speech. Investigating the learning of Russian through an L2-L1 paired associate learning task, Rodgers (1969) found an effect for PoS, with 58% of target nouns successfully learned to the level of meaning recall compared to 41% of adjectives, 23% of adverbs, and 21% of 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). This is because the noun advantage reported in some studies may have stemmed from confounding variables such as imageability or concreteness (Peters, 2020). These two intralexical variables have been found to impact the acquisition of L2 lexis, with higher levels of imageability and concreteness associated with increased learning gains (de Groot, 2006; de Groot & Keijzer, 2000; Puimège & Peters, 2019; van Zeeland & Schmitt, 2013b). Moreover, imageability and concreteness are related to PoS, with nouns more imageable/concrete than other parts of speech (Peters, 2020). Importantly, none of the studies reviewed in this section controlled for concreteness while manipulating PoS; thus, there is a need for further research in this area. In contrast, the investigation presented in this chapter manipulated PoS while controlling for concreteness.

4.2.2 The Effect of Word Length on Learning Burden

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, in a study of Bulgarian learners of English, 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 (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 upon 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 (Willis & Ohashi, 2012). The research presented in this chapter examined the role of word length while controlling for potentially confounding variables such as L2 frequency.

Moreover, as discussed in Section 2.1.4, much of the research to date has used learning gains to determine burden. A frequency-based metric represents a more reliable approach to determining learning burden than the methods adopted by the majority of studies reviewed in this section, and thus there is a need for research employing such a metric to verify the effect of PoS and word length. Importantly, the present study used a frequency-based approach to measuring learning burden.

4.2.3 The Effect of PoS and Word Length on Lexical Decay

As discussed in Section 2.2.3, while some studies have looked at the effect of contextual (Ellis & Beaton, 1993) and interlexical factors (de Bot & Stoessel, 2000; de Groot, 2006; de Groot & Keijzer, 2000; Lotto & de Groot, 1998; Tonzar, Lotto, and Job, 2009) on decay, only a few studies have targeted intralexical factors. 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) investigated 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. For instance, Willis and Ohashi (2012) investigated the extent to which word length, as measured by number of letters, phonemes, and syllables, affected Japanese learners' knowledge of English lexical items. They administered the Vocabulary Size Test (Nation & Beglar, 2007) and analysed the extent of knowledge considering numerous intralexical and interlexical factors, of which length was one. They found a greater probability of student knowledge if items were shorter. However, given the relationship between length and frequency, with more frequent items likely to be comparatively short (Peters, 2020), this study may have confounded word length with L2 frequency. Thus, we currently know little about how word length impacts the loss of language knowledge. The study presented in this chapter sought to address this gap.

4.2.4 The Relationship between Learning Burden and Vocabulary Loss

As argued in Section 2.3, there is a common assumption that items posing a higher initial learning burden are also those that are most easily forgotten (Webb & Nation, 2017), suggesting a positive relationship between decay and burden. The studies that have been conducted in this area generally support this position (Bahrick & Phelps, 1987; de Groot & Keijzer, 2000); however, such investigations may have been influenced by intersessional exposure to the target items and asymmetrical patterns of lexical retrieval during the learning phase. Crucially, both of these factors may impact the process of lexical decay, strengthening knowledge of items encountered during the retention interval and items that were retrieved more often during the period of acquisition. Therefore, more evidence is required to gain a better understanding of the role learning burden plays in the process of lexical decay. The current study employed a comparatively brief retention interval and low-frequency words to reduce the likelihood of exposure to the target items during the retention interval. It also employed a dropout learning design, removing items from the target item pool once they had been learned. This prevented items learned earliest receiving additional opportunities for retrieval, which may have biased the retention data.

4.2.5 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 chapter aimed to address these lacunae.

The following three research questions were examined:

1. To what extent do PoS and word length affect the learning burden of L2 lexis?
2. To what extent do PoS and word length affect the decay of L2 lexis?
3. To what extent does learning burden affect 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 the number of times a participant saw an item to be measured, and exposures prior to lexical encoding (i.e., learning burden) to be differentiated 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.

4.3 Methodology

4.3.1 Participants

Forty-eight English learners from different backgrounds and 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 EAP course at a UK tertiary institution at the time of data collection and had B2 proficiency as shown by their entry examinations. Additionally, proficiency was measured by a self-report rating scale on which learners indicated their perceived proficiency on a ten-point scale (one = extremely poor, almost no knowledge; ten = extremely good, almost native like). The mean scores were as follows: reading 6.22 (*SD* = 1.21), writing 5.89 (*SD* = 1.19), listening 6.24 (*SD* = 1.49), and speaking 5.04 (*SD* = 1.33). The learners had lived in an English L1 environment for an average of 1.23 months (*SD* = 1.40), had first contacted English at an average age of 8.5 years old (*SD* = 3.62), and had been learning English for an average of 9.19 years (*SD* = 4.99 years). Participation in the study was voluntary and learners did not receive compensation for their participation. Students gave their informed consent (Appendix 3) and were told they could withdraw at any time. The study was run in line with the approved ethical requirements (Appendix 5).

4.3.2 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 (number of letters). Subsequently, nouns were categorised as either abstract or concrete. This initial judgment was verified by a group scaling procedure with seven speakers of English as a first language (L1). This procedure was based on the methodology of Spreen and Scqziulz (1966). Participants were asked to rate each item on a scale from highly concrete to highly abstract. Key terminology was explained, and examples of highly concrete and highly abstract items provided. Only items evaluated as concrete were included in the pool of potential target items (see Appendix 6). Items were then deleted from this pool if the participants' teacher considered it likely that they were known to the students, 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 4.1.

Table 4.1

A Description of the Target Items (Study 1)

Word length	Number of nouns	Number of verbs	Total
3	4	4	8
6	4	4	8
8	4	4	8
11	4	4	8
Total	16	16	32

The 32 target items were separated into four decks. This was done to make the learning task less arduous and enable words to be learned and recycled in a systematic manner. Each deck contained eight items, one noun and one verb of each word length (three, six, eight, and eleven letters). The decks were labelled so that learners navigated through the learning task in a consistent order.

4.3.3 Definitions

Because the participants did not share the same L1, the meanings of the target items were disambiguated using English definitions taken from a monolingual learner's dictionary of English. As such they were written in controlled language and 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 Table 4.2.

Table 4.2

The Target Items and Definitions (Study 1)

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 embarrass someone who is speaking
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 you have already eaten back into your mouth

4.3.4 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). As explained in Section 3.4, the software *Anki* was chosen as it best met the functionality required by the studies in this thesis.

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 the number of exposures to be examined 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.

4.3.5 Language Background Questionnaire

A language background questionnaire was used to obtain information about the participants. This instrument included items related to biographic information such as gender, age, L1, course of study, length of English study, age of first contact, length of residency in an English L1 context, proficiency (IELTS), and the presence of an additional L2. Participants also rated their proficiency in the four skills using a 10-point rating scale. This information was selectively incorporated as covariates in the statistical modelling. The questionnaire is presented in Appendix 7.

4.3.6 *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 (see Webb, 2008 for an example of a non-cued test of form recall). Thus, this instrument is a more difficult 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. Several approaches to lenient scoring have been adopted in the literature. For instance, Thomas and Dieter (1987) scored responses at three levels of accuracy: W (word spelled correctly); W+(W-1) (words with one letter incorrect); and W+(W-1)+F (words with one letter incorrect or word fragments). However, this scoring system does not account for differential word lengths; one wrong letter in a three-letter word equates to 33% of the item spelled incorrectly, while one letter of an eleven-letter item equates to only 9% of the item spelled incorrectly. Therefore, this study adopted the above scoring method adjusted for word length. 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 using the following formula (WL = word length): W + (W-(WL*.33). In practice, this accepted one letter wrong for three-letter items, two letters wrong for six-letter items, three letters wrong for eight-letter items, and four letters wrong for eleven-letter items. Thus, responses were scored dichotomously at two levels of leniency: strict (correct = fully correct spelling) and lenient (correct = two-thirds of an item spelled correctly).

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. Lastly, an *I don't know* option was included. There has been considerable investigation into the effect of an *I don't know* option on multiple choice vocabulary tests (see Lucovich, 2014; Zhang, 2013). This instrument utilised an *I don't know* option to minimise guessing, but participants were told that there was no penalty for an incorrect guess and were encouraged to select an answer choice if they had any notion of the correct answer.

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. The test is included in Appendix 8.

4.3.7 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, the language background questionnaire was comprehensible, 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 completed the language background questionnaire. Learners were given a USB flash drive containing the learning software and programmed with daily tasks for the participants to complete. This was distributed and collected at the start and end of each experimental session to avoid participants checking the items outside the experimental sessions. The participants studied the vocabulary at their own pace during class using identical computers. They were supervised throughout, and I was on hand to answer any questions that arose. The learning procedure lasted for two sessions, lasting approximately one hour each. 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.

4.3.8 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 previous knowledge of target items was omitted from analysis on a by-learner basis. Statistical modelling was conducted with items seen three or more times, and thus only items previously unknown to the participants were considered. 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 three levels: items learned to the level of strict form recall, lenient form recall, and 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 strict form recall, lenient form recall, and form recognition. The fixed effects were word length, PoS, L1, and a vocabulary measure (the amalgamated scores from 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 the decay of learned knowledge was appropriately isolated, only items for which knowledge had been demonstrated on the immediate test were included. This prevented the analysis of decay being confounded 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 the number of exposures needed for learning (i.e., learning burden) as a covariate in the analysis of decay, it was possible to determine whether items that were harder to learn were more likely to be forgotten (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 PoS (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). Consensus has yet to form on the optimal method of model selection with mixed-effect models (Gries, 2013). This study employed a backward elimination method. 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 then fitted with all potential covariates listed above. The best-fitting model was determined by a process of backward elimination. This involved removing one covariate from the model at a time and analysing its effect on model fit using likelihood ratio tests. This was done by comparing the fit of model iterations that were identical save for one covariate. If removing a covariate resulted in a non-significant X^2 statistic, that covariate was concluded not to impact model fit and was excluded from subsequent iterations of the model. In this way, it was possible to examine the impact of each covariate on model fit and determine the most parsimonious model.

The model selection process is illustrated in Table 4.3, using the model for lexical decay at the level of form recognition. The process began with computation of Model 1 which contained the fixed and random effect structures outlined above. Model 2 was made by removing one fixed effect, total IELTs score. Comparing Model 1 to Model 2 therefore examined the effect of this variable on model fit (indicated by AIC). No statistically significant effect was found, indicating that removal of the IELTs variable did not negatively impact model fit. This process of backward elimination continued with the deletion of self-reported proficiency (Model 3) and the vocabulary measurement (Model 4). On removal of L1 (Model 5) and, subsequently, frequency of exposure (Model 6) a significant effect was found. Thus, Model 4 marked the point at which no further fixed effects could be removed from the model without significantly increasing the AIC. Model 4 was therefore selected as the best-fitting model for lexical decay at the level of form recognition. This process was followed for the analysis of learning burden, with separate models calculated for frequency of exposure at the three levels of measurement sensitivity.

Table 4.3

Summary of Generalised Logistic Mixed-Effect Model Comparisons for Lexical Decay (Form Recognition – Study 1)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Length; PoS; L1; Exposures; Proficiency (self-report); Proficiency (IELTS); Vocabulary	By-participant random slopes for target variables + by-item random intercept.	1621.4	NA	NA	NA
2	Model 1 – Proficiency (IELTS)	Same as Model 1	1619.8	-1.6	X^2 (1) = 0.31	*P* = 0.5
3	Model 2 – Proficiency (Self-report)	Same as Model 1	1618.5	-1.3	X^2 (1) = 0.70	*P* = 0.42
4	Model 3 – Vocabulary	Same as Model 1	1617.3	-1.2	X^2 (1) = 0.83	*P* = 0.36
5	Model 4 – L1	Same as Model 1	1619.0	1.7	X^2 (6) = 13.68	*P* = 0.03*
6	Model 4 - Exposures	Same as Model 1	1629.2	10.2	X^2 (1) = 13.91	*P* = 0.000***

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

4.4 Results

The immediate and delayed tests contained eight non-target items included to measure a test effect. Comparing the scores on these items across the two tests, it was possible to determine the effect of the immediate test on learner knowledge. The results showed that no words were learned at the level of strict form recall or lenient form recall, and less than one item was learned on average at the level of form recognition (M = 0.77, SD = 1.07). Thus, the results suggested that the first administration of the test led to minimal learning.

The remainder of this section presents the results of statistical modelling conducted to delineate the effect of PoS and word length on learning burden (research question 1), their impact on lexical decay (research questions 2), and the effect of learning burden irrespective of the target variables on lexical decay (research question 3).

4.4.1 Learning Burden

Table 4.4 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.

Inspection of the immediate test scores showed that considerable learning occurred over the treatment and that more learning took place when knowledge was considered at the form recognition level than at the form recall level, with an average of 13.31 and 15.27 words learned at the level of strict and lenient form recall respectively, and 26.27 words learned to the level of form recognition. The descriptive statistics (Table 4.4) also indicated that more learning took place when form recall knowledge was measured using the more inclusive lenient metric than when the strict index was employed. A further general observation with regard to learning gains is that there was considerable individual variation, particularly when knowledge was considered at the form recall level. For example, the standard deviations for the strict and lenient conditions were approximately half as large as the means.

In terms of learning burden, the data showed that approximately six exposures were required on average to acquire the target items. There was a tendency for longer words to pose more burden than shorter words; for instance, an average of 4.72 exposures were needed to learn three-letter items to the level of form recognition, while 5.69 exposures were required on average to learn eleven-letter items to the same level of measurement sensitivity. This pattern was repeated on both levels of form recall. In contrast, the effect of PoS on burden was less clear, although verbs required slightly more exposures to be learned than nouns. Nouns needed an average of 5.76, 5.87, and 5.14 exposures to be learned to the level of strict form recall, lenient form recall, and form recognition respectively. In contrast, verbs required an average of 6.09, 6.18, and 5.38 exposures to be acquired at the equivalent levels of lexical mastery.

These initial observations were tested using statistical modelling. MEMs were fitted using the two intralexical factors and several covariates (see Section 4.3.8) as predictors and the frequency of exposure needed for learning (i.e., learning burden) as the outcome. Separate models were computed for the three levels of measurement sensitivity.

Table 4.4

Mean Immediate Test Scores and Learning Burden (Strict and Lenient Form Recall and Recognition) by PoS and Word Length (SD in Brackets)

		Strict form recall		Lenient form recall		Form recognition	
Length	PoS	Score	Exposures	Score	Exposures	Score	Exposures
3	noun	2.23 (1.31)	5.23 (0.65)	2.31 (1.32)	5.25 (0.63)	3.33 (0.78)	4.75 (0.36)
	verb	2.15 (1.15)	5.05 (0.13)	2.35 (1.18)	5.00 (0.12)	3.38 (0.87)	4.69 (0.18)
	total	4.38 (2.14)	5.14 (0.45)	4.67 (2.13)	5.13 (0.44)	6.73 (1.32)	4.72 (0.27)
6	noun	1.52 (1.22)	5.50 (0.64)	1.71 (1.25)	5.53 (0.54)	3.17 (0.97)	4.90 (0.23)
	verb	1.77 (1.28)	5.78 (0.69)	1.88 (1.23)	5.67 (0.66)	3.29 (0.90)	5.13 (0.54)
	total	3.29 (2.22)	5.64 (0.64)	3.58 (2.22)	5.60 (0.56)	6.46 (1.66)	5.02 (0.41)
8	noun	1.65 (1.26)	5.91 (0.92)	1.94 (1.28)	6.03 (0.84)	3.31 (0.78)	5.38 (0.41)
	verb	1.58 (1.29)	6.63 (1.16)	1.79 (1.27)	6.84 (1.06)	3.38 (0.89)	5.85 (0.30)
	total	3.23 (2.29)	6.27 (1.04)	3.73 (2.33)	6.44 (0.93)	6.69 (1.39)	5.61 (0.42)
11	noun	1.33 (1.17)	6.40 (0.94)	1.85 (1.25)	5.87 (0.83)	3.48 (0.77)	5.53 (0.31)
	verb	1.08 (1.20)	6.89 (0.96)	1.44 (1.30)	7.20 (1.06)	2.92 (1.01)	5.85 (0.11)
	total	2.42 (2.18)	6.65 (0.92)	3.29 (2.32)	6.93 (0.93)	6.40 (1.45)	5.69 (0.28)
Total	noun	6.73 (4.06)	5.76 (0.85)	7.81 (4.30)	5.87 (0.84)	13.29 (2.60)	5.14 (0.45)
	verb	6.58 (3.72)	6.09 (1.05)	7.46 (3.86)	6.18 (1.13)	12.98 (2.67)	5.38 (0.59)
	total	13.31 (7.78)	5.92 (0.96)	15.27 (8.16)	6.02 (0.99)	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. Score refers to the score on the immediate test.

Strict Form Recall

MEMs were fitted to determine the effect of the target variables on learning burden. At the level of written form recall, the best-fitting model showed that there was a significant effect for length, with longer words requiring more exposures to be learned than shorter words (t = 2.31, p < .03). No statistically significant effects were found for PoS (t = 0.09, p = .93), showing that nouns and verbs posed the same level of burden. The interaction between PoS and word length was also found to be non-significant (t = -0.31, p = .76), indicating that the effect for word length was consistent across nouns and verbs. Finally, no significant effects were found for any other covariate considered. The results are presented in Table 4.5.

Table 4.5

Fixed and Random Effects for Selected Strict Recall Model, Learning Burden (Study 1)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
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

Random Effects		
Parameter	Variance	*SD*
Item	.01	.06
Participant	.01	.09
Word Length	.02	.12
PoS	.01	.08
Length x Class	.01	.07

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

Lenient Form Recall

At the level of lenient form recall, the best-fitting model showed that PoS (t = 0.12 p = .91) had no effect on learning burden, but there was a statistically significant effect for word length (t = 2.31, p = .03). The interaction between length and PoS was not statistically significant (t = -0.25, p = .80) and no effects were found for any of the covariates included in the analysis. The MEM results are presented in Table 4.6.

Table 4.6

Fixed and Random Effects for Selected Lenient Recall Model, Learning Burden (Study 1)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	1.41	0.08	17.81	<.001
Word length	0.12	0.05	2.31	.03*
PoS	0.01	0.11	0.12	.91
Length*Class	-0.02	0.06	-0.25	.80
Random Effects				
Parameter	Variance	*SD*		
Item	.01	.06		
Participant	.01	.11		
Word Length	.04	.19		
PoS	.02	.14		
Length x Class	.01	.12		

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

Form Recognition

At the level of written form recognition, the best-fitting MEM indicated that word length significantly impacted learning burden (t = 3.12, p = .003), with longer items associated with more burden. However, in line with the form recall analysis, no statistically significant effect was found for PoS (*t* = 0.39, *p* = .70), suggesting that nouns and verbs posed a similar level of burden. The interaction between PoS and word length was also non-significant (*t* = -0.66, *p* = .51), showing that the word length effect was not moderated by PoS. Lastly, no significant effects were found for any of the covariates included. The results are reported in Table 4.7.

Table 4.7

Fixed and Random Effects for Selected Recognition Model, Learning Burden (Study 1)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
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
Random Effects				
Parameter	Variance	*SD*		
Item	.01	.04		
Participant	.01	.12		
Word Length	.03	.16		
PoS	.01	.08		
Length x Class	.01	.08		

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

4.4.2 Lexical Decay

With regard to the second research question, the descriptive statistics for the delayed test are given in Table 4.8. The table presents the mean items recalled per word length and PoS on the delayed test. 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.

The descriptive statistics showed that there was considerable decay of accrued knowledge. This loss was especially severe for strict form recall knowledge, where 92.5% of the learned knowledge was found to have decayed. There was marginally less decay when the lenient scoring condition was considered, with 88.41% of learned knowledge forgotten over the four-week retention interval; however, this figure still represents substantial loss. Less decay was found at the level of form recognition, with 44.5% of accrued knowledge forgotten over the retention interval. Another general observation is that there was considerable individual variance in the amount of lexical knowledge retained. For instance, the standard deviation for total retention at the form recognition level was approximately half the size of the mean. In terms of the two target intralexical variables, more nouns were retained than verbs; for instance, at the level of form recognition, 7.83 nouns were retained on average, compared to 6.75 verbs. Word length seems to have had little impact on decay, with broadly similar delayed test scores for the various lengths.

Statistical modelling was conducted with the delayed test data to determine the effect of the two target intralexical factors on the decay that occurred. GLMERs were employed as the outcome variable, responses on the delayed test, were binomial, with an item on the test answered either correctly or incorrectly. PoS, word length, and the frequency of exposure needed to learn an item, in addition to several covariates (see

Section 4.3.8), were used as predictors. Of note is that by including the frequency of exposure in the decay analysis, it was possible to consider the role of learning burden in lexical decay (research question 3). Three models were computed in total: one at each level of measurement sensitivity.

Table 4.8

Mean (SD in brackets) Delayed Test Scores Relative to Learning by PoS and Word Length (Study 1)

Length	PoS	Strict form recall	Lenient form recall	Form recognition
3	noun	0.25 (0.48)	0.40 (0.62)	2.00 (1.22)
	verb	0.21 (0.50)	0.42 (0.75)	1.96 (1.10)
	total	0.46 (0.71)	0.75 (0.99)	3.96 (2.00)
6	noun	0.10 (0.37)	0.18 (0.45)	1.67 (1.23)
	verb	0.17 (0.47)	0.24 (0.53)	1.63 (1.27)
	total	0.27 (0.73)	0.38 (0.80)	3.29 (2.23)
8	noun	0.08 (0.28)	0.35 (0.61)	1.98 (1.22)
	verb	0.10 (0.37)	0.18 (0.45)	1.73 (1.27)
	total	0.18 (0.44)	0.49 (0.73)	3.71 (2.16)
11	noun	0.08 (0.28)	0.26 (0.44)	2.19 (1.18)
	verb	0.00 (0.00)	0.03 (0.17)	1.44 (1.15)
	total	0.08 (0.28)	0.26 (0.44)	3.63 (2.11)
Total	noun	0.52 (0.79)	1.04 (1.23)	7.83 (3.87)
	verb	0.48 (0.91)	0.77 (1.25)	6.75 (3.83)
	total	1.00 (1.41)	1.77 (2.06)	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.

Form Recall

The scores from the strict and lenient form recall tests 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 models for the retention of form recall knowledge. To begin to understand the role of learning burden on the retention of form recall knowledge (research question 3), descriptive statistics are reported for the strict and lenient conditions (Table 4.9). 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.

Looking at Table 4.9, the relative retention figures suggest that at the level of strict recall, words that were acquired more easily were better recalled. For instance, more of the words that were seen only three times during the learning procedure were recalled than words seen six times, which in turn were recalled more frequently than words seen eight times. However, there were exceptions to this pattern, (e.g., items learned with nine exposures), and so it is difficult to state with confidence the relationship between learning burden and decay. Similarly, at the level of lenient form recall, items viewed least frequently were better retained; yet this was not consistently the case. For instance, items that were seen only four times during the learning procedure were recalled with similar likelihood as items seen three times as often. Thus, based on the findings presented in Table 4.9, it seems that although no clear pattern emerges, the relative retention figures suggest 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 needed to confirm this initial observation.

Table 4.9

Number of Items Learned (Immediate Test) and Retained (Delayed Test) on the Measures of Strict and Lenient Form Recall by Frequency of Exposure (Study 1)

	Strict form recall			Lenient form recall		
Exposure freq.	Items learned	Items retained	Relative retention	Items learned	Items retained	Relative retention
3	73	9	0.12	79	16	0.20
4	187	15	0.08	216	28	0.13
5	97	9	0.09	110	15	0.14
6	75	5	0.07	80	8	0.10
7	48	5	0.10	56	6	0.11
8	35	1	0.03	38	1	0.03
9	31	4	0.13	37	6	0.16
10	14	0	0	16	0	0
11	11	0	0	16	2	0.13
12	10	0	0	11	2	0.18
13	6	0	0	9	0	0

Form recognition

Moving on to the analysis of decay at the level of written form recognition, the results indicated that neither word length (z = -1.36, p = .18) nor PoS (z = -0.6, 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 (z = 1.1, p = .27) (see Table 4.10 for the coefficients of the random effects and the fixed effect structures of the best-fitting model). Thus, the results suggested that nouns and verbs, and words of differing lengths, were equally likely to decay over the four-week retention interval. Secondary effects were found for L1, with L1 speakers of Mandarin suffering less decay than other participants.

Table 4.10

Fixed and Random Effects for Selected Form Recognition Model, Lexical Decay (Study 1)

Fixed Effects

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

Random Effects

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

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

In response to the third research question, learning burden was found to significantly affect decay ($z = 3.73$, $p < .001$), with more exposures (i.e., a higher learning burden) associated with the mitigation of loss of form recognition knowledge. To explore this effect in more detail, Table 4.11 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 4.11

Number of Items Learned (Immediate Test) and Retained (Delayed Test) on the Measure of Form Recognition by Frequency of Exposure (Study 1)

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

4.4.3 Summary of the Findings for Learning Burden and Lexical Decay

The results of the learning burden and decay analysis are summarised in Table 4.12. Overall, the findings show a consistent picture: PoS had no significant effect on burden or decay, length impacted burden but not decay, there were no interactions between PoS and length on burden or decay, and increased burden was associated with less decay of form recognition knowledge.

Table 4.12

Summary of the Findings for Learning Burden and Lexical Decay (Study 1)

	Strength of knowledge	PoS	length	PoS *length	burden	L1
Learning burden	Strict form recall	X	✓	X		
	Lenient form recall	X	✓	X		
	Form recognition	X	✓	X		
Decay	Form recognition	X	X	X	✓	✓

Note. ✓ represents a statistically significant effect and x represents a non-statistically significant effect.

4.5 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.

4.5.1 RQ1: To What Extent Do PoS and Word Length Affect the Learning Burden of L2 Lexis?

Before discussing the effect of the target intralexical variables, some general comments about the learning gains and learning burden observed in this study are given. First, the findings showed that an impressive number of L2 target items could be learned in a relatively brief period using electronic flashcard software. This supports the calls for employing such software to develop lexical knowledge when learning a foreign language

(Nation, 2013; Schmitt, 2000; Webb & Nation, 2017). Second, more target items were learned to the level of form recognition than form recall. This was expected given the hierarchical relationship that research has reported between recall and recognition, with the former found to be more challenging than the latter (Laufer & Goldstein, 2004). Third, the target items required an average of approximately six exposures across the two learning sessions to be acquired. However, considerable individual variance was observed in the extent of learning burden posed by the target items, with some items posing more burden than others and some learners needing more exposures than others to learn the target items. This finding echoes that of the pilot study (see Chapter 3) and supports Higa's (1965) notion of learning burden as a by-item by-learner variable.

Moving onto the role 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 (Baddeley, Thomson, & Buchanan, 1975; Gerganov & Taseva, 1982). No effect for PoS was found on learning burden. This contradicts the findings of previous studies that have found nouns to pose a lighter learning burden than verbs (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 (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 (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, with some studies finding a noun advantage (Ellis & Beaton, 1993), some reporting no effect for PoS (the present study), and others finding an advantage for verbs (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.

4.5.2 RQ2: To What Extent do PoS and Word Length Affect the Decay of L2 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 showed 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 (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 accrued 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 (González-Fernández & Schmitt, 2019; 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 it was not possible to perform statistical analyses. 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 items 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, 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 measure, 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.

4.5.3 RQ3: To What Extent Does Learning Burden Affect the Decay Process (Irrespective of the Target Intralexical Variables)?

The results showed that the number of exposures needed for learning was negatively related to decay 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 were developing knowledge of the form-meaning link seems to have led 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.

4.6 Limitations

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 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 nature of the participant group, 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 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 would likely 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 and in some contexts target items are indeed not often recycled, there may be occasions where learners would engage with items more frequently than the retention interval used here and, in these cases, those extra exposures would affect the decay rates reported. This speaks to the need for future research to employ different retention intervals in classroom settings.

4.7 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.

Chapter 5: Study 2

The Effect of Code of Meaning Presentation and Perceived Usefulness on the Learning Burden and Decay of L2 Vocabulary Knowledge

5.1 Introduction

Study 1 showed that lexical knowledge decays and that this loss is impacted by the number of exposures needed during the learning stage, but not by word length or PoS. However, in addition to the target intralexical factors investigated in that study, there are numerous candidate factors that influence the process of lexical acquisition (see Section 2.1) which might have an equivalent or differential effect on the process of lexical decay. Section 2.1.2 showed how learner variables seem to affect vocabulary learning. One such learner variable is the perceived usefulness of target items. This factor is important as it is likely to influence learner motivation, which has been shown to strongly impact the acquisition process (Dörnyei, 2005). Currently, little is known about the impact of this affective factor on the learning process (Peters & Webb, 2018), and almost nothing is known about how it affects the process of decay. One goal of this study, therefore, was to determine the effect of perceived item usefulness on the learning burden and decay of L2 vocabulary knowledge.

Additionally, Section 2.1 showed that contextual factors can influence the efficacy of pedagogical treatments. One such contextual factor is the language in which word meaning is presented. Studies have generally found that using L1 equivalents to present word meaning leads to more learning that using L2 definitions (Lado et al., 1967; Laufer & Shmueli, 1997; Mishima, 1967); however, as will be shown in this chapter, these studies have not measured learning burden, nor have they explicitly considered the decay of learned knowledge. Therefore, it is currently unclear how meaning presentation code impacts both learning burden and the loss of intentionally acquired lexical knowledge. Consideration of meaning presentation code may also help to clarify the incongruity between the non-effect for PoS reported in Study 1 (Chapter 4) and the results of previous research. This is because studies that have shown an effect for PoS on learning burden have done so by utilising L1-L2 word pairs (Ellis & Beaton, 1993), while Study 1 employed L2 definitions. Thus, a secondary goal of the current study was to determine whether the non-effect for PoS found in Study 1 was due to the manner in which word meaning was presented.

Of the contextual variables that might affect the efficacy and/or efficiency of intentional L2 vocabulary learning, the manner of target language meaning presentation has been repeatedly investigated. However, despite research targeting the mode of both meaning and form presentation, including the manner of form presentation (Lado et al., 1967), the use of pictorial support (Lado et al., 1967; Yoshii, 2006), the order of form and meaning presentation (Griffin & Harley, 1996), the mode (i.e., spoken or written) of meaning presentation (Mishima, 1967), and the language of meaning disambiguation (Laufer & Shmueli, 1997), it seems that teaching practices have not been greatly affected.

One area that illustrates this lack of impact, and that is of particular relevance to the study described in this chapter, is the code used to present item meaning (i.e., L1 or L2). The use of the L1 in the L2 classroom has been a controversial issue for many years. This controversy stems, in part, from paradoxical assumptions: considerable interaction in and with the L2 is necessary for the development of competent language use; while the L1 can facilitate expedited language acquisition and alleviate language learning anxiety (R. Ellis & Shintani, 2013). It is an issue that still receives national (e.g., MEXT, 2013), institutional (e.g., Berger, 2011), practitioner (Edstrom, 2006), and student (Prodromou, 2002) consideration, and thus has wide-ranging implications. In fact, the use of the L1 in the L2 classroom has been described as "the most important theoretical and pedagogic question facing both the research and practitioner communities today" (Macaro, 2011, quoted in R. Ellis & Shintani, 2013, p. 225). However, R. Ellis and Shintani (2013, p. 245) recognise that "there is a conspicuous lack of research... [on the] effect (facilitative or debilitative) that use of the L1 has on actual learning". In light of this lacuna, this study explored the efficacy of L1 and L2 meaning presentation codes. It did this by looking at the impact of meaning presentation code on the learning burden, operationalised as frequency and time of exposure, and decay, operationalised by comparison of an immediate test with a delayed test, of intentionally learned L2 vocabulary knowledge.

5.1.1 Terminology

One goal of this study was to determine the effect of meaning presentation code on the learning burden and retention of L2 lexis. The two codes employed in this study were L1 and L2. Some researchers have recognised that the use of these terms is flawed; for example, Hall and Cook (2012) rightly state that classifying the language of study as the L2 suggests that learners did not speak any additional languages prior to learning the foreign language of interest. That is to say, the L2 may in fact be a learner's L3, L4 or L5. Instead of L2, therefore, they recommend the use of *new language* to denote the language of study and *own language* to refer to the L1. Other terms that have been utilised to describe the L1 include *native language* and *mother tongue*. The former of these is an ambiguous term (e.g., Davies, 2003), while the latter is often inaccurate "for the obvious reason that many people's 'mother tongue' is not their mother's 'mother tongue'!" (Hall & Cooks, 2012, p. 274). Therefore, while aware of the criticism surrounding the use of the L1 and L2, this chapter utilises this terminology (See R. Ellis & Shintani, 2013 for similar use).

5.2 Background

This section is organised according to the two experimental foci of this chapter, meaning presentation code and perceived target-item usefulness.

5.2.1 Meaning-Presentation Code

L1 Use in the Language Classroom

The belief that the L2 is best taught via the L2 (the *monolingual assumption*; Hall & Cook, 2012) dates at least from the nineteenth century and the introduction of the Direct Method (R. Ellis & Shintani, 2013). This was introduced in reaction to grammar-translation, and in contrast to that approach, stipulated that only the target language be used in the classroom. In addition, the monolingual assumption has been propagated by the commercial success of the Berlitz Schools (founded in 1878; Berlitz, 2009) and the associated Berlitz Method (Hall & Cooks, 2012). This insisted, and still insists in language schools around the world (Berlitz, 2009), on exclusive L2 use in the language classroom.[1] Since these early proponents of monolingual instruction, language teaching methods have

[1] The Berlitz method of immersive language teaching was reportedly started when a monolingual French teacher covered classes for the founder of the company, Maximillian Berlitz. After six weeks of sickness the owner returned to the classroom and witnessed considerable linguistic improvement from the student body. This led to the prioritisation of monolingual language teaching and the development of the Berlitz Method (Berlitz, 2009).

varied greatly in the extent to which the L1 is seen as facilitating or inhibiting language acquisition (R. Ellis & Shintani, 2013). This is illustrated by Table 5.1. As can be seen, there has been little agreement among methodologists about the role the L1 should play in SLA. The pendulum has swung back and forth between the two poles for the last hundred years, and while there seems to have been a move away from the maximal position in which the target language is used throughout (Barker, 2003), history gives us no reason to trust the permanency of this current stance. Furthermore, as will be shown below, this current theoretical position does not necessarily equate with institutional or classroom practice.

Although current instructional methodologies seek, in general, to utilise the L1 as a pedagogical resource (particularly for less proficient learners), many educational stakeholders (e.g., boards of education, teaching institutions, materials writers, teachers, parents, and students) remain committed to monolingual teaching (Hall & Cook, 2012). For example, at the level of national education policy, a continued preference for exclusive L2 use can be seen in some parts of the world. In Japan, for example, the Ministry of Education, Culture, Sports, Science, and Technology released guidelines for senior high-school English language instruction that required maximal L2 use in the classroom. It stated that "in order to expand opportunities for students to come into contact with English, and make classes into sites of real communication, classes are to be conducted in English" (MEXT, 2013). Another example comes from South Korea, where the 7th National Educational Curriculum (MOE, 1997) recommends that teachers exclusively employ L2 instruction in primary education, while the Teaching English in English scheme suggests teachers should maximise L2 use where possible at the secondary level (see Chung & Choi, 2016).

Table 5.1

The Role of the L1 in Different Language Teaching Methods (Adapted from R. Ellis & Shintani, 2013)

Approaches favouring exclusive L2 use	Approaches favouring use of the L1
Direct Method Vocabulary is taught through mime and ostensive definition	*Grammar Translation* L2 vocabulary introduced via L1 equivalents.
Audiolingual Method Inductive acquisition through drilling and performing dialogues scripted in L2.	*Community Language Learning* The teacher invites students to say something on a topic that interests them in their L1 and then translates it into the L2 and asks the students to repeat it. In this way, a dialogue in the L2 is built up through translation.
Total Physical Response Vocabulary taught inductively by requiring learners to demonstrate comprehension of commands by performing actions.	*Bilingual methods* The teacher uses the L1 to support students' repetition of L2 dialogues until they are able to perform them in the L2 without it. The students are only allowed to use the L2.
Situational Language Teaching Vocabulary controlled through reference to frequency counts and taught inductively through situational exercises.	*Translanguaging Approach* The learner receives information in one language receptively and is then asked to produce it in the other language (i.e., input and output are conducted in different languages). It is an approach used for students who are already bilingual.
Silent way The teacher is totally silent and elicits production in the L2 form from learners by means of various artefacts (e.g., Cuisenaire Rods and colour-coded pronunciation charts). No recourse to the L1 is allowed.	*Two-way Immersion Programmes* Students are taught the same subject content in two languages (i.e., their home language and the L2), which are kept separate. These programmes aim at supporting the students' L1 while developing the L2.
Natural approach Vocabulary learned incidentally through comprehensible L2 input supplied by the teacher and teaching materials.	
Content-based instruction Vocabulary is learned inductively while learner focus is on the acquisition of knowledge of content areas such as mathematics, science, or geography.	

At the institutional level, this preference is also evident. For example, private language schools often advertise maximal use of the L2 as a method of attracting students. NOVA, a private English education firm in Japan, markets their service as being "in the target language only in order to be fully immersive" (NOVA, n.d.). At the secondary level, there are also reports of institutional support for immersive L2 environments. For instance, in the Hong Kong context, Chung (2018), quoting a teacher, reports a situation in which "when the principal...hears you speak in Chinese, he will send you an email reminding you that it's important to speak in English" (p. 505). In addition to such institutional pressure, Glick (2003) reports that some schools prefer teaching staff who do not speak the learners' L1 to ensure maximal L2 classroom use. Lastly, it is also common for tertiary institutions to prescribe English use in the L2 classroom (Berger, 2011). For example, the student handbook of one such institution states that "your English language program classes are 100% English" (Ono & Jordan, 2018, p. 1), while another reports that "as a rule, class activities will be conducted in English" (EEC, 2018, p. 1).

However, while it may be the case that policy makers often expect exclusive use of the target language, it is also the case that practitioners with knowledge of the learner L1 often do not heed such expectations (Edstrom, 2006; Song & Andrews, 2009). Research has shown, for example, that some teachers employ the L1 for tasks such as clarifying teaching points, classroom management, and checking comprehension (Edstrom, 2006). Figures for the percentage of L1 used in L2 classes vary greatly between contexts (27% in the Iranian context [Samar & Moradkhani, 2014], 32.1% of university-level EFL classrooms [Duff & Polio, 1990], and as much as 80% in the tertiary context in China [Cai, 2011]). In addition, it seems that many practitioners underestimate their use of L1 (Song & Andrews, 2009), so following prescribed percentages of L1 use (e.g., Atkinson, 1989; Kalivoda, 1990), is almost impossible in practice. Rather, considering the stigma surrounding L1 use in TESOL (Hall & Cook, 2012), it is interesting to consider how such use is perceived by the practitioners themselves.

Even if practitioners do utilise the L1, which is commonly found (Fortune, 2012), they may feel guilty about this usage (Barker, 2003; Cook, 2001; Harbord, 1992), portraying it negatively (Edstrom, 2006) with terms such as "unfortunate" and "regrettable" (Macaro, 2006, p. 68). This may stem not only from national and institutional policy (see above), but also from the paucity of coverage the L1 receives in teacher training programmes (Atkinson, 1989) and teaching guides (R. Ellis & Shintani, 2013). For example, R. Ellis and Shintani (2013) report "the overriding assumption in the published teacher guides...is that language teaching should be entirely - or almost entirely - L2-based" (p. 228). Thus, although it is clear that many teachers utilise the L1 in the classroom (Song & Andrews, 2009), it is also clear that the misalignment of such use with stakeholder policy and internationally published materials leads to self-admonishment.

L1 Use in Vocabulary Learning

Although excessive L1 use can limit opportunities for target language input and use (Hall & Cooks, 2012), research has shown that judicious use of the L1 can positively impact vocabulary learning. One strand of research in this area has considered the effect of codeswitching (CS) on lexical acquisition. For example, Tian and Macaro (2012) investigated the influence of CS on the acquisition of English vocabulary in the Chinese context. Participants (n=117), tertiary-level Chinese learners of English, were presented with target items disambiguated through either L1 (codeswitches) or L2 explanation. The study took place over a number of sessions, each of which was immediately followed by an immediate test. Results showed that teacher codeswitching led to more learning than L2 explanation.

This finding is supported by Fatemi and Barani (2014), who also investigated the effect of codeswitching on instructed lexical acquisition. Their Iranian participants (n=60) studied the target language over the course of a semester via L2 description or codeswitching. The results showed a significant difference between the two groups, with those who had recourse to the L1 outperforming those who had engaged with the target items through English. However, the lack of control, as well as methodological and analytical underreporting make it difficult to interpret meaningfully the findings of the latter study.

Lee and Macaro (2013) also reported an advantage for CS over L2 meaning disambiguation. In addition to investigating the impact of codeswitching on foreign language lexical acquisition, they also considered learner age as a covariate. Young beginner learners and more proficient adult Korean learners of English engaged with target items through either the L1 or the L2. Knowledge of the target items was measured immediately following instruction and after a three-week delay. Instruments included a meaning recall tool based on the Vocabulary Knowledge Scale (VKS; Paribakht & Wesche, 1997). Comparison of the immediate test with a pretest showed that the CS condition led to significantly better gains than the L2 condition for both age groups. Furthermore, this advantage was generally maintained for the delayed test, with CS outperforming L2 explanation. Consideration of the effect sizes associated with these various results showed that the young inexperienced learners benefitted to a greater extent from the provision of L1 than the older, more proficient learner group. Similar findings are reported by Lee and Levine (2018), who investigated the influence of code of meaning disambiguation and learner proficiency on listening comprehension and vocabulary learning. They found that CS led to improved vocabulary learning. Furthermore, this effect was found to be particularly beneficial for intermediate learners. Thus, these studies suggest not only that CS is an effective strategy for ad hoc meaning disambiguation in linguistically homogenous classes, but also that, contrary to TESOL lore, exclusive L2 use from the earliest stages of learning may well be less effective than utilisation of learner L1.

Codeswitching studies, however, do not typically employ the degree of experimental control common to investigations of vocabulary learning. For example, variables such as time on task, frequency of exposure, lexical frequency of the defining language, and saliency of phonological production are typically not controlled (or at least not commonly reported). Furthermore, studies that have included a delayed test have not employed a consistent retention interval for all target items (cf. Lee & Macaro, 2013). For instance, Tian and Macaro (2012) introduced their target items in six sessions each separated by seven days. Two weeks after the last session, a delayed test was administered, which measured knowledge of all target items; however, as the items were not recycled in each session, the retention interval was asymmetrical as items studied earlier had a longer retention interval than items studied in later sessions. Thus, we currently do not have a clear picture of the long-term benefit of codeswitching as a method of meaning disambiguation.

Another research strand in this area relates to dictionary use. Studies have shown that learners prefer to use bilingual dictionaries rather than their monolingual equivalents (Atkins & Varantola 1998; Bensoussan, Sim, & Weiss 1984; Carter & McCarthy 1988; Lew 2004; Piotrowski 1989; Schmitt 1997). However, English language teachers typically prefer learners to use a monolingual dictionary (Hunt 2009; Thompson 1987). In fact, bilingual dictionaries are often perceived as appropriate for only the earliest stages of language learning, while TESOL orthodoxy contends that as a learner’s proficiency increases, progressively greater use of monolingual reference tools be made (Tomaszczyk, 1983).

Studies have shown, however, that both bilingual dictionaries (Hulstijn, Hollander, & Greidanus, 1996; Knight, 1994; Luppescu & Day, 1993; Watanabe, 1997) and monolingual dictionaries (Hayati & Fattahzadeh, 2006) can lead to lexical acquisition. Yet, the extent to which this occurs seems to be contingent on how skilled individual learners are at using these reference tools (Laufer, 1994).

The effect of gloss language on vocabulary learning has also been considered. Gablasova (2015) investigated the relationship between the language in which technical vocabulary was encountered and the semantic completeness of the meaning that was learned. Participants saw target lexis in either an L1 (Slovak) or L2 (English) text and subsequently completed an oral meaning recall instrument. She found that incidental learning occurred in both conditions, but the meanings provided after the L1 input were more complete and more precise than those given after the L2 input. Jacobs, Dufon, and Fong (1994), investigating anglophone learners of Spanish, also found that both L1 and L2 glosses led to learning at the written form recall level, compared to a control. Learner proficiency seems to be a key covariate, moderating the effectiveness of the different codes. For example, Miyasako (2002) found that L2 glosses were associated with greater gains with more proficient learners, while L1 glosses were more effective for learners of lower proficiency.

Studies have also shown that the L1 can be an effective vehicle for communicating the meaning of target items during intentional learning activities (Grace, 1998; Lotto & de Groot, 1998; Prince, 1996; Ramachandran & Rahim, 2004). In particular, three studies are pertinent to the research presented in this chapter, comparing different codes of meaning presentation on the extent of learning. These have reported an advantage for the use of L1.

Lado, Baldwin, and Lobo (1967) investigated the effect of multiple methods of meaning and item presentation on learning to the level of meaning recall and meaning recognition. In a series of small experiments involving twenty to thirty participants each, they manipulated variables such as the manner in which the target item was presented (i.e., the written or the spoken form), the manner in which the meaning was conveyed (i.e., L1 equivalent or L2 delineation), and the manner in which the meaning was presented (i.e., written or spoken, with or without pictorial support). Items were studied in blocks of 100 and presented using slides and/or recorded speech depending on the particular criterion of the experiment. Knowledge of the target language was assessed with a meaning production task and a multiple-choice meaning recognition task. The results showed that learners could acquire a large number of items receptively in a short space of time: one experiment led to the acquisition of ninety-five items in forty-five minutes. A further finding was that L1 meaning presentation consistently produced greater learning gains than L2 delineation.

Mishima (1967) investigated the effect of five modalities on the learning of Japanese by native English speakers. Her participants (n = 50) studied 100 low-frequency foreign language items presented with L1 definitions, L1 translations, L2 definitions, L2 synonyms, or in a defining context. The items were divided into the various conditions considering intralexical factors such as PoS and number of syllables. The participants were divided into three groups: beginner, intermediate, and advanced, to determine the possible interaction of proficiency with learning condition. The target form was initially presented in the top half of a slide, with the meaning following below. Participants were given five seconds to learn each pair in the L1 and L2 synonym conditions, ten seconds in the L2 definition condition, and thirty seconds in the context-inference condition. Items were studied twice over the course of one session after which a meaning recall instrument was administered and relative gains calculated. Results showed that for the beginner and intermediate learners the L1 conditions were superior to the L2 treatments. Thus, using an L1 equivalent to disambiguate target item meaning led to greater relative learning than utilising L2 definitions or L2 synonyms. This general trend was also found with the advanced group, although a t-test revealed no significant difference between the L1 equivalents and L2 definitions. Mishima (1967) explained her findings with reference to the processing capacity required to decode L2 definitions for learners of low-medium ability. Once learners develop fluent language processing of the definitional vocabulary, the difference between the conditions was less apparent. However, it is important to note that even at advanced levels of proficiency, there was no learning advantage for using L2 to disambiguate word meaning. Mishima's (1967) findings need to be considered in light of the small number of participants in each group (11, 25, and 14 respectively), the lack of experimental control (e.g., meaning was read to beginner learners, while presented visually to others), and the lack of environmental validity (i.e., the length of presentation was controlled by the researcher, meaning that the retrieval frequency of the items may have been unequal; see Section 2.1.4). Moreover, and of particular relevance to this thesis, as a delayed test was not administered, it was not possible to determine the effect of the learning conditions on the pattern of forgetting.

Laufer and Shmueli (1997) investigated the impact of task and language of presentation on learning and retention of foreign language vocabulary. As such, it is, to the best of my knowledge, the only study that has actually compared different codes of meaning presentation while considering learning gains *and* retention. Five groups of L1 Hebrew high-school students studied 20 target items in one of five conditions: list, sentence, text, elaborated text, and control. The list group studied the target items using word pairs (time on task ten minutes). The sentence condition involved word pairs and an example sentence (time on task ten minutes). The text condition involved target items embedded in a text with glosses (time on task fifty-five minutes), and in the elaborated text condition, the text was adapted so that synonyms or appositive explanation supported target item disambiguation (time on task fifty-five minutes). In the control group, the target items were presented without meaning disambiguation. In each condition, half the target items were presented in L1 and half in L2. In all conditions except the control, initial learning was followed by a cloze exercise in which the target items were provided. This consolidation activity was followed by a multiple-choice test, which was repeated five weeks later. The results showed that the list and sentence presentation conditions were more effective than the text and elaborated text conditions. Also, words that were presented in L1 were better learned and better retained than words presented in L2. The descriptive statistics also suggested that less decay occurred when words were presented in the list condition in L1 (Table 5.2). However, the role of the cloze task in increasing engagement with the target items and supporting previous learning is unclear. The cloze task offered participants an opportunity to strengthen knowledge, but crucially this was only of benefit to those items initially learned. Therefore, this methodology may have led to the development of lexical knowledge of items with the lightest learning burdens and potentially biased the posttest data.

In sum, methodologists disagree about the role of the L1 in instructed SLA; however, researchers have shown that teachers do utilise the common language to disambiguate various aspects of vocabulary knowledge. Research has generally found this to be an effective pedagogical strategy, but it is currently unclear what effect the code of presentation has on the learning burden and the subsequent decay of learned lexical knowledge.

Table 5.2

The Decay of Target Items by Condition and Presentation Code (Max Score = 10) (Laufer & Shmueli, 1997)

		T1		T2	
Condition	Language	*M*	*SD*	*M*	*SD*
List	L1	8.26	2.13	8.21	1.69
	L2	6.68	2.08	6.84	2.14
Sentence	L1	9.12	1.15	8.23	1.43
	L2	7.94	1.37	6.84	1.83
Text	L1	7.20	2.33	6.60	2.84
	L2	6.65	2.11	5.45	2.42
Elaborated text	L1	8.12	1.81	7.05	2.68
	L2	7.20	1.55	5.48	2.11

Theoretical Underpinnings

The Types of Processing – Resource Allocation (TOPRA) model (Barcroft, 2015) is relevant to the use of different codes of meaning-presentation. This model recognises that, while learning vocabulary, a learner's limited processing capacity can be allocated to the acquisition of word form, word meaning, and form-meaning mapping. Barcroft (2015) suggests that greater focus on one of these may facilitate learning of that area, but that this occurs at the expense of the other two areas. For example, extensive attention to word meaning may result in enhanced word meaning knowledge but will impair the development of word-form knowledge and form-meaning mapping. This model is of significance as the choice of L1/L2 meaning presentation code will likely influence the type of processing required of a learner. An L2 definition is generally longer and more syntactically complex than an L1 equivalent. For example, the average length of the L1 equivalents used in the study presented in this chapter is 3.00 logographs, and the majority are single nouns or verbs. In contrast, the average definition length is 8.30 words, and these are often complex noun phrases or extended verb phrases (e.g., containing adjectival/adverbial prepositional phrases). Moreover, L2 items are likely to pose a greater processing challenge to the majority of learners, especially those with lower L2 proficiency (Mishima, 1967).

Applying the TOPRA model, it can be suggested that the use of L1 equivalents would result in comparatively little finite processing capacity being spent on semantic processing, allowing more to be used for form learning. In contrast, the use of the L2 code is likely to be more demanding as semantic processing will require a greater share of a learner's cognitive resources. Therefore, this model predicts that L2 definitional use will result in a heavier learning burden than the use of L1 equivalents: items presented with L1 equivalents will require fewer exposures/less time to be learned than those presented with L2 definitions because comparatively more processing resources can be spent on the task of form learning.

However, while the TOPRA model can be used to predict the effect of code of meaning presentation on learning burden, it cannot be used to determine the influence of this variable on decay. This is because learning burden is operationalised in this thesis according to the frequency-based criterion. Put in the terminology of the TOPRA model, successive exposures to meaning and form will, presumably, allow a learner, who initially devoted cognitive resources to semantic processing at the detriment of form learning, to shift processing distribution towards word form. Thus, all items will likely have received sufficient form-learning attention to reach the task criterion upon successful completion of the learning procedure. Therefore, the TOPRA model applied to the methodology employed in this thesis predicts that items disambiguated with L1 equivalents will be acquired more quickly than those explained by L2 definitions. However, it is currently unclear how this can be applied to the maintenance of that accrued knowledge.

5.2.2 The Effect of Perceived Usefulness

Study 1 employed low-frequency items. These items were also utilised in the investigation reported in this chapter. Using low-frequency items was a necessary selection criterion to exclude words familiar to the participants prior to learning. However, the use of such items may have impacted learner attitudes towards learning. That is, if learners perceived the target items to be of little use, they may have made limited effort to learn them. One goal of the present study was to determine the effect of learner perception of item usefulness on learning burden and decay.

Theoretical explanations of foreign language vocabulary learning suggest that items perceived to be more useful will be comparatively better learned and retained. The Levels of Processing Hypothesis (Craik & Lockhart, 1972) predicts a positive relationship between processing depth and memory trace persistence. Craik and Lockhart (1972) explain that "memory trace persistence is a function of depth of analysis, with deeper levels of analysis associated with more elaborate, longer lasting, and stronger traces" (p. 675). It might be the case that the perceived usefulness of lexical items is associated with levels of processing. This may be the case because deeper levels of engagement are related with semantic processing, which is likely to result from a learner determining that a particular lexical item is useful to him/her; that it fills a gap in a learner's L2 lexicon. Similarly, the Involvement Load Hypothesis (ILH), Hulstijn and Laufer's (2001) attempt at operationalising the Levels of

Processing Hypothesis, maintains that usefulness is positively correlated with learning. The authors suggest an involvement construct consisting of *need*, *search*, and *evaluation*. The first component, *need*, most relevant to the present study, is described as follows:

> The need component is the motivational, noncognitive dimension of involvement. Two degrees of prominence were suggested for need: moderate and strong. Need is moderate when it is imposed by an external agent. An example is the need to use a word in a sentence that the teacher has asked for. Need is strong when it is intrinsically motivated, that is, self-imposed by the learners, for instance, by the decision to look up a word in an L1–L2 dictionary when writing a composition. (Hulstijn & Laufer, 2001, p. 543)

Thus, the ILH predicts that lexical items required or chosen for study will be better learned than items judged to be unimportant. Furthermore, the distinction between the degree of need is determined by the agent selecting the items. Thus, the ILH conceptualises *need* as a task-induced variable. Studies that have adopted such a definition have produced mixed findings. Those that have manipulated relevance (used synonymously with *need* in the literature; see Peters, 2007) through a teaching exercise have generally found that words that need to be understood to facilitate accurate responses to comprehension questions are more frequently looked up in a dictionary (Hulstijn, 1993; Laufer & Levitzky-Aviad, 2006). This increased engagement has been shown to result in more learning and retention. For example, Peters (2007), investigating L1 Dutch learners of German, presented participants with comprehension questions after reading a passage containing target vocabulary. These questions necessitated engagement with some target vocabulary. Such target items were categorised as *relevant*, while those not related to the comprehension questions were classed as not relevant to the task. Digital logs of look up behaviour showed that word relevance was associated with more frequent dictionary use. Furthermore, a series of tests measuring the form-meaning link showed that relevance led to more learning and retention. Similarly, Peters, Hulstijn, Sercu, and Lutjeharms (2009) found that task-induced word relevance affected look-up behaviour and promoted word retention at the level of meaning recall.

However, research designs that have not employed comprehension questions to manipulate engagement with target vocabulary have found no effect for word relevance. For example, Peters and Webb (2018) considered the effect of word relevance on the incidental acquisition of vocabulary from viewing. They also operationalised relevance as a task-induced variable; rating lexical items in a documentary as more or less necessary for comprehension of that documentary. Three raters familiar with TESOL scored items against a seven-point scale. The mean of these individual ratings was then included in a repeated measures logistic regression to determine the effect of word relevance on learning. Analysis showed that relevance did not significantly impact the learning of items: items rated as highly relevant were as likely to be learned as those rated less relevant.

However, it might be argued that the notion of *need* in the ILH, and its operationalisation *relevance,* is undertheorized. This is the case because within each dimension of the *need* construct there is likely to be considerable variance in terms of perceived usefulness. Items prescribed for learning by a teacher will not be considered equally useful for task completion and will differentially align to the longer-term learning goals of the students. Essentially, although the ILH conceptualises *need* as a purely task-related variable, it may be more accurate to label it a by-item by-learner task-related variable. For example, even if Learners X and Y study vocabulary using the same task, and thus are faced with the same involvement load, the extent to which the individual target items resonate with each learner or are considered to be important to task completion is likely to differ. Thus, while perhaps useful as a rather blunt pedagogical tool, *a priori* analysis of task-induced involvement that does not consider the individual learner, the individual items, and the interaction of learner and item is likely to misconstrue the involvement load as it is actually experienced by the learners.

Similarly, Peters and Webb's (2018) understanding of the relevance construct is problematic. First, as *comprehension* varies with the cultural and linguistic background of the individual learner (i.e., learners can have a different understanding of the same text) (see Ertmer & Newby, 1993), the use of a group scaling procedure to determine word relevance may have resulted in ratings not representative of individual learner perception. Second, in the norming portion of the study, as the rating procedure took place after the entire documentary was viewed, the raters could consider the relevance of each item retrospectively in light of successful global comprehension. However, participants needed to construct an understanding of the documentary online (i.e., while viewing) and thus may not have been immediately aware of the contribution each item made to overall meaning. That is, item relevance may only have become clear upon successful global comprehension. In sum, previous research and theory conceptualises relevance/need as a task-induced variable; however, this ignores the important role of the learner.

To the best of my knowledge, no research to date has considered the influence of word relevance as a learner-related variable. To highlight the different construct underlying this understanding of *need*/*relevance*, this study employed the term *usefulness*. Thus, one of the goals of this study was to determine the role of perceived usefulness on the learning burden and retention of L2 vocabulary. The perceived usefulness of a target item may affect the level of learner engagement in the learning process. *Engagement* is a heuristic term that covers variables such as attention, manipulation, time on task, and frequency of exposure (Schmitt, 2008). It has been argued that engagement is positively correlated with vocabulary learning (Schmitt, 2008) and retention (Schmitt, 2010). Four years before the use of the term by Schmitt (2008), it was described by Fredricks, Blumenfeld, and Paris (2004). Their analysis of the construct into three facets is particularly beneficial to the discussion of perceived usefulness in this chapter. They suggest that engagement consists of *behaviourial* (the actions undertaken by the learners), *emotional* (the level of enjoyment felt), and *cognitive* (the effort expended) aspects. The perceived usefulness of target items is likely to impact the latter two elements of this framework: items perceived to be more useful to a learner may be processed with greater enjoyment and with more effort. This deeper engagement may increase the depth of the memory trace associated with such items (Craik & Lockhart, 1972), expediting acquisition (Schmitt, 2008) and improving retention (Schmitt, 2010). Therefore, theoretical descriptions of the vocabulary acquisition process suggest that items perceived to be useful to a learner will be acquired more quickly and retained more fully.

5.3 The Study

In general, studies have reported an advantage for L1 over L2 meaning presentation in intentional learning. However, little research consideration has been given to the decay of knowledge learned in L1 or L2 conditions. Additionally, the methodologies employed in some studies to date mean that we do not have a clear understanding of the effect of meaning presentation code on learning burden, supporting the need for further investigation. The study reported in this chapter targeted these lacunae by considering the impact of meaning presentation code on the learning burden and decay of L2 lexical knowledge at the level of form recall and recognition. Furthermore, minimal research attention has been given to the effect of usefulness as a by-learner by-item variable on lexical acquisition. A secondary aim was therefore to disambiguate the impact of perceived usefulness on the learning burden and decay of vocabulary knowledge. Lastly, as with Study 1, this investigation sought to understand the relationship between learning burden and decay, irrespective of the effect of the target variables.

Following the procedure used in Study 1, learning burden was operationalised as the number of exposures to criterion. Additionally, due to a change in the data analysis procedure (i.e., Study 1 used a Microsoft Excel macro to calculate exposure frequency while this study used a SQL join clause to combine tables in a relational database), this study was also able to consider the duration of exposure. This provided increased detail on the learning process and addressed one of the limitations of Study 1 (i.e., the number of exposures to criterion did not provide detail on learner behaviour during each exposure). Thus, in this study learning burden was operationalised as both the frequency and time of exposure.

The study adopted the following research questions:

1. To what extent do meaning presentation code and perceived target item usefulness affect the learning burden of L2 vocabulary knowledge?
2. To what extent do meaning presentation code and perceived target item usefulness affect the decay of L2 vocabulary knowledge?

3. To what extent does learning burden affect the decay process (irrespective of the target variables)?

To answer these questions, 76 learners studied 32 target L2 items using electronic flashcard software. Word meaning was disambiguated in two conditions, L1 equivalent and L2 definition, with 39 and 37 participants per group respectively. This was followed by the immediate test measuring form recall and recognition, and a subsequent delayed test administered after a one-week interval. After completing the delayed test, participants undertook a final learning protocol and evaluated the usefulness of each target item. This latter sequence was adopted both so that learners did not assign usefulness ratings based solely on the extent of target item retention, and also so that the act of assigning usefulness ratings did not bias the retention of the target items. Decay was determined by comparing attainment and retention scores.

5.4 Methodology

5.4.1 Participants

Given the manipulation of meaning presentation code, it was necessary for the participants to share the same L1. In the UK academic context, L1 Chinese learners make up the largest single group of international students[2] (HESA, 2017). As this thesis hopes to draw impactful pedagogical implications for the UK HE context, Chinese L1 students were chosen as the target population of this study. However, to prevent contact with the target items outside the parameters of the experiment, this study was conducted with learners of English not resident in the L2 context. All learners were enrolled on postgraduate degrees in a

[2] The Higher Education Statistics Agency report that more than one quarter of international students in UK universities in the academic year 2016-2017 reported to come from China or Hong Kong.

Transnational Educational (TNE) context at the time of participation: they were students at two UK universities, but studying in China. Their degree programme included instruction of their disciplinary content in both L1 and L2, and additional EAP language classes taught in the L2. In total, 82 learners took part in the study (59 female, 23 male; mean age = 22.26 years (*SD* = 2.25), range = 9 years). However, six participants were removed from the sample as they either did not follow instructions, or their electronic data were corrupted. Thus, the final sample was 76 learners. These learners had studied English for an average of 13.36 years (*SD* = 2.97 years), had never lived in an English L1 environment, and their average first contact with English was at 8 years old (*SD* = 2.42 years).

Learner proficiency was initially determined by consideration of language proficiency test score (e.g., IELTS, TOEFL). However, some learners had either not taken such a test or not taken one for some time, and thus it was not possible to utilise these third-party assessment results. Research has shown that vocabulary size can be used to predict language proficiency (Laufer & Aviad–Levitzky, 2017). Thus, the VST was used to determine the relative proficiency of the learners. The results of the VST showed the mean score was 63.46 words (*SD* = 10.04). For the study, the participants were split into two groups, one group (n = 39) saw target item meaning via L1 equivalents and the other (n = 37) via L2 definitions. A t-test was conducted to ensure that the proficiency of the L1 and L2 groups was comparable. No statistically significant difference was found (t = -1.23, p = .22). Also, the learners completed a self-report measure of proficiency in the four skills. On a ten-point scale (one = extremely poor, almost no knowledge; ten = extremely good, almost native like) the mean scores were as follows: reading 6.22 (*SD* = 1.14), writing 5.20 (*SD* = 1.12), listening 5.41 (*SD* = 1.47), and speaking 5.04 (*SD* = 1.33). The study was run in accordance with the ethical approval (Appendix 9) and participants completed a consent form prior to taking part in the study (Appendix 3)

5.4.2 Instruments

Target Items

The items employed were those used for Study 1 (see Section 4.3.2). Therefore, items (N = 32) were selected to manipulate word length and word class, and to control for potentially confounding factors such as word frequency, concreteness, morphological transparency, referent familiarity, and orthographic neighbourhood size. To determine whether or not the target items were likely to be known to the target population, twenty-five learners similar in proficiency, L1, and from the same learning context as the target population completed a checklist test on Microsoft Excel containing the target items (thirty-two target items and eight items included to measure a test effect), twenty-four plausible nonwords taken from Meara (1992), and ninety items randomly sampled from the most frequent 5,000 words in English (Nation, 2012) (Appendix 10). The latter were included to maintain learner motivation during completion of the instrument, while the nonwords were used to control for a guessing effect. Data of participants who reported knowledge of three or more nonwords were deleted from further analysis. This led to six deletions and thus a final pilot cohort of nineteen. The results, presented in Table 5.3, showed that although some items were known to some learners, in general there was minimal knowledge of the target items and thus no changes were made. This initial *a priori* item piloting was supported by *a posteriori* control for item knowledge using both the frequency of exposure during the learning process and a final survey targeting perceived prior knowledge of the target items. The data for the *a posteriori* control is discussed in Section 5.4.4.

Table 5.3

Results of the Item Piloting in Study 2. Figures Indicate the Percentage of Participants who Reported Knowing an Item.

Item	Checklist	Item	Checklist
nab	5.26%	mackerel	0.00%
asterisk	0.00%	truncate	0.00%
venerate	15.79%	cinder	5.26%
spinster	0.00%	marten	0.00%
expedite	5.26%	irk	0.00%
archipelago	0.00%	harpsichord	0.00%
bop	0.00%	decant	5.26%
keg	0.00%	cauldron	5.26%
regurgitate	0.00%	tenacity	5.26%
prance	0.00%	clique	5.26%
imp	0.00%	bib	5.26%
camaraderie	0.00%	scintillate	0.00%
pontificate	5.26%	kip	5.26%
zealot	5.26%	heckle	5.26%
contraption	0.00%	frolic	0.00%
tic	0.00%	antagonise	10.53%
zap	5.26%	orb	0.00%
matriculate	5.26%	condominium	0.00%
jangle	15.79%	ambulate	15.79%
conflate	5.26%	voyeur	0.00%

The L2 definitions were based on those used in Study 1 but amended so that the PoS of each target item was explicitly stated. This involved the inclusion of *noun* or *verb* in parentheses before the target definition. This alteration was made to investigate whether the non-significant result for PoS in the first study could be explained by the lack of word class saliency in the item presentation method. As this study sought to determine the potentially differential effect of L1 and L2 meaning presentation on learning and decay, L1 equivalents were also employed. As with the L2 condition, PoS was indicated by the inclusion of 名词 (noun) or 动词 (verb) in parentheses before the L1 equivalent.

The method of L1 equivalent selection was as follows. The target items were initially translated into Mandarin Chinese by a native speaker highly proficient in English. These translations were then checked by a second native speaker of Mandarin Chinese and any differences between the two translations were discussed. The translators had access to English-Chinese dictionaries throughout this process. Finally, all equivalents were piloted with participants similar in language and educational background to the participants of the main study to confirm that the L1 equivalents would likely be known to the population. There were four items for which the translators felt additional disambiguating information was necessary. This was included in parentheses after the L1 equivalent (see Appendix 11 for target items, L1 equivalents, and L2 definitions).

Flashcards and Decks

Flashcards were set up productively, as with Study 1. Learners saw the meaning of a target item and were tasked with producing the target form. After responding to a target meaning or inputting no response, learners were presented with feedback on their production and the correct form. Based on this feedback, they rated their production using one of three options: *Easy* if an item was known prior to the study; *Good* for a correct response; *Again* for an incorrect response. Learners in the L1 condition saw target meanings presented via an L1 equivalent and learners in the L2 condition saw L2 definitions. Other than this difference, flashcard construction was identical. As with Study 1, the target items were organised into four decks of eight items, with one verb and noun of each word length

(3, 6, 8, and 11 letters) in each deck. The order in which the decks were presented to learners was fixed so that the L1 and L2 groups were exposed to the decks in a consistent manner.

Attainment and Retention Instruments

The measurement instrument was based on the tool developed for Study 1. Thus, it consisted of measures of form recall and form recognition. A detailed description of this instrument is given in Section 4.3.6. Of note is that, as with Study 1, the recall instrument was scored at two levels of sensitivity: strict (W) and lenient (W + (W-(WL*.33) (see Section 4.3.6).

The instrument differed from that employed in Study 1 in that the form recall and recognition stimulus was modified to reflect the meaning-presentation condition of the learner. Thus, learners who used L1 equivalents during the learning stage saw the same L1 equivalents on the testing instrument, while participants who used L2 definitions to learn the target vocabulary were presented with L2 definitions as stimuli. Additionally, the PoS tag included in the learning stage was employed in the measurement tool. Of note is that the distractors on the form recognition instrument were of the same word class as the key, so the impact of including this PoS information on any guessing is likely to have been minimal.

Background questionnaire

As with Study 1, a language background questionnaire was used to understand biographical information and L2 proficiency. This information was used in the statistical modelling of learning burden and lexical decay. The questionnaire from Study 1 was used.

Vocabulary Size Test

Vocabulary size was measured by the bilingual Chinese-English Vocabulary Size Test (Nation & Beglar, 2007). It has been argued that this lacks validation evidence (Schmitt et al., 2019) and does not include sufficient test items to accurately represent knowledge of the population of words it attempts to measure (Gyllstad et al., 2015); however, as vocabulary size correlates with general proficiency tests (Alderson, 2005) it is increasingly being used to show the relative proficiency level of a learning group (see Gyllstad et al., 2015). The Chinese-English Vocabulary Size Test is a four-option multiple-choice instrument. Test takers see an L2 item presented individually and in a non-defining sentence and are tasked with choosing the correct L1 equivalent. The authors developed the test by first determining the L2 forms associated with each distractor on the monolingual VST. These were then translated into Chinese to produce the bilingual version. Thus, the test is identical in format to the VST, but with the key and distractors presented via L1 lexical items. The test was administered via Microsoft Excel to allow for automatic score calculation and preclude the need for internet access (see Figure 5.1)

Figure 5.1

An Example of the Chinese-English Bilingual VST Used in Study 2

3	shoe: Where is your **shoe**?		
		a	父或母
		b	钱包
		c	钢笔
		d	鞋子

Survey of Prior Knowledge, Intersessional Exposure, and Perceived Target Item Usefulness

Prior knowledge of the target items, intersessional interaction, and perceived target item usefulness were assessed via a self-report questionnaire. The instrument presented each target item with its meaning according to the learning condition of the learner; thus, learners in the L1 condition saw the L1 equivalents (see Appendix 12), and those in the L2 group saw L2 definitions (see Appendix 13). For prior knowledge and intersessional exposure, participants checked items that they judged to be known before the learning procedure or to have been seen during the retention interval. Perceived usefulness was determined with a by-item six-point Likert scale measure, with one indicating *not useful at all* and six representing *very useful.* An even number of response options was chosen to prevent participants opting for the noncommittal middle option (see Wang, Hempton, Dugan, & Komives, 2008). The instrument was administered via Microsoft Excel.

5.4.3 Procedure

The study consisted of three sessions, lasting approximately one hour each. On the first day, the parameters of the study were explained to the learners, who then provided their written consent to participate. They completed the approved consent form (see Appendix 3) and the study was run in accordance with the ethics approval. Participants then studied the thirty-two target items using the electronic flashcard software *Anki,* following the same procedure outlined in Section 4.3.7. The target items were presented in four sets of eight items (a noun and a verb at each target word length), with presentation order randomised. As with Study 1, each learner interacted with each target item as often as necessary to reach the learning criterion, which was set to two consecutive accurate productions of the target form. Learners rated whether their production was correct or incorrect by clicking *Good* or *Again* respectively. Learners were instructed only to use a third rating option, *Easy*, for items known prior to beginning the study. This process was thoroughly explained to the learners prior to starting the experiment and learner behaviour was carefully monitored throughout. Learners interacted with the electronic flashcard software on their own laptops using USBs that were distributed before each session. These contained the learning tool and all measurement instruments. The first session concluded with administration of the language background questionnaire (Appendix 7).

The second session took place after an interval of one-day. It consisted of the relearning of all target items using the same learning procedure, the first administration of the measurement instrument, and administration of the VST. All measures were completed on computer, with participants using the same USB flash drive as session one (these were distributed immediately before each session and collected upon completion to prevent learners studying the target items outside the experimental sessions). Completion of this stage marked the last interaction with the target items before the delayed test. As with session one, the researcher monitored the students to ensure that no items were written down. One participant was found to be recording target items to study between sessions, and her data were excluded from analysis.

The final session was conducted one week after session two, representing a retention interval of seven days. This span was chosen to avoid the floor effect on the delayed test for productive knowledge found in Study 1. Additionally, studies that have considered lexical acquisition (e.g., Laufer & Girsai, 2008) and the effect of factors on learning burden and decay (e.g., de Groot, 2006) have employed a similar retention interval. This third session included the delayed test, a relearning protocol, the self-report measures of prior knowledge and intersession exposure, and the Likert scale measure of target item usefulness. Participants first completed the delayed test. This was the same instrument as was employed for the immediate test, but item order was varied to control for an order effect. They then completed the relearning protocol, which involved using the flashcard software once again to study the 32 target items. This was done to ensure that all target items were learned prior to completing the measure of perceived target item usefulness, so that judgements on that instrument represented perceived usefulness rather than reflecting target item knowledge. Once finished, participants completed the survey of prior knowledge, intersessional exposure, and perceived target item usefulness, after which all USBs were collected.

5.4.4 Analysis

Analysis began by calculating frequency and time of exposure. This was achieved by accessing the SQLite learner database file and performing a join between three tables: *revlog*, *cards*, and *notes*. As with Study 1, items that received only two exposures were removed from further analysis. Two exposures equated to a participant evaluating a target item as *Easy* (in the Anki verbiage) on both days of the learning procedure. Participants were told only to use this evaluation if an item was known to them prior to starting the learning procedure. Thus, two exposures was taken as a proxy for prior knowledge. Inspection of the frequency of exposure data showed that no items met the criterion for exclusion. Prior knowledge was also determined by a self-report questionnaire (Appendix 12 and Appendix 13); items which learners reported to know were removed from further analysis. In total 3.2% of data points were removed. Thus, the learning data only included those items unknown prior to the treatment as indicated by both the frequency of exposure data set and the self-report checklist.

Following initial trimming, winsorizing was conducted (see Field, 2018) with the frequency and time of exposure data sets to control for outliers (taken as $z \geq 3.29$) (see Field, 2018). This procedure involves substituting any outlier with the highest or lowest value that is not an outlier. Thus, frequency and time of exposure scores that exceeded 3.29 standard deviations from the mean were replaced with the numeric equivalent of $z = 3.29$. The effect of this procedure is evident from a comparison of the descriptive statistics prior to and after adjustment (see Table 5.4).

Table 5.4

The Effect of Winsorizing on the Frequency and Time of Exposure Datasets (Study 2)

	M	*SD*	Min	Max	Range
Frequency of exposure					
Before winsorizing	9.74	5.01	3	65	62
After winsorizing	9.42	3.84	3	26	23
Length of exposure (secs)					
Before winsorizing	110.30	75.52	15.69	744.28	728.59
After winsorizing	107.50	67.15	15.69	561.03	545.34

Potential learning from the testing instrument was measured by analysis of the eight items included to measure the test effect. These items were not studied in the initial two learning sessions but included on the immediate and delayed tests. Improvement between the two administrations of the testing instrument thus represented the confounding effect of the instrument on learner knowledge. The results indicated that minimal learning took place from the instrument. There was more learning at the level of written form recognition (*m* = 16.13% of items) than the level of form recall (strict: *m* = 1.16%; lenient: *m* = 1.56%); however, considering the measurement instrument employed a multiple-choice item format, this increase could stem from a guessing effect.

To determine the effect of code of meaning presentation on learning burden (Research Question 1) a series of linear mixed effects models were calculated using lme4 (Bates et al., 2015) in the statistical package *R* (R Core Team, 2016). Prior to modelling, t-tests were conducted at each level of measurement sensitivity to check the equity of the learning gains (i.e., the scores on the immediate test) in the two conditions. No statistically significant differences were found (strict recall – t (72) = 1.05, p = .30, two-tailed, L1 – 25.33 [5.61], L2 – 23.69 [7.81]; lenient recall – t (54.71) = 1.26, p = .21, L1 – 27.72 [4.60], L2 – 25.86 [7.61]; recognition – t (43.03) = 1.76, p = .09, L1 – 31.33 [1.38], L2 – 30.20 [3.58]).

Learning burden was operationalised as the frequency and time of exposure needed for a target item to reach criterion. Thus, the outcome variables for statistical modelling were these two indices of learning burden. Visual inspection of residual plots and boxplots showed that there were no obvious deviations from homoscedasticity or normality. Separate models were conducted for frequency and length of exposure at each condition (strict form recall, lenient form recall, and form recognition), with six models computed in total (two outcome variables x three conditions). Only items for which knowledge had been demonstrated on the immediate test at the relevant condition were included in the analysis. For example, analysis of the frequency of exposure data for the strict recall condition only considered items for which strict form recall knowledge had been demonstrated on the immediate test.

Model fitting began with the calculation of a core model in which frequency/time of exposure was predicted by meaning presentation code, perceived usefulness, and the interaction of these two fixed effects. Following this, a maximal model was calculated which contained all potential covariates: meaning presentation code, perceived item usefulness, vocabulary size, knowledge of an additional language, age, word length, and PoS. Additionally, the following interactions were considered: code*perceived usefulness, word length*PoS, code*word length, and code*PoS. By-item and by-participant random slopes for usefulness were employed. The inclusion of PoS and word length allowed for confirmation of the results from Study 1, while their potential relationship with meaning presentation code was investigated by including the interactions between the intralexical variables and code. The best-fitting model was determined by backward elimination. This involved removing one covariate from the model at a time and analysing its effect on model fit using likelihood ratio tests. Any covariate whose removal resulted in a non-significant X^2 statistic was excluded from subsequent iterations of the model. This process was followed until the most parsimonious model was determined. Separate models were calculated for frequency of exposure and time of exposure at each of the three levels of measurement sensitivity. This model-selection process was the same as that employed in Study 1 (example provided in Section 4.3.8). Details of the model selection procedure are given in Appendix 14.

The effect of the code of meaning presentation and perceived item usefulness (Research Question 2) and learning burden (Research Question 3) on the decay of target language knowledge was determined through computation of mixed-effect logistic regression models. Models were computed for both levels of form-recall sensitivity and for form recognition. Each analysis only considered items for which knowledge was displayed at the equivalent level on the immediate test. For example, modelling of retention data at the strict form-recall level was limited to those items that were accurately produced on the recall instrument of the immediate test. GLMERs were employed because the outcome variable, scores on the delayed test, was binomial: correct or incorrect. Prior to model computation, continuous variables were tested for linearity of the logit (see Field, 2018). The assumption of linearity was not met for some covariates which were omitted from the model selection procedure. The following covariates were included: meaning presentation code, perceived usefulness, the interaction between meaning presentation code and usefulness, PoS, word length, learner proficiency, the number of exposures, and the time of exposure. Once again, including PoS and length allowed for confirmation of the results of Study 1. Furthermore, including indices of learning burden (i.e., frequency and time of exposure) in the decay analysis allowed me to measure its role in the decay process (Research Question 3). The best-fitting model was once again determined by a backwards elimination procedure that utilised likelihood ratio tests. Details of the model-selection procedures are given in Appendix 15. The analysis employed the BOBYQA algorithm and iteration number were adjusted to 100,000 to avoid convergence failure (see Singmann, 2014).

5.5 Results

This section first introduces details of the models fitted to analyse the effect of meaning presentation code and perceived usefulness on learning burden (Research Question 1). It then considers the models fitted to delineate the effect of meaning presentation code, perceived usefulness (Research question 2), and learning burden (Research Question 3) on decay.

5.5.1 Learning Burden

Learning burden was operationalised as the frequency and time of exposure needed to learn a target item. Descriptive statistics for these two metrics are presented in Table 5.5 at the three levels of measurement sensitivity. Additionally, Table 5.5 presents the mean scores on the immediate test at each level of measurement.

The data showed that learners needed, on average, approximately nine exposures and between 97 and 106 seconds, depending on the level of measurement strictness, to learn the target items. More exposures and time were needed to learn items at the level of form recognition than lenient form recall, which in turn, required more exposures and time than strict form recall. This finding does not indicate that developing form recall knowledge was less burdensome than form recognition knowledge; rather, the greater average burden at the form recognition level was likely a function of the larger learning gains at that level. Form recognition is a weaker strength of mastery than form recall (Laufer & Goldstein, 2004), and recognition instruments are therefore more sensitive to partial knowledge than measures of form recall (Schmitt, 2010). The current study found more learning occurred at the recognition than the recall level, indicating that some items were only partially acquired. The increased mean frequency and length of exposure at the recognition level might indicate that items only partially acquired were also those that posed the heaviest burden during the period of learning. Therefore, knowledge of the most burdensome items was perhaps best captured by the form recognition instrument. Table 5.5 also indicates that there was considerable individual variance in the extent of burden experienced, as shown by the large standard deviations for frequency and time of exposure.

Table 5.5

Mean Scores for the Immediate Test (Score), Frequency of Exposure (Freq.), and Time of Exposure (Time) by Mode of Form Presentation. Standard Deviations are Presented in Brackets (Study 2)

	L1 condition			L2 condition			Total		
	Score	Freq.	Time	Score	Freq.	Time	Score	Freq.	Time
Strict form recall	25.33 (1.76)	8.88 (3.34)	93.84 (60.93)	22.41 (2.51)	8.89 (3.39)	102.52 (58.15)	23.91 (2.19)	8.88 (3.36)	97.80 (59.84)
Lenient form recall	27.72 (1.46)	9.06 (3.54)	97.18 (62.93)	24.46 (2.44)	9.05 (3.48)	106.14 (61.50)	26.13 (2.04)	9.06 (3.51)	101.26 (62.44)
Form recognition	31.33 (0.51)	9.29 (3.81)	100.04 (67.31)	28.57 (1.99)	9.47 (3.89)	113.84 (66.92)	29.99 (1.50)	9.37 (3.85)	106.44 (67.48)

Note. Maximum score = 32.

The effect of the target variables on the learning burden was investigated by statistical modelling. First, the results relevant to frequency of exposure are presented at the three levels of measurement sensitivity. Findings from the time of exposure modelling are then presented, also at the three levels of mastery considered.

Frequency of Exposure

Strict Form Recall

Table 5.6 presents the analysis of frequency of exposure data in the strict recall condition. The best-fitting model showed there was no statistically significant effect for the code of meaning presentation (t = -0.03, p = .66: L1 - 8.88 [3.34], L2 - 8.89 [3.39]) or the perceived usefulness of a target item (t = -1.51, p = .13). Significant effects were found for the intralexical variables word length (t = 6.57, p < .001) and PoS (t = 2.46, p = .02), indicating that longer words and verbs required more exposures to be learned. Additionally, a statistically significant effect for learner vocabulary size was found (*t* = -0.30, *p* = .03), which showed that participants with comparatively larger vocabulary sizes required fewer exposures to learn the target items.

Table 5.6

Fixed and Random Effects for Frequency of Exposure, Strict Form Recall Condition (Study 2)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	2.93	0.58	5.06	<.001***
Code of meaning	-0.03	0.06	-0.44	.66
Usefulness	-0.04	0.02	-1.51	.13
Code * Usefulness	0.04	0.03	1.37	.17
Vocabulary	-0.30	0.14	-2.17	.03*
Length	0.25	0.04	6.57	<.001***
PoS	0.09	0.04	2.46	.02*
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	.03	.18	
	Usefulness	<.01	.01	
Participant	Intercept	.01	.11	
	Usefulness	<.01	.03	

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

Lenient form recall

The results of the best-fitting MEM (see Table 5.7) showed there was no statistically significant effect for the target variable meaning presentation code (*t* = -0.52, *p* = .61; L1 = 9.06 [3.54], L2 = 9.05 [3.48]), nor for perceived target item usefulness (*t* = -1.42, *p* = .16).

Significant effects were found for word length (t = 6.56, p < 0.001) and PoS (t = 2.28, p = .03). This showed that participants required more exposures to learn longer words and verbs. Additionally, a statistically significant effect was found for learner vocabulary size (t = -2.34, p = .02); participants with comparatively large vocabularies needed fewer exposures to acquire the target language.

Table 5.7

Fixed and Random Effects for Frequency of Exposure, Lenient Form Recall (Study 2)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	3.05	0.59	5.19	<.001***
Code of meaning	-0.03	0.06	-0.52	.61
Usefulness	-0.03	0.02	-1.42	.16
Code * Usefulness	0.05	0.03	1.45	.15
Vocabulary	-0.33	0.14	-2.34	.02*
Length	0.25	0.04	6.56	<0.001***
PoS	0.08	0.04	2.28	0.03*

Random Effects			
Parameter		Variance	*SD*
Item	Intercept	.04	.19
	Usefulness	<.01	<.01
Participant	Intercept	.01	.12
	Usefulness	<.01	.04

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

Form Recognition

Finally, Table 5.8 presents analysis of the form recognition data. The best-fitting model indicated that the code of meaning presentation did not have a statistically significant effect (*t* = 0.59, *p* = .56; L1 = 9.29 [3.81], L2 = 9.47 [3.89]). No statistically significant effect was found for perceived item usefulness either (*t* = -0.98, *p* = .33). Statistically significant effects were found for the two intralexical factors, word length (*t* = 6.95, *p* < 0.001) and PoS (*t* = 2.33, *p* = 0.03), with longer words and verbs needing more exposures to be learned. The effect for learner vocabulary size was also found to be statistically significant (*t* = -2.35, *p* = 0.02); learners with comparatively larger vocabularies needed fewer exposures to learn the target vocabulary.

Table 5.8

Fixed and Random Effects for Frequency of Exposure, Form Recognition Condition (Study 2)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	3.13	0.64	4.89	<.001***
Code of meaning	-0.04	0.07	-0.59	.56
Usefulness	-0.02	0.02	-0.98	.33
Code * Usefulness	0.05	0.03	1.51	.13
Vocabulary	-0.36	0.15	-2.35	.02*
Length	0.08	0.04	6.95	<.001***
PoS	0.09	0.03	2.33	.03*
Code*Length	-0.04	0.02	-1.71	.09
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	.04	.21	
	Usefulness	<.01	.02	
Participant	Intercept	.01	.11	
	Usefulness	<.01	.03	

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

Time of Exposure

This section now considers analysis of time of exposure at the three levels of measurement sensitivity.

Strict form recall

The best-fitting model (see Table 5.9) indicated a statistically significant effect for the code of meaning presentation (t = 6.31, p < 0.001), with words presented in the L1 condition requiring an average of 93.84 seconds (SD = 60.93 seconds) to learn compared to a mean of 102.52 seconds (SD = 58.15 seconds) in the L2 condition. No statistically significant effect was found for perceived item usefulness (t = -0.84, p = .40). Statistically significant effects were found for the participant variables vocabulary size (t = -0.88, p < 0.001) and age (t = 0.92, p < .001): learners with larger vocabularies and younger learners needed less time to learn the target items. In addition, a significant effect was found for word length (t = 11.13, p < .001), with longer items requiring more time to be learned. Significant effects were also found for the interactions between PoS and code of meaning presentation (t = -2.76, p = .01). There was also a significant interaction between word length and code. The first interaction showed that the noun advantage was more pronounced in the L1 condition and the second interaction indicated that the L1 advantage was greater for shorter words than it was for longer words.

Table 5.9

Fixed and Random Effects for Length of Exposure, Strict Form Recall Condition (Study 2)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	10.52	0.95	11.06	<.001***
Code of meaning	0.59	0.09	6.31	<.001***
Usefulness	-0.03	0.03	-0.84	.40
Code * Usefulness	0.03	0.04	0.62	.54
Vocabulary	-0.88	0.15	-5.95	<.001***
Age	.92	0.24	3.86	<.001***
Length	0.83	0.08	11.13	<.001***
PoS	0.11	0.07	1.57	.13
Code * Length	-0.22	0.03	-6.57	<.001***
Code * PoS	-0.09	0.03	-2.76	.01**
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	<.01	.20	
	Usefulness	<.01	<.01	
Participant	Intercept	<.01	.18	
	Usefulness	<.01	.34	

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

Lenient form recall

Next, the results relating to the total time of exposure needed for items to be learned at the level of lenient form recall are considered (Table 5.10). The best-fitting model indicated that there was a statistically significant effect for code of meaning presentation ($t = 6.47$, $p < .001$), with items presented in the L1 condition needing, on average, 97.18

seconds to be learned (*SD* = 62.93), but an average of 106.14 seconds (*SD* = 61.50) in the L2 condition. No statistically significant effect was found for the perceived usefulness of the target items (t = -0.86, p = .39). Additional significant effects were found for the intralexical variable word length (t = 11.39, p < .001), showing that shorter words needed less time to be learned than longer items. No effect was found for PoS (t = 1.43, p = .16). The results also showed significant interactions between code and length (t = -7,12, p <.001), and code and PoS (t = -2.73, p = .01). The former interaction showed that the effect for code was particularly pronounced with shorter words, suggesting that the added burden of longer items nullified the L1 advantage. With regard to the latter interaction, descriptive statistics showed that the noun advantage was only prevalent in the L1 condition. Lastly, there were significant effects for learner vocabulary size (t = -6.28, p < 0.001) and age (t = 3.80, p < .001). Again, students with more vocabulary knowledge and younger participants learned the target items more quickly.

Form Recognition

The best-fitting model (Table 5.11) showed a statistically significant effect for meaning presentation code (t = 6.76, p < 0.001). Items presented in the L1 condition needed an average of 100.04 seconds to learn (*SD* = 67.31), compared to 113.84 seconds (*SD* = 66.92 seconds) in the L2 condition. No statistically significant effect was found for perceived item usefulness (t = -0.70, p = .49). The learner-related variables age (t = 3.48, p = .001) and vocabulary size (t = -6.23, p < .001) were found to have a statistically significant effect on the learning burden. Older learners and those with comparatively small vocabulary sizes needed more time to learn the target language. Additionally, word length was found to impact learning burden (t = 11.60, p < 0.001), and a statistically significant interaction between code and length was also found (t = -6.82, p < .001). Descriptive statistics indicated that the advantage for L1 presentation was greater for shorter items. Furthermore, although there was no discrete effect for PoS (*t* = 1.48, *p* = .15), a statistically significant interaction between code and PoS (*t* = -2.63, *p* = .01) showed that the difference between the learning burden of nouns and verbs was greater in the L1 condition than the L2 condition, indicating that a noun advantage was limited to meaning presentation via the L1.

Table 5.10

Fixed and Random Effects for Length of Exposure, Lenient Form Recall (Study 2)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	10.70	0.94	11.36	<.001***
Code of meaning	0.60	0.09	6.47	<.001***
Usefulness	-0.03	0.03	-0.86	.39
Code * Usefulness	0.04	0.04	0.74	.46
Vocabulary	-0.91	0.14	-6.28	<.001***
Age	0.89	0.23	3.795	<.001***
Length	0.86	0.08	11.39	<.001***
PoS	0.10	0.07	1.43	.16
Code * Length	-0.23	0.03	-7.12	<0.001***
Code * PoS	-0.09	0.03	-2.73	0.01**

Random Effects			
Parameter		Variance	*SD*
Item	Intercept	<.01	.20
	Usefulness	<.01	<.01
Participant	Intercept	<.01	.19
	Usefulness	<.01	<.01

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

Table 5.11

Fixed and Random Effects for Length of Exposure, Form Recognition Condition (Study 2)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	10.91	0.97	11.23	<.001***
Code of meaning	0.62	0.09	6.76	<.001***
Usefulness	-0.02	0.03	-0.70	.49
Code * Usefulness	0.04	0.04	0.92	.36
Vocabulary	-0.92	0.15	-6.23	<.001***
Age	0.84	0.24	3.48	.001**
Length	0.84	0.07	11.60	<.001***
PoS	0.10	0.07	1.48	.15
Code * Length	-0.21	0.03	-6.82	<.001***
Code * PoS	-0.08	0.03	-2.63	.01**
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	<.01	.20	
	Usefulness	<.01	<.01	
Participant	Intercept	<.01	.18	
	Usefulness	<.01	.02	

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

Summary of the findings for learning burden

Table 5.12 presents a summary of the findings for frequency and time of exposure. Overall, the results demonstrate that there was a consistent effect for code of meaning presentation, with L2 meaning presentation associated with increased time of exposure. In contrast, no effect was found when burden was operationalised via frequency of exposure.

The perceived usefulness of the target items was found to have no effect on burden. Secondary effects were found for the learner-related variables vocabulary size and learner age on the time of exposure, and vocabulary size on frequency of exposure. Younger learners and those with larger vocabulary sizes experienced less burden. Secondary effects were also found for the intralexical factors word length (frequency and time of exposure) and PoS (frequency of exposure), with nouns and shorter items posing the lightest learning burden. The interactions between meaning presentation code and the intralexical variables demonstrated that these effects were more prominent when word meaning was presented via L1 equivalents, indicating that L1 meaning presentation was particularly effective for items with a lighter learning burden.

Table 5.12

Summary of the Results for Learning Burden (Frequency of Exposure = Freq.; Time of Exposure = Time) by Strength of Knowledge (Study 2)

Strength	Metric	Code	Usefulness	PoS	Length	Vocab size	Age	PoS * code	Length * code
strict recall	freq.	X	X	✓	✓	✓			
	time	✓	X	X	✓	✓	✓	✓	✓
lenient recall	freq.	X	X	✓	✓	✓			
	time	✓	X	X	✓	✓	✓	✓	✓
form recog.	freq.	X	X	✓	✓	✓			
	time	✓	X	X	✓	✓	✓	✓	✓

Note. ✓ represents a statistically significant effect and X represents a non-statistically significant effect.

5.5.2 Decay

Descriptive statistics for the decay of accrued knowledge are presented by level of measurement sensitivity and condition in Table 5.13. Only items for which knowledge was demonstrated on the immediate test at the equivalent level of measurement sensitivity were included in the decay analysis. Table 5.13 also presents the proportion of items answered correctly on the immediate test maintained to the delayed test.

Overall, Table 5.13 shows that there was minimal decay at the level of form recognition, with around 91% of learned items maintained to the delayed test. Greater decay occurred at the form recall level, with an average of 52% and 45% of accrued knowledge lost over the one-week retention interval at the strict and lenient levels respectively. Additionally, the data showed that there was considerable individual variance, with large standard deviations observed at all three levels of measurement sensitivity. Furthermore, the descriptive statistics suggested that there was less decay when meaning was presented via L1 equivalents than L2 definitions. Statistical modelling was conducted to confirm this initial observation. Results are presented by measurement strictness.

Table 5.13

Mean Retention of Items from Immediate to Delayed Test and Proportion of Learned Items Retained by Meaning Presentation Code (SD in Brackets)

	L1 condition		L2 condition		Total	
Strength	Mean	Proportion	Mean	Proportion	Mean	Proportion
strict form recall	13.54 (7.03)	.51 (.25)	10.03 (5.56)	.43 (.21)	11.83 (7.50)	.48 (.23)
lenient form recall	16.51 (8.29)	.58 (.28)	12.65 (6.74)	.51 (.24)	14.63 (8.27)	.55 (.24)
form recognition	29.56 (13.11)	.94 (.42)	25.35 (12.03)	.88 (.39)	27.51 (8.00)	.91 (.19)

Strict Form Recall

The best-fitting model (Table 5.14) showed that the effect for meaning presentation code approached statistical significance (z = -1.75, p = .08), with an average of 13.54 (SD = 7.03) target items retained in the L1 condition, and 10.03 (SD = 5.56) items in the L2 condition, indicating a tendency for items disambiguated with L1 equivalents to be better retained than those presented with L2 definitions. No statistically significant effect was found for either perceived usefulness (z = 0.69, p = .49) or the interaction between perceived usefulness and meaning presentation code (z = -0.17, p = .87). Statistically significant effects were found for both indices of learning burden; with the frequency (z = -4.48, p <.001) and time of exposure (z = -4.47, p < .001) negatively associated with decay. Thus, items that posed the lightest learning burden were least likely to decay. With regard to the intralexical factors considered, there was a statistically significant effect for PoS (z = -2.72, p = .01), indicating that nouns were significantly better retained than verbs. No effect was found for word length.

Lenient Form Recall

The best-fitting model (Table 5.15) showed that there was a statistically significant effect for meaning presentation code (z = -6.82, p = .03), with less decay in the L1 condition than the L2. An average of 16.51 (SD = 8.29) words were retained in the L1 condition and 12.65 words (SD = 6.74) in the L2 condition. No statistically significant effect was found for perceived usefulness (z = 0.29, p = .78) or the interaction between perceived usefulness and code (z = -0.80, p = .94). This indicates that perceived usefulness did not impact retention, nor was any impact of usefulness moderated by meaning presentation code. Statistically significant effects were found for both learning burden measurements (frequency of exposure: z = -3.72, p <.001; time of exposure: z = -3.70, p < .001), indicating that lighter learning burdens were associated with more retention. A secondary effect was found for PoS, with retention of nouns superior to verbs ($z = -3.01$, $p = .002$). No statistically significant effect was found for word length.

Table 5.14

Fixed and Random Effects for Strict Form Recall, Lexical Decay (Study 2)

Fixed Effects				
Parameter	Estimate	*SE*	*Z*	*p*
Intercept	15.27	2.12	7.21	<.001***
Code of meaning	-1.85	1.06	-1.75	.08
Usefulness	0.15	0.21	0.69	.49
Code * Usefulness	-0.05	0.30	-0.17	.87
PoS	-0.55	0.20	-2.72	.01**
Learning burden (frequency)	-1.87	0.42	-4.48	<.001***
Learning burden (time)	-1.00	0.22	-4.47	<.001***
Learning burden (frequency)*Code	0.78	0.45	1.72	.09
Learning burden (time)*Code	-0.05	0.30	-0.17	.87

Random Effects			
Parameter		Variance	*SD*
Item	Intercept	0.16	.39
	Usefulness	<.01	.05
Participant	Intercept	1.45	1.20
	Usefulness	<.01	.02

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

Table 5.15

Fixed and Random Effects for Lenient Form Recall, Lexical Decay (Study 2)

Fixed Effects				
Parameter	Estimate	*SE*	*Z*	*P*
Intercept	14.25	2.24	6.36	< .001***
Code of meaning	-6.82	3.15	-2.16	.03*
Usefulness	0.06	0.21	0.29	.78
Code * Usefulness	-0.02	0.29	-0.8	.94
PoS	-0.50	0.17	-3.01	.002**
Learning burden (frequency)	-1.57	0.42	-3.72	< .001***
Learning burden (time)	-0.91	0.25	-3.70	< .001***
Code*Learning burden (frequency)	-0.03	0.62	-0.05	.06
Code*Learning burden (time)	-.58	0.35	1.65	.10
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	0.17	0.41	
	Usefulness	<0.01	0.05	
Participant	Intercept	1.28	1.13	
	Usefulness	0.02	0.15	

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

Form Recognition

The best-fitting model (Table 5.16) showed that there was no statistically significant effect for meaning presentation code ($z = -1.45$, $p = .15$). Neither was there a significant effect found for perceived usefulness ($z = -0.23$, $p = .82$) or the interaction between code and perceived usefulness ($z = -0.15$, $p = .88$). This showed that the perception of item usefulness did not affect the likelihood of retention. An effect was also found for frequency of exposure (z = -1.97, p = .05), but no effect was found for time of exposure. This indicated that when burden was operationalised by frequency of exposure, items with a lower learning burden were better retained. Secondary effects were again found for PoS (z = -2.28, p = .02), showing that nouns were more likely to be retained than verbs. No effect was found for word length. The interaction between PoS and meaning presentation code approached statistical significance (z = 1.93, p = .06). Examination of the descriptive statistics revealed that the noun advantage was greater in the L1 condition.

Table 5.16

Fixed and Random Effects for Form Recognition, Lexical Decay (Study 2)

Fixed Effects				
Parameter	Estimate	*SE*	*Z*	*p*
Intercept	-6.58	3.63	1.81	.07
Code of meaning	-1.81	1.25	-1.45	.15
Usefulness	-0.17	0.73	-0.23	.82
Code * Usefulness	-0.12	0.77	-0.15	.88
PoS	-1.08	0.47	-2.28	.02*
Learning burden (frequency)	-1.11	0.56	-1.97	.05*
Learning burden (time)	0.12	0.38	0.32	.75
Code * PoS	0.95	0.49	1.93	.06
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	<.01	<.01	
	Usefulness	.31	.55	
Participant	Intercept	6.50	2.55	
	Usefulness	.07	.27	

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

Summary of decay findings

The results, summarised in Table 5.17, show that the effect of meaning presentation code varied with the level of measurement sensitivity, with form recall knowledge suffering less decay in the L1 condition than the L2 condition. In contrast, no effect was found at the level of form recognition. Neither the perceived usefulness of a target item, nor the interaction between usefulness and meaning presentation code had a significant impact on the extent of decay that occurred. The learning burden was generally found to impact decay, with items that needed fewer exposures and less time during the initial period of learning less likely to suffer decay. This shows that the learning burden was negatively associated with the decay process. Secondary effects were consistently reported for the intralexical variable PoS, with nouns better retained than verbs. No effect was found for word length, nor the learner-related variable vocabulary size, on decay at any level of measurement sensitivity.

Table 5.17

Summary of the Results for Decay by Strength of Knowledge (Study 2)

Strength of knowledge	Code	Perceived usefulness	Code * usefulness	Word length	PoS	Vocabulary size	Burden (freq.)	Burden (time)	Code * PoS
Recall (strict)	Δ	X	X	X	✓	X	✓	✓	X
Recall (lenient)	✓	X	X	X	✓	X	✓	✓	X
Recognition	X	X	X	X	✓	X	✓	X	Δ

Note. ✓ represents a statistically significant effect, Δ represents an effect approaching statistical significance, and x represents a non-statistically significant effect.

5.6 Discussion

This study has shown that a considerable number of words can be learned in a relatively brief period using flash card software. This reflects literature to date that has found intentional learning activities such as flashcards foster expeditious learning (Nakata, 2016; Rodgers, 1969). More learning occurred (see Table 5.5) when word knowledge was considered at the form recognition level than at the form recall level. This finding is in line with previous research that has demonstrated a hierarchical relationship between recognition and recall knowledge (Laufer & Goldstein, 2004). With regard to learning burden, the results show that an average of approximately nine exposures lasting around 100 seconds were needed for target items to be encoded. However, considerable individual and item variance was found in the amount of burden experienced. This supports Higa's (1965) conceptualisation of burden as a by-item by-learner phenomenon.

The findings also showed that the accrued knowledge was relatively robust over the one-week retention interval, particularly when knowledge was measured at the level of form recognition. Overall, less decay was found in this study than in Study 1, which employed a longer retention interval, suggesting that the length of the retention interval impacted the extent of decay observed. In the current study, less decay was found at the level of form recognition than form recall. To date, studies have shown that form recall knowledge is harder to acquire than form recognition knowledge (González-Fernández & Schmitt, 2019; Laufer & Goldstein, 2004). This study, in line with the results of Study 1, also suggests that knowledge of form recall is more susceptible to loss than form recognition knowledge.

This study was conducted to delineate the effect of two target variables, meaning presentation code and perceived item usefulness, on the extent of burden experienced during learning and the amount of decay that took place. The remainder of the discussion section considers these variables, organised according to the research questions.

RQ1: To What Extent Do Meaning Presentation Code and Perceived Target Item Usefulness Affect the Learning Burden of L2 Vocabulary Knowledge?

Before discussing learning *burden*, results related to learning gains are discussed. The results of t-tests fitted to the immediate test data showed that there was no statistically significant difference in the number of items learned on the L1 and L2 conditions. This finding seems to contradict the L1 advantage reported in previous investigations (Lado *et al*., 1967; Laufer & Shmueli, 1997; Mishima, 1967); however, key methodological differences explain this apparent contradiction. Unlike prior research in this area, this study did not directly (e.g., by stipulating a common frequency of exposure) or indirectly (e.g., by limiting the time given to learning) control learner interaction with the target items. Learners were able to determine the frequency and time for which the target items were studied on a by-item basis. This study shows, therefore, that when learners have greater control over their learning behaviour and are allowed to see target items as often, and as long, as necessary for learning to occur, the advantage for L1 meaning disambiguation is less well-defined. Indeed, given the nature of the methodology employed in the current study, it is unsurprising that there was no significant difference in learning gains between the conditions; participants were able to engage with comparatively burdensome items more frequently, which increased the probability of acquisition. Therefore, any difference between the two conditions would be reflected not in the learning gains, but in the learning burden.

Learning burden was operationalised in terms of frequency and time of exposure. The data from these two metrics showed that the impact of meaning presentation code depended on which of these measures was considered. With regard to time of exposure, a significant effect was found for meaning presentation code, with items disambiguated by L1 equivalents requiring less time to be learned than those presented with L2 definitions. In contrast, no significant difference was found between the two conditions in terms of frequency of exposure. Therefore, for the population considered in this study, the use of L2 definitions led to longer, but not more frequent exposure to the target lexical items. There are several explanations for this finding. First, the L2 condition may have facilitated greater task engagement than the L1 condition, which was manifested in greater time of exposure. This explanation, however, seems unlikely given that increased engagement is associated with more learning (Schmitt, 2008) and no significant difference was found between the learning gains of the L1 and L2 conditions. A more likely explanation relates to the different processing loads and definition lengths associated with the two conditions. Although the definitions were written in controlled vocabulary, it is likely that the L2 condition made greater processing demands on the learners, which may have led to comparatively protracted engagement with each item. In addition, due to the construction of the definitions, indeed the nature of L2 definitions in general, there was more, and grammatically more complex, content to decode in the L2 condition. For example, the definition employed to disambiguate the target item *kip* was as follows: *to sleep somewhere that is not your home*. In contrast, the same item was delineated by two logographs (睡觉) in the L1 condition. This pattern of more, and more syntactically complex, language is indicative of all target L1 equivalents and L2 definitions, which contained an average of 3.0 logographs and 8.30 words respectively. The longer time of exposure was therefore somewhat unsurprising given the learners were required to process more input.

Perceived usefulness was included as a fixed effect in the analysis of both indices of learning burden (i.e., frequency and time of exposure). Although there was a trend for words evaluated as more useful to require fewer exposures and less time to learn, no statistically significant effect was found on either metric of burden. Previous studies have reported mixed results for the effect of task-induced word relevance, with some suggesting increased relevance is associated with more learning (Peters, 2007) and others finding little effect (Peters & Webb, 2018). The current study considered the extent to which an item was perceived to be useful as dependent on the interaction between a learner and an item, with some learners finding some items more useful than other items, and some items being perceived as more useful by some learners than by their peers. This conceptualisation of usefulness as a by-learner by-item factor differed from much of the literature to date. This different operationalisation likely explains why the findings of the current study seemingly contradict those of some studies to date (e.g., Peters, 2007). This study shows that when perceived usefulness is considered at the learner and item level, it does not impact the learning burden of L2 vocabulary. This conclusion is limited, however, to the specific instructional treatment employed in this study.

In addition to findings related to meaning presentation code and perceived usefulness, secondary effects were found for a number of factors (see Table 5.12). The results showed that learners with larger vocabulary sizes acquired items more expeditiously than those with smaller vocabulary sizes. This indicates a Matthew Effect (Stanovich, 1986) which, in SLA, refers to patterns in learning in which early higher achievement leads to quicker rates of subsequent achievement, with more proficient learners able to make faster learning gains than lower proficiency learners. Previous learning studies have reported a Matthew Effect for L1 reading (Stanovich, 1986), learner motivation (Williams et al., 2002), and general L2 development (Lamb, 2011). The present investigation adds to this literature, showing that vocabulary size impacted the frequency and time of exposure, with a negative correlation between vocabulary size and learning burden. However, this does not mean that vocabulary size alone impacted the extent of learning burden experienced. It may be that vocabulary size acted as proxy for superordinate variables such as general language knowledge or language learning aptitude. Thus, it is unlikely the case that a singular focus on increasing vocabulary size at the expense of developing other areas of language knowledge would increase learning speed to the extent reported in this study.

As with Study 1, a significant effect was found for word length on learning burden. Unsurprisingly, longer words were associated with a heavier learning burden than shorter items. This finding is in line with previous literature that has reported a word-length effect (Baddeley, Thomson, & Buchanan, 1975; Ellis & Beaton, 1993; Gerganov & Taseva-Rangelova, 1982; Tehan & Tolan, 2007; Willis & Ohashi, 2012). Furthermore, the results of the time of exposure analysis showed a significant interaction between length and meaning presentation code. Inspection of the descriptive statistics showed that although longer items were equally challenging for all learners, shorter items were comparatively easier to learn when presented with L1 rather than L2 meaning disambiguation. This finding suggests that the L1 advantage was particularly salient with items that posed the lowest learning burden.

A significant effect was also found for PoS on the frequency of exposure needed to learn the target items, with nouns posing a lighter learning burden than verbs. This finding is in line with previous research (Ellis & Beaton, 1993; Horst & Meara, 1999; Rodgers, 1969), although it contradicts the results of Study 1, which reported no effect for PoS on frequency of exposure. This divergence can be explained by methodological differences between Studies 1 and 2. Firstly, in Study 1, learners were required to infer the PoS from definitions, while PoS was explicitly stated on the flashcards in the current study. This likely increased the saliency of the PoS, amplifying the effect it had on learning burden. Furthermore, Study 1 employed L2 definitions, while this study additionally used L1 equivalents. To investigate the moderating influence of meaning presentation code on the effect of PoS, the interaction between PoS and code was included in statistical modelling. A statistically significant effect was found for this interaction on learning burden (time of exposure), with nouns posing a lighter learning burden than verbs in the L1 condition than the L2 condition. This finding implies that the non-significant results found for PoS may be partly attributable to the manner in which word meaning was conveyed. Interestingly, studies that have reported a noun advantage have tended to employ L1 equivalents to delineate word meaning, further supporting this interpretation. Thus, the findings of this study suggest that PoS is more salient when items are accompanied by L1 equivalents rather than L2 definitions, even when the PoS is explicitly stated. Further research employing tests of grammatical aspects of word knowledge in addition to measures of the form-meaning link are needed to investigate this inference.

RQ2: To What Extent Do Meaning Presentation Code and Perceived Target Item Usefulness Affect the Decay of L2 Vocabulary Knowledge?

Firstly, although unsurprising given that vocabulary acquisition studies typically report that scores almost inevitably fall between immediate and delayed tests (Schmitt, 2010), this study found that target language knowledge decayed over the one-week retention interval. Furthermore, decay was found to be greater at the level of form recall than form recognition. The proportion of learned knowledge that was subsequently retained at the level of written-form recognition was considerably higher (M = .92, SD = .23) than at the level of either strict (M = .47, SD = .34) or lenient (M = .55, SD = .34) recall. This supports studies that have demonstrated a hierarchical relationship between recall and recognition (see González-Fernández & Schmitt, 2019; Laufer & Goldstein, 2004), in which knowledge at the form recall level implies knowledge at the level of form recognition. Moreover, the two measurement tasks employed for the recall and recognition instruments differed in the extent to which they tapped into partial knowledge. The recall instrument largely excluded partial knowledge (this was especially the case when recall was graded using the strict criterion), while the recognition instrument recognised partial knowledge to a greater degree. Thus, this investigation found that language knowledge decayed, and that this decay was more acute at the level of form recall. This differential pattern of loss can be ascribed to the relationship between recall and recognition knowledge and/or the level of challenge posed by the instruments.

Moving on to consider the effect of meaning presentation code on this decay, in terms of form recall, L1 meaning disambiguation facilitated retention to a greater extent than definition via the L2. This effect was significant in the more inclusive lenient scoring procedure and just failed to reach statistical significance when recall was judged strictly. Taken together, the results suggest a retention advantage when meaning is presented via L1 equivalents. To date, research has shown that using L1 equivalents fosters more (Lado *et al.*, 1967; Laufer & Shmueli, 1997; Mishima, 1967) and more expeditious (this study, see Section 5.5.1) learning. The results of this study now suggest that L1 meaning disambiguation may also facilitate greater retention at the level of form recall. This finding supports those who advocate (e.g., Nation, 2013) judicious use of the L1 in the L2 classroom. There are, of course, numerous reasons why practitioners might stipulate a policy of exclusive L2 use, as discussed earlier in this chapter; however, if the goal of instruction is the expeditious acquisition and retention of form recall knowledge, there seems little evidence to justify prohibiting the use of L1 equivalents to disambiguate item meaning.

At the level of form recognition, the data also suggest that items were better retained in the L1 condition, although this finding failed to reach statistical significance. This result needs to be considered in light of the brevity of the retention interval. Minimal decay was found at the level of form recognition, and it may be that the one-week interval was too short for sufficient decay of recognition knowledge to occur. Future research should employ longer retention intervals to allow for greater decay to take place prior to statistical modelling.

The advantage for the L1 can be explained in terms of the differential allocation of cognitive resources in the two conditions. It has been argued that learners have limited capacity with which to process a novel word's form and meaning (see Barcroft, 2015). The extent to which a learner can devote attentional resources to word form is inherently related to the resources spent processing word meaning. Increasing the load associated with the learning of meaning necessarily limits the attentional resources that can be devoted to form learning. This is to say that increased attention to word meaning results in better knowledge of meaning and impoverished knowledge of form, while increased attention on the learning of word form results in improved word-form knowledge and impoverished knowledge of word meaning (Barcroft, 2015). It is perhaps the case that presenting word meaning via the L2 required participants to dedicate more attentional resources to word meaning than when presentation was made via the L1. Therefore, it follows that due to the increased focus on word meaning, learners in the L2 condition had less attentional resources available to devote to the learning of word form. Crucially, the vocabulary tests employed in the current study targeted knowledge of form. Thus, the L1 advantage may stem from the manner in which attentional resources are divided in the two conditions, with L1 meaning presentation allowing for more processing of word form than L2 meaning disambiguation. If this were the case, we might expect there to be a greater effect for meaning presentation code with learners of lower proficiency, who would need to dedicate comparatively more resources to process the L2 definitions. Therefore, replications with learners of diverse proficiencies are needed.

The data also indicated that perceived item usefulness and the probability of successful retention were positively associated; however, the best-fitting models for strict recall, lenient recall, and form recognition indicated there was no significant effect for perceived usefulness on the decay of lexical knowledge. The data showed that the decay of items judged to be less useful was as likely as that of items considered to be more useful. To the best of my knowledge, no study to date has examined the impact of perceived usefulness as a by-item by-learner variable on the decay of accrued knowledge. Interestingly however, theoretical perspectives suggest that usefulness should positively impact both learning and retention, as learners are likely to be more engaged (see Blumenfeld & Paris, 2004; Schmitt, 2008), and employ deeper processing strategies (see Craik & Lockhart, 1972), when learning items that are perceived to be more useful. This incongruity can be explained by the learning task employed. In this investigation, learners were required to learn each item in the same manner, needing to accurately produce each item to complete the learning task. This means that independent of perceived usefulness, items were broadly dealt with in the same manner. It may be that these task constraints mitigated the impact of perceived usefulness on the retention of accrued knowledge. That is, although perceived usefulness may have impacted the extent of cognitive engagement with the learning process to some extent, due to the prescriptive nature of the task, it did not directly affect learning behaviour nor the retention of knowledge. Thus, these findings cannot be generalised to more open or indeed dissimilar learning tasks without further investigation.

Additionally, this study found an effect for PoS on decay, with verbs more likely to decay than nouns. This finding differs from Study 1, where no effect for PoS was found. This difference may stem from changes to the design of the flashcards; the current study included direct reference to the PoS of each target item, making the PoS more salient. This increased saliency may have impeded the decay found in Study 1. Furthermore, the current study employed a one-week retention interval whereas Study 1 used a retention interval of four weeks. The increased duration of the retention interval in Study 1 may have provided more opportunity for nouns to decay than the present study. That is to say, the findings of the two studies may not be contradictory, but actually tap into two discrete points of the decay process. A further explanation, however, relates to the increased ease with which nouns were learned. The data show that items with a lower learning burden were better retained. They also revealed that nouns were less burdensome than verbs. Therefore, the increased retention of nouns may stem from their lower learning burden. Previous research has also shown a retention advantage for nouns (Ellis and Beaton, 1993). These studies have attributed the noun advantage to increased imageability, with the image evoking potential of the stimulus credited with predicting retention (Paivio, 1963). The current study used concrete nouns which were likely to be more imageable than the target verbs; thus, the increased imageability of the nouns used in this study may have led to greater retention. The logical next step, therefore, would be to replicate the current study employing both concrete and abstract nouns to clarify the source of the retention advantage for nouns.

RQ3: To What Extent Does Learning Burden Affect the Decay Process (Irrespective of the Target Variables)?

The results of this study indicated that an increased burden during the learning procedure was associated with higher rates of decay. Learning burden was operationalised as both the frequency and time of exposure necessary for acquisition. Significant effects were found on retention for both indices at the strict and lenient form-recall level, and for frequency of exposure at the level of form-recognition. This finding is particularly interesting as the majority of studies to date have reported a positive effect for time on task and frequency of exposure on vocabulary learning and retention. For example, Peters (2014) found that in an instructed SLA context, items that occurred more frequently in the instructional activities were better retained. Similarly, Laufer and Rozovski-Roitblat (2015) found that increased encounters with target items in a reading and focus on forms task positively impacted meaning retention. Furthermore, a positive relationship between frequency of exposure and learning has also been found with incidental contexts (Uchihara, Webb, & Yanagisawa, 2019).

The apparent incongruity between those investigations and the current study likely stems from differences in the methodologies employed. In this investigation, items were removed from the learning procedure after successful retrieval. This design differentiated exposures during the process of encoding from retrievals of learned knowledge. This is important as frequency of retrieval has been shown to positively impact retention (Folse, 2006; Nakata, 2016). Previous studies have generally stipulated a common frequency of exposure for all target items; for instance, Peters (2014) and Laufer and Rozovski-Roitblat (2015) used such a design. As previously argued (see Section 2.1.4), prescription of frequency and/or time of exposure may bias retention data as items with a lower learning burden receive more retrievals than items with a higher learning burden, potentially strengthening the knowledge of items with a lower learning burden, making their decay less likely. This study indicated that when the number of retrievals is controlled, items that pose the lightest learning burden are those items least likely to decay. This finding illustrates that repeated exposure to target items can be indicative of increased burden, which this study has shown to be positively correlated with the probability of decay. There are important implications of this finding both for classroom pedagogy and for the design of CALL software. These are discussed in Chapter 7.

5.7 Limitations

There were several limitations of this study. First, this study targeted only one aspect of word knowledge, the form meaning link. The instruments used measured knowledge of word form in response to a given meaning, but no explicit measures of meaning recall or recognition were included. This was the case because a measure of meaning would have necessarily led to additional exposures to the target form and encounters in test settings are likely to be particularly salient, foster deeper engagement, and impact memory trace strength (Schmitt, 2010). As a result, inclusion of a meaning-focused instrument may have impacted the decay process of word form knowledge. Future research is therefore needed to determine the effect of meaning presentation code on the recall and recognition of target item meaning. This is particularly important as one potential benefit of L2 definitions is that they can convey more nuanced target item meanings and lead to more semantically complete representations of target items (see Gablasova, 2015).

Second, this study measured word knowledge using discrete instruments of form recall and form recognition. Written form recall is generally associated with producing target items in writing; however, it is important to recognise that written form recall is not the same as free production (Schmitt, 2019). Therefore, based on the data presented here, it is not possible to claim that the participants would be able to accurately use the accrued target items in their own writing or speaking. Yet, it is equally important to recognise that this limitation is not unique to the study presented in this chapter; in fact, it applies to many, if not the majority of, investigations of L2 vocabulary learning. Thus, while it is true that the findings outlined in this chapter cannot be generalised to free production without further research, it is also the case that vocabulary research more generally needs to determine the extent to which certain treatments lead to actual language use.

Third, the target items employed were concrete nouns and verbs. It is possible that such items were particularly suited to disambiguation via L1 equivalents. Different target items may be more suited to meaning presentation via extended definitions; for instance, items that do not have an equivalent in the L1 or that possess more abstract qualities. Importantly, however, such definitions could be presented in the L1 as, for example, is the case with bilingualized dictionaries (Laufer, 1994). Further research comparing L1 equivalents, L2 definitions, and L1 definitions is needed to determine the relative effectiveness of L1/L2 definitions. Such research would also show whether the effect for code of meaning presentation reported in the present study was due to the different codes (i.e., L1/L2) or the format of meaning presentation (i.e., equivalents/definitions).

Additionally, this study is, to the best of my knowledge, the first time by-item by-participant perceived usefulness has been considered in an investigation of vocabulary learning. The results showed that usefulness did not affect the learning burden and decay. However, as discussed in Section 5.6, this finding may have stemmed from the controlled nature of the instructional activity that required learners to successfully respond to all of the flashcards in order to finish the learning phase, potentially overriding the effect of perceived usefulness. Therefore, further research with learning activities that offer students more choice over the words with which they engage is necessary before the findings of the current investigation can be generalised more broadly.

Lastly, due to the nature of the educational setting it was not possible to cross-reference the target items to the teaching materials used while students completed the study. It was therefore not possible to ensure that no exposure to the target items during the retention interval occurred. To compensate, a self-report checklist tool was included in the survey conducted after the delayed test, which asked participants to mark items to which they had been exposed between the immediate and delayed tests. Any items reported to have been seen during the retention interval were removed from analysis. Thus, it was unlikely that intersessional exposure to the target items impacted the results of the decay analysis.

5.8 Conclusion

The study presented in this chapter sought to investigate the effect of meaning presentation code and perceived usefulness on the learning burden and decay of L2 lexical items. Firstly, the results showed that a considerable number of L2 words can be learned in a relatively brief period and maintained over a one-week retention interval to the level of form recognition. This first point confirms the findings from previous studies showing the short and longer-term effectiveness of intentional vocabulary learning. With regard to the target variables, meaning presentation code was found to impact the time, but not the frequency, of exposure needed to learn the target items, with words presented alongside L1 equivalents requiring less time to be learned. Meaning presentation code also impacted target item retention. The data showed that items disambiguated via L1 equivalents were more likely to be retained than those presented with L2 definitions. Thus, the use of L1 equivalents led to more expeditious learning and facilitated retention to a greater extent than L2 definitions.

The results also showed that perceived item usefulness did not significantly affect either learning burden or decay. This study was one of the first to have investigated usefulness as a by-item by-learner variable, and we therefore know little about the effect of this factor on vocabulary acquisition. Future research is thus needed to explore the effect that this conceptualisation of usefulness has on other types of vocabulary learning activities. For instance, it may be the case that usefulness appears as a predictor of learning burden and decay in other, more incidental, learning conditions where learners have more control of how they engage with the target items.

Perhaps the most surprising finding was that items seen less often and for shorter periods during the learning phase were better retained after the one-week retention interval, indicating that learning burden was positively associated with decay. Importantly, this finding differs from Study 1, which showed that higher burden was associated with less decay of form recognition knowledge. This was likely caused by the differing retention interval lengths.

This study included PoS and word length as covariates in the analysis of word length and decay. The findings showed that both intralexical variables had an effect on the learning burden and that PoS impacted decay. This finding differs from Study 1, which found no effect for PoS on either burden or decay. In the current study, PoS was included on the flashcards, increasing its saliency. This may have led to increased noticing of item PoS than in Study 1, impacting patterns of learning burden and decay.

Overall, this study has added to our knowledge of learning burden, decay, and the relationship between burden and decay. It has shown that some variables have an impact on the learning burden of L2 vocabulary but do not affect the decay of learned knowledge, while other factors impact both learning burden and decay. This inconsistency speaks to the importance of investigating the comparative effect of variables on burden and decay. The study presented in this chapter looked at meaning presentation code and perceived item usefulness, while Study 1 investigated two intralexical factors. However, numerous potentially impactful variables that warrant further investigation remain. In the next chapter, two such variables, the mode of form presentation and language learning aptitude are considered.

Chapter 6: Study 3

The Effect of Mode of Form Presentation and Language Learning Aptitude on the Learning Burden and Decay of L2 Vocabulary Knowledge

6.1 Introduction

The previous chapter dealt with two ways in which word meaning can be communicated to learners, i.e., through L1 equivalents and L2 definitions. This is an important area as language learning textbooks typically focus on developing knowledge of word meaning (Schmitt, 2010). There are, of course, other aspects involved in developing L2 word knowledge, with perhaps the most crucial being word form (Barcroft, 2015). The importance of word form tends to be overlooked by teachers and textbook writers (Horst *et al*., 2010), although developing word form knowledge can be problematic for learners (Schmitt, 2010) and insufficient focus on word form can be detrimental to retrieval (Barcroft, 2002, 2015). Thus, word form is key to effective L2 lexical instruction and an area that, due to this importance, warrants greater investigation. In addition, a more thorough understanding of this area may have important implications for the teaching of vocabulary and the design of vocabulary learning software. For instance, a better understanding of the effect of different form presentation modes on learning burden and patterns of decay would allow learning platforms to alter the design of activities to maximise learning and minimise loss. The study presented in this chapter investigated the effect of one factor relevant to word form, presentation mode (i.e., simultaneous spoken and written form presentation or written form only), on learning burden and decay.

Furthermore, one of the findings common to the studies presented in Chapters 4 and 5 is the large variance between the learners both in terms of learning burden and decay. By incorporating both learner and item into statistical models, the analysis has been able to respond to this variance; however, as of yet, this variance remains unexplored. Therefore, the current study investigated the effect of a learner variable, language learning aptitude (LLA), on the learning burden and any decay that occurred. Individual differences such as LLA play a crucial role in SLA (Williams & Burden, 1997), but have received little attention in studies of vocabulary learning.

This study considered the effect of form presentation mode, LLA, and the interaction between mode of form presentation and LLA on learning burden and decay. As with Study 2, learning burden was operationalised using two metrics: the number of exposures needed to learn a target item and the time taken to learn a target item. Few studies have overtly investigated the role of LLA in vocabulary learning and, to the best of my knowledge, no study has considered the effect of LLA on the learning burden of foreign language lexis, the decay of foreign language vocabulary knowledge, or the effect of form-presentation mode on these two phenomena.

6.2 Background

6.2.1 Form-Presentation Mode

Word form is a superordinate term consisting of the spoken and the written forms. The spoken form relates to the pronunciation of a word, while the written form relates to its spelling (Nation, 2001). Whether the pronunciation or spelling is prioritised largely depends on the learning context and the needs of the learner. Some educational contexts (e.g., those adopting a grammar-translation method) and some learners (e.g., those needing to understand English journal articles) may give precedence to the written form, while other educational contexts (e.g., those adopting an audiolingual method) and learners (e.g., those needing to interact orally in English) might exclusively focus on the spoken form. Independent presentation of the written or the spoken form is termed unimodal form presentation, while simultaneous presentation of both the written and spoken form is referred to as bimodal form presentation.

Although there may be situations that favour unimodal form presentation, bimodal presentation is typically recommended (Nation, 2005; Thornbury, 2002; Scrivener 2011; Ur, 2010). Furthermore, the Certificate in English Language Teaching to Adults (CELTA), a popular teaching training course offered by Cambridge, instructs trainee teachers to present written and spoken aspects of word form together (see Thornbury & Watkins, 2007). Therefore, it seems that we have a putative understanding in TESOL that bimodal form presentation is superior to unimodal form presentation.

One assumption underpinning this belief is that spoken and written form knowledge develop in parallel; however, this is not always the case. This lack of correspondence is partly because some words occur more frequently in one modality than the other. For example, all things being equal, better written-form knowledge would be expected if an item appeared more commonly in writing than in speech; while better spoken-form knowledge would be expected if an item was more common to spoken than written language. A further reason relates to intralexical factors which can make a word form harder to learn in one modality. Written word-form difficulty can vary according to several factors;

for example, the number of letters, similarity to known L1 or L2 written forms, and orthographic typicality (Schmitt, 2010). Numerous factors can also impact the difficulty of spoken forms; for instance, the number of phonemes, the number of syllables, pronunciation and intonation variability, similarity to known L1 or L2 spoken forms, and phonological typicality (Schmitt, 2010). Crucially, due to these intralexical factors, it is unlikely that written-form burden ever exactly corresponds to spoken-form burden. Thus, it can be expected that, in some cases, form knowledge will lag in one modality.

Another assumption underpinning this belief is that knowledge of the form in one modality can be strengthened by interaction with both modes. Psycholinguistic studies have reported priming across modalities (Bird & Williams, 2002), implying that processing of the written form can impact the representation of the spoken form; however, few vocabulary learning studies have considered this area. Thus, the extent to which the learning burden of the spoken or written form can be eased if learners interact with both types of word form is currently unclear.

The Effect of Form Presentation Mode on Vocabulary Learning

Studies have shown that both unimodal written (Ebbinghaus, 1885; Elgort, 2011; Elgort & Piasecki, 2014; Laufer & Shmueli, 1997; Nakata, 2015) and unimodal spoken (Bisson et al., 2015; Rodgers, 1969) form presentation can lead to vocabulary learning. Comparatively few studies have investigated bimodal form presentation, but those that have suggest it can facilitate vocabulary learning (Sandberg et al., 2011; Webb & Chang, 2012). Therefore, it seems that unimodal and bimodal form presentation both result in vocabulary learning. An obvious question to ask therefore, is which form presentation method results in more learning. Surprisingly, there have been few systematic comparisons of unimodal and bimodal form presentation methods in intentional learning contexts. However, the limited literature in this area generally finds that bimodal form presentation leads to more learning.

Lado, Baldwin, and Lobo (1967) investigated the effect of unimodal and bimodal form presentation on intentional vocabulary learning. Additionally, they considered the order of the components in the bimodal condition (i.e., spoken form before written form, written form before spoken form, and simultaneous presentation of the spoken and written forms). Twenty-two American students learned 100 Spanish words via one of five conditions (i.e., spoken only, written only, spoken before written, written before spoken, and simultaneous spoken and written), with 20 items presented in each condition. After two learning sessions in which participants were exposed to both form and meaning with slides and synchronised tapes for presentation of the spoken form, a meaning recall test was administered and results were compared to an identical pre-test to determine the extent of learning that occurred. Results showed the spoken only condition was least effective and the simultaneous spoken and written condition was most effective. There was little difference between the written only, written before spoken, and spoken before written conditions. Although the results suggested that bimodal form presentation fostered more learning than unimodal form presentation, the lack of counter-balancing means that there might have been a confounding item effect.

Similarly, Hill (1994) investigated the impact of form presentation mode on intentional vocabulary learning. Her L1 Cantonese learners studied 30 target words in one of two conditions, with 28 participants in each group. In the control group, participants saw the written form, the phonetic transcription, an L1 equivalent, and an L2 definition for each target item. In the experimental group, learners were also aurally presented with the spoken form and listened to this recording repeatedly. Spoken and written form recall tests were administered both immediately after the learning phase and a week later. A comparison of these tests revealed no significant difference between the two conditions on the immediate test of written form recall; however, on the delayed instrument, there was less decay in the bimodal condition than the unimodal condition. With regard to spoken form recall, unsurprisingly, greater learning and retention was reported in the bimodal condition than the unimodal condition.

Studies that report a superiority for bimodal input typically explain this advantage from one of several positions. The first explanation relates to learner preference for bimodal form presentation. For example, Tragant Mestres and Pellicer-Sánchez (2019) recognised that learners typically prefer bimodal input (in an incidental context) and suggest that this preference might positively impact motivation and, in turn, learning. Second, it has been suggested that bimodal presentation allows learners to engage with the input according to their preferred learning style. Tragant et al. (2016) suggested that the use of spoken and written modes allow learners to approach a task according to their individual strengths; some relying on written input with others leaning on aural input.

Thus, there is some evidence pointing to a positive effect of bimodal form presentation for vocabulary learning. However, this claim is based on limited evidence and the studies that have been conducted have focused on learning gains, with very few focusing on lexical decay. Additionally, to the best of my knowledge, no vocabulary learning study has considered the impact of form presentation mode on learning burden. One study, Tinkham (1997), did use a frequency-based criterion to determine the impact of spoken unimodal form presentation and written unimodal form presentation on the learning burden of nonwords. He reported that items presented in the written mode required fewer exposures to reach the learning criterion (set at two accurate retrievals) than the items presented via the spoken mode. However, this investigation did not delineate the comparative effect of unimodal and bimodal form presentation on learning burden; thus, this question remains unanswered to date. Given that electronic flashcard software allow for the inclusion of both spoken and written forms (Nakata, 2008), functionality that is typically praised in the literature (Chien, 2015; Nakata, 2008), it is important to systematically investigate the effect this multimodal functionality has on the learning burden of foreign language vocabulary. The findings of such research may inform CALL and MALL instructional design.

6.2.2 Individual Differences

Language teachers typically observe differences between individual learners. That is, despite having access to the same input, it is common for learners to differ both in the speed of intake and the extent of linguistic mastery achieved. This asymmetry can partly be explained by individual differences (ID). ID is the branch of psychology that seeks to determine the most parsimonious description of individual variation, explaining both how and why individual variance occurs (Cooper, 2002). ID are defined as "characteristics or traits in respect of which individuals may be shown to differ from each other" (Dörnyei, 2005, p. 1).

ID is a superordinate term encompassing variables that relate to learner characteristics such as motivation, personality, anxiety, and language learning aptitude (Ehrman & Oxford, 1995), as well as demographic factors such as age and gender (Dörnyei, 2005). It is an important consideration for research, given that it can help to explain why a treatment is successful with some learners and not others, meaning research findings can be more effectively deployed to both teaching and theory-building (Cooper, 2002). Thus, it has been argued that a better understanding of this ID-treatment interaction is important for both pedagogy and educational theory (Robinson, 2002).

Numerous ID have been researched in the field of SLA and many have been found to relate to language learning success (Dörnyei, 2005). Due to the considerable influence on attainment, calls have been made for more research of ID (Skehan, 2002) and there has been somewhat of a resurgence of research in the area in recent years. However, to date comparatively few studies have explored ID in the area of vocabulary research, and, to the best of my knowledge, no study has investigated the impact of ID on the learning burden and decay of lexical knowledge. Of course, it would be unrealistic to explore all ID in one study, thus the research reported in this chapter focused on only one: language learning aptitude. This is because LLA is a key variable in SLA (Skehan, 1989) that has received considerable attention (Ehrman & Oxford, 1995; Saito, 2017), and has been found to influence the acquisition of aspects of language knowledge (Saito, Suzukida, & Sun, 2019). This attention, however, does not extend to studies of lexical acquisition as LLA has yet to be sufficiently explored in vocabulary research.

6.2.3 Language Learning Aptitude

Language learning aptitude is "a specific talent for learning foreign languages which exhibits considerable variation between individual learners" (Dörnyei & Skehan, 2003, p. 590). It relates to the time needed to complete learning tasks, with more able learners needing less time to complete learning tasks than those with lower language learning aptitude. It also impacts the extent of acquisition that occurs given equitable learning and motivational conditions; meaning that comparatively able learners can acquire more language knowledge than their less able peers in a given timeframe (Carroll, 1990). As it impacts the rate and extent of learning that occurs, it is an important variable in SLA (Skehan, 1989).

For many years aptitude research was somewhat marginalised within SLA as it was largely associated with structural rather than communicative approaches and thus considered to be less relevant to the modern L2 classroom (Skehan, 2002). In recent years though, there has been a renaissance of research in this area (Wen et al., 2017), due, in part, to the availability of convenient measurement instruments (e.g., Meara, 2005) and advances in cognitive psychology (Granena, 2013). However, while some areas of SLA (e.g., pronunciation; see Saito, 2017) are increasingly using aptitude to explain variance in levels of attainment, aptitude is not often considered in studies of vocabulary learning. Before discussing the research connecting aptitude with SLA and vocabulary acquisition, the construct and measurement of aptitude is discussed.

Aptitude is generally understood not as a single mass, but as a multi-componential construct (Granena, 2013; Sternberg, 2002). Indeed, aptitude can be considered a superordinate term to describe an amalgam of different cognitive and perceptual abilities. This multi-componential view of aptitude is rooted in the work of Carroll and Sapon (1959) and their aptitude framework manifested in the Modern Language Aptitude Test (MLAT). This instrument is the result of research that administered numerous tests at the start of a course of study and correlated the results against attainment at the end. Tests that did not correlate significantly or those that displayed high intercorrelation with other instruments were omitted from analyses. Subsequent factor analysis suggested a four-dimensional model of aptitude:

Phonetic coding ability	-	the capacity to distinguish specific sounds, to associate sounds with their written symbols, and to retain these associations.
Grammatical sensitivity	-	the ability to recognise the grammatical role of words in a sentence.
Inductive language learning ability	-	the capacity to infer structural rules from language input
Associative memory capacity	-	the ability to associate a form with a label and retain this association.

The MLAT is undoubtedly the most influential aptitude instrument created to date (Skehan, 1989) and, in fact, despite its age is still widely employed (Sasaki, 1996). However, it is by no means the only measurement of aptitude available to language teaching professionals; several testing batteries have been produced in the years since the MLAT was first developed. Examples here include Hi-LAB (Linck et al., 2013) and LLAMA (Meara, 2005). Due largely to developments in cognitive psychology and differing test purposes, these instruments often differ in the specific cognitive variables included in their aptitude constructs. For instance, Hi-LAB assesses working memory, associative memory, long-term memory retrieval, implicit learning ability, processing speed, and auditory perceptual acuity, while the LLAMA battery includes measures of associative memory, phonological short-term memory, phonetic coding ability, and grammatical inference. The choice of aptitude instrument is typically driven by participant variables (e.g., proficiency, L1) and procedural limitations. For example, some batteries predict initial learning rates (e.g., MLAT; Carroll & Sapon, 1959), while others determine advanced attainment (e.g., Hi-LAB; Linck et al., 2013). Furthermore, some instruments can be used by learners from a variety of L1 backgrounds (e.g., LLAMA; Meara, 2005), whereas others cannot (e.g., LAT; Meara, Milton, & Lorenzo-Dus, 2003). In terms of procedural limitations, instruments vary in how they are administered, their availability, and their cost. For example, the MLAT is a paper-based test available on request for a fee. In contrast, LLAMA (2005) is a free computer-based test. The research presented in this chapter employed the LLAMA battery as it was appropriate for the proficiency of the participants, is language-independent, computer-based, and freely available. The specific design characteristics of this battery are described later in the chapter.

Research has shown that aptitude is related to language learning in L1 (Dörnyei, 2005; Engel de Abreu & Gathercole, 2012) and L2 (Ehrman & Oxford, 1995; Granena, 2013; Li, 2016; Ortega, 2009), and studies have found a connection between aptitude and specific features of L2 proficiency; for instance, pronunciation (Granena & Long, 2012; Saito, 2019; Saito, Suzukida, and Sun, 2019) and grammar (Granena, 2014; Yalçın & Spada, 2016) have received particular attention.

However, fewer studies have investigated the relationship between aptitude and L2 lexical acquisition. Granena and Long (2012) investigated the effect of age of onset, length of residence in the L2 context, and aptitude on the attainment of L2 phonology, vocabulary, and morphosyntax. Aptitude was measured with the LLAMA (Meara, 2005), while the three linguistic domains were determined with a discrete measurement battery. The results indicated that, in general, aptitude was positively associated with the attainment of lexical knowledge in late-onset bilinguals, with a statistically significant correlation of .59. In particular, phonological short-term memory and phonetic coding ability were most strongly correlated with vocabulary learning. In her meta-analysis, Li (2016) found positive correlations between vocabulary learning and overall foreign language aptitude ($r = .15$). Additionally, she reported the impact of discrete aspects of the LLA complex on vocabulary learning; for example, moderate correlations were found between vocabulary acquisition and associative memory capacity ($r = .20$) and phonetic coding ability ($r = .38$). Furthermore, Engel de Abreu and Gathercole (2012), investigating foreign language acquisition with L1 Luxemburgish children aged seven to eight years old, found that phonological short-term memory (PSTM) (not considered in Li's meta-analysis) was also moderately associated with foreign language vocabulary learning.

Thus, research has demonstrated that greater aptitude is associated with more successful L2 acquisition and various aspects of the aptitude complex are more or less strongly associated with the learning gains of different linguistic domains. Studies to date suggest that aptitude is positively correlated with vocabulary acquisition; however, due to the paucity of studies linking aptitude with lexical acquisition, more evidence is needed to substantiate this claim. Motivation for the focus on aptitude in this chapter stems from limitations not only in the quantity, but also the nature of the research conducted to date. Investigations have generally looked at the relationship between vocabulary knowledge and aptitude, using this comparison to infer the effect of aptitude on the process of vocabulary learning (e.g., Granena & Long, 2012). In such studies, a positive relationship between vocabulary knowledge and aptitude is used to argue that higher levels of aptitude led to faster learning. However, as correlations of attainment and aptitude may be impacted by confounding factors such as the amount of effort expended by a learner, it is important also to investigate the relationship between aptitude and the effort required to learn L2 lexical items. The research presented in this chapter considered the relationship between components of aptitude and learning burden.

While some research attention has been given to the effects of LLA on vocabulary learning, it may be the case that LLA also impacts a learner's ability to retain knowledge. This points towards a potential relationship between LLA and lexical decay; however, research evidence in this respect is scarce. One of the few studies to be conducted in this area investigated the attrition of L1 morphosyntactic knowledge (Bylund et al., 2009). A comparison of aptitude data, elicited from the LAT (Meara, Milton, & Lorenzo-Dus, 2003; a precursor to LLAMA) with grammaticality judgement tasks indicated that aptitude was negatively associated with the maintenance of L1 morphosyntactic knowledge; attriters with above-average aptitude showed better retention of L1 structures than those with comparatively lower levels of aptitude. Based on these data, the researchers suggested that aptitude can offset the impact of reduced target-language exposure and argued that, given this effect, "it may be valuable to continue exploring the role of aptitude in attrition" (Bylund et al., 2009, p. 459). Although investigating implicit knowledge and a type of language loss that qualitatively differs from decay, one of the goals of this chapter is to investigate whether language aptitude (and its various subcomponents) plays a similarly preventative role in the decay of lexical knowledge with high levels of aptitude inhibiting the process of knowledge deterioration.

The above discussion has demonstrated several uses of LLA research. First, aptitude profiles can be used to explain differential attainment. Such analysis has been extensively conducted in some areas of SLA (e.g., pronunciation; Saito & Hanzawa, 2016), and aptitude has been shown to influence learning processes and attainment. However, few vocabulary studies have included aspects of LLA in assessment batteries or as covariates in analyses; therefore, we currently know little about how different aptitude components affect vocabulary learning and retention. Second, there have been calls to align teaching materials with learner aptitude profiles (Robinson, 2002). It has been claimed that better alignment fosters learner autonomy (Ehrman & Oxford, 1995) and expedites linguistic development (Skehan, 1989). However, to the best of my knowledge, no vocabulary learning study to date has attempted to understand how aspects of the aptitude construct interact with contextual factors such as the mode of form presentation (i.e., spoken/written) to help or hinder the learning process. The study discussed in this chapter sought to meet these lacunae by investigating three aspects of language aptitude (associative memory capacity, phonological short-term memory, and phonetic coding ability) as covariates in statistical analysis. This allowed consideration of whether learners with specific cognitive characteristics benefit (i.e., a lighter learning burden and less decay) more from bimodal form presentation than other learners.

6.2.4 This Study

The above literature review has shown that little is currently known about the effect of form presentation mode on the learning and decay of foreign language vocabulary. Additionally, it has been demonstrated that few intentional vocabulary learning studies have considered the role of aptitude, or the interaction of aptitude with form-presentation mode, in these processes. The study presented in this chapter addressed these issues by considering the following research questions:

1. To what extent does form presentation mode (i.e., unimodal or bimodal) affect the
    a. learning burden of L2 vocabulary?
    b. decay of foreign language lexical knowledge?
2. To what extent do aspects of the aptitude complex (i.e., associative memory capacity, PSTM, and phonetic coding ability) influence
    a. the learning process?
    b. the decay process?
3. Irrespective of form presentation mode, what is the role of learning burden in the decay process?

Learners of English studied 32 target vocabulary items using flashcard software, with target word form presented in one of two conditions: unimodal written and bimodal. Learners saw a target L2 definition and were tasked with producing the appropriate written form. After production, they rated their output as accurate or inaccurate. They could also indicate prior knowledge of target items at this stage. Keystroke logging software was used to allow post-hoc verification of this self-report evaluation. The effect of the form-presentation manipulation on learning burden (i.e., the number of exposures and the time of exposure needed to reach criterion) was calculated by computing mixed-effects models, while its impact on any decay was determined by computing logistic mixed-effects models. Aspects of LLA were measured using the LLAMA (Meara, 2005) battery and were included in statistical modelling. As with previous studies presented in this thesis, the role of learning burden irrespective of form presentation mode on the maintenance of target item knowledge was also considered by including the frequency and time of exposure necessary for learning as covariates in the decay analysis. This delineated the extent to which items that were easier to learn were better retained.

6.3 Methodology

6.3.1 Participants

The research presented in this thesis aimed to draw pedagogically meaningful implications for English language learners at tertiary institutions in the UK. Students from China, Vietnam, and Thailand represent a considerable proportion of this cohort (HESA, 2017). As such, the participants (N = 65) recruited for this study were L1 Chinese (Mandarin and Cantonese) (n = 33), Vietnamese (n = 23), and Thai (n = 9) students enrolled at a British university. Initially, 65 learners were recruited; however, three learners did not attend all data collection sessions, two learners were observed photographing the target words to allow for intersessional revision, and the data of two students were corrupted. Therefore, the final sample was 58 learners (11 males, 47 females). They had lived in an English-speaking environment for an average of 4.69 months (*SD* = 5.62 months), had been learning English for 11.75 years on average (*SD* = 5.67 years), and the average first contact with English was at 8.24 years old (*SD* = 3.12 years).

At the time of data collection, all learners were enrolled on either a pre-sessional or an in-sessional EAP course. As pre-sessional students aim to develop the requisite English proficiency to matriculate to university while in-sessional students have already attained or surpassed this level, proficiency varied between the learners. L2 vocabulary size has been shown to strongly correlate with L2 proficiency (Laufer & Aviad–Levitzky, 2017). Therefore, to control for learner proficiency, the vocabulary size test was administered and the scores included as a covariate in statistical analyses, as was also the case with Study 2. The results of the VST (levels 1-8) showed that the mean score was 68.866 words (*SD* = 12.281). Additionally, learners completed a self-report questionnaire that targeted perceived proficiency in each of the four skills. On a ten-point scale (one = extremely poor, almost no knowledge; ten = extremely good, almost native like) the mean scores were as follows: reading 6.47 (*SD* = 1.43); writing 5.24 (*SD* = 1.32); listening 6.50 (*SD* = 1.55); speaking 6.21 (*SD* = 1.45). The study was run in accordance with ethical approval (Appendix 16) and participants completed the approved consent form before taking part in the study (Appendix 3).

6.3.2 Instruments

Vocabulary Size Test

Vocabulary size was measured with bilingual versions of the Vocabulary Size Test (Nation & Beglar, 2007). Chinese L1 learners used the bilingual Chinese-English VST as was the case with Study 2. Vietnamese L1 learners used the bilingual Vietnamese-English VST (Nguyen & Nation, 2011) and Thai learners the bilingual Thai-English VST (Nirattisai & Palanukulwong, 2016). Like the original monolingual version, these bilingual versions contain 140 items, with ten items at each of the first fourteen 1000-word family frequency bands. The test embeds target items in a non-defining context and measures written-meaning recognition knowledge with a four-option multiple choice format. Although the validity arguments of the VST require greater exposition (see Schmitt, Nation, & Kremmel, 2020), as does the equivalence of the different forms, bilingual forms were chosen for two reasons. First, using the L1 precluded participants needing to understand complex grammatical structures such as relative clauses, which are often used to present word meaning in the monolingual form (e.g., *a weapon that explodes*; distractor for target item *Jug*) (Nguyen & Nation, 2011). This avoided the introduction of construct irrelevant variance (i.e., L2 grammatical competence). Second, the bilingual VSTs are cognitively less burdensome and are thus quicker to complete (Nguyen & Nation, 2011). This was beneficial given the extended nature of the study. Furthermore, although the equivalence of the bilingual versions has not been investigated in depth, there is similarly no research showing that the monolingual version functions equitably for the various populations (see Walker, 2011, for an outline of Differential Item Functioning). Therefore, it could not be assumed that use of the original VST would have led to a more robust comparison.

The three versions of the VST contain the same target items. Additionally, all tests employ the same distractor selection procedures utilising the original monolingual form. First, the L2 item being defined by a distractor on the monolingual English VST was determined. This word was then translated into the L1 with either an L1 equivalent or definition. Therefore, the tests were identical except for the language in which the key and distractors were written (see Figure 6.1). Unlike the Chinese and Vietnamese tests, the Thai version includes an *I don't know* option. This was excluded in the present study as it may have affected test-taker behaviour, reducing wild guessing and guessing based on partial knowledge. Neither of these effects is inherently problematic per se, but from a methodological perspective it would have produced results that could not have been compared to those from the Chinese or Vietnamese versions. The study employed a shortened version of the two instruments, which has been shown to effectively discriminate learners of different proficiencies (Nguyen & Nation, 2011) and provide sufficient variance for statistical analysis (see Study 2). The first eight levels were employed.

Figure 6.1

Example Items from the Vietnamese, Chinese, and Thai Versions of the VST Used in Study 3

3. Candid: Please be **candid**
   a. cẩn thận (careful)
   b. thông cảm (sympathetic)
   c. công bằng (equal)
   d. thẳng thắn (candid)

3. Candid: Please be **candid**
   a. 小心的 (careful)
   b. 表示同情的 (sympathetic)
   c. 公平的 (equal)
   d. 直率的 (candid)

3. Candid: Please be **candid**.
   a. ระมัดระวัง (careful)
   b. แสดงความเห็นอกเห็นใจ (sympathetic)
   c. ใหค้วามยุติธรรมกบั ทั้งสองฝ่าย (equal)
   d. พดู ในสิ่งที่คุณคิดจริงๆ (candid)

The Learning Software and Target Items

Like in Studies 1 and 2, *Anki* was used to present the target items to the learners. Importantly, *Anki* allows items to be presented in both written and spoken form, facilitating the comparison of unimodal and bimodal conditions. This software was loaded onto USB flash drives to allow for the collection of frequency and time of exposure data on a by-learner by-item basis. The target items were identical to those employed for the research described in Studies 1 and 2. There were 32 items in total that varied for word length (3, 6, 8, and 11 letters) and word class (nouns and verbs). Section 4.3.2 provides a detailed description of the selection criteria employed. Prior to the study, two L1 Chinese, Vietnamese, and Thai speakers highly proficient in English confirmed that none of the target items in this study are loanwords in these three languages. Due to the multilingual nature of the participants, this study used L2 definitions to convey target item meaning. These were the same definitions employed in Study 2. The definitions were graded so that the vast majority of the definitional language came from the most frequent 2,000 word families. Although the VST is a test of size and does not indicate coverage of any specific level, given the results (M = 6,886 word families), it is highly probable that all learners knew all of the definitional vocabulary. To confirm this assumption, at the end of the experiment, learners were asked whether they had any difficulty understanding the definitions. No students reported difficulty. The target items were embedded in four decks of eight words.

Half of the target forms were presented unimodally (i.e., the written form) and half bimodally (i.e., the written and spoken form). The spoken forms were recorded by a professional voice actor and subsequently checked for intelligibility by an L1 speaker of English. In the unimodal condition, learners were presented with the written form only, while in the bimodal condition, the written form was followed by the spoken form after an interval of 400 ms. The length of delay was considered sufficiently brief to discourage learners from attempting to produce the spoken form prior to hearing it, and long enough to allow interaction with the target form prior to aural presentation. Initial piloting with ten participants of similar characteristics to the participants of the main study confirmed that this was a suitable length of delay.

The flashcards were organised so that all cards in one deck were presented either unimodally or bimodally. A within-participant counter-balanced design was employed. Half the participants learned the target words in each deck via unimodal form presentation and half through bimodal form presentation. Decks were ordered so that the form presentation condition alternated between decks. Furthermore, the order in which the learners engaged with the decks varied by session. This is illustrated in Figure 6.2. 32 learners were randomly assigned to Group A and 26 to Group B. The imbalance in group size was caused by participant attrition in Group B.

As in Studies 1 and 2, the flashcards were set up productively so that learners were tasked with producing the target form to match a given meaning. After production, learners saw the correct form and evaluated their production using three options: *Again, Good,* and

Easy. *Again* was used when a learner had incorrectly produced or did not know a target form. This resulted in a target item being recycled within a period of one minute. *Good* was used when a learner had accurately produced a target form. This option resulted in a target item being recycled within a ten-minute period. *Easy* was used for items known to a learner prior to starting the experiment. In reality, we would expect a learner to select *Again* initially and then *Good* as their knowledge of a target item developed. An extensive description of the learning software is given in Chapter 3. Throughout the experiment, learners were monitored, and I was on hand to answer any questions about the learning software.

Figure 6.2

Counter-Balanced Form Presentation in Study 3

	Group A		Group B	
Session 1:	Deck	Condition	Deck	Condition
	1	unimodal	1	bimodal
	2	bimodal	2	unimodal
	3	unimodal	3	bimodal
	4	bimodal	4	unimodal
Session 2:	Deck	Condition	Deck	Condition
	3	unimodal	3	bimodal
	4	bimodal	4	unimodal
	1	unimodal	1	bimodal
	2	bimodal	2	unimodal

Keystroke Logging

An important methodological limitation of Studies 1 and 2 was that while learning target items on the *Anki* platform, production was evaluated as correct or incorrect by the learners themselves. This was a limitation because this self-evaluation could have been wrong. In order to overcome this limitation, in the present study I used keystroke logging software to check the accuracy of these self-evaluations. Keystroke logging software records writing activity as text is composed on a computer (Spelman Miller & Sullivan, 2006). The data produced by keystroke logging software include all keyboard actions (i.e., scrolling, deletion, cursor navigation, and the production of text) and mouse clicks. In addition, the temporal distribution of these activities is recorded, facilitating analyses of pause duration and fluency of lexical production. This software has been extensively used with L1 and L2 writing research to unobtrusively observe cognitive aspects of written language production. This is because it produces detailed data that allow researchers to closely monitor and reconstruct computer-based writing processes in an ecologically valid manner (Leijten & Van Waes, 2006).

Several logging software are used in the literature. Some are associated with specific research areas (e.g., Translog for research in translation studies) and/or methodologies (e.g., Scriptlog is often used in combination with eye-tracking). This study used keystroke logging to record participant interaction with the flashcard software. As such, it was important that the chosen platform function outside the word-processing space. The vast majority of logging programmes are limited to proprietary word-processing environments (Spelman Miller & Sullivan, 2006) and thus were unsuitable for this experiment. One programme that can operate outside word-processing environments is Inputlog 5.2 (Leijten & Van Waes, 2013). This platform was selected to record keyboard and mouse actions while learners were interacting with the flashcards.

To date, Inputlog has been used to investigate numerous aspects of writing. For example, it was used to investigate pauses and linguistic processing in video game written communication (Chukharev-Hudilainen, 2014), to determine the effect of task on the linguistic complexity and behaviour of L2 writers (Révész et al., 2017), and to investigate the effect of electronic outlining on the writing of argumentative essays (de Smet et al., 2014). To the best of my knowledge however, no study to date has used keystroke loggers in collaboration with flashcard software. Importantly, the use of Inputlog did not obtrude on the learning process; thus, the ecological validity of the study was maintained.

The primary research goal in utilising keystroke logging software was to validate the self-report evaluation participants conducted on their output within the flashcard software. After typing a response to a definitional stimulus, participants rated their production as correct or incorrect using the keys *1*, *2, or 3*, proxies for the *Again*, *Good*, and *Easy* rating buttons. Keystroke logging software was employed to assess the reliability of this self-evaluation by comparing learner evaluation (i.e., *1* or *2*) to actual production (as recorded by the keystroke logging software). For example, Figure 6.3, an example output, demonstrates the following process: the learner pressed ENTER to see the target definition; the learner typed *ponifica* in response to the definition for the item *pontificate*; the learner pressed ENTER for a second time to reveal the correct form; the learner evaluated his/her production as inaccurate (i.e., *1*). Thus, in this instance, this learner correctly evaluated his/her production (i.e., the output was inaccurate and was rated as such).

Figure 6.3

Example Output (Trimmed) of Inputlog in Study 3

#Id	Event Type	Output	StartTime	EndTime	ActionTime	PauseTime
142	keyboard	RETURN	387217	387297	80	4424
143	keyboard	p	387424	387505	81	207
144	keyboard	o	388105	388224	119	681
145	keyboard	n	388497	388601	104	392
146	keyboard	i	388769	388921	152	272
147	keyboard	f	388921	389024	103	152
148	keyboard	i	389281	389393	112	360
149	keyboard	c	389505	389673	168	224
150	keyboard	a	389793	389936	143	288
151	keyboard	RETURN	389985	390105	120	192
152	keyboard	1	392248	392305	57	2263

A random sample of ten participants from the study was used to confirm the validity of the self-report evaluation data, involving 3104 observations. For each learner, the key-stroke logging data was manually inspected and compared to the self-report evaluations. A correct evaluation was awarded a score of one and an incorrect evaluation a score of zero. The mean accuracy of the self-report evaluations in this data set was .99 (*SD* = .01). This showed that, in most cases, learners accurately evaluated their production. This finding supported the validity of the methodology employed in this study; in particular, it suggested that the self-report procedure adequately differentiated exposures during the learning process from retrievals of learned knowledge, allowing a dynamic approach to exposure frequency while controlling the number of retrievals.

Attainment and Retention Instruments

The assessment battery was identical to that used for Studies 1 and 2, and is described in detail in Section 4.3.6. Measures of form recall and form recognition were administered immediately after the second learning session (i.e., the immediate test).

Learners saw the target definitions and were tasked with producing or recognising the written form. Each instrument contained 40 items: the 32 target items and eight additional items included to measure a test effect. The test was adaptive; any item correctly produced on the form recall instrument was omitted from the measure of form recognition. This was to avoid giving positive feedback on learned knowledge that might have biased the retention data. Thus, the assessment battery contained 80 items in total, but participants only saw all items if they incorrectly produced all 40 items on the form-recognition instrument. The assessment battery was administered again two weeks after the immediate test. The order in which the target items were presented was varied to prevent a sequence effect.

Language Learning Aptitude

Three tests from the LLAMA battery (Meara, 2005) were employed. This tool is based on the MLAT (Meara, 2005) which was developed by Carrol and Sapon (1959), adopting a componential understanding of LLA. It includes instruments assumed to measure associative memory capacity, grammatical inferencing, phonetic coding ability, and PSTM. The last of these measures was not included on the MLAT but was added in light of research demonstrating the relationship between sound recognition ability and language learning (Engel de Abreu & Gathercole, 2012; Service, 1992).

Three tests, LLAMA_B, LLAMA_D, and LLAMA_E were employed in this study and are described below. LLAMA_B is a measure of associative memory capacity and is claimed to test vocabulary learning ability (Meara, 2005). It assesses the capacity to ascribe novel meanings to novel forms. Learners have two minutes to examine 20 novel forms and their object referents. The forms are taken from a Central American language and are thus unlikely to be known to most learners. The learning procedure involves clicking on an object to reveal the associated written form in the centre of the screen. Learners have two minutes to interact with all 20 target items. Upon completion of the learning period, learners are then presented with a written form and are required to select the associated character from the 20 on screen. When an item is selected, test-takers hear a sound indicating the accuracy of their selection. Test-takers earn five points for each correct answer; thus, the maximum score is 100. Figure 6.4 shows the interface for LLAMA_B.

LLAMA_B is based on the MLAT paired-associate instrument; however, there are several key differences that affect test difficulty. LLAMA_B does not present items linearly but requires learners to click on a shape to reveal the related form; novel meanings are not English equivalents like the MLAT but characters unknown to test-takers; the test uses all 20 target items as distractors unlike the multiple-choice format of the MLAT. These differences suggest that LLAMA_B is a more difficult test than the MLAT paired-associate instrument. Indeed, piloting in preparation for this study showed that the latter was more likely to result in a ceiling effect. Thus, the LLAMA_B was chosen to measure associative memory capacity.

Figure 6.4

The LLAMA_B interface (Study 3)

LLAMA_D is assumed to measure PSTM. This aspect was not considered in older LLA testing batteries (e.g., the MLAT); however, its inclusion is justified in the LLAMA suite as follows:

> a learner who is able to recognise repeated stretches of sound is more likely to notice small variations in speech, and this makes it easier for them to isolate the individual words and variants of these words that signal morphology (Rogers et al., 2017, p. 51)

Thus, LLAMA_D assesses test-takers ability to differentiate sounds that they have heard previously from novel forms. There are two parts to this test. In Part One, test-takers hear ten novel forms taken from a British Columbian Indian language. These forms are unlikely to be known to participants. In Part Two, participants hear twenty forms, some of which are repetitions of the sounds played in Part One while others are previously unheard items. Test-takers decide on a by-item basis whether a form was played in Part One. Judgements are made by selecting a smiley face for a repetition or a neutral face for a previously unheard form. Five points are awarded for a correct answer. The maximum score for the LLAMA_D is 100. The interface for the LLAMA_D is presented in Figure 6.5.

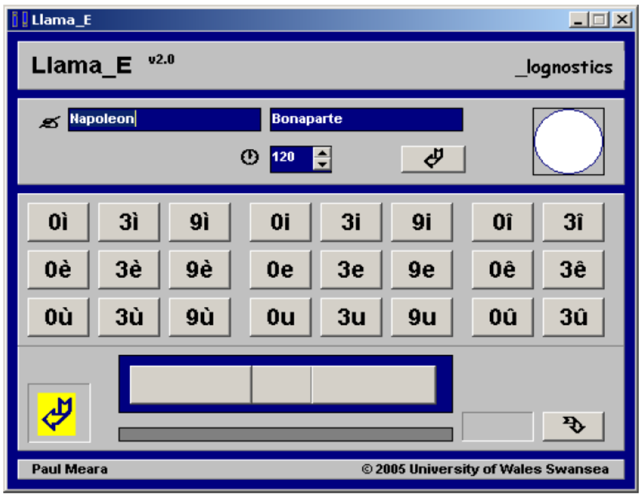
Figure 6.5

The Interface for LLAMA_D (Study 3)

Finally, LLAMA_E is a sound-symbol correspondence task assumed to measure phonetic coding ability. This instrument is based on the phonetic coding measure on the MLAT. Test-takers see an interface consisting of 24 buttons with alphabetic script in unusual combinations. During the two-minute learning phase, they click on buttons to hear the spoken monosyllabic form associated with the orthographic label. They are tasked with encoding the sound-symbol correspondences. During the testing phase, test-takers hear disyllabic combinations of the labels and, at the same time, they are presented with two orthographical possibilities from which to choose the matching written form. There are 20 items in total and a maximum score of 100. Figure 6.6 presents the interface for LLAMA_E.

Figure 6.6

The Interface for LLAMA_E (Study 3)



Researchers rightfully recognise that a validity argument has yet to be sufficiently made for this tool. For example, Skehan (personal communication) emphasised the necessity of reliability and validity analyses to allow for confident interpretation of test results. Furthermore, the developer of the LLAMA suite, Paul Meara (2005), himself stated that "The [tests are] exploratory versions of on-going research and they should NOT be used in high-stakes situations where accuracy and reliability are at a premium" (p. 21). The initial validation work that has been conducted showed that the battery seems robust but that scores are influenced by two learner variables, age and L2 status (Rogers, see Meara & Mirelpeix, 2016; Rogers, Meara, Barnett-Legh, Curry, & Davie, 2017). Importantly, neither of these variables are relevant to the sample in this study. Furthermore, for learner L1 and the number of foreign languages known, two learner variables that are pertinent to the sampling criteria employed in this study, Rogers *et al*. (2016) found no effect.

Despite the need for more validation evidence, the aptitude battery has been widely used (Granena & Long, 2012; Saito, 2017). In fact, Rogers *et al.* (2017) report that the LLAMA suite had 700 citations on Google Scholar in 2013 and has been used to answer a variety of questions. Therefore, although more evidence is needed to clarify the construct validity of this tool (Rogers et al., 2017), it is widely accepted and has been extensively used by the SLA community. For these reasons, it was adopted for use in this study.

Specifically, tests B, D, and E were employed. These tests were chosen as their associated constructs are directly relevant to lexical acquisition and the methodology employed in this study. LLAMA_B, for example, is considered to be a measurement of intentional vocabulary learning (Meara, 2005), while LLAMA_D and LLAMA_E consider a learner's capacity to remember sound and associate it with written symbols. As this study presented target items in unimodal and bimodal conditions, it was hypothesised that the two aspects of LLA relevant to spoken input may impact attainment in the bimodal condition. In contrast, the other test in the LLAMA suite, LLAMA_F, relates to inferencing grammatical patterns and was not relevant to this research. The three tests were loaded onto each participant's USB flash drive and administered electronically. A document explaining the software, a printed copy of which was also provided to each learner (see

Appendix 18), and an instructional video were included on the USB. Participants watched the video pertaining to a specific test (i.e., B, D, or E) immediately before completing that test. The researcher monitored closely, and participants were able to ask any questions prior to starting a test.

Language Background Questionnaire

As with previous studies described in this thesis, the participants completed a language background questionnaire (see Appendix 17). This was a similar instrument to that described in Section 4.3.5, but an additional question relating to the frequency of flashcard use was added for this study. This latter item facilitated consideration of familiarity with flashcard software as a covariate in the analysis of learning burden and decay.

Target Language Survey

As with previous studies, after the final test had been administered participants completed a survey in which prior knowledge of, and intersessional contact with, the target language was reported. Items which were reported known or seen between sessions were deleted from statistical analysis.

6.3.3 Procedure

Participant recruitment was conducted via email using a convenience sampling procedure: learners who fulfilled the sampling criteria (i.e., L1, place of study) were contacted. Participation in the study was entirely voluntary and learners received no compensation for participation. Prior to beginning the study, learners received a description of the project and were given the opportunity to ask questions. They then completed consent forms in line with the ethical approval granted (see Appendix 16). The study took place over three sessions. Prior to the first session, flash drives were prepared with the learning software, Inputlog, the VST, the immediate test, the delayed test, the survey instrument, LLAMA B, LLAMA D, and LLAMA E, and the LLAMA instructional video installed. As with Studies 1 and 2, tests were in a password protected folder to prevent learners from accessing them before the instruments were administered. During the first session, participants completed the background questionnaire and then USBs were distributed. The participants were inducted into the use of the learning software and had an opportunity to ask questions. They were informed about the immediate and delayed posttests, and the different modes of form presentation. Following the induction, participants completed a practice set of items, once again having the opportunity to ask questions. After all questions had been answered, the keystroke logging software was activated, and participants studied the 32 target items in four decks of eight cards on *Anki*. They used headphones to listen to the spoken form of the target items. Individual computers were separated from neighbouring machines by partitions to ensure no distracting interaction between participants. There was no time limit and learners worked at their own pace. The researcher monitored student activity to ensure the intended procedure was followed. After completing the learning task, participants returned their USBs.

Session Two took place the next day. Participants collected their USBs, worked through some practice items on the learning software, and had the opportunity to ask procedural questions. Subsequently, Inputlog was activated and participants relearned the same 32 items from Session One on *Anki*. Again, items were organised into four decks of eight words. After the Session Two learning tasks were finished, the immediate test was administered. This marked the last interaction with the target items prior to the delayed test in Session Three.

Session Three took place two weeks later. The length of the retention interval was selected as such durations are common in vocabulary learning studies (deHaan et al., 2010; Grace, 2000). Additionally, as the studies presented thus far in this thesis have investigated retention intervals of four weeks and one week respectively, employing an intermediary duration allowed for the impact of retention interval length to be considered during the general discussion (see Chapter 7). Furthermore, it was hoped that the brevity of the interval would mitigate a floor effect on the recall instrument (found when a four-week retention interval was employed), while ensuring that sufficient decay occurred to enable analysis. Session Three began with administration of the delayed test, with data saved to flash drives. There was no time limit and participants worked at their own pace. Following completion of this instrument, the survey was conducted to determine prior knowledge of the target items and any intersessional item engagement. Next, learners completed the VST, choosing the form according to their L1. Finally, the instructional LLAMA video was shown, and learners worked through the three tests. Each session lasted approximately one hour, but learners completed the activities at their own pace.

6.3.4 Data Analysis

As with Study 2, learning burden was measured in two ways: the number of exposures needed to learn an item and the total time that participants studied a target item. These metrics were pulled from the learning software by performing a join between three tables in the SQLite database file of each learner. This file was automatically produced by the learning software. The various data sets (i.e., frequency of exposure, time of exposure, keystroke logging, immediate test, delayed test, target language survey) were concatenated and organised by learner. Items for which a learner indicated prior knowledge or intersessional exposure on the target language survey, or items which the frequency of exposure data showed had been viewed only twice were removed. The former procedure resulted in the deletion of 87 data points. The latter protocol was adopted as two exposures equated to a learner selecting *Easy* during both Session One and Session Two, indicating either prior knowledge (learners were instructed to select *Easy* if they were familiar with an item) or disengagement with the learning process. Seven data points were removed as a result of this procedure. Following this initial trimming, the number and time of exposure data was winsorised to mitigate the impact of outliers (Field, 2012). This process is outlined in Section 5.4.4. The effect of winsorising in indicated in Table 6.1.

Table 6.1

Descriptive Statistics to Illustrate the Effect of Winsorising on the Learning Burden Data (Study 3)

	M	SD	Min	Max	Range
Frequency of exposure					
Before winsorizing	8.89	5.026	3	39	36
After winsorizing	8.78	4.519	3	25	22
Length of exposure (secs)					
Before winsorizing	105.98	81.55	13.96	585.34	571.38
After winsorizing	104.95	77.37	13.96	374.31	360.35

The immediate and delayed tests contained eight items included to measure a test effect. These data were compared to determine the amount of learning that occurred from interaction with the target language on the immediate test. The results of this comparison are presented in Table 6.2. The results suggested that minimal learning resulted from the immediate test.

Table 6.2

Mean Scores (SD in brackets) for Items Included to Measure Learning from the Immediate Test (Study 3)

	Items answered correctly (max = 8)		
	Strict form recall	Lenient form recall	Form recognition
Immediate test	0.02 (0.13)	0.02 (0.13)	2.24 (1.84)
Delayed test	0.03 (0.18)	0.07 (0.32)	3.04 (1.82)

Data analysis was conducted in the statistical platform *R* (2016). Prior to statistical modelling, all continuous variables were log transformed. As with Study 2, to determine the effect of the target variables on learning burden a series of linear mixed effects models were computed using lme4 (Bates et al., 2015). Separate models were calculated for the two measures of learning burden (number of exposures and time of exposure) needed for learning to occur at three degrees of word-form knowledge: strict recall, lenient recall, and recognition. Thus, six models were constructed in total (two outcome variables x three conditions). Only the frequency and length data of items for which knowledge was demonstrated on the immediate test were included; for example, the lenient recall analysis only considered the learning burden of items for which knowledge was demonstrated at the level of lenient form recall. The most parsimonious model was determined by a backward elimination procedure (see Section 4.3.8). Details of the model selection process are given in Appendix 19.

To answer Research Questions 1a and 2a, the following regressors were fitted to models of the frequency and time of exposure data sets: form presentation mode, PoS, word length, the VST score, L1, L3, familiarity with flashcards, group, LLAMA B, LLAMA D, LLAMA E. Additionally, the interactions between form presentation mode and the three dimensions of aptitude were included in analysis. This facilitated consideration of the possible moderating role of aptitude on the efficacy of form presentation mode; for example, this allowed me to determine whether learners with a specific aptitude construction benefitted more than other learners from the presence of bimodal form presentation. Details of the specific models fitted to the two outcome variables (i.e., number and time of exposure) at each strength of knowledge (i.e., strict and lenient form recall, and form recognition) are given in the next section.

To answer Research Questions 1b and 2b, Generalised Logistic Mixed Effects Models were fitted to the delayed test data set. As with Studies 1 and 2, logistic models were fitted because the outcome variable was binomial: a correct or an incorrect answer on the delayed test. Separate models were fitted at the level of strict form recall, lenient form recall, and form recognition. Importantly, only those items shown to have been learned on the immediate test were included in the analysis. The best-fitting models were again determined by backward-elimination involving comparison of iterative models to establish the most parsimonious model. The following covariates were included in the maximal model: form presentation mode, PoS, word length, VST score, L1, L3, familiarity with flashcards, group, LLAMA B, LLAMA D, LLAMA E. Interactions between aptitude and form presentation mode were also included in analysis. Additionally, indices of learning burden were included in the decay analysis. Consideration of the number of exposures and the time of exposure needed for learning in the models of decay allowed me to determine whether words that were quicker to learn were also those better retained. In doing so, I was able to answer Research Question 3. Details of the model selection process are provided in Appendix 20.

6.4 Results

This section presents a description of the models computed to delineate the effect of form presentation mode (Research Question 1a) and language learning aptitude (Research Question 2a) on learning burden. It then introduces the models computed to understand the role of form presentation mode (Research Question 1b), aptitude (Research Question 2b), and learning burden (Research Question 3) on the decay of target language knowledge.

6.4.1 Learning Burden

Results are presented by strength of knowledge: strict written form recall, lenient written form recall, and written form recognition. Learning burden was considered using two metrics, the frequency of exposure and the time of exposure, the analysis of which are presented separately.

Descriptive statistics for the frequency and time of exposure needed to learn the target language at the three levels of measurement sensitivity are presented by condition in Table 6.3. The data showed that learners needed, on average, approximately eight exposures to the target items to learn them and that this process took between 93 and 104 seconds depending on the level of knowledge required by the immediate test. The data also showed that more exposures were needed for items learned to the level of written form recognition than written form recall. Intuitively, we expect more exposures to be needed for items to be learned to the level of form recall as this is a more challenging component (Laufer & Goldstein, 2004). However, given this higher level of challenge, form recognition instruments are more sensitive to partial form knowledge than measures of form recall, which require a greater level of mastery. The data showed that more items were learned to the level of form recognition than form recall (see Score in Table 6.3) and that words with the heaviest learning burden were more likely to be learned to the level of form recognition than form recall (see Frequency and Time in Table 6.3). Thus, the data showed that the form recognition instrument better captured knowledge of the most burdensome words. One more general observation with regard to Table 6.3 is that there seemed to be considerable individual variance, as indicated by the high standard deviations, particularly in terms of the time needed for learning, in both the unimodal and bimodal conditions.

Table 6.3

Mean Scores for the Immediate Test (Score), Frequency of Exposure (Freq.), and Time of Exposure (Time) by Mode of Form Presentation. Standard Deviations Are Presented in Brackets (Study 3)

	Unimodal			Bimodal			Total		
	Score	Freq.	Time	Score	Freq	Time	Score	Freq	Time
Strict form recall	11.25 (4.07)	8.03 (3.55)	93.79 (66.87)	11.07 (4.16)	7.90 (3.37)	92.64 (63.56)	22.33 (7.71)	7.97 (3.47)	93.22 (65.25)
Lenient form recall	12.09 (3.88)	8.37 (3.98)	99.58 (73.58)	11.97 (3.97)	8.08 (3.59)	95.79 (66.39)	24.05 (7.29)	8.23 (3.79)	97.69 (70.12)
Form recognition	14.66 (2.38)	8.86 (4.49)	106.88 (79.63)	14.05 (3.06)	8.54 (4.26)	102.61 (74.32)	28.71 (5.09)	8.69 (4.78)	104.78 (79.05)

Inspection of the LLA results indicated that the data were not normally distributed. This is a common issue with the LLAMA instrument and can be solved by splitting participants into two groups (Saito, personal communication). Numerous possibilities for determining the cut scores for this split were considered. For example, Meara (2005) provides bands to indicate a poor, average, and good score. However, converting a band into a single score can be problematic as arbitrary decisions as to the exact placement of the cut score need to be made. For instance, one would need to decide whether the cut score should be at the lowest threshold, the upper threshold, or the middle point of a band in order to determine a specific cut score. Another possibility was using the median score from the sample to split the participants; however, this option was rejected as, without reference to norming data, a median split can lead to unreliable and ungeneralizable findings. Finally, the norming data from Bokander and Bylund (2019) was used to perform the split. They administered the LLAMA instrument to 350 participants with many different L1s represented. They report the following mean scores: LLAMA B = 50.13; LLAMA D = 29.66; LLAMA E = 69.29. These means were used to split the participants into two groups. Learners who scored above the normed mean score presented in Bokander and Bylund (2019) on a specific instrument were placed into one group (High) and those who scored below the mean score were placed into another group (Low). Separate splits were performed for each aspect of aptitude, meaning that a learner could be in the high group for one element and the low group for another.

Table 6.4 displays descriptive statistics for LLAMA B (associative memory capacity), Table 6.5 shows LLAMA D (PSTM), and Table 6.6 presents the data for LLAMA E (phonetic coding ability). Results are presented according to the two form presentation conditions to facilitate consideration of a possible interaction between form presentation mode and aptitude. An initial inspection of the results suggests that the learners with higher levels of aptitude learned the target items more quickly than those with lower levels of aptitude. This seems to have particularly been the case with associative memory capacity; for example, learners with low associative memory capacity required 1.62 times the exposure needed by learners with high associative memory capacity to acquire items to the level of strict form recall (similar figures were also found for lenient form recall and form recognition).

Table 6.4

Mean Scores for the Immediate Test (Score), Frequency of Exposure (Freq.), and Time of Exposure (Time) by Mode of Form Presentation and Associative Memory Capacity. Standard Deviations Are Presented in Brackets (Study 3)

		Unimodal			Bimodal			Total		
	LLAMA B	Score	Freq.	Time	Score	Freq.	Time	Score	Freq.	Time
Strict form recall	High	12.60 (3.12)	6.61 (2.32)	66.09 (39.34)	12.40 (3.55)	6.52 (1.75)	68.72 (40.69)	25.00 (6.35)	6.57 (2.08)	67.40 (40.07)
	Low	10.55 (4.37)	8.93 (3.88)	111.21 (75.38)	10.37 (4.33)	8.77 (3.83)	107.69 (70.43)	20.92 (8.06)	8.85 (3.85)	109.46 (72.41)
Lenient form recall	High	13.15 (2.98)	6.64 (2.30)	67.20 (41.40)	13.25 (3.21)	6.65 (1.96)	71.07 (44.06)	26.40 (5.92)	6.64 (2.15)	69.15 (42.83)
	Low	11.52 (4.20)	9.40 (4.40)	119.02 (81.50)	11.29 (4.20)	8.97 (4.06)	111.05 (72.98)	22.82 (7.70)	9.19 (4.24)	115.08 (77.44)
Form recognition	High	14.60 (2.93)	6.74 (2.30)	70.43 (42.31)	14.40 (3.33)	6.75 (2.13)	74.72 (47.85)	29.00 (6.19)	6.74 (2.88)	72.58 (47.92)
	Low	14.68 (2.08)	9.95 (4.93)	125.73 (87.50)	13.87 (2.93)	9.52 (4.78)	117.85 (81.48)	28.55 (4.49)	9.74 (5.26)	121.89 (87.11)

Table 6.5

Mean Scores for the Immediate Test (Score), Frequency of Exposure (Freq.), and Time of Exposure (Time) by Mode of Form Presentation and PSTM. Standard Deviations Are Presented in Brackets (Study 3)

		Unimodal			Bimodal			Total		
	LLAMA D	Score	Freq.	Time	Score	Freq.	Time	Score	Freq.	Time
Strict form recall	High	12.06 (3.55)	7.60 (3.12)	87.44 (58.00)	11.61 (3.65)	7.80 (3.25)	90.31 (60.90)	23.68 (6.62)	7.69 (3.18)	88.85 (59.42)
	Low	10.33 (4.49)	8.62 (3.98)	102.32 (76.45)	10.44 (4.48)	8.03 (3.52)	95.61 (66.80)	20.78 (8.66)	8.32 (3.78)	98.95 (71.83)
Lenient form recall	High	12.90 (3.21)	7.78 (3.35)	90.39 (61.32)	12.61 (3.57)	7.84 (3.29)	91.23 (61.16)	25.52 (6.47)	7.81 (3.32)	90.81 (61.20)
	Low	11.15 (4.40)	9.15 (4.58)	111.79 (85.85)	11.22 (4.33)	8.39 (3.94)	101.67 (72.26)	22.37 (8.31)	8.77 (4.29)	106.71 (79.47)
Form recognition	High	14.90 (2.23)	8.19 (3.83)	96.61 (68.21)	14.52 (3.05)	8.32 (4.07)	98.77 (70.42)	29.42 (4.91)	8.25 (4.37)	97.68 (71.42)
	Low	14.37 (2.56)	9.65 (5.05)	119.07 (89.92)	13.52 (3.03)	8.81 (4.52)	107.34 (78.69)	27.89 (5.26)	9.24 (5.19)	113.36 (86.75)

Table 6.6

Mean Scores for the Immediate Test (Score), Frequency of Exposure (Freq.), and Time of Exposure (Time) by Mode of Form Presentation and Phonetic Coding Ability. Standard Deviations Are Presented in Brackets (Study 3)

		Unimodal			Bimodal			Total		
	LLAMA E	Score	Freq.	Time	Score	Freq.	Time	Score	Freq.	Time
Strict form recall	High	11.86 (3.78)	7.52 (3.10)	82.52 (57.52)	11.72 (3.82)	7.51 (2.94)	84.54 (57.17)	23.58 (7.19)	7.51 (3.02)	83.52 (57.09)
	Low	10.27 (4.42)	9.00 (4.10)	115.12 (78.06)	10.00 (4.56)	8.67 (3.97)	108.17 (71.92)	20.27 (8.25)	8.84 (4.05)	111.69 (75.19)
Lenient form recall	High	12.56 (3.55)	7.66 (3.25)	85.37 (60.50)	12.58 (3.37)	7.65 (3.19)	87.17 (60.68)	25.14 (6.47)	7.65 (3.22)	86.27 (60.56)
	Low	11.32 (4.34)	9.65 (4.79)	125.39 (87.12)	10.95 (4.72)	8.91 (4.14)	111.99 (73.39)	22.27 (8.31)	9.28 (4.51)	118.79 (80.95)
Form recognition	High	14.89 (2.15)	8.20 (4.00)	93.39 (68.86)	14.42 (2.88)	7.96 (3.64)	94.13 (67.71)	29.31 (4.86)	8.09 (4.25)	93.76 (70.13)
	Low	14.27 (2.73)	9.96 (5.03)	129.69 (90.76)	13.46 (3.31)	9.54 (5.01)	117.47 (82.72)	27.73 (5.42)	9.76 (5.45)	123.75 (89.73)

These initial observations were verified by statistical modelling. Mixed effects models were fitted using the target variables and numerous covariates as predictors, and frequency/time of exposure as outcomes. The models relevant to the frequency of exposure data are presented first, followed by the models fitted to the time of exposure data.

Frequency of Exposure

Table 6.7 presents the analysis of the frequency of exposure data in the strict written form recall condition. The best-fitting model showed that there was a statistically significant effect for the mode of form presentation (t = -2.41, p = .02), with items in the bimodal condition (7.90 [3.37]) requiring fewer exposures than those in the unimodal condition (8.03 [3.55]). In terms of language learning aptitude, a statistically significant effect was found for associative memory capacity (LLAMA B) (t = -3.06, p = .003) showing that learners in the high group required fewer exposures (6.57 [2.08]) to learn the target items than those in the low group (8.85 [3.85]). With regard to PSTM (LLAMA D) and phonetic coding ability (LLAMA E), there was a trend for more apt participants to learn the target vocabulary in fewer exposures (indicated by the negative t score); however, these effects were not statistically significant (PSTM: t = -1.72, p = .09; Phonetic coding ability: t = -1.50, p = .14). There was also a statistically significant interaction between mode of form presentation and PSTM (t = 2.30, p = .02). Comparison of the descriptive statistics for the high and low groups showed a greater difference between the frequency of exposure needed for items learned in the unimodal condition (High: 7.60 [3.12]; Low: 8.62 [3.98]) and those learned in the bimodal condition (High: 7.80 [3.25]; Low: 8.04 [3.52]). Of the covariates considered, statistically significant effects were found for word length, PoS, and vocabulary size. This showed, respectively, that longer words required more exposures than shorter words, verbs required more exposures than nouns, and learners with larger vocabularies needed fewer exposures to learn the target items than learners with comparatively smaller vocabulary sizes.

Table 6.8 details the best-fitting model for lenient written form recall. Overall, the findings matched those reported for strict form recall. A statistically significant effect was found for the mode of form presentation (t = -2.72, p = .007) indicating that items learned in the bimodal condition (8.08 [3.56]) required fewer exposures that those learned in the unimodal condition (8.37 [3.98]). There was also a statistically significant effect for associative memory capacity (t = -2.96, p = .004); the High group (6.64 [2.15]) needed fewer exposures to learn the target items than the Low group (9.19 [4.24]). No statistically significant effects were found for either PSTM (t = -1.79, p = .08) or phonetic coding ability (t = -1.57, p = .12); however, there was a significant interaction between PSTM and form presentation mode (t = 2.16, p = .03). Upon inspection of the descriptive statistics, the reason for this is the greater difference between the scores of the High (7.78 [3.35]) and Low (9.15 [4.58]) PSTM groups in the unimodal condition than the bimodal condition (High: 7.84 [3.29]; Low: 8.39 [3.94]. Additional statistically significant effects were found for word length (longer items required more exposures than shorter items), PoS (verbs needed more exposures than nouns), and learner vocabulary size (vocabulary size negatively correlated with frequency of exposure).

Table 6.7

Fixed and Random Effects for the Selected Model - Frequency of Exposure, Strict Condition (Study 3)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	4.20	0.64	6.57	<.001***
Mode of form presentation	-0.07	0.03	-2.41	.02*
LLAMA B	-0.19	0.06	-3.06	.003**
LLAMA D	-0.09	0.06	-1.72	.09
LLAMA E	-0.09	0.06	-1.50	.14
Familiarity with flash cards	-0.12	0.06	-1.91	.06
Word length	0.22	0.03	6.85	<.001***
PoS	0.09	0.03	2.96	.006**
Vocabulary size	-0.55	0.15	-3.70	<.001***
LLAMA B * Mode of form presentation	-0.01	0.03	0.01	.99
LLAMA D * Mode of form presentation	0.07	0.03	2.30	.02*
LLAMA E * Mode of form presentation	0.03	0.03	0.94	.35
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	0.01	0.08	
Participant	Intercept	0.03	0.19	

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

Table 6.8

Fixed and Random Effects for the Selected Model - Frequency of Exposure, Lenient Condition (Study 3)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	4.28	0.69	6.23	<.001***
Mode of form presentation	-0.08	0.03	-2.72	.007**
LLAMA B	-0.21	0.07	-2.96	.004**
LLAMA D	-0.11	0.06	-1.79	.08
LLAMA E	-0.10	0.06	-1.57	.12
Familiarity with flash cards	-0.12	0.07	-1.76	.09
Word length	0.24	0.03	7.00	<.001***
PoS	0.09	0.03	2.68	.01*
Vocabulary size	-0.58	0.16	-3.56	<.001***
LLAMA B * Mode of form presentation	-0.01	0.03	0.41	.68
LLAMA D * Mode of form presentation	0.07	0.03	2.16	.03*
LLAMA E * Mode of form presentation	0.03	0.03	0.89	.37
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	0.01	0.08	
Participant	Intercept	0.04	0.20	

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

Table 6.9 presents the analysis of the frequency of exposure data for items learned to the level of written form recognition. The best-fitting model indicated similar findings to the results of written form recall. Statistically significant effects were found for mode of form presentation ($t = -2.59$, $p = .009$) and associative memory capacity ($t = -2.58$, $p = .01$). The results indicated that the bimodal condition (8.54 [4.26]) necessitated fewer exposures that the unimodal condition (8.86 [4.49]) and learners with comparatively low associative memory capacity needed more exposures to learn the target items (High: 6.74 [2.88]; Low:

9.74 [5.26]). However, no statistically significant effects were found for PSTM (t = -1.04, p = .30) and phonetic coding ability (t = -0.54, p = .59). A statistically significant interaction was found between PSTM and form presentation mode (t = 2.39, p = .02); there was a greater difference between the average frequency of exposure between the High (8.19 [3.83]) and Low (9.65 [5.05]) groups in the unimodal condition than in the bimodal condition (High: 8.32 [4.07]; Low: 8.81 [4.52]). Secondary findings include the impact of word length (longer items required more exposures), PoS (verbs required more exposures), and vocabulary size (learners with smaller vocabulary sizes needed more exposures).

Table 6.9

Fixed and Random Effects for the Selected Model - Frequency of Exposure, Recognition Condition (Study 3)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	4.17	0.77	5.44	<.001***
Mode of form presentation	-0.07	0.03	-2.59	.009**
LLAMA B	-0.21	0.08	-2.58	.01*
LLAMA D	-0.07	0.06	-1.04	.30
LLAMA E	-0.04	0.07	-0.54	.59
Word length	0.24	0.03	6.57	<.001***
PoS	0.09	0.04	2.54	.02*
Vocabulary size	-0.59	0.18	-3.18	.002**
LLAMA B * Mode of form presentation	0.02	0.03	0.66	.51
LLAMA D * Mode of form presentation	0.07	0.03	2.39	.02*
LLAMA E * Mode of form presentation	0.01	0.03	0.66	.51
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	0.01	0.09	
Participant	Intercept	0.06	0.24	

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

Time of Exposure

This section now considers the factors impacting the time of exposure needed to learn target items at the three levels of measurement sensitivity.

Table 6.10 presents the analysis of the time of exposure data at the level of strict written form recall. The best-fitting model showed there was no statistically significant effect for form presentation mode (t = -1.53, p = .13), with learning in the unimodal condition taking an average of 93.80 seconds (*SD* = 66.87 seconds) and 92.64 seconds (*SD* = 63.56 seconds) in the bimodal condition. A significant effect was found for associative memory capacity (t = -2.97, p = .004), indicating that learners with higher associative memory capacity (67.398 seconds [67.40 seconds]) required less time to learn a target item than those with lower associative memory capacity (109.46 seconds [72.41 seconds]). The data showed a trend for PSTM and phonetic coding ability to negatively associate with time of exposure; however, neither effect was statistically significant (PSTM: t = -1.36, p = .18; Phonetic coding ability: t = -0.92, p = .36). There were no significant interactions between aspects of the aptitude complex and form presentation mode, although the interaction involving PSTM approached significance (t = 1.87, p = .06). Inspection of the descriptive statistics revealed that this interaction stemmed from a greater difference between the scores of the High (87.44 seconds [58.00 seconds]) and Low (102.32 seconds [76.45 seconds]) groups in the unimodal condition than in the bimodal condition (High: 90.31 seconds [60.90 seconds]; Low: 95.61 seconds [66.80 seconds]). Additionally, significant effects were found for several covariates: word length (longer words took longer than shorter words), vocabulary size (a negative association with time of exposure), and L1 (L1

Chinese learners took longer than L1 Vietnamese learners; there was no difference between L1 Chinese and L1 Thai learners, or L1 Vietnamese and L1 Thai learners).

Table 6.10

Fixed and Random Effects for the Selected Model - Time of Exposure, Strict Condition (Study 3)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	14.66	0.85	17.19	<.001***
Mode of form presentation	-0.06	0.04	-1.53	.13
LLAMA B	-0.26	0.09	-2.97	.004**
LLAMA D	-0.10	0.08	-1.36	.18
LLAMA E	-0.07	0.08	-0.92	.36
Word length	0.62	0.07	8.30	<.001***
L1	-0.23	0.08	-2.77	.008**
	-0.13	0.01	-1.28	.21
Vocabulary size	-0.99	0.20	-4.90	<.001***
LLAMA B * Mode of form presentation	-0.05	0.05	0.27	.79
LLAMA D * Mode of form presentation	0.07	0.04	1.87	.06
LLAMA E * Mode of form presentation	0.04	0.04	0.80	.42
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	0.06	0.25	
Participant	Intercept	0.13	0.19	

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

Table 6.11 shows the best-fitting model for lenient written form recall. Overall, the details of this model correspond with the model at the level of strict form recall. No effect was found for form presentation mode (t = -1.62, p = .11) (Unimodal: 95.79 [66.39]; Bimodal: 99.58 seconds [73.58 seconds]). There was a statistically significant effect for associative memory capacity (t = -2.84, p = .006) with the High group (69.15 seconds [42.83

seconds]) learning the target items more quickly than the Low group (115.08 seconds [77.44 seconds]). Neither PSTM (*t* = -1.43, *p* = .16) nor phonetic coding ability (*t* = -1.05, *p* = .30) were found to affect the time of exposure. Furthermore, there were no significant interactions between the indices of aptitude and form presentation mode. Secondary findings include significant effects for word length (longer words required more time than shorter words), L1 (Chinese L1 learners needed more time than Vietnamese learners. No differences were found between L1 Chinese and L1 Thai, or L1 Vietnamese and L1 Thai), and vocabulary size (a negative association between vocabulary size and time of exposure).

Table 6.11

Fixed and Random Effects for the Selected Model - Time of Exposure, Lenient Condition (Study 3)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	14.79	0.88	16.83	<.001***
Mode of form presentation	-0.07	0.04	-1.62	.11
LLAMA B	-0.26	0.09	-2.84	.006**
LLAMA D	-0.11	0.08	-1.43	.16
LLAMA E	-0.09	0.08	-1.05	.30
Word length	0.63	0.07	8.51	<.001***
L1	-0.24	0.09	-2.81	.007**
	-0.15	0.11	-1.37	.18
Vocabulary size	-1.03	0.21	-4.89	<.001***
LLAMA B * Mode of form presentation	0.03	0.04	0.73	.47
LLAMA D * Mode of form presentation	0.07	0.04	1.65	.10
LLAMA E * Mode of form presentation	0.02	0.05	0.48	.63
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	0.07	0.26	
Participant	Intercept	0.04	0.19	

Table 6.12 shows the analysis of the time of exposure data for written form recognition. The best-fitting model largely confirmed the findings of the form recall analysis; albeit with two key differences. The first difference related to form presentation mode, where a significant effect was found (t = -1.98, p = .05). This indicated that items presented in the bimodal condition (102.61 seconds [74.32 seconds]) required less time to learn than items presented in the unimodal condition (109.88 seconds [79.63 seconds]). An effect was again found for associative memory capacity (t = -2.53, p = .01), with learners in the Low group (121.84 seconds [87.11 seconds]) requiring more time than learners in the High group (72.58 seconds [47.92 seconds]). No significant effects were found for the other aspects of aptitude considered (PSTM: t = -0.06, p = .53; phonetic coding ability: t = -0.05, p =.62). The second difference is the significant interaction between PSTM and form presentation mode (t = 2.20, p = .03). Inspection of the descriptive statistics showed that this interaction stemmed from a greater difference between High (96.61 seconds [68.21 seconds]) and Low (119.07 seconds [89.92 seconds]) groups in the unimodal condition compared to the bimodal condition (High: 98.77 seconds [70.42 seconds], Low: 107.34 seconds [78.69 seconds]); in turn, this suggested that the Low group benefitted more from the bimodal presentation than the High group. Secondary findings include significant effects for L1 (Chinese L1 learners needed more time than Vietnamese participants. No other differences were found), vocabulary size (learners with larger vocabulary sizes needed less time), and word length (shorter words needed less time than longer words).

Table 6.12

Fixed and Random Effects for the Selected Model - Time of Exposure, Recognition Condition (Study 3)

Fixed Effects				
Parameter	Estimate	*SE*	*t*	*p*
Intercept	14.80	0.93	15.90	<.001***
Mode of form presentation	-0.08	0.04	-1.98	.05*
LLAMA B	-0.25	0.10	-2.53	.01*
LLAMA D	-0.05	0.08	-0.63	.53
LLAMA E	-0.04	0.08	-0.50	.62
Word length	0.64	0.08	8.08	<.001***
L1	-0.28	0.09	-3.01	.004**
	-0.09	0.12	-0.80	.43
Vocabulary size	-1.05	0.22	-4.71	<.001***
LLAMA B * Mode of form presentation	-0.03	0.04	0.63	.53
LLAMA D * Mode of form presentation	0.09	0.04	2.20	.03*
LLAMA E * Mode of form presentation	0.04	0.04	0.80	.42
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	0.04	0.21	
Participant	Intercept	0.08	0.28	

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

Summary of the Findings for Learning Burden

Table 6.13 presents a summary of the findings for frequency and time of exposure. Overall, there was a relatively consistent effect for form presentation mode, with bimodal form presentation posing less burden than unimodal presentation. This was largely the case when burden was considered as the number of exposures needed to learn target items. Associative memory capacity impacted burden with lower levels of associative memory related to more burden. No effects were found for PSTM or phonetic coding ability. There was also a relatively consistent interaction between PSTM and mode of form presentation. This interaction suggested that learners with low PSTM gained more from bimodal form presentation than those with high PSTM capacity. Consistent secondary effects were found for word length (longer words were more burdensome that shorter words) and vocabulary size (learners with larger vocabularies experienced less burden relative to those with smaller vocabularies). PoS impacted the number of exposures needed to learn the target items (nouns were easier to learn than verbs) and L1 Chinese learners needed more time to learn target items than the other language groups. No effect was found for familiarity of using flashcards.

Table 6.13

A Summary of the Findings for Learning Burden by Strength of Knowledge (Study 3)

			Aptitude			Aptitude * Mode							
Strength of knowledge	Metric of burden	Mode of form presentation	B	D	E	B	D	E	Word length	PoS	Vocab size	L1	Use of flashcards
Strict form recall	Frequency of exposure	✓	✓	x	x	x	✓	x	✓	✓	✓		△
	Time of exposure	x	✓	x	x	x	△	x	✓		✓	✓	
Lenient form recall	Frequency of exposure	✓	✓	x	x	x	✓	x	✓	✓	✓		x
	Time of exposure	x	✓	x	x	x	x	x	✓		✓	✓	
Form recognition	Frequency of exposure	✓	✓	x	x	x	✓	x	✓	✓	✓		
	Time of exposure	✓	✓	x	x	x	✓	x	✓		✓	✓	

6.4.2 Decay

Descriptive statistics for the decay of knowledge at the three levels of measurement sensitivity are presented in Table 6.14. This shows the mean number of items retained on the delayed test per condition. Additionally, the proportion of items retained relative to learning is given. Importantly, analysis at each level only considered items for which knowledge was demonstrated at the equivalent level on the immediate test. For example, analysis of strict written form recall data only included those items accurately produced on the immediate test. Overall, the table demonstrates that there was less decay at the level of form recognition than form recall (strict and lenient), with less than 30% of learned items successfully recalled at the level of strict form recall but more than 90% successfully answered on the test of written form recognition. Furthermore, the table shows that there was considerable individual variation in the amount of loss that occurred; for example, on the strict written form recall measure, the standard deviations are approximate to the means, indicating substantial variance among the participants.

Table 6.14

Descriptive Statistics (Means with SD in Brackets) Relevant to the Decay of Knowledge by Condition and Measurement Strictness (Study 3)

	Unimodal		Bimodal		Total	
	M (SD)	Rel. to learning	M (SD)	Rel. to learning	M (SD)	Rel. to learning
Strict form recall	3.59 (3.09)	.29 (.23)	3.68 (3.08)	.31 (.25)	7.02 (5.83)	.30 (.22)
Lenient form recall	5.14 (3.46)	.40 (.24)	5.29 (3.76)	.42 (.26)	10.43 (6.73)	.40 (.24)
Form recognition	13.27 (3.16)	.90 (.15)	13.02 (3.47)	.93 (.12)	26.29 (6.30)	.92 (.12)

The descriptive statistics for the three aspects of aptitude are presented in Tables 6.15 - 6.17. Again, these tables show the mean number of items retained and the proportion of learned items for which knowledge was demonstrated on the delayed test. In general, it can be observed that in all cases the High groups outperformed the Low groups with higher retention relative to learning figures throughout. However, there was still considerable variance in the amount of decay that occurred.

Table 6.15

Descriptive Statistics (Means with SD in Brackets) for the Decay of Knowledge by Strength of Associative Memory Capacity (Study 3)

		Unimodal		Bimodal		Total	
	LLAMA B	M (SD)	Rel. to learning	M (SD)	Rel. to learning	M (SD)	Rel. to learning
Strict form recall	High	5.15 (2.99)	.41 (.23)	5.10 (2.85)	.42 (.23)	10.25 (5.42)	.42 (.23)
	Low	2.72 (2.81)	.22 (.20)	2.89 (2.71)	.24 (.23)	5.32 (5.35)	.23 (.19)
Lenient form recall	High	6.70 (3.32)	.50 (.21)	7.00 (3.34)	.53 (.23)	13.70 (5.97)	.52 (.20)
	Low	4.28 (3.30)	.34 (.25)	4.33 (3.70)	.35 (.27)	8.61 (6.62)	.34 (.24)
Form recognition	High	13.80 (3.11)	.95 (.08)	13.85 (3.33)	.97 (.06)	27.65 (6.23)	.96 (.06)
	Low	12.97 (4.27)	.89 (.26)	12.56 (4.43)	.88 (.24)	25.53 (8.42)	.90 (.24)

Table 6.16

Descriptive Statistics for the Decay of Knowledge by Strength of PSTM. Standard Deviations Are Given in Brackets (Study 3)

		Unimodal		Bimodal		Total	
	LLAMA D	M (SD)	Rel. to learning	M (SD)	Rel. to learning	M (SD)	Rel. to learning
Strict form recall	High	4.13 (3.32)	.32 (.21)	3.87 (3.13)	.32 (.24)	7.74 (6.07)	.32 (.21)
	Low	2.96 (2.78)	.25 (.24)	3.46 (3.10)	.29 (.25)	6.19 (5.53)	.27 (.23)
Lenient form recall	High	5.67 (3.71)	.42 (.23)	5.57 (4.10)	.42 (.28)	11.23 (7.27)	.42 (.24)
	Low	4.54 (3.36)	.38 (.28)	4.96 (3.51)	.40 (.27)	9.50 (6.44)	.39 (.26)
Form recognition	High	13.63 (3.37)	.92 (.19)	12.21 (4.30)	.95 (.19)	27.33 (7.06)	.94 (.18)
	Low	12.85 (4.55)	.88 (.25)	12.23 (4.33)	.90 (.22)	25.07 (8.65)	.89 (.23)

Table 6.17

Descriptive statistics (means with SD in brackets) for the decay of knowledge by strength of phonetic coding ability (Study 3)

		Unimodal		Bimodal		Total	
	LLAMA E	M (SD)	Rel. to learning	M (SD)	Rel. to learning	M (SD)	Rel. to learning
Strict form recall	High	4.11 (3.34)	.32 (.22)	4.22 (3.21)	.35 (.24)	8.11 (6.14)	.33 (.21)
	Low	2.71 (2.52)	.24 (.23)	2.76 (2.72)	.24 (.24)	5.23 (4.91)	.24 (.22)
Lenient form recall	High	5.71 (3.58)	.44 (.24)	5.83 (3.70)	.46 (.26)	11.54 (6.72)	.44 (.23)
	Low	4.19 (3.28)	.34 (.26)	4.38 (3.90)	.34 (.29)	8.57 (6.88)	.34 (.26)
Form recognition	High	13.54 (3.71)	.91 (.22)	13.46 (3.95)	.94 (.19)	27.00 (7.47)	.92 (.20)
	Low	12.81 (4.34)	.90 (.23)	12.29 (4.45)	.92 (.23)	25.10 (8.43)	.91 (.23)

Results are presented by level of measurement sensitivity: strict written form recall, lenient written form recall knowledge, and written form recognition. The best-fitting model for strict form recall is delineated in Table 6.18. This shows that there was no effect for mode of form presentation ($z = 0.39$, $p = .70$) on the decay of learned lexical knowledge; that is, words learned in the bimodal condition underwent equivalent decay to those learned in the unimodal condition. Nor was any statistically significant effect observed for the aspects of aptitude considered, associative memory capacity ($z = -0.61$, $p = .54$), PSTM ($z = 1.19$, $p = .24$), and phonetic coding ability ($z = -0.39$, $p = .70$). This indicated that, for each aspect, the High group was as likely to suffer loss as the Low group. Furthermore, none of the interactions included in the model affected decay. Thus, it seems that aptitude did not impact the probability of decay at the level of strict form recall. However, statistically significant effects were returned for both indices of learning burden (frequency of exposure: $z = -4.34$, $p < 0.001$; time of exposure: $z = -2.36$, $p = .02$). This indicated that items that were seen more times or for a longer length of time were more likely to be forgotten; showing that the higher the learning burden, the more likely the loss of knowledge. Of the covariates considered, effects were found for PoS, with knowledge of verbs suffering more decay than knowledge of nouns, and L1, with Chinese L1 learners of English suffering more loss than either the Vietnamese L1 or Thai L1 participants.

Table 6.18

Fixed and Random Effects for the Selected Model - Decay, Strict Condition (Study 3)

Fixed Effects				
Parameter	Estimate	*SE*	*z*	*p*
Intercept	8.12	2.20	3.69	<.001***
Mode of form presentation	0.12	0.30	0.39	.70
LLAMA B	-0.18	0.30	-0.61	.54
LLAMA D	0.32	0.27	1.19	.24
LLAMA E	-0.11	0.29	-0.39	.70
LLAMA B * Mode of form presentation	0.02	0.31	0.07	.95
LLAMA D * Mode of form presentation	-0.27	0.29	-0.93	.35
LLAMA E * Mode of form presentation	0.16	0.32	0.48	.63
L1	1.27	0.25	5.04	<.001***
	1.12	0.31	3.62	<.001***
PoS	-0.52	0.20	-2.64	.008**
Frequency of exposure	-1.72	0.40	-4.34	<.001***
Time of exposure	-0.56	0.24	-2.36	.02*
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	0.15	0.38	
Participant	Intercept	0.29	0.53	

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

Details of the model fitted to the lenient written form recall data are presented in Table 6.19. Overall, the findings are comparable to the analysis of strict written form recall. No effect was found for form presentation mode (z = 0.33, p = .74) showing that condition did not predict the decay of knowledge at this level of measurement sensitivity. Similarly, no effects were found for associative memory capacity (z = -0.54, p = .59), PSTM (z = 1.41, p = .16), or phonetic coding ability (z = -0.48, p = .63). Additionally, none of the interactions between mode and aptitude were shown to predict the decay of target language knowledge. In terms of learning burden, frequency of exposure (z = -4.42, p <.001) and time of exposure (*z* = -2.61, *p* = .008) both significantly impacted the decay of lexical knowledge, with words that were more burdensome to learn more vulnerable to loss. Secondary effects were found for PoS (knowledge of nouns showed less decay than knowledge of verbs) and L1 (Chinese L1 learners suffered significantly more decay than L1 Vietnamese and L1 Thai students).

Table 6.19

Fixed and Random Effects for the Selected Model - Decay, Lenient Condition (Study 3)

Fixed Effects				
Parameter	Estimate	*SE*	*z*	*p*
Intercept	8.59	2.17	3.95	<.001***
Mode of form presentation	0.10	0.30	0.33	.74
LLAMA B	-0.16	0.29	-0.54	.59
LLAMA D	0.38	0.27	1.41	.16
LLAMA E	-0.14	0.28	-0.48	.63
LLAMA B * Mode of form presentation	-0.06	0.30	-0.17	.86
LLAMA D * Mode of form presentation	-0.29	0.29	-1.00	.32
LLAMA E * Mode of form presentation	0.22	0.32	0.71	.48
L1	1.32	0.25	5.29	<.001***
	1.15	0.31	3.76	<.001***
PoS	-0.50	0.20	-2.56	.01*
Frequency of exposure	-1.72	0.39	-4.42	<.001***
Time of exposure	-0.61	0.23	-2.61	.008**
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	0.15	0.39	
Participant	Intercept	0.29	0.54	

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

The analysis relevant to the decay of written form recognition knowledge is presented in Table 6.20. The best fitting model echoed the pattern of results of the written form recall data (strict and lenient). There was no effect for form presentation mode (z = 0.33, p = .74), indicating that lexical knowledge acquired in the bimodal condition was as vulnerable to decay as that learned via unimodal form presentation. Aptitude was also found to have no statistically significant effect on decay (associative memory capacity: z = -0.07, p = .94; PSTM: z = 1.45, p = .15; phonetic coding ability: z = 0.51, p = .61). Furthermore, none of the interactions between mode and aptitude were significant. However, learning burden was found to significantly effect decay (frequency of exposure: z = -3.72, p <.001; time of exposure: z = -2.16, p = .03); the words that posed more burden during the learning procedure were found to be more vulnerable to loss. Of the covariates considered, significant effects were found for PoS, with knowledge of verbs displaying more decay than knowledge of nouns, and L1, with Chinese L1 learners of English suffering more loss than the other two groups.

Table 6.20

Fixed and Random Effects for the Selected Model - Decay, Recognition Condition (Study 3)

Fixed Effects				
Parameter	Estimate	*SE*	*z*	*p*
Intercept	7.02	2.37	2.97	.003**
Mode of form presentation	0.09	0.29	0.33	.74
LLAMA B	-0.03	0.35	-0.07	.94
LLAMA D	0.46	0.32	1.45	.15
LLAMA E	0.17	0.33	0.51	.61
LLAMA B * Mode of form presentation	0.05	0.30	0.17	.87
LLAMA D * Mode of form presentation	-0.24	0.28	-0.87	.39
LLAMA E * Mode of form presentation	0.06	0.31	0.22	.83
L1	1.42	0.32	4.49	<.001***
	1.38	0.39	3.53	<.001***
PoS	-0.63	0.22	-2.78	.005**
Frequency of exposure	-1.50	0.40	-3.72	<.001***
Time of exposure	-0.55	0.26	-2.16	.03*
Random Effects				
Parameter		Variance	*SD*	
Item	Intercept	0.25	0.50	
Participant	Intercept	0.67	0.82	

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

Summary of the Findings for Lexical Decay

In sum, the findings present a remarkably consistent picture of the decay process and the various factors affecting it. It seems that neither mode of form presentation, aptitude, nor the interactions between mode and aptitude impacted the loss of intentionally acquired lexical knowledge. Crucially, some of these target variables were found to affect the learning burden (see Table 6.21), which, in turn, indicates that several factors had one effect on learning burden and a different effect on decay. Additionally, the burden posed by a lexical item consistently predicted decay, with results showing that it was the words that were hardest to learn that were most likely to be forgotten. Secondary findings included a consistent effect for PoS and L1. Knowledge of nouns generally suffered less decay than that of verbs and the lexical knowledge of Chinese L1 learners of English was more likely to decay than that of L1 Vietnamese and L1 Thai learners. These findings are summarised in Table 6.21.

Table 6.21

Summary of the Findings for Lexical Decay by Strength of Knowledge (Study 3)

		Aptitude			Aptitude * Mode						
Strength of knowledge	Mode of form presentation	B	D	E	B	D	E	PoS	L1	Frequency of exposure	Time of exposure
Strict form recall	X	X	X	X	X	X	X	✓	✓	✓	✓
Lenient form recall	X	X	X	X	X	X	X	✓	✓	✓	✓
Form recognition	X	X	X	X	X	X	X	✓	✓	✓	✓

6.5 Discussion

Results are discussed in relation to the research questions this study set out to answer. Although the main aim of the study was to look at the effect of various factors on learning burden and decay, it is also interesting to make some general observations about learning gains, as presented in Table 6.3. The results show that a large number of foreign language lexical items were efficiently learned using electronic flashcards. This finding supports the other investigations presented in this thesis and also research on the educational benefits of flashcards stretching back over 100 years (see Nation, 2013 for a summary). These studies suggest that flashcards are particularly useful for developing form-meaning knowledge of L2 lexis (Webb & Nation, 2017) and it was to this end that they were effective in the current study. As it seems to result in expeditious learning, teachers should encourage flashcard use, particularly during the early stages of word learning when knowledge of the form-meaning link is likely to be prioritised (Schmitt, 2000).

A second general observation relates to the number of exposures needed for learning. Studies to date have aimed to determine a specific number of exposures at which learning occurs. This is perhaps a more common research area for studies within the incidental learning paradigm (Saragi, Nation & Meister, 1978, Uchihara, Webb & Yanagisawa, 2019, Webb, 2007), yet it is a question that investigations of deliberate learning have also considered. Webb and Nation (2017) summarise this research, stating that around seven repetitions seem to be required for the acquisition of target items in a deliberate learning context. The results of this study support this assertion; they show that, on average, eight exposures were needed to learn target items. Importantly however, the results also support Huckin and Coady (1999) by showing the impossibility of establishing a number of repetitions beyond which learning is guaranteed. Some items required only three exposures to be learned while others needed thirty-nine. Although it would be theoretically possible to recognise the upper limit of this variance as a repetition frequency beyond which learning occurs, such a number is prohibitively large and pedagogical unhelpful. In general, the results indicate considerable individual and item variance, illustrating that learning burden depends on numerous intralexical, interlexical, and contextual factors, and the learning process is impacted by learner differences. Given this variance, specifying a frequency of exposure at which learning occurs seems a Sisyphean task and any figure will inevitably overestimate or underestimate the challenge posed by certain items for certain learners.

With regard to the loss of knowledge, target item knowledge was found, unsurprisingly, to decay over the two-week retention interval, supporting Studies 1 and 2, as well as the previous literature (de Groot & Keijzer, 2000; Ellis & Beaton, 1993). Greater loss was found at the level of written form recall than written form recognition. This was expected given the hierarchical nature of recall and recognition (see Gonzalez-Fernandez & Schmitt; Laufer & Goldstein, 2004), and the level of difficulty associated with the test item types used for each measure (i.e., production vs multiple choice). However, it is a reminder that knowledge at the recall level is more susceptible to loss than at the recognition level. In fact, there was minimal loss of knowledge at the level of form recognition, with fewer than 10% of learned items forgotten over the two-week interval. In comparison, considerable loss of knowledge at the form recall level occurred; knowledge of 70% and 60% of the target items was found to have decayed at the levels of strict and lenient form recall respectively. This finding reinforces the importance of research in this area as a greater understanding of the factors affecting decay may lead to pedagogical treatments to prevent such dramatic loss. This study explicitly investigated two factors. The results relevant to these are now discussed.

To What Extent Does Form Presentation Mode (i.e., Unimodal or Bimodal) Affect the Learning Burden of L2 Vocabulary?

The results show that the learning burden was impacted by form presentation mode when it was considered as frequency of exposure. At all levels of measurement sensitivity, learners experienced more learning burden in the unimodal condition than in the bimodal condition. When learning burden was considered as time of exposure, the results were not as consistent. At the level of form recall (strict, lenient) there was a tendency for bimodal presentation to expedite learning, but no statistically significant effects were found. At the level of form recognition, however, bimodal presentation led to faster learning than unimodal presentation. Overall, it seems safe to conclude that form presentation mode does impact learning burden, with co-presentation of the written and spoken forms easing the learning burden. These findings align with previous research; for example, Lado *et al*. (1967) also reported an advantage for bimodal form presentation. Hill (1994) however found no difference between unimodal and bimodal conditions. The divergence between her findings and the present study likely stems from methodological differences. In her unimodal group, Hill presented the phonetic transcription as well as the written form. This allowed learners to indirectly access the spoken form of the target items, confounding the effect of form presentation mode. Therefore, the current findings are likely to better reflect the contribution of the form presentation variable to L2 vocabulary learning. In doing so, they provide evidence to support the advice common in practitioner discourse: teachers should endeavour to present the spoken and written forms of target items together.

As outlined in Section 6.2.1, numerous explanations for this bimodal advantage have been proposed. For example, some have attributed it to different learning styles and others to an increase in motivation. The participants in this study informally reported that bimodal form presentation positively impacted their engagement with the target items, with engagement argued to be a key factor in L2 vocabulary learning (Schmitt, 2008). They also commented that bimodal presentation supported their learning as it provided them with two representations of the form, creating a stronger memory trace and allowing them to draw upon either mode during retrieval. While such anecdotal data is certainly not without limitations, it perhaps suggests that, for various reasons, learners interacted with the target items in the bimodal condition in a qualitatively different manner. This different interactional pattern may explain the learning advantage found for bimodally presented items. In turn, this speaks to the need for future qualitative enquiry and the potential for such research to delineate the reasons for the bimodal advantage observed in the current study.

To What Extent Does Form Presentation Mode (i.e., Unimodal or Bimodal) Affect the Decay of Foreign Language Vocabulary?

No effect was found for form presentation mode on the decay of foreign language vocabulary knowledge. This indicates that words learned in the unimodal and bimodal conditions were equally likely to be forgotten. This finding was consistent across the three levels of measurement sensitivity considered and demonstrates that the factors affecting vocabulary learning do not necessarily impact the process of decay in the same manner; for example, form presentation mode was found to affect the learning burden but did not have an effect on decay. Thus, with regard to this specific target variable, results suggest that although the speed at which memory traces are developed in the two conditions vary, once knowledge has been developed, they decay in a similar manner. To the best of my knowledge, only one study to date has considered, albeit indirectly, the role of form presentation mode in the decay of vocabulary knowledge. Hill (1994) reported that there was less decay in a bimodal condition than a unimodal condition; however, as stated in the previous section, the nature of the unimodal condition in that study differed from the present investigation, presenting the spoken form indirectly to learners. Therefore, the current study is likely to better reflect the effect of form presentation mode on patterns of lexical decay.

Bimodal presentation led to faster learning and did not have a detrimental effect on maintenance of learned knowledge. Therefore, materials writers and teachers should be encouraged to utilise this mode of form presentation to lighten the learning burden of L2 lexical items. For teachers, this might include saying the target words aloud when writing them on the board, including the spoken form when developing electronic materials on educational platforms such as *Anki*, and encouraging learners to investigate the spoken form when searching for words in an electronic dictionary.

To What Extent Do Aspects of the Aptitude Complex (i.e., Associative Memory Capacity, PSTM, and Phonetic Coding Ability) Influence the Learning Process?

Analysis showed that only one aspect of aptitude had an impact on the learning process. Associative memory capacity was found to significantly affect the frequency and time needed for learning to take place: learners with high associative memory capacity needed fewer exposures/less time to learn the target items. Consideration of the descriptive statistics (see Table 6.4) suggests that participants in the Low group required approximately 1.4 times the exposures and 1.65 times the time to learn the target items than participants in the High group (these figures are produced by averaging the scores at the three levels of measurement sensitivity). Thus, it seems that at least one aspect of aptitude plays a crucial role in learning L2 vocabulary, making the learning process faster. This finding partially supports the results of Granena and Long (2012) who report a positive relationship between aptitude and lexical attainment in late-onset bilinguals, and Li (2016), whose meta-analysis showed positive correlation between associative memory capacity and lexical knowledge. Crucially, however, the current study goes some way to explaining the findings of this previous research. Previous studies have investigated the relationship between aptitude and attainment, while the current study considered the association between aptitude and learning burden. In showing that learners with higher associative memory capacity learned L2 lexical items faster, this study suggests that this increased learning efficiency (and not for example, better retention) leads to the differential attainment reported in previous studies.

For the two other aspects of aptitude considered in this study, no main effects were found. This is somewhat surprising given that previous research (Li, 2016; Engel de Abreu & Gathercole, 2012) reported moderate correlations between these cognitive variables and attainment. However, there was a statistically significant interaction between PSTM and form presentation code, showing that learners with low PSTM learned the target items more quickly in the bimodal condition than in the unimodal condition. For learners with higher levels of PSTM, there was little difference. This is, to the best of my knowledge, the first study to have considered the interaction of language learning aptitude and contextual factors on the learning of L2 vocabulary. The significant findings point to the need for more research in this area. This is particularly the case because such research could potentially inform the design of CALL and MALL software, allowing for the parameters of learning activities to be adjusted to meet individual learner profiles. In order to achieve this goal, however, more research on the role individual differences such as aptitude play in the L2 vocabulary learning process is needed. The current study has demonstrated how such studies might be conducted and I hope it can set an example for further research in this area.

To What Extent Do Aspects of the Aptitude Complex (i.e., Associative Memory Capacity, PSTM, and Phonetic Coding Ability) Influence the Decay Process?

No effect was found for any aspect of the aptitude complex on the decay process. Additionally, none of the interactions between aptitude and mode were found to be significant. This means that although language learning aptitude was found to impact the learning burden of L2 vocabulary knowledge, it did not affect the loss of that knowledge over a two-week period. One previous study in this area concluded that the L1 attrition of morphosyntactic knowledge was moderated by learners' aptitude profiles (Bylund, Abrahamsson, & Hyltenstam, 2009) and so, on initial inspection, contradicts the findings of this study. However, Bylund *et al.* (2009) investigated the attrition of implicit knowledge whereas the current study looked at the decay of explicit knowledge. Thus, we would not necessarily expect synergy between the research findings.

Stepping back, we might question why we would expect aptitude to impact the maintenance of lexical knowledge. Research has shown aptitude to positively correlate with attainment. Attainment is a multivariate construct consisting of different memory processes included encoding, consolidation, and storage (Baddeley, 2014). This study has shown that some aspects of the aptitude complex are involved in encoding memory traces; learners with higher associative memory capacity encoded lexical items more quickly than others. This advantage would allow such learners to acquire more target items in a given period than less apt peers. Thus, research that reports positive correlations between aptitude and attainment (Li, 2016) might actually tap into differential levels of encoding. The current study shows that while encoding may be impacted by elements of the aptitude complex, consolidation and retrieval may not be. To explain the variance in decay, it may therefore be necessary to consider the variables that impact consolidation and retrieval. The literature on semantic memory and retrieval may be particularly useful here. The former refers to our cognitive system for storing information about the world, including our vocabulary knowledge, whereas the latter refers to the retrieval of that knowledge (Baddeley, 2014).

An alternative explanation for the nonsignificant effect of aptitude on decay relates to weaknesses in the LLAMA instrument. As recognised in Section 6.3.2, despite extensive use in SLA research, a validity argument has yet to be fully realised for the LLAMA battery. In particular, the instrument as a whole and the various components have undergone minimal assessment of internal validity (Bokander & Byland, 2019). More worryingly, recent studies that have begun to consider the validity of LLAMA B, D, and E suggest it has insufficient internal validity. For example, Bokander and Byland (2019) assessed the validity of LLAMA with 350 participants based in diverse contexts. Using classical item analysis, Rasch analysis, and principal component analysis, they found that LLAMA D had low internal consistency (α = .54), while LLAMA B (α = .81) and E (α = .74) had more acceptable internal consistency. They also found that of the components, only LLAMA B had acceptable levels of discrimination. Additionally, they highlighted issues relating the construct validity of LLAMA D and E, suggesting that the former may tap into implicit language learning aptitude (also, see Granena, 2013) and the latter is open to test strategies and thus suffers from construct irrelevance.

While the LLAMA test battery was considered the most appropriate for the current study for the reasons explained in Section 6.3.2, it would be interesting to validate these results with studies using other instruments, such as MLAT or HiLAB. Furthermore, while the role of aptitude has received minimal attention in studies of vocabulary learning and thus the current study makes a useful initial contribution, the results need to be "treat[ed] with appropriate carefulness" (Bokander & Bylund, 2019, p. 1). This is especially the case with regard to PSTM and phonetic coding ability, the components of aptitude targeted by LLAMA D and E respectively. As Bokander and Bylund (2019) found that LLAMA B functioned sufficiently, conclusions in that regard can be less hedged: associative memory capacity was found to affect the learning process but not impact the process of decay. That is, learners with comparatively high associative memory capacity were able to learn the target items more quickly; however, higher associative memory capacity did not safeguard knowledge from decay.

Irrespective of Form Presentation Mode, What Is the Role of Learning Burden in the Decay Process?

The results showed that a higher learning burden was associated with greater decay. The words that needed more exposures/time during the learning process were also those that were more likely to be forgotten. This finding mirrors the results of Study 2. Previous research has also found this relationship; for example, de Groot and Keijzer (2000), investigating the learning and forgetting of translation pairs using non-words found that "words that were easiest to learn in the first place also left more permanent traces in memory" (p. 23). In the field of vocabulary studies, an increase in frequency of exposure is typically considered beneficial. For instance, Webb and Nation (2017) state that "repetition is important in learning, and the more repetitions there are, the more likely learning is to occur" (p. 67). The current study shows that this is clearly the case: for certain items and for some learners, a large number of exposures and a considerable amount of time was needed for learning to take place. Therefore, providing the opportunities to engage with items as many times as learners need increases the probability that a learner will sufficiently engage with a word to overcome its learning burden. However, the results also show that the items requiring more exposures to be learned were also those most likely to be forgotten, suggesting that a heavier learning burden is associated with a higher likelihood of decay.

6.6 Limitations

There were inevitably limitations associated with this study. First, this investigation was only able to consider form-meaning aspects of word knowledge. While this is perhaps the most crucial aspect of vocabulary knowledge (Schmitt, 2010) and reflects the most common use of flashcards (Webb & Nation, 2017), there are clearly other aspects of word knowledge that need to be learned if an item is to be used accurately (Nation, 2013). This study, and indeed this thesis, represents an initial foray into lexical decay and therefore it is, perhaps, unavoidable that focus be trained on form-meaning aspects of word knowledge; however, future studies that investigate a broader spectrum of word knowledge aspects are needed. Furthermore, at the level of form, this study focused solely on knowledge of the written form. As such, it is likely to have ignored gains in spoken form knowledge, particularly in the bimodal condition. The decision was taken as it would not have been possible to conduct a spoken form recognition test without disambiguating the spoken representation of known written forms in the unimodal condition. However, it is important to recognise that the instrumentation used may have missed some word knowledge acquired. This would further support the use of bimodal form presentation; not only was it found to facilitate expedited acquisition of the written form but may potentially lead to increased gains in spoken form.

Third, in the bimodal condition learners were only able to listen to the spoken form once per exposure; while they were able to read a written form multiple times per exposure. It may well be that providing more opportunities for learners to interact with the spoken form would have changed the nature of the results (see Hill, 1994). However, this decision was largely determined by the parameters of the learning software employed: *Anki* does not allow learners to listen repeatedly to the spoken form. Thus, while this may be considered a limitation, it is certainly the case that the methodology was environmentally valid. Furthermore, although this study showed that bimodal form presentation can expedite learning, the specific contribution of the spoken form to easing the learning burden is unclear. In the bimodal condition, the spoken form was presented with the written form for each exposure. However, it might be that the presence of the spoken form is particularly beneficial during the initial exposures; alternatively, perhaps it is after the initial stages, when learners have partial knowledge of the written form that the spoken form supports learning. This is, of course, an empirical question and warrants future investigation.

6.7 Conclusion

The study presented in this chapter investigated the role form presentation mode played in the learning burden and decay of L2 vocabulary in an instructional context. It showed that when word form was presented simultaneously in the written and spoken modes, learners encountered less burden than when presented in the written mode only; however, there was no difference in the amount of decay suffered in both conditions. This shows that although the two conditions impacted the speed of learning, once learned, the knowledge fostered in both conditions behaved in a similar manner. Thus, this suggests that teachers and learners should utilise bimodal form presentation where possible; it has been shown to lead to less burdensome learning and does not result in greater loss.

Additionally, some aspects of language learning aptitude were found to impact the learning burden of L2 vocabulary. This confirms what teachers have long seen in the classroom: certain learners acquire target items more quickly than others. This study has also shown that certain aspects of the aptitude complex interact with contextual factors. However, aptitude, nor any interactions between the aspects considered and the learning activity, did not impact the decay of target items. This suggests that there may be other cognitive variables that affect the forgetting of L2 vocabulary and speaks to the importance of further research in this area. More research is also needed on the relationship between learner-related factors and other contextual variables.

One key finding from this study was the role of learning burden in the decay process. Words that were harder to learn were also those most likely to be forgotten. This finding, in addition to others relating to the effect of mode of form presentation and aptitude on decay, indicates the value of research in this area. A better understanding of lexical decay can help the development of effective teaching and research design; these pedagogical implications are considered in the next chapter.

Chapter 7: General Discussion and Conclusion

7.1 Introduction

This thesis set out to better understand L2 lexical decay. In addition, it aimed to investigate the effect of various factors on the learning burden of L2 lexis and the effect of burden on the decay process. As explained in Chapter 1, these foci were motivated by my classroom observations. As was also outlined in Chapter 1, the research presented in this thesis was conducted with the belief that a better understanding of decay can help inform L2 pedagogy. However, in order to develop teaching resources to prevent loss, we first need to know a great deal more about the role different variables play in learning burden and lexical decay. To this end, this thesis has considered intralexical (i.e., PoS and word length), contextual (i.e., code of meaning presentation and mode of form presentation), and learner-related (i.e., perceived target item usefulness and language learning aptitude) factors in three experimental studies. This chapter begins by briefly summarising each of these studies in turn and then considers findings that were inconsistent across the three studies, exploring possible reasons for variance. Discussion then turns to the effect of the retention interval length on decay and an exploratory cross-study analysis is presented using data from Studies 1-3 to begin to understand this area. After this, factors beyond the experimental parameters of this thesis that future research might explore are examined. Implications of this thesis for research and pedagogy are then introduced before a final conclusion draws the thesis to a close.

7.2 Summary of the Findings from the Three Experimental Studies

The first study investigated the effect two intralexical factors, PoS and word length, had on the learning burden and decay of accrued knowledge, as well as the effect of learning burden on decay, over a four-week retention interval. The results showed that the learners required an average of approximately six exposures to learn the target items. There was a significant effect for word length, with longer words requiring more exposures than shorter words. No effect was found for PoS. This showed that word length but not PoS impacted the learning burden. With regard to decay, there was considerable loss at the level of form recall. In fact, a floor effect at the recall level prevented statistical analysis. Knowledge at the level of form recognition also decayed, but not as sharply. At this level, neither of the two intralexical variables affected decay; however, a significant effect was found for learning burden, showing lexical items that required a greater number of exposures during the learning procedure were better retained at the level of form recognition. A secondary effect for L1 was also reported (Mandarin Chinese L1 learners experienced less loss than other participants).

The second study looked at the effect of a contextual variable (i.e., code of meaning presentation: L1 equivalent or L2 definition), and an individual factor (i.e., perceived target item usefulness), on learning burden and decay over a one-week retention interval. The results showed that an average of approximately nine exposures or around 97 seconds was needed to learn items to the level of strict form recall. Mixed-effects modelling showed that the meaning presentation code impacted the learning burden: target items presented with L1 equivalents required less time to learn than items presented with an L2 definition. No effect was found for perceived usefulness on learning burden. Secondary effects were found for vocabulary size (a larger vocabulary size was associated with faster learning), word length (positively associated with learning burden), and PoS (nouns were learned faster than verbs). With regard to decay, a significant effect was found for meaning presentation code on the loss of knowledge at the level of form recall, with items that had been presented with an L1 equivalent more resilient. No effect was found for perceived usefulness, which means that this factor affected neither learning nor decay. A significant effect was found for learning burden on decay, with the target items that posed more burden during the learning procedure suffering the most loss. Additionally, a secondary effect was found for PoS, showing that fewer nouns were forgotten than verbs.

The final study was conducted to assess the impact of another contextual factor (i.e., form presentation mode: unimodal or bimodal) and a further learner variable (i.e., language learning aptitude) on the learning burden and decay processes. As with Studies 1 and 2, this study also targeted the influence of learning burden on the loss of intentionally learned lexical knowledge. One novel feature of this study was the use of key-stroke logging software to verify learner self-report evaluations during the learning process. This study found that the learners required an average of approximately eight exposures or 93 seconds to learn the target items to the level of strict form recall. The bimodal condition was found to pose less learning burden than the unimodal condition when learning burden was measured by number of exposures. Associative memory capacity was also found to impact the learning process; learners with higher scores in this area required fewer exposures to learn the target items. Additionally, an interaction between mode and PSTM was found; learners with comparatively poor PSTM gained more from the bimodal condition. Secondary effects for vocabulary size, word length, and PoS were also found, supporting the findings of Study 2. An additional effect was found for L1. In terms of decay, neither mode of form presentation nor any of the three aspects of aptitude was found to have an effect. However, an effect was found for learning burden, with the items that posed more burden during the learning process more likely to be forgotten, again supporting the findings of Study 2. Secondary effects were found for PoS (more decay of verbs occurred than nouns) and L1.

The experimental studies presented in this thesis confirmed that there is indeed lexical decay and that it seems to be more pronounced for form recall than form recognition components. Some factors were found to impact the process of lexical decay (e.g., the code of meaning presentation), while others had little impact on the loss of accrued knowledge (e.g., language learning aptitude, perceived target item usefulness, word length, and form presentation mode). In relation to learning burden, the investigations found that the extent of burden varied by learner and item, demonstrating that some words were harder to learn than others and some learners experienced more burden than their peers. Some target variables were found to impact burden (e.g., word length, meaning presentation code, form presentation mode), whereas other factors did not (e.g., perceived item usefulness). However, the effect of other target, and secondary, variables was less consistent across the three studies. For instance, seemingly contradictory findings were found for the role of learning burden on the decay process, PoS, and the covariate vocabulary size. These points are explored in the following section.

In Chapter 2, two tables presented summaries of previous findings relating to learning burden and decay. These tables can now be updated to reflect the findings of Studies 1-3. Tables 7.1 and 7.2 represent a synthesis of the findings of this thesis (in red) with the literature to date (in black).

Table 7.1

Factors Affecting the Learning Burden of L2 Vocabulary – Updated in Light of the Research in this Thesis

Factor type	Facilitating factors	Difficulty-inducing factors	No effect
Intralexical	• concreteness • more imageable • less structurally complex • congruent sound-script relationship • inflexional regularity • derivational regularity • morphological regularity • one meaning, one form • PoS (nouns) • length (shorter words)	• abstractness • less imageable • more structurally complex • incongruent sound-script relationship • inflexional complexity • derivational complexity • deceptive morphological transparency • polysemy • PoS (verbs) • length (longer words)	
Interlexical	• cognateness • orthographic wordlikeness • phonotactic typicality • high L1 frequency	• non cognateness • orthographic non-wordlikeness • phonotactic atypicality • low L1 frequency	
Contextual	• high frequency of occurrence • informative context • presentation with semantically dissimilar items • meaning presentation code (L1) • form presentation mode (bimodal)	• low frequency of occurrence • uninformative context • presentation of items in semantic sets • meaning presentation code (L2) • form presentation mode (unimodal)	
Moderating learner-related	• higher language learning aptitude	• lower language learning aptitude	• perceived usefulness

Note. red represents findings from this thesis; black represents findings from the literature

Table 7.2

Factors Affecting Lexical Loss – Updated in Light of the Research in this Thesis

Variable type	Factors associated with less loss	Factors associated with more loss	Factors without a clear effect
Intralexical	• word class (nouns) • concreteness • imageability	• word class (verbs) • abstractness • non-imageability	• word length
Interlexical	• phonotactical typicality • high L1 frequency • high L2 frequency • cognateness • formal similarity to L1	• phonotactical atypicality • low L1 frequency • low L2 frequency • non-cognate • formal dissimilarity to L1	
Contextual	• intentional learning activity • use of keyword method • spaced repetition • extended spacing between repetitions • L1 meaning presentation	• incidental learning activity • using rote learning • massed repetition • brief intervals between repetitions • L2 meaning presentation	• form presentation mode
Learner-related	• high L2 proficiency • not young learner • L2 literate • more motivated to learn • more motivated to retain • light learning burden	• low L2 proficiency • young learner • L2 illiterate • less motivated to learn • less motivated to retain • heavy learning burden	• vocabulary size • aptitude • perceived usefulness

Note. red represents findings from this thesis; black represents findings from the literature

7.2.1 Inconsistent Findings across the Three Experimental Studies

One aim of this thesis has been to disambiguate the role of learning burden in the decay process. Study 1 found that a heavier burden during the learning process led to less loss at the level of form recognition after a retention interval of 4 weeks. Interestingly, Studies 2 and 3 found that a heavier burden was associated with more loss at both the form recall and form recognition level after retention intervals of one and two weeks respectively. These findings can be explained by comparing the average learning gains in each study. Fewer target items were learned in Study 1 than in Studies 2 and 3, and there were smaller standard deviations in the later studies than the first. More specifically, the data show that fewer items associated with a high burden (e.g., orthographically longer words) were learned in Study 1 than in Studies 2 and 3. The frequency of exposure data further demonstrated this conclusion as learners needed fewer exposures in Study 1 than Studies 2 and 3 to learn the same target items. Remembering frequency of exposure was only calculated for learned items (i.e., a correct response on the immediate test), a lower average frequency of exposure implies that fewer burdensome items were learned in Study 1 than later studies. Thus, Study 1 did not demonstrate the same range of burden as Studies 2 and 3. Putting these findings together, two out of three studies found that higher burden was associated with more decay. Moreover, these findings were reported at both the recall and recognition level. Therefore, it can be concluded that burden negatively correlates with retention. Such a conclusion is in line with the findings of previous literature (Bahrick & Phelps, 1987; de Groot & Keijzer, 2000).

Study 1 explicitly targeted the effect of PoS and word length on learning burden and decay. Additionally, these intralexical variables were considered as covariates in Studies 2 and 3. There was a consistent finding in relation to word length, with all three studies finding that length had a statistically significant effect on learning burden, but no effect on lexical decay. However, the effect for PoS was less stable. Study 1 reported no statistically significant effect for PoS on either burden or decay. In contrast, a statistically significant noun advantage was found in Study 2 when learning burden was considered as the frequency and time of exposure. A significant effect was also found for decay, with verbs more likely to be forgotten than nouns. In Study 3, a noun advantage was found on the number of exposures needed for learning but not for length of exposure. A statistically significant effect was also found in the decay analysis, with knowledge of verbs shown to be more vulnerable than knowledge of nouns. The key question then is what explains the difference between Study 1, and Studies 2 and 3? One explanation relates to how the meanings were conveyed to the learners. In Study 1, the PoS was only implied through the syntactic construction of the definitions, whereas the later studies employed grammatical metalanguage on the flashcards to make the PoS more salient. This salience may explain the different findings. Overall, it can be concluded that if learners are aware of the PoS (i.e., a target item is embedded in context or metalanguage is used), verbs are associated with heavier burden and greater decay. This finding supports previous literature, which has generally found that nouns are learned more easily (Ellis & Beaton, 1993; Horst & Meara, 1999; Rodgers, 1969; van Zeeland & Schmitt, 2013b), and are better retained (Ellis & Beaton, 1993), than verbs.

The role of vocabulary size also differed across the three studies. This variable was included as a covariate in burden and decay analysis in all of the investigations. For practical reasons, an amalgamation of the first three levels of the VLT was employed in Study 1, while bilingual versions of the VST were used in Studies 2 and 3. Effects were found for vocabulary size on learning burden in these latter studies, but no effect was found for vocabulary size on burden in Study 1. The non-effect in Study 1 is perhaps attributable to the instrument. The VLT is a diagnostic measure of vocabulary knowledge at different frequency bands rather than an explicit test of vocabulary size. Studies 2 and 3 may, therefore, better reflect the impact that vocabulary size has on learning burden and decay. These studies found that learners with larger vocabularies faced less learning burden, demonstrating a Matthew Effect. This effect hypothesises that the rich ger richer while the poor get poorer (Stanovich, 1986). In this thesis, findings show the *rich* learners with comparatively large vocabularies got richer, in that they learned novel L2 lexical items more quickly than learners with smaller vocabulary sizes. Interestingly, however, no effect for vocabulary size was found on decay in any study, suggesting that, to continue Stanovich's metaphor, the rich and poor were equally likely to lose their accumulated wealth. In drawing this conclusion, it is important to recognise the limited range of vocabulary sizes in the participant samples from Studies 2 and 3. Therefore, to confirm this finding, future research with learners of diverse proficiencies is needed.

7.3 The Role of the Retention Interval Length on Decay

Another factor that warrants further investigation is the effect of the length of the retention interval on decay. As discussed in Chapter 2, research has shown that the length of the retention interval impacts the amount of loss that occurs. For example, studies in the attrition context indicate that substantial loss takes place initially, but that some lexical knowledge is incredibly resilient to forgetting and can be demonstrated as long as 50 years after instruction (Bahrick, 1984). Another robust finding from the literature is that productive vocabulary knowledge decays comparatively quickly but that word knowledge is often maintained when measured by recognition instruments. Thus, it seems that the length of the retention interval and the nature of the assessment battery are key variables in studies of lexical decay and are likely to have a direct bearing on the extent of loss reported on delayed assessments. For instance, shorter retention intervals and recognition instruments may result in less reported decay than longer retention intervals and the use of recall instruments. However, the length of retention intervals and the manner in which knowledge of target items is measured varies greatly between studies, making comparison of experimental findings difficult. In fact, there is currently little agreement on the ideal retention interval length and, in lieu of principled justifications, experimental design is largely governed by practical concerns. On this subject, Schmitt (2010, p. 157) stated that:

> a delayed posttest of three weeks should be indicative of learning which is stable and durable. If this three-week ideal cannot be met for practical reasons, I would suggest that any delayed posttest of less than one week is likely to be relatively uninformative and should be avoided if possible. But whatever interval is practical, delayed posttests should be included in all acquisition research designs.

While I would agree on the necessity of delayed posttests, partly to allow for consideration of decay, it is important to recognise that presently we know little about how vocabulary knowledge fostered by different pedagogical conditions degrades as a function of time. Therefore, it is currently challenging to recommend a specific timeframe for delayed posttests to be conducted. In fact, in lieu of empirical support, Schmitt (2010) based his advisory three-week retention interval on the opinion of vocabulary research experts. This advice requires empirical validation.

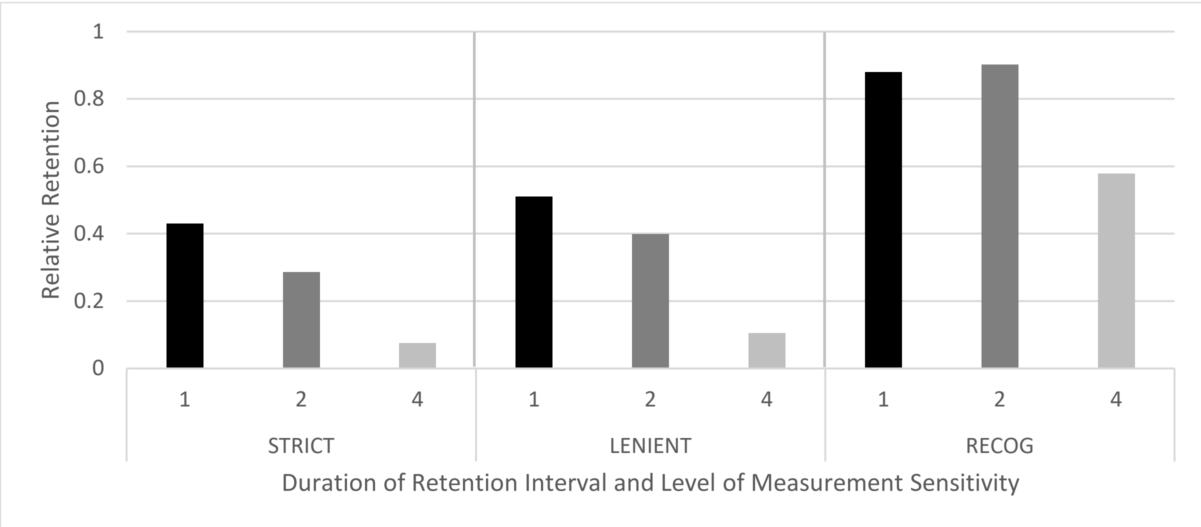
The focus of this thesis has been the effect of different factors on the decay process, not the relationship between the duration of the retention interval and decay. However, the methodological design of the three studies does allow for exploratory investigation of this area. This is because the three studies have employed identical target items and learning software. Furthermore, although the focus of each study has differed, one condition in each study has remained unchanged: unimodal (i.e., written form only) form presentation paired with L2 definitions. Additionally, although there have been some important differences between the participants of the three studies (e.g., Study 2 was conducted with a monolingual cohort in an EFL context while Studies 1 and 3 were carried out with multilingual cohorts in an ESL environment), the samples were broadly similar in age and were all pre-sessional/in-sessional EAP students at a university at the time of investigation. Crucially, the three investigations have employed different retention intervals, 4 weeks, 1 week, and 2 weeks respectively, facilitating a cross-study comparison of the results to provide initial exploratory evidence for the role of time on the decay process.

This comparison is presented in Figure 7.1. The data represent the mean proportion of knowledge retained compared to learning from Studies 1, 2, and 3 and are presented by level of measurement sensitivity. The order in which the data are presented does not match the order in which the studies were conducted. Study 2 (i.e., a one-week retention interval) is presented first, followed by Study 3 (i.e., a two-week retention interval), and then Study 1 (i.e., a four-week retention interval). Overall, the data show that the aspect of vocabulary knowledge assessed has a great deal of influence on the amount of decay that is reported. For example, there was consistently less decay at the level of form recognition than form recall at all lengths of retention interval. Furthermore, the data show that even after a retention interval as short as one or two weeks, there was considerable decay at the level of form recall, while little decay occurred at the level of form recognition.

The data also suggest that the length of the retention interval is a crucial factor in decay research. At the form recall level, decay was observed even after a brief retention interval of one week. Furthermore, there was more decay after a two-week than a one-week interval, and, in turn, considerably more decay after a four-week than a two-week interval. At the level of form recognition, there was little difference between the decay of lexical knowledge after one-week and two-week retention intervals. In contrast, substantial decay seems to have occurred between two and four weeks. Overall, the data suggest that particularly at the level of form recognition, lexical knowledge does not seem to decay in a purely linear manner as a function of time. In fact, at this level there seems to be a threshold between two and four weeks where significantly more loss occurs than before.

Figure 7.1

A Cross-Study Comparison (Studies 1, 2, and 3) to Determine the Effect of the Retention Interval Duration on Decay at Three Levels of Measurement Sensitivity



This comparison suggests, therefore, that Schmitt's (2010) recommendation holds weight. It is valuable for vocabulary acquisition studies to include delayed posttests even if these are conducted only one week after the onset of the retention interval; however, the results also suggest that such investigations should include recall instruments to better capture the loss that occurs during this period. This is the case because there is likely to be minimal decay at the level of form recognition over a brief retention interval. Furthermore, considerable loss seems to occur between two and four weeks, indicating that Schmitt's

(2010) three-week recommendation may well demonstrate durable learning. However, to test this assertion, longitudinal research is needed to track the degradation of learning beyond this three-week period. Such a study should disambiguate whether knowledge retained at the three-week threshold is subsequently maintained for an extended period.

7.4 Future Research Directions

7.4.1 Measuring the Multifaceted Nature of Vocabulary Knowledge

The measurement battery employed in this thesis required learners to produce a target form in response to a given meaning and to recognise the target form in a multiple-choice task. Thus, it measured written form recall and recognition. This focus on the form-meaning link is common in studies of vocabulary learning (Schmitt, 2019) and a crucial initial aspect in the incremental development of word knowledge (Barcroft, 2015). However, it is important to consider the extent to which the results from these instruments can be generalised to the actual tasks that learners frequently perform (Schmitt, 2019). Such consideration is important because learners do not typically study foreign language lexis for the sake of merely increasing their vocabulary knowledge, but rather to facilitate effective language use. That is, vocabulary study is a means to an end rather than typically the end itself. Language use can be understood in terms of the four skills, with receptive vocabulary knowledge often associated with reading and listening and productive vocabulary knowledge related to writing and speaking. Therefore, written form recall is often thought to relate to the ability to use a target item in writing. This is because when learners write, they typically have a meaning in mind and need to recall the associated form, a task which is cognitively similar to a form recall measure. However, given the hierarchical nature of productive and receptive knowledge (see Laufer & Goldstein, 2004), such a test is also associated with the ability to comprehend an item when reading.

However, the extent to which written form recall instruments actually predict the productive use of target words in authentic contexts is an open question. Minimal research has been conducted in this area and it is difficult to conceive of methodologies that would tap into such use. This is because a lack of target item use does not necessarily indicate a learner lacks knowledge of a target item. It may be the case that a learner chooses not to use a target item in preference to a semantically similar word (Schmitt, 2019). Thus, with regard to the interpretation of form recall data in this thesis, I do not claim that knowledge displayed on the measurement instruments equated to actual language use; however, it seems reasonable to conclude that the written form recall measure indicated that a target item was likely to be understood when encountered in a reading passage and may have shown that a target item could possibly have been used in writing.

The transferability of the results from the written form recognition test to language use is similarly complex. This thesis, like many studies of vocabulary learning, employed a multiple-choice measurement format, which tasked participants with choosing the correct form from four options. Such a task does not reflect typical behaviour when learners are using language in an authentic environment and thus such tasks are limited in the information they provide about a learner's ability to use the target language (Schmitt, Nation, & Kremmel, 2020). Therefore, results from the written form recognition test should not be transferred to contexts of language use; instead, they should be interpreted as representing one of the earliest stages in the incremental development of vocabulary knowledge. Thus, successfully answering an item on the written form recognition instrument does not abstract to language use but shows that the learner's knowledge is on the developmental pathway that will ultimately facilitate reading comprehension and potentially allow for accurate productive use.

However, the appropriate use of vocabulary involves more than knowledge of the form-meaning link (Nation, 2013), the aspect of vocabulary knowledge investigated in this thesis. There is now consensus in the field of vocabulary studies that vocabulary knowledge is a multi-componential construct (Nation, 2013). Nation's (2013) word knowledge framework (see Section 1.2) suggests that lexical knowledge can be analysed into form (e.g., the spoken form, knowledge of word parts), meaning (e.g., multiple meaning senses, associations), and use (e.g., collocational knowledge, register awareness). Studies have shown that these aspects do not develop in parallel (Gonzalez-Fernandez & Schmitt, 2019; Webb, 2005) with knowledge of the form-meaning link among the earliest aspects typically acquired. This means that the results from this thesis cannot be extended to aspects of knowledge beyond form and meaning.

As was the case with the studies presented in this thesis, vocabulary research has tended to investigate the acquisition of different aspects of word knowledge separately (e.g., collocations: Peters, 2016; associations: Qian, 2002). This has facilitated understanding of how these individual aspects of word knowledge develop; however, there have been calls for research to examine multiple aspects concurrently and to consider the interactions between different knowledge aspects in order to truly understand the nature of vocabulary development (Milton & Fitzpatrick, 2014; Schmitt, 2019). Some recent studies have responded to this call and are beginning to delineate the relationships between aspects of word knowledge (Gonzalez-Fernandez & Schmitt, 2019). In a similar vein, future decay research might track the loss of different word knowledge aspects over a period of time. Such a study would allow us to determine if certain word knowledge aspects are more vulnerable to being forgotten than others and whether an increasingly multi-componential understanding of target items reduces the probability of loss.

The studies presented in this thesis considered knowledge at the form recall and recognition level, in line with the majority of studies in the field (see Nation & Webb, 2011) However, it is important to note that while considering three levels of measurement sensitivity allowed this thesis to tap into different strengths of target item knowledge, some partial knowledge may have been overlooked. Capturing this weaker knowledge may necessitate the use of highly sensitive methodologies, such as lexical priming. By comparing reaction times in such methodologies, it is possible to determine whether partial knowledge of items has been acquired (McRae & Boisvert, 1998) and consider the extent to which newly learned words have been integrated into the mental lexicon (see Elgort, 2011). Importantly, neither of these indices are available with tests of form recall or recognition. Thus, although it was not practical to incorporate priming methodology into the research presented in this thesis, future studies should consider such a design to track the decay of learned knowledge in an increasingly nuanced manner.

7.4.2 The Role of Fluency in Learning Burden and Lexical Decay

Other variables that may predict decay and thus warrant investigation in future studies are the fluency of lexical access and production. In relation to the use of a word, lexical knowledge entails being able to fluently access that knowledge in recognition and also when involved in production. In terms of the speed with which the mental representation of a target item is accessed, by utilising flashcard software, the studies in this thesis were able to determine the number of times and length each learner saw each target item. Additionally, learners' evaluations of their production were used to differentiate exposures prior to encoding from the retrieval of encoded knowledge. This offered a fine-grained picture of the learning process not typically seen in vocabulary learning studies. However, with this methodology, it was not possible to differentiate the duration of various behaviours such as reading a definition and accessing the lexical representation of a target form, and producing a target written form and evaluating the accuracy of production. Thus, although the three studies represent considerable methodological innovation and allow for a relatively detailed understanding of learning burden and decay, additional nuance may further illuminate important aspects of these processes.

In order to obtain such process data, methods that allow for the online measurement of learning behaviour are needed. Study 3 employed keystroke logging software to verify the learner self-report data. This verification involved manually inspecting the logging data to determine if learner production matched a given evaluation. However, keystroke logging software additionally return a range of indices that can be used to calculate numerous metrics of lexical production: the time between meaning presentation and the onset of lexical production; the time between meaning presentation and the termination of lexical production; the total time of lexical production; the number and duration of mid-word pauses; and the number of revisions. These measures could be considered as covariates in decay analysis to determine whether items produced more fluently prior to a retention interval are better retained. To facilitate analysis of this sort in the future, I have written a Microsoft Excel macro to automatically concatenate the keystroke logging data with the frequency of exposure data produced by the flashcard software (see Figure 7.2 for a sample). Online data of this sort is rarely considered in intentional vocabulary learning studies. This is despite the potentially impactful implications findings might have for the development of CALL and MALL software; for example, if it were found that certain metrics of fluency predicted decay, these could be written into the recycling algorithms of software to promote maximally efficient learning.

Figure 7.2

An Example of the Combined Frequency of Exposure and Keystroke Logging Data from Study 3

Item	Time of ex.	Keystrokes	Cleaned output	BWP	WWP	Revisions	Evaluation length
irk	8.421	ircBACKBACKrk	irk	4777	8730	2	2359
archipelago	14.46	arciBACKhipelago	archipelago	4959	15303	1	1480
asterisk	8.676	asterisk	asterisk	2341	2530	0	2053
scintillate	23.563	sBACKscintillate	scintillate	3562	8270	1	2655

Note. BWP = Before word pause; WWP = Within word pause

7.4.3 Other Factors Affecting Learning Burden and Lexical Decay

The literature suggests that more factors increase the learning burden (see Laufer 1997; Schmitt, 2010; Webb & Nation, 2017) than were able to be empirically investigated in this thesis. Thus, it is still an open question whether these further variables, in addition to affecting burden, impact the process of lexical decay. One or more of these factors may contribute to burden and decay and so warrant empirical investigation. Four factors in particular should be considered: L1 frequency, the extent of initial learning, compound learning burden, and learning strategies.

The first variable is L1 frequency. This is the frequency of the corresponding L1 equivalent for a target L2 lexical item. Research has found this to impact the learning burden of foreign language lexical items (see Peters, 2020). For example, in an early study, Chapman and Gilbert (1937) investigated the impact of referent familiarity on the learnability of L2 items. They measured the learning and retention of 24 L2 Hindustani nouns paired either with known (n = 12) or unknown (n = 12) English equivalents. They found that L2 words paired with known L1 equivalents were both learned more quickly, and the knowledge acquired was less susceptible to decay than those paired with unknown L1 equivalents. Furthermore, a series of studies by de Groot and colleagues (de Groot, 2006; de Groot & Keijzer, 2000; Lotto & de Groot, 1998) found that L1 frequency had an impact on learning and also correlated with decay, showing that the less frequent the L1 referent, the more likely the L2 item was to be forgotten. This finding has been attributed to a concept familiarity effect because some concepts are encountered more often, they become more entrenched in the mental lexicon and connections to novel L2 forms are developed more easily (Peters, 2020). This may be an important variable for future studies to consider.

A further variable that deserves research attention is the effect of the extent of initial learning on decay. Studies that have looked at language attrition have suggested that the extent of loss is proportionate to initial learning. For instance, the critical threshold hypothesis holds that more language knowledge is associated with less loss (Neisser, 1984). Evidence for this hypothesis comes from Hansen (1999), who investigated the attrition of Japanese negation patterns by L1 English former missionaries. She reported that the amount of attrition was a function of the time spent in the L2 environment, concluding that the more L2 knowledge learners had upon repatriation, the less they lost. There are implications of this research for investigations of lexical decay.

The tests of attainment in this thesis showed that not all participants learned each of the 32 target items during the learning treatment. On the one hand, this finding justifies the inclusion of such tests rather than relying on the learning data extracted from the flashcard software. By only considering words for which knowledge was demonstrated on the immediate test, I was able to remove items that were only held in working memory for the duration of the learning procedure. As a result, the studies avoided confounding the loss of lexical knowledge with the lack of acquisition. Thus, the findings are likely to more accurately represent the actual loss of deliberately learned L2 lexical knowledge. On the other hand, it is possible that the amount of target items learned may have impacted the maintenance of that knowledge. It seems intuitively appealing to claim that the decay experienced by learners who initially develop knowledge of more words may differ from those who learn fewer target items. Some studies have considered the impact of the volume of target items on learning burden. For example, Nakata and Webb (2016) found no difference in learning gains between target items organised into sets of four, ten, or twenty words. However, we currently know little about how the size of learning gains impacts decay patterns. To answer this question, in addition to studies that deliberately manipulate the number of target items and measure learning burden and decay, future research might also consider including the number of target items learned as a covariate in models fitted to data from delayed tests.

A further area that requires research attention relates to the manner of item presentation. Although the target items were presented in sets (as is common practice both in teaching and research), learning burden was calculated separately for individual members of a set. However, it is possible that the learning of one word was influenced by other items both within a set and/or the wider pool of target items. For example, it might be the case that the learning burden of a target item was impacted by the difficulty of the items directly preceding it, with a burdensome word potentially increasing the burden of subsequent items. This thesis randomised the order of target item presentation in part to control for this compound burden but did not explicitly investigate its effect. In fact, to the best of my knowledge, to date minimal research has considered this factor despite potentially impactful findings which might offer advice regarding the construction of target item sets. This may be an important area for future research.

Another area that requires further investigation is the role of strategies employed during learning on the decay process. In the investigations presented in this thesis, learners used flashcard software to study the target items. This led to a consistent approach to learning and controlled for task-related variables on the learning and decay processes. While it is the case that this software encouraged a specific learning behaviour, it is also the case that individual learners might have adopted unique strategies (e.g., aural repetition, the keyword method, etc.). Such strategies have been associated with learning gains (Ellis & Beaton, 1993) and predicted to impact decay (Webb & Nation, 2017). Crucially, such learning behaviour is likely to vary by learner, context, and target item (Gu, 2020) as the form and/or meaning of a word may encourage the use of a specific strategy. For example, in Study 3, a learner reported that she used the first letter of a target item, *orb*, to remember its meaning, positing that this strategy helped her learn and retain formal knowledge of this item. Another learner from Study 3 described how the meaning of the target item *regurgitate* provoked a physical reaction that reflected the word's meaning, increasing her engagement with the item and promoting learning. Crucially, such imagery and embodiment depend to a large extent on the target item and the learner and are therefore impossible to control. This thesis was able to mitigate this issue by employing sufficiently large learner samples, yet to obtain a comprehensive understanding of the factors affecting decay, it may be necessary to consider data at this nuanced level. To do this, case studies employing stimulated recall which correlate item level learning behaviour against indices of learning burden and decay are needed.

7.5 Implications

This section discusses pedagogical and research implications of this thesis.

7.5.1 Pedagogy

The time that learners have available for language study is finite and thus every pedagogical decision is associated with an opportunity cost; a focus on one area comes at the direct expense of another. It is, therefore, vital to make sure that the time devoted to vocabulary study is fully utilised. Maximum value can be extracted from vocabulary learning and teaching by, *inter alia*, considering lexical frequency and learner needs when selecting target items (Webb & Nation, 2017), providing varied repetitions (Barcroft & Summers, 2005), and introducing derivations (Schmitt, 2000) and vocabulary learning strategies (Nation, 2008) when teaching target items. Additionally, one central tenet of this thesis has been that a better understanding of lexical decay can improve the efficacy of language pedagogy. That is, clarifying the manner and speed of foreign language lexical decay, as well as the variables associated with it, might lead to the development of more robust learning procedures. One simple strategy would be to provide richer instruction or reduce the retention intervals for items most vulnerable to decay. For example, this thesis has shown that when comparatively brief retention intervals have been employed (i.e., Studies 2 and 3), PoS affected the speed at which language knowledge decayed, with knowledge of verbs decaying faster than knowledge of nouns. This finding suggests that verbs and nouns might need different pedagogical treatment, with the former receiving richer instruction and more frequent recycling to ensure equitable long-term retention.

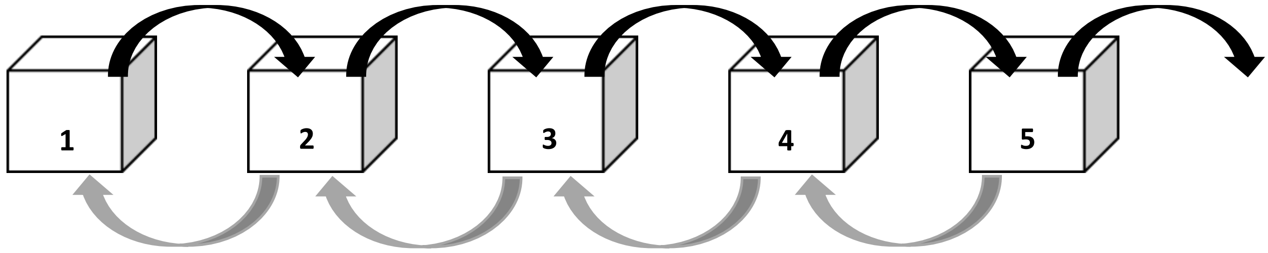
Additional pedagogical implications relate to the findings of individual studies. Practitioners are often wary of employing the L1 in the L2 classroom (R. Ellis & Shintani, 2013), despite repeated calls to embrace L1 equivalents to convey word meaning (Folse, 2004; Laufer & Shmueli, 1997). Study 2 supports these calls, suggesting that, where appropriate, word meaning should be presented via the L1 as this can lighten the learning burden and reduce decay. Study 3 showed that simultaneous presentation of the spoken and written form lightened the learning burden without increasing the probability of decay. Thus, teachers, materials designers, and EdTech developers should adopt bimodal presentation of target items where possible. Studies 2 and 3 showed that vocabulary size was negatively correlated with learning burden, indicating that lower-level learners may need richer instruction to facilitate learning. Conversely, more ambitious targets can be set for advanced learners who also may not require such rich instruction to learn L2 vocabulary. However, as Study 3 showed that individual factors such as aptitude can impact learning burden, practitioners should consider setting minimal rather than maximal goals for learners. This would ensure the development of weaker students without limiting the growth of stronger learners. Lastly, Study 1 showed that even after an interval of 4 weeks, there was considerable retention of the target items at the level of form recognition. Furthermore, some attrition research has reported a relearning advantage for seemingly forgotten lexical knowledge (de Bot & Stoessel, 2000). Thus, it is important that students do not become too downcast when they experience lexical decay, but rather see it as part of the lexical development process.

Another finding relevant to Studies 2 and 3 was that increased learning burden was associated with faster decay. This finding was consistent whether burden was considered as frequency or time of exposure and suggests the words that learners find difficult to learn should receive richer instruction and/or be recycled more frequently to mitigate this loss. Crucially however, while there are certain variables that have been argued to impact learning burden in a consistent manner across a population (see Laufer, 1997; Schmitt, 2010), burden is essentially a by-learner by-item phenomenon (Higa, 1965); thus, the burden posed by target items will vary greatly for different learners. It is, therefore, challenging for material designers and teachers to modify instruction or recycling patterns pre-emptively for an entire cohort, although this should be attempted where a clear relationship exists between a certain variable and a pattern of decay (e.g., PoS). Instead, or perhaps in addition, learners should be told about the relationship between burden and decay so that they can adapt their learning behaviour on an ad hoc basis to increase the probability of retention. Thus, in addition to factoring in, anticipatorily, learning burden at a curricular level, encouraging the development of strategies to respond to burden as experienced is vital.

Such a bottom-up response to learning burden might involve modifying current recommendations for best practice. For instance, flashcards and spaced repetition are important vocabulary learning practices (Nation, 2013; Webb & Nation, 2017). However, research (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Nakata, 2015) and recommendations (Webb & Nation, 2017) for spacing tend to treat all items equally. The studies in this thesis suggest that for maximum efficacy, spacing patterns may need to vary as a function of burden. One approach to the systematic spacing of flashcards is the Leitner System (see Figure 7.3).

Figure 7.3

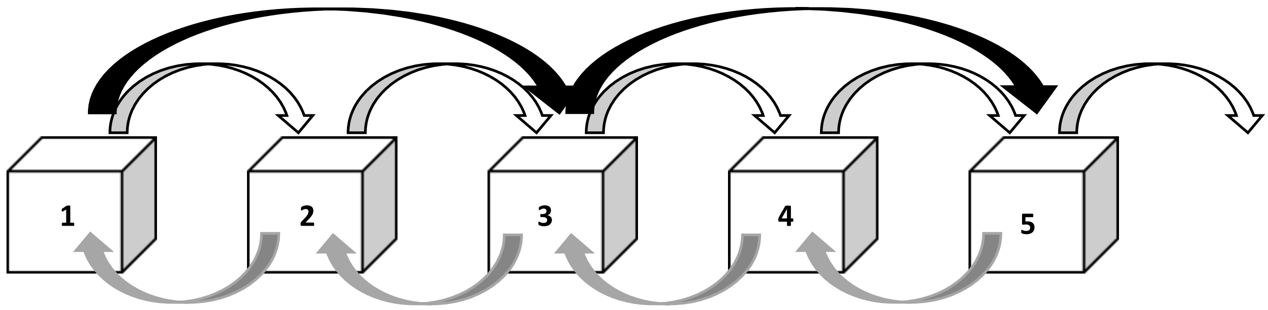
A Visual Representation of the Leitner System



This system of spaced repetition was developed in the 1970s to improve the efficacy of learning. Learners interact with one box at a time. Flashcards are moved to a subsequent box if they are successfully retrieved or demoted if they cannot be retrieved (Edge, Fitchett, Whitney, & Landay, 2012). The interval between repetitions increases as the words are promoted to increasingly higher boxes. This system represents a user-friendly method of operationalising spaced repetition and also ensures that words for which knowledge decays receive more encounters. However, given the findings of this thesis, there are two disadvantages of this system. First, words that are learned in the first instance are not differentiated according to burden. In the Leitner System, the effect of burden is considered retrospectively, by demoting words that cannot be retrieved and thus were also likely to have posed an initially high learning burden. One method of proactively responding to burden and thus avoiding unnecessary demotion would be to employ different degrees of promotion based on perceived burden (see Figure 7.4): words that are learned comparatively easily (represented by the black arrows) could be promoted faster than those which induced more burden (represented by the white arrows). This differential promotion would avoid the unnecessary repetition of known items, allowing learners to devote more time to those items most at risk of decay (i.e., the items that posed the greatest initial burden).

Figure 7.4

A Visual Representation of the Revised Leitner System



Secondly, in the Leitner system, words that cannot be recalled are demoted so that they receive more repetition. However, in addition to this additional exposure, increased richness of form/meaning presentation is likely to engender deeper memory traces for these problematic items. This rich exposure could be achieved in several ways: adding images (Bisson et al., 2015), adopting the keyword method (Ellis & Beaton, 1993), using L1 equivalents (Study 2, this thesis), etc. Not only would these additions likely decrease the probability of decay in and of themselves, but the act of locating this information and adapting the flashcards is also likely to result in increased engagement, which should further promote retention (Schmitt, 2008).

Importantly, these two suggestions can also be adopted in a CALL/MALL context. Language learning software and smartphone applications for deliberate vocabulary acquisition are increasingly common and popular study tools (Nakata, 2011). Vocabulary learning programmes such as *Anki* and *Word Engine* employ spacing algorithms to recycle target items. However, to minimise decay, such software should also consider factors which this research has shown to impact loss, such as PoS and code of meaning presentation, in their recycling algorithms. Additionally, such software could provide increasingly rich learning contexts for target items that pose greater learning burden. This means that recycling patterns and the nature of target item presentation would be adaptive to the learning behaviour of individual users. Thus, rather than use group norming data as a metric of burden, language learning applications need to interpret burden based on the frequency/time of exposure relative to the norms of each learner. In order to achieve this, platforms could initially collect baseline data on each user using target words manipulated for several intralexical factors. This could then be used to evaluate the burden of future learning, with words posing more burden recycled more quickly. Such calibration would need to be iterative as this thesis has shown learning burden to be influenced by learner proficiency. Adopting such computer adaptive spacing algorithms will likely expedite the learning process and facilitate more robust vocabulary acquisition.

7.5.2 Research

In the field of vocabulary studies, we have several fundamental axioms. For example, researchers agree that vocabulary learning is incremental (Schmitt, 2000), that frequency of exposure positively correlates with the probability of learning (Peters, 2014), and that learning is affected by several factors (Schmitt, 2010). The studies presented in this thesis contribute, or have the potential to contribute, to each of these agreed principles. Firstly, the development of word knowledge is commonly conceptualised as an incremental journey from no knowledge through partial knowledge to complete mastery. While this notion is no doubt accurate, it is important to stress that lexical acquisition is not a unidirectional process. Although it is tempting to visualise lexical acquisition as piecemeal progression with periods of growth and stability (illustrated in Figure 7.5), this thesis has clearly shown that backsliding occurs, confirming that lexical acquisition is non-linear (Figure 7.6).

Figure 7.5

An Illustration of a Linear Lexical Acquisition Process

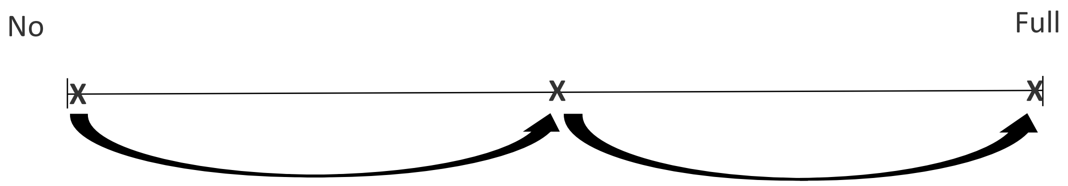
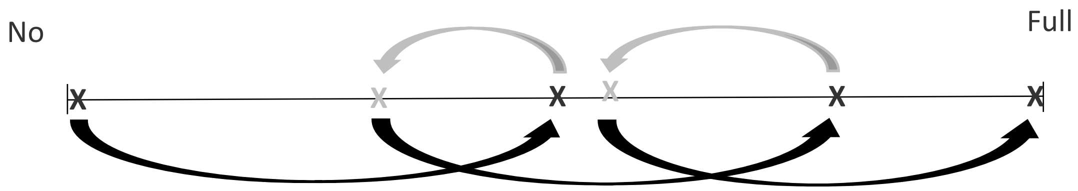


Figure 7.6

An Illustration of a Non-Linear Lexical Acquisition Process



These graphic representations are purely hypothetical and developmental patterns will differ by item and learner: one size will not fit all. This is the case because the learning burden of an item and learner variables affect the speed of lexical development (Webb & Nation, 2017). However, this is also the case because, as this thesis has shown, learning burden impacts the extent of backsliding that occurs, with some lexical items more susceptible to decay than others. Thus, just as the journey from zero knowledge to complete mastery varies by item and learner, so does the loss of knowledge that moves in the other direction. Partial and complete mastery are sometimes understood as receptive and productive knowledge. Schmitt (2019) argued that "learning most words to receptive mastery is relatively easy; it is enhancing that knowledge to productive mastery which is the real challenge" (p. 264). This thesis and the literature on L2 attrition suggest that receptive mastery is also relatively robust whereas productive mastery is more susceptible to decay. Therefore, it seems that productive mastery is difficult to develop and hard to maintain.

Another axiom of vocabulary studies is that frequency of exposure is positively associated with learning (Peters, 2014). Each target item presents a unique learning challenge; some words require more engagement and richer instruction to be acquired, while others can be learned with comparative ease. It is clear, therefore, that an increased number of exposures to a target item enhances the probability of sufficient engagement being expended for learning to occur. In this sense, frequency of exposure is positively associated with learning. However, this thesis has shown that in the specific learning procedure investigated which adopted a dynamic approach to frequency of exposure (i.e., learners could choose the number of times they could see the words), the number of repetitions needed for learning also indicates the learning burden. Words that need to be seen more often and for longer pose a greater learning challenge to a student. Seen in this light, an increased number of exposures is not inherently positive, but rather a sign of difficulty. Furthermore, Studies 2 and 3 show that the need for increased exposure to target vocabulary during the learning stage indicates a higher likelihood of decay. This finding requires us to re-evaluate the role frequency of exposure plays in the learning process when activities such as those employed in this thesis are used, framing the need for increased exposure as a metric of burden and a predictor of decay.

Finally, for some time the field of vocabulary studies has recognised that lexical acquisition is a complex construct affected by several factors (see Laufer, 1997; Schmitt, 2010). Intralexical, interlexical, contextual, and learner-related variables impact the learning process, making research into these factors a key area of investigation. Such research has disambiguated the effect of certain variables (e.g., concreteness, cognateness, word frequency, PoS) on the learning process. However, as has been argued throughout this thesis, lexical decay should be considered as a part of the development process. Therefore, in order to develop a complete understanding of the effect of these variables on vocabulary development, research needs to consider their effect on both the learning and decay processes. This thesis has shown that some variables have an effect on learning and decay, while other factors have an effect on one but not the other. Thus, we cannot necessarily assume the effect of a factor on decay from its relationship with learning and thus explicit measurement is required to fully understand the role numerous variables play in lexical development.

Schmitt (2019) argued that one of the most pressing needs for the field of vocabulary studies is the delineation of a practical model of vocabulary acquisition, detailing the development of lexical knowledge from no knowledge to productive mastery. However, the process of lexical development consists not only of sustained attainment, but also decay; thus, such a model needs to include the backsliding of knowledge and the various factors which affect this process in order to be considered comprehensive. Interestingly, the few models that have been proposed to date make little allowance for decay (e.g., Brown & Payne, cited in Hatch & Brown, 1995). To better understand the non-linearity of lexical development, studies should, where possible, employ longitudinal designs, tracking the nature of lexical knowledge at multiple points. This is partially the case because longitudinal studies better account for individual variability, which has been shown to impact developmental patterns (e.g., associative memory capacity, Study 3). However, it is also the case that cross-sectional designs may confuse the non-acquisition of word knowledge with the decay of that knowledge. That is, descriptive studies utilising cross-sectional designs may find word knowledge to differ by item, concluding that some words, or word knowledge aspects, are acquired before others. Such a conclusion though, discounts the possibility that item knowledge was once developed but subsequently decayed. Thus, inferring an order of acquisition from such cross-sectional studies may prove problematic.

In addition to advocating longitudinal methodologies, this thesis has three further implications for research design. Firstly, the three studies presented in this thesis employed electronic flashcard software both to introduce and to measure the learning of target items. A comparison of different platforms showed that *Anki* was best suited for this research (see Chapter 3). One of the main reasons for this selection was that *Anki* allows access to learning information at the learner and item level, facilitating measurement of frequency and time of exposure, and disambiguates exposures during the encoding process from retrievals of encoded knowledge. Thus, this software provides a nuanced view of learning rarely seen in classroom-based vocabulary studies while ensuring ecological validity. Furthermore, as exposures can be differentiated from retrievals, such software allow researchers to control the frequency of retrieval rather than the frequency of exposure. There are two advantages here: a) as increased retrieval frequency is associated with the development of deeper memory traces (Baddeley, 1990), controlling retrieval frequency ensures that any loss that occurs is a result of natural forgetting patterns rather than an artefact of methodological design; and b) as proficiency, aptitude, and other individual factors are likely to affect the frequency of exposure needed to overcome the learning burden, adopting a dynamic approach to frequency of exposure allows all learners to see all items as often as they need for learning to occur, limiting the confounding effect of learner variables on the data. However, although such software provides informative data, future research should include keystroke logging tools to capture online features of learner production (e.g., revisions, time to onset of lexical production) and develop automatic scoring protocols to eliminate the need for learner self-report.

Second, as a result of using electronic flashcard software and adopting a dynamic approach to frequency of exposure, this thesis was able to employ a frequency-based approach to the measurement of learning burden. As discussed in Section 2.1.4, this is preferable to a gains-based metric as a measure of burden. However, as gains are more easily measured, the gains-based approach tends to predominate in the literature (e.g., Ishii, 2015). Furthermore, of the studies that have adopted a frequency-based approach, few have managed to do so in an ecologically valid manner (see Tinkham, 1993, 1997; Waring, 1997). The methodology employed in this thesis both adopted a frequency-based metric of burden and did so using a common learning tool, thus maintaining ecological validity. Additionally, the frequency-based approach facilitated consideration of learning burden as a covariate in the decay analysis; something that, to the best of my knowledge, has not been considered in any research to date. Thus, future research targeting learning burden should, where possible, adopt a frequency-based approach and consider employing electronic flashcard software to achieve this end.

The final methodological feature of note relates to the use of adaptive tests. Vocabulary tests provide particularly salient exposures to target items (Schmitt, 2010) and thus can impact the extent to which target items are learned. Controlling for this test effect is especially important when considering lexical decay because any exposure outside the learning treatment may bias the decay data. In other words, repeated measurement on a test of attainment may strengthen encoded knowledge making demonstration of that knowledge on a test of retention more likely. To mitigate this testing effect, this thesis employed an adaptive battery measuring written form recall and recognition. This design utilised the implicational association between recall and recognition, where knowledge at the level of form recall assumes knowledge at the level of form recognition (Gonzalez-Fernandez & Schmitt, 2019). Future research in this area should likewise take steps to minimise the testing effect, potentially by adopting similar adaptive instruments.

7.6 Conclusion

As this chapter has demonstrated, there are numerous avenues for future studies of learning burden and lexical decay and the methodology employed in this thesis may provide direction for the design of these investigations. Yet, although there is much still to be done, the studies in this thesis have made a considerable contribution to our understanding of these areas. Independently, the three research investigations have illuminated the effect of several key variables on burden and decay. Together, the findings demonstrate that some factors can have one effect on learning burden and a quite different effect on decay, while other variables impact both in a similar manner. These findings help to explain the causes of learning burden and lexical decay and, with regard to the latter, to move us beyond both the general assumption that second language knowledge simply decays as a function of time and platitudes such as "*use it or lose it".*

In doing so, this thesis has demonstrated that decay should be considered as part of the process of lexical development. To date, research has understood this lexical development in terms of learning. The majority of acquisition studies do not consider the backsliding of knowledge and the models of lexical development we have (e.g., Brown & Payne, 1994, cited in Hatch & Brown, 1995) generally ignore lexical decay. This thesis has demonstrated that, in addition to learning, lexical decay should be included as a measure of pedagogical efficacy. Furthermore, it has shown that any comprehensive theory of lexical development needs to include explanation of decay and the various features that impact it. Achieving this feat will not be an easy endeavour; L2 vocabulary acquisition is incredibly complex and adding a further metric only adds to the complexity of the construct. However, if we truly want to understand the phenomenon of lexical development and make impactful suggestions for language pedagogy, it is a complexity we need to embrace.

REFERENCES

Abbasian, R., & Khajavi, Y. (2010). Lexical attrition of general and special English words after years of non-exposure: The case of Iranian teachers. *English Language Teaching, 3*(3), 47-53.

Al-Hazemi, H. (2000). Lexical attrition of some Arabic speakers of English as a Foreign language: A study of word loss. *The Internet TESL Journal, VI*(12).

Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. Continuum.

Atkins, B. T. S., & Varantola, K. (1998). Language learners using dictionaries: The final report on the EURALEX/AILA research project on dictionary use. In B. T. S. Atkins (Ed.), *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators.* Max Niemeyer Verlag.

Atkinson, D. (1989). The mother tongue in the classroom: A neglected resource? *ELT Journal, 41*(4), 241-247.

Baddeley, A. D. (1990). *Human Memory and Theory and Practice*. Lawrence Erlbaum Associates.

Baddeley, A. D. (2013). *Essentials of Human Memory*. Psychology Press.

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.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113*(1), 1-29.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of Foreign Language Vocabulary and the Spacing Effect. *Psychological Science, 4*(5), 316-321.

Bahrick, H. P., & Phelps, E. (1987). Rentention of Spanish Vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(2), 344-349.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning, 52*(2), 323-363.

Barcroft, J. (2015). *Lexical Input Processing and Vocabulary Learning*. John Benjamins.

Barcroft, J., & Rott, S. (2010). Partial Word Form Learning in the Written Mode in L2 German and Spanish. *Applied Linguistics, 31*(5), 623-650.

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition*, 387-414.

Bardovi-Harlig, K., & Stringer, D. (2010). Variables in Second Language Attrition. *Studies in Second Language Acquisition, 32*(01), 1.

Bardovi-Harlig, K., & Stringer, D. (2013). The lexicon in second language attrition: What happens when the cat's your tongue? In J. Altarriba & L. Isurin (Eds.), *Memory, Language, and Bilingualism* (pp. 291-319). Cambridge University Press.

Barker, D. (2003). Why English Teachers in Japan Need to Learn Japanese. *The Language Teacher, 2*.

Bartolotti, J., & Marian, V. (2017). Orthographic Knowledge and Lexical Form Influence Vocabulary Learning. *Appl Psycholinguist, 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)*.

Beaton, A., Gruneberg, M., & Ellis, N. (1995). Retention of foreign vocabulary learned using the keyword method: a ten-year follow-up. *Second Language Research, 11*(2), 112-120.

Bell, H. (2009). The messy little details: a longitudinal case study of the emerging lexicon. *Lexical processing in second language learners*, 111-127.

Bensoussan, M., Sim, D., & Weiss, R. (1984). The Effect of Dictionary Usage on EFL Test Performance Compared kk Student and Teacher Attitudes and. *Reading in a Foreign Language, 2*(2), 262.

Berger, M. (2011). English-Only Policy for All? Case of a University English Class in Japan. *Polyglossia, 20*, 27-43

Berko-Gleason, J. (1982). Insights from Child Language Acquisition for Second Language Loss. In R. D. Lambert & B. F. Freed (Eds.), *The Loss of Language Skills* (pp. 13-23). Newbury House.

Berlitz. (2009). *History. The Berlitz Story.* http://www.berlitz.de/en/berlitz_company/tradition/history/

Berman, R. A., & Olshtain, E. (1983). Features of first language transfer in second language attrition. *Applied Linguistics, 4*(3), 222-234.

Bert, W., Kees de, B., & Theo van, E. (Eds.). (1986). *Language Attrition in Progress*. Foris.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4), 509.

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *Quarterly Journal of Experimental Psychology, 68*(7), 1306-1326.

Bokander, L., & Bylund, E. (2020). Probing the Internal Validity of the LLAMA Language Aptitude Tests. *Language Learning, 70*(1), 11-47.

Bot, K. d., Clyne, M., & van Els, T. J. M. (1989). Dimensions of Language Attrition. *ITL - International Journal of Applied Linguistics, 83/84*.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136-163.

Burgess, S., R., & Murray, A., B. (2014). Use of traditional and smartphone app flashcards in an introductory psychology class. *Journal of Instructional Pedagogies 13*. 1-7.

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2009). The Role of Language Aptitude in First Language Attrition: The Case of Pre-pubescent Attriters. *Applied Linguistics, 31*(3), 443-464.

Cai, G. (2011). *The tertiary English language curriculum in China and its delivery: A critical study.* Open University, UK

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. S. Parry & C. W. Stansfield (Eds.), *Language in education: Theory and practice 74. Language aptitude reconsidered* (pp. 11-27). Prentice Hall Regents/Prentice Hall.

Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test*. Psychological Corporation.

Carter, R., & McCarthy, M. (1988). *Vocabulary and language teaching*. Longman London.

Catalán, R. M. J. (2003). Sex differences in L2 vocabulary learning strategies. *International Journal of Applied Linguistics, 13*(1), 54-77.

Ceci, S. J., & Howe, M. J. A. (1978). Semantic knowledge as a determinant of developmental differences in recall. *Journal of Experimental Child Psychology, 26*(2), 230-245.

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.

Chapman, F. L., & Gilbert, L. C. (1937). A study of the influence of familiarity with English words upon the learning of their foreign language equivalents. *Journal of Educational Psychology, 28*(8), 621-628.

Chien, C. W. (2015). Analysis the Effectiveness of Three Online Vocabulary Flashcard Websites on L2 Learners' Level of Lexical Knowledge. *English Language Teaching, 8*(5), 111-121.

Chukharev-Hudilainen, E. (2014). Pauses in spontaneous written communication: A keystroke logging study. *Journal of Writing Research, 6*(1), 61-84.

Chung, E. (2018). Revisiting Second Language Vocabulary Teaching: Insights from Hong Kong In-Service Teachers. *Asia-Pacific Education Researcher, 27*(6), 499-508.

Chung, J., & Choi, T. (2016). English Education Policies in South Korea: Planned and Enacted. In R. Kirkpatrick (Ed.), *English Language Education Policy in Asia*. Springer.

Cohen, A. D. (1974). The Culver City Spanish Immersion Program: How does summer recess affect Spanish speaking ability? *Language Learning, 24*, 55-68.

Cohen, A. D. (1975). Forgetting a Second Language. *Language Learning, 25*(1), 127-138.

Cohen, A. D. (1989). Attrition in the Productive Lexicon of Two Portuguese Third Language Speakers. *Studies in Second Language Acquisition, 11*, 135-149.

Cole, R. D. (1929). The effect of a summer vacation on students' knowledge of French. *The Modern Language Journal, 14*(2), 117-121.

Cook, V. (2001). Using the first language in the classroom. *Canadian Modern Language Review, 57*(3), 402–423.

Cooper, C. (2002). *Individual Differences*. Routledge.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narritive and Meta-Analytic Review. *Review of Educational Research, 66*(3), 227-268.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of Processing: A Framework for Memory Research. *Journal of Learning and Verbal Behavior, 11*, 671-684.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research, 28*(3), 369-382.

Dahlen, K., & Caldwell-Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal, 97*(4), 902-916.

Davies, A. (2003). *The native speaker: Myth and reality*. Multilingual Matters.

de Bot, K. (2004). Special Issue on language attrition. *International Journal of Bilingualism, 8*(3).

de Bot, K., & Stoessel, S. (2000). In Search of Yesterday's Words: Reactivating a Long-Forgotten Language. *Applied Linguistics, 21*(3), 333-353.

de Bot, K., & Weltens, B. (1995). Foreign Language Attrition. *Annual Review of Applied Linguistics, 15*, 151-164.

de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets Of Speaking Proficiency. *Studies in Second Language Acquisition, 34*(1), 5-34.

de Smet, M. J. R., Brand-Gruwel, S., Leijten, M., & Kirschner, P. A. (2014). Electronic outlining as a writing strategy: Effects on students' writing products, mental effort and writing process. *Computers & Education, 78*, 352-366.

deHaan, J., Reed, W. M., & Kawada, K. (2010). The Effect of Interactivity with a Music Video Game on Second Langage Vocabulary Recall. *Language Learning and Technology, 14*(2), 74-94.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum. .

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Blackwell.

Duff, P., & Polio, C. (1990). How much foreign language is there in the foreign language classroom? . *The Modern Language Journal, 74* (2), 154–166.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychol Sci, 18*(1), 35-39.

Ebbinghaus, H. (1885). *Über das Gedchtnis. Untersuchungen zur experimentellen Psychologie*. Duncker & Humblot.

Edge, D., Fitchett, S., Whitney, M., & Landay, J. (2012). MemReflex: adaptive flashcards for mobile microlearning. Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services,

Edstrom, A. M. (2006). L1 Use in the L2 Classroom: One Teacher's Self-Evaluation. *Canadian Modern Language Review, 2*(09), 275-292.

EEC. (2018). English Syllabus. Ehime University.

Ehrman, M. E., & Oxford, R. L. (1995). Cognition Plus: Correlates Of Language Learning Success. *The Modern Language Journal, 79*(1), 67-89.

Elgort, I. (2011). Deliberate Learning and Vocabulary Acquisition in a Second Language. *Language Learning, 61*(2), 367-413.

Elgort, I., & Piasecki, A. E. (2014). The effect of a bilingual learning mode on the establishment of lexical semantic representations in the L2. *Bilingualism: Language and Cognition, 17*(3), 572-588.

Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology, 46A*(3), 533-558.

Ellis, R., & Shintani, N. (2013). *Exploring Language Pedagogy through Second Language Acquisition Research*. Routledge.

Engel de Abreu, P. M. J., & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology, 104*, 974-986.

Ertmer, P. A., & Newby, T. J. (1993). Behaviorism, Cognitivism, Constructivism: Comparing Critical Features from an Instructional Design Perspective. *Performance Improvement Quarterly, 6*(4), 50-72.

Fatemi, A. H., & Barani, G. (2014). The Impact of Teachers' Code-Switching on the Vocabulary Learning of Iranian University EFL Learners. *Journal of Language Sciences & Linguistics, 2*(4), 91-98.

Field, A. (2018). *Discovering Statistics Using R*. Sage.

Folse, K. (2004). *Vocabulary myths*. University of Michigan Press.

Folse, K. S. (2006). The Effect of Type of Written Exercise on L2 Vocabulary Retention. *TESOL Quarterly, 40*(2), 273-293.

Fortune, J. (2012). The forbidden fruit: using the mother tongue in a Bogota university EFL programme. *Columbia Applied Linguistics Journal, 14*(2), 70-87.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research, 74*(1), 59-109.

Freed, B. F. (1982). Language Loss: Current Thoughts and Future Directions. In R. D. Lambert & B. F. Freed (Eds.), *The Loss of Language Skills* (pp. 1-6). Newbury House.

Fukazawa, S. (1984). A study of English retention by Japanese college students over summer vacation. *The Chugoku Academic Society of English Language Education Bulletin, 14*, 29-31.

Gablasova, D. (2015). Learning technical words through L1 and L2: Completeness and accuracy of word meanings. *English for Specific Purposes, 39*, 62-74.

Gardner, R. C. (1982). Social factors in language retention. In R. D. Lambert & B. Freed (Eds.), *Loss of Language Skills*. newbury House.

Gardner, R. C., Lalonde, R. N., Moorcroft, R., & Evers, F. T. (1987). Second Language Attrition: The Role of Motivation and Use. *Journal of Language and Social Psychology, 6*(1), 29-47.

Gardner, R. C., & Lysynchuk, L. M. (1990). The Role of Aptitude, Attitudes, Motivation, and Language Use on Second-Language Acquisition and Retention. *Canadian Journal of Behavioural Science, 22*(3), 254-270.

Gardner, R. C., & Maclntyre, P. D. (1991). An Instrumental Motivation in Language Study: Who Says It Isn't Effective? *Studies in Second Language Acquisition, 13*, 57-72.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought, and culture* (pp. 301-334). Erlbaum. Retrieved from http://hdl.handle.net/2142/17514

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.

Glick, C. (2003). Considerations for Securing an English Teaching Position at a Japanese University (Part 2). *The Language Teacher, 26*(9), 7-9.

González-Fernández, B., & Schmitt, N. (2019). Word Knowledge: Exploring the Relationships and Order of Acquisition of Vocabulary Knowledge Components.

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.

Grace, C. (1998). Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning-level CALL software. *The Modern Language Journal, 82*(4), 533-544.

Grace, C. A. (2000). Gender Differences: Vocabulary Retention and Access to Translations for Beginning Language Learners in CALL. *The Modern Language Journal, 84*(2), 214-224.

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105-130). John Benjamins.

Granena, G. (2014). Language Aptitude and Long-term Achievement in Early Childhood L2 Learners. *Applied Linguistics, 35*(4), 483-503.

Granena, G., & Long, M. H. (2012). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*(3), 311-343.

Gries, S. (2013). *Statistics for Linguistics with R, A Practical Introduction*. De Gruyter Mouton.

Griffin, G., & Harley, T. (1996). List learning of second language vocabulary. *Applied Psycholinguistics, 17*(4), 443-460.

Groot, A. M. B. d. (2006). Effects of Stimulus Characteristics and Background Music on Foreign Langanguage Learning and Forgetting. *Language Learning, 56*(3), 463-506.

Groot, A. M. B. d., & 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.

Gu, P. Y. (2020). Strategies for Learning Vocabulary. In W. Stuart (Ed.), *The Routledge Handbook of Vocabulary Studies*. Routledge.

Gyllstad, H., Vilkaite, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL International Journal of Applied Linguistics, 166* 276–303.

Hall, G., & Cooks, G. (2012). Own-language use in language teaching and learning: state of the art. *Language Teaching, 45*(3), 271-308.

Hansen, L. (1999). Not a total loss: The attrition of Japanese negation over three decades. In L. Hansen (Ed.), *Second Language Attrition in Japanese Contexts* (pp. 142-153). Oxford University Press.

Hansen, L. (Ed.). (1999). *Second language attrition in Japanese contexts*. Cambridge University Press.

Hansen, L. (2011). The acquisition, attrition and relearning of mission vocabulary. In M. S. Schmid & W. Lowie (Eds.), *Modelling bilingualism: From structure to chaos* (pp. 115–134). John Benjamins.

Hansen, L. (2012). Second Language Vocabulary Attrition. In C.A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*, Wiley Blackwell.

Hansen, L., & Chantrill, C. (1999). Literacy as a second language anchor: Evidence from L2 Japanese and L2 Chinese. Representation and process: Proceedings of the 3rd Pacific Second Language Research Forum, Tokyo: Aoyama Gakuin University.

Hansen, L., & Chen, Y.-L. (2001). What counts in the acquisition and attrition of numeral classifiers? *23, 1*, 90-110.

Hansen, L., & Reetz-Kurashige, A. (1999). Investigating second language attrition: An introduction. In L. Hansen (Ed.), *Second language Attrition in Japanese Contexts* (pp. 3-20). Oxford University Press.

Hansen, L., Umeda, Y., & McKinney, M. (2002). Savings in the Relearning of Second Language Vocabulary: The Effects of Time and Proficiency. *Language Learning, 52*(4), 653-678.

Harbord, J. (1992). The use of the mother tongue in the classroom. *ELT Journal, 46*(4), 350– 355.

Hashemi, M. R., & Gowdasiaei, F. (2005). An Attribute-Treatment Interaction Study: Lexical-Set versus Semantically-Unrelated Vocabulary Instruction. *RELC Journal, 36*(3), 341-361.

Hatch, E., & Brown, C. (1995). *Vocabulary, Semantics, and Language Education*. Cambridge University Press

Hayashi, G. (2011). Language Attrition of Japanese Returnee Students. *Accents Asia [Online], 4*(1), 20-35.

Hayati, M., & Fattahzadeh, A. (2006). The effect of monolingual and bilingual dictionaries on vocabulary recall and retention of EFL learners. *The Reading Matrix, 6*(2).

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*(2), 303-317.

HESA. (2017). *HE Student Data | HESA*. Retrieved 8th February from https://www.hesa.ac.uk/data-and-analysis/students

Higa, M. (1965). The psycholinguistic concept of "difficulty" and the teaching of foreign language vocabulary. *Language Learning, 15*(3-4), 167-179.

Hill, M. M. (1994). A Word in your Ear: vocabulary acquisition by Chinese tertiary students. . In N. Bird, P. Falvey, A. B. M. Tsui, D. M. Alison, & A. McNeill (Eds.), *Language and Learning*. Government Printer.

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.

Horst, M., White, J., & Bell, P. (2010). First and second language knowledge in the language classroom. *International Journal of Bilingualism, 14*(3), 331-349.

Hoshino, Y. (2010). The Categorical Facilitation Effects on L2 Vocabulary Learning in a Classroom Setting. *RELC Journal, 41*(3), 301-312.

Huckin, T., & Coady, J. (1999). Incidental Vocabulary Acquisition in a Second Language. *Studies in Second Language Acquisition, 21*, 181-193.

Hulstijn, J. H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *The Modern Language Journal, 77*(2), 139-147.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal, 80*(3), 327-339.

Hulstijn, J. H., & Laufer, B. (2001). Some Empirical Evidence for the Involvement Load Hypothesis in Vocabulary Acquisition. *Language Learning, 51*(3), 539–558.

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.

Hutz, M. (2004). Is there a natural process of decay? A longitudinal study of language attrition. . In M. S. Schmid, B. Kopke, M. Keijzer, & L. Weilemar (Eds.), *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues* (pp. 189-206). John Benjamins.

Ishii, T. (2015). Semantic connection or visual connection: Investigating the true source of confusion. *Language Teaching Research, 19*(6), 712-722.

Jacobs, G. M., Dufon, P., & Fong, C. H. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading, 17*(1), 19-28.

Jahangard, A. (2007). Which word type (technical or general) are more difficult to retain by the Iranian school learners? *The Asian ESP Journal, 3*(2), 6-23.

Kalivoda, T. B. (1990). Teaching Grammar in the Target Language. *Hispania, 73*(1), 267-269.

Kaneko, M. (2017). Vocabulary Size Targets for the TOEIC Test. *JACET Journal, 61*, 57-67.

Karpicke, J. D., & Roediger III, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science, 319*, 966-968.

Knight, S. (1994). Dictionary, the tool of last resort effects on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal, 78*(3), 285-298.

Kramer, B., Matsuo, T., McLean, S., & Cornwell, S. (2019). Measuring Attrition of L2 Receptive Vocabulary Knowledge Over the Summer Vacation. *Vocabulary Education and Research Bulletin, 8*(1), 17-22.

Kuhberg, H. (1992). Longitudinal L2-attrition versus L2-acquisition, in three Turkish children - empirical findings. *Second Language Research, 8*(2), 138-154.

Lado, R., Baldwin, B., & Lobo, F. (1967). *Massive vocabulary expansion Expansion in a Foreign Language Beyond the Basic Course: The Effects of Stimuli, Timing, and Order of Presentation.* U.S. Department of Health, Education, and Welfare.

Lamb, M. (2011). A Matthew Effect in English language education in a developing country context. . In *Dreams and Realities: Developing Countries and the English Language* (pp. 186 - 206). The British Council.

Lambert, R., & Moore, S. (1986). Problem areas in the study of language attrition. In B. Weltens, K. D. Bot, & T. V. Els (Eds.), *Language attrition in progress* (pp. 177-186). Foris.

Lambert, R. D., & Freed, B. F. (1982). Preface. In R. D. Lambert & B. F. Freed (Eds.), *The Loss of Language Skills*. Newbury House Publishers.

Laufer-Dvorkin, B. (1991). *Similar Lexical Forms in Interlanguage*. Narr.

Laufer, B. (1990). Words you know: how they affect the words you learn. In J. Fisiak (Ed.), *Further insights into contrastive linguistics* (pp. 573-593). Benjamins.

Laufer, B. (1994). Monolingual, Bilingual and 'Bilingualised' Dictionaries: Which are More Effective, for What and for Whom? Proceedings of the 6th EURALEX International Congress,

Laufer, B. (1997). What's in a word that makes it hard or easy? Intralexical factors affecting the difficult of vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition, and Pedagogy* (pp. 140-155). Cambridge University Press.

Laufer, B. (2006). Comparing Focus on Form and Focus on FormS in Second-Language Vocabulary Learning. *The Canadian modern language review, 63*(1), 149-166.

Laufer, B. (2017). From word parts to full texts: Searching for effective methods of vocabulary learning [editorial]. *Language Teaching Research, 21*(1), 5-11.

Laufer, B., & Aviad–Levitzky, T. (2017). What Type of Vocabulary Knowledge Predicts Reading Comprehension: Word Meaning Recall or Word Meaning Recognition? *The Modern Language Journal, 101*(4), 729-741.

Laufer, B., & Girsai, N. (2008). Form-focused Instruction in Second Language Vocabulary Learning: A Case for Contrastive Analysis and Translation. *Applied Linguistics, 29*(4), 694-716.

Laufer, B., & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language Learning, 54*(3), 399-436.

Laufer, B., & Levitzky-Aviad, T. (2006). Examining the effectiveness of 'bilingual dictionary plus'–a dictionary for production in a foreign language. *International Journal of Lexicography, 19*(2), 135-155.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabualry size and readin comprehension. *Reading in a Foreign Language, 22*(1), 15-30.

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

Laufer, B., & Shmueli, K. (1997). Memorizing New Words: Does Teaching Have Anything to Do with It? *RELC Journal, 28*(1), 89-107.

Lee, J. H., & Levine, G. S. (2018). The effects of instructor language choice on second language vocabulary learning and listening comprehension. *Language Teaching Research*, 24(2), 250-272.

Lee, J. H., & Macaro, E. (2013). Investigating Age in the Use of L1 or English-Only Instruction: Vocabulary Acquisition by Korean EFL Learners. *Modern Language Review, 97*(4), 887-901.

Leijten, M., & Van Waes, L. (2006). Inputlog: New Perspectives on the Logging of On-Line Writing Processes in a Windows Environment. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer Keystroke Logging: Methods and Applications*. Elsevier.

Leijten, M., & Van Waes, L. (2013). Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes. *Written Communication, 30*(3), 358-392.

Leki, I., & Carson, J. (1994). Students' perception of EAP writing instruction and writing needs across the disciplies. *TESOL Quarterly, 28*(1), 81-101.

Lew, R. (2004). How do Polish learners of English rate bilingual and monolingual dictionaries? Proceedings of the 11th EURALEX International Congress, Lorient, France.

Li, S. (2016). The Construct Validity of Language Aptitude: A Meta-Analysis. *Studies in Second Language Acquisition, 38*(4), 801-842.

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & Doughty, C. J. (2013). Hi-LAB: A New Measure of Aptitude for High-Level Language Proficiency. *Language Learning, 63*(3), 530-566.

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.

Luppescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning, 43*(2), 263-279.

Macaro, E. (2006). Codeswitching in the L2 classroom: A communication and learning strategy. . In E. Llurda (Ed.), *Non-native language teachers: Perceptions, challenges and contributions to the profession.* (pp. 63–84). Springer.

Marefat, H., & Rouhshad, A. (2007). Second language attrition: Are different nouns equally likely to be lost? *Porta Linguarum, 8*(junio), 85-98.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(3), 558.

Meara, P. (1992). *EFL Vocabulary Tests*. Centre for Applied Linguistics Studies.

Meara, P. (2004). Modelling vocabulary loss. *Applied Linguistics, 25*(2), 137-155.

Meara, P. (2018). Mapping Vocabulary Research. *Journal of Foreign Language Education and Technology, 3*(1), 280-308.

Meara, P., Milton, J., & Lorenzo-Dus, N. (2003). *Language Aptitude Tests*. Express.

Meara, P., & Mirelpeix, I. (2016). *Tools for Researching Vocabulary*. Multilingual Matters.

Meara, P. M. (2005). *Llama Language Aptitude Tests*. Lognostics.

Medler, D. A., & Binder, J. R. (2005). *MCWord: An On-Line Orthographic Database of the English Language*

Mehotcheva, T. H. (2010). *After the fiesta is over: Foreign Language Attrition of Spanish in Dutch and German Erasmus students* [PhD Thesis]. University of Groningen/Pompeu Fabra University.

Mext (2013). *English Education Reform Plan Corresponding to Globalisation*. From http://www.mext.go.jp/en/news/topics/detail/__icsFiles/afieldfile/2014/01/23/1343591_1.pdf.

Mishima, T. (1967). An experiment comparing five modalities of conveying meaning for the teaching of foreign language vocabulary. *Dissertation Abstracts, 27*.

Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students? . *Language Education & Technology, 39*, 1-20.

MOE. (1997). 차 초등교육과정 – 영어 . *[7th Primary Curriculum – English]*. Ministry of Education.

Moorcroft, R., & Gardner, R. C. (1987). Linguistic factors in second-language loss. *Language Learning, 37*(3), 327-340.

Morshedian, M. (2008). The role of initial English as foreign language proficiency in lexical attrition/retention of Iranian learners: Is productive or the receptive word knowledge of learned nouns mre likely to be lost? *The Linguistics Journal, 3*(1), 75-95.

Murakami, A. (2016). Modeling Systematicity and Individuality in Nonlinear Second Language Development: The Case of English Grammatical Morphemes. *Language Learning, 66*(4), 834-871.

Murtagh, L., & Slik, F. v. d. (2004). Retention of Irish skills: A longitudinal study of a school-acquired second language. *International Journal of Bilingualism, 8*(3), 279-302.

Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: implications from cognitive psychology research for optimal spaced learning. *ReCALL, 20*(1), 3-20.

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: a critical investigation of flashcard software. *Computer Assisted Language Learning, 24*(1), 17-38.

Nakata, T. (2015). Effects of Expanding and Equal Spacing on Second Language Vocabulary Learning: Does Gradually Increasing Spacing Increase Vocabulary Learning? *Studies in Second Language Acquisition, 37*(677 – 711).

Nakata, T. (2016). Does Repeated Practice Make Perfect? The Effects of Within-Session Repeated Retrieval on Second Language Vocabulary Learning. *Studies in Second Language Acquisition*, 1-27.

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.

Nakata, T., & Webb, S. (2016). Does studying vocabulary in smaller sets increase learning. *Studies in Second Language Acquisition, 38*(3), 523-552.

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. Heinle and Heinle.

Nation, I. S. P. (2006). How Large a Vocabulary Is Needed For Reading and Listening. *The Canadian modern language review, 63*(1), 59-82.

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques.* Heinle, Cengage Learning.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Nation, I. S. P., & Webb, S. (2011). *Researching and Analying Vocabulary*. Heinle, Cengage Learning.

Nation, P. (2001). *Learning vocaublary in another language*. Cambridge University Press.

Nation, P. (2005). Teaching Vocabulary. *Asian EFL Journal, 7*(3).

Nation, P. (2012). *The BNC/COCA word family lists*. https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/information-on-the-BNC_COCA-word-family-lists

Nation, P. (2013). *Learning Vocabulary in Another Language*. Cambridge University Press.

Nation, P. (2020). The Different Aspects of Vocabulary Knowledge. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies*. Routledge.

Neisser, U. (1984). Interpreting Harry Bahrick's discovery: What confers immunity against forgetting? *Journal of Experimental Psychology: General, 113*(1), 32-35.

Nguyen, L. T. C., & Nation, P. (2011). A Bilingual Vocabulary Size Test of English for Vietnamese Learners. *RELC Journal, 42*(1), 86-99.

Nirattisai, S., & Palanukulwong, T. (2016). A new measure for Thai students' vocabulary size: the English-Thai version of vocabulary size test. *Journal of English Studies, 14*, 155-181.

NOVA. (n.d.). *About Us*. Retrieved 8th February from http://nova-holdings.jp/teachinjapan/aboutus.html

Olshtain, E. (1989). Is Second Language Attrition the Reversal of Second Language Acquisition? *Studies in Second Language Acquisition, 11*, 151-165.

Olshtain, E., & Barzilay, M. (1991). Lexical retrieval difficulties in adult language attrition. In H. W. Seliger & R. M. Vago (Eds.), *First Language Attrition* (pp. 139-150). Cambridge University Press.

Ono, L., & Jordan, J. (2018). *ELP Student Handbook*. Kwansei Gakuen University.

Ortega, L. (2009). *Understanding second language acquisition*. Hodder Education.

Paechter, M., Luttenberger, S., Macher, D., Berding, F., Papousek, I., Weiss, E. M., & Fink, A. (2015). The Effects of Nine-Week Summer Vacation: Losses in Mathematics and Gains in Reading. *EURASIA Journal of Mathematics, Science & Technology Education*.

Paivio, A. (1963). Learning of adjective-noun paired associates as a function of adjective-noun word order and noun abstractness. . *Canadian Journal of Psychology/Revue canadienne de psychologie, 17* (4), 370-379.

Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. In B. Kopke, M. S. Schmid, M. Keijzer, & S. Dostert (Eds.), *Language attrition: Theoretical perspectives* (pp. 121-133). John Benjamins.

Paribakht, T. S., & Wesche, M. (1996). Enhancing vocabulary acquisition through reading: A hierarchy of text-related exercise types. *The Canadian modern language review, 52*(2), 155-178.

Park, E. S. (2018). Language Attrition. In J. I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching*. John Wiley & Sons.

Pellicer-Sánchez, A. (2016). Incidental L2 Vocabulary Acquisition from and while Reading: An Eye-Tracking Study. *Studies in Second Language Acquisition*, 38 (1), 97-130.

Pellicer Sanchez, A. (2017). *Incidental learning of L2 collocations: A classroom study*. Language Teaching Research, 21 (3), 381-402.

Peters, E. (2007). Manipulating L2 learners' online dictionary use and its effect on L2 word retention. *Language Learning & Technology, 11*(2), 36-58.

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. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research, 20*(1), 113-138.

Peters, E. (2020). Factors Affecting the Learning of Single-Word Items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies*. Routledge.

Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German Vocabulary Through Reading: The Effect of Three Enhancement Techniques Compared. *Language Learning, 59*(1), 113-151.

Peters, E., & Webb, S. (2018). Incidental Vocabulary Acquisition through Viewing L2 Television and Factors that Affect Learning. *Studies in Second Language Acquisition, 40*(3), 551-577.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1-28.

Pimsleur, P. (1967). A Memory Schedule. *Modern Language Journal, 51*(2), 73-75.

Piotrowski, T. (1989). Monolingual and bilingual dictionaries: fundamental differences. In M. Tickoo (Ed.), *Learners' dictionaries: State of the art, RELC Anthology Series 23* (pp. 72-83). SEAMEO RELC.

Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *Modern Language Journal, 80*, 478–493.

Prodromou, L. (2002). *From mother tongue to other tongue*. https://www.teachingenglish.org.uk/article/mother-tongue-other-tongue

Puimège, E., & Peters, E. (2019). Learners' English Vocabulary Knowledge Prior to Formal Instruction: The Role of Learner-Related and Word-Related Variables. *Language Learning, 69*(4), 943-977.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536.

Quizlet. (2018). *About Quizlet*. Retrieved 04/02/2018 from https://quizlet.com/mission

Ramachandran, S. D., & Rahim, H. A. (2004). Meaning Recall and Retention: The Impact of the Translation Method on Elementary Level Learners' Vocabulary Learning. *RELC Journal, 35*(2), 161-178.

Révész, A., Kourtali, N.-E., & Mazgutova, D. (2017). Effects of Task Complexity on L2 Writing Behaviors and Linguistic Complexity. *Language Learning, 67*(1), 208-241.

Richards, J. C. (1976). The Role of Vocabulary Teaching. *TESOL Quarterly, 10*(1), 77-89.

Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and vocabulary. *ITL - International Journal of Applied Linguistics, 169*(1), 191-211.

Rodgers, T. S. (1969). On Measuring Vocabulary Difficulty. *IRAL, VII*(4), 327-343.

Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association, 1*(1), 49-60.

Royer, J. M. (1973). Memory effects for test-like-events during acquisition of foreign language vocabulary. *Psychological Reports, 32*, 195-198.

Russell, R. A. (1999). lexical Maintenance and Attrition in Japanese as a Second Language. In L. Hansen (Ed.), *Second Language Attrition in Japanese Contexts* (pp. 114-141). Oxford University Press.

Saelinger, M. I. (1928). *The effects of summer vacation upon the retention of the elementary school subjects*. Literary Licensing.

Saito, K. (2017). Effects of Sound, Vocabulary, and Grammar Learning Aptitude on Adult Second Language Speech Attainment in Foreign Language Classrooms. *Language Learning, 67*(3), 665-693.

Saito, K. (2019). The role of aptitude in second language segmental learning: The case of Japanese learners' English /ɹ/ pronunciation attainment in classroom settings. *Applied Psycholinguistics, 40*(1), 183-204.

Saito, K., & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics, 37*(4), 813-840.

Saito, K., Suzukida, Y., & Sun, H. (2019). Aptitude, experience, and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition, 41*(1), 201-225.

Samar, R. G., & Moradkhani, S. (2014). Codeswitching in the Language Classroom: A Study of Four EFL Teachers' Cognition. *RELC Journal, 45*(2), 151-164.

Sandberg, J., Maris, M., & de Geus, K. (2011). Mobile English learning: An evidence-based study with fifth graders. *Computers & Education, 57*(1), 1334-1347.

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System, 6*(2), 72-78.

Sasaki, M. (1996). *Second Language Proficiency, Foreign Language Aptitude, and Intelligence*. Lang.

Scherer, G. A. C. (1957). The forgetting rate in learning German. *The German Quarterly, 30*(4), 275-277.

Schmid, M. S. (2011). *Language Attrition*. Cambridge University Press.

Schmid, M. S. (2012). The impact of age and exposure on bilingual development in international adoptees and family migrants: A perspective from Holocaust survivors. *Linguistic Approaches to Bilingualism, 2*(2), 177-208.

Schmidt, A. (1991). Language attrition in Boumaa Fijian and Dyirbal. In H. W. Seliger & R. M. Vago (Eds.), *First language attrition*, 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. (2000). *Vocabulary in Language Teaching*. Cambridge University Press.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363.

Schmitt, N. (2010). *Researching Vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning, 64*(4), 913-951.

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching, 52*, 261-274.

Schmitt, N., Nation, P., & Kremmel, B. (2019). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, 1-12.

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.

Scrivener, J. (2011). *Learning teaching: The essential guide to English language teaching*. Oxford: Macmillan Education.

Service, E. (1992). Phonology, working memory and foreign language learning. *Quarterly Journal of Experimental Psychology, 45*, 21-50.

Singmann, H. (2014). *Re. Warning messages from mixed model (glmer) [Online discussion post]*. https://stats.stackexchange.com

Skehan, P. (1989). *Individual Differences in Second Language Learning*. Edward Arnold.

Skehan, P. (2002). Theorising and Updating Aptitude. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning*. John Benjamins.

Smythe, P. C., Jutras, G. C., Bramwell, J. R., & Gardner, R. C. (1973). Second Language Retention over Varying Intervals. *The Modern Language Journal, 57*(8), 400-405.

Snow, M. A., Padilla, A. M., & Campbell, R. N. (1988). Patterns of Second Language Retention of Graduates of a Spanish Immersion Program. *Applied Linguistics, 9*(2), 182-197.

Song, Y., & Andrews, S. J. (2009). *The L1 in L2 learning: teachers' beliefs and practices*. LINCOM Europa.

Spelman Miller, K., & Sullivan, K. P. H. (2006). Keystroke Logging: An Introduction. In K. Sullivan & E. Lindgren (Eds.), *Computer Key-Stroke Logging and Writing : Methods and Applications*. Elsevier.

Spreen, O., & Scqziulz, R. W. (1966). Parameters of Abstraction, Meaningfulness, and Pronunciability for 329 Nouns. *Journal of verbal learning and verbal behaviour, 5*, 459-468.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 22*, 360-407

Sternberg, R. J. (2002). The theory of successful intelligence and its implication to language-aptitude testing. In Robinson (Ed.), *Individual Differences and Instructed Language Learning*. John Benjamins.

Stroud, R. (2014). Student engagement in learning vocabulary with CALL CALL Design: Principles and Practice; Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands.

Sundqvist, P., & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System, 51*, 65-76.

Suzuki, Y., Nakata, T., & Dekeyser, R. (2019). Optimizing Second Language Practice in the Classroom: Perspectives from Cognitive Psychology [https://doi.org/10.1111/modl.12582]. *The Modern Language Journal, 103*(3), 551-561.

Taura, H. (2001). L2 attrition & maintenance observed in Japanese returnee students in terms of productive oral proficiency. *Journal of Fukui Medical University, 2*(1), 71-79.

Team, R. C. (2014). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Austria.

Tehan, G., & Tolan, G. A. (2007). Word length effects in long-term memory. *Journal of Memory and Language, 56*(1), 35-48.

Thomas, M. H., & Dieter, J. N. (1987). The positive effects of writing practice on integration of foreign words in memory. *Journal of Educational Psychology, 79*(3), 249-253.

Thompson, G. (1987). Using Bilingual Dictionaries. *ELT Journal, 41*(4), 282-286. https://eric.ed.gov/?id=EJ361119

Thornbury, S. (2002). *How to teach vocabulary*. Longman.

Thornbury, S., & Watkins, P. (2007). *The CELTA course: Certificate in English Language Teaching to Adults trainer's manual*. Cambridge University Press.

Thorndike, E. L. (1908). Memory of Paired Associates. *Psychological Review, 15*, 122-138.

Tian, L., & Macaro, E. (2012). Comparing the effect of teacher codeswitching with English-only explanations on the vocabulary acquisition of Chinese university students: A Lexical Focus-on-Form study. *Language Teaching Research, 16*(3), 367-391.

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System, 21*, 371-380.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second language Research, 13*(2), 138-163.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Tomaszczyk, J. (1983). On bilingual dictionaries: the case of bilingual dictionaries for foreign language learner. In R. R. K. Hartmann (Ed.), *Lexicography: Principles and Practice* (pp. 41-51). Academic Press.

Tomiyama, M. (1999). The first stage of second language attrition: A case study of a Japanese returnee. In L. Hansen (Ed.), *Second Language Attrition in Japanese Contexts* (pp. 59-79). Oxford 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. https://eric.ed.gov/?id=EJ850413

Tragant, E., Muñoz, C., & Spada, N. (2016). Maximizing Young Learners’ Input: An Intervention Program. *Canadian Modern Language Review, 72*(2), 234-257.

Tragant Mestres, E., & Pellicer-Sánchez, A. (2019). Young EFL learners' processing of multimodal input: Examining learners’ eye movements. *System, 80*, 212-223.

Tseng, W.-T., & Schmitt, N. (2008). Towards a model of motivated vocabulary learning: A structural equation modeling approach. *Language Learning, 58*(2), 357-400.

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta‐analysis of correlational studies. *Language Learning, 69*(3), 559-599.

Ur, P. (2010). *Teaching vocabulary*. Cambridge University Press.

Van Els, T. (1986). An overview of European research on language attrition. In B. Weltens, K. d. Bot, & T. V. Els (Eds.), *Language Attrition in Progress* (pp. 3-18). Foris Publications.

van Zeeland, H., & Schmitt, N. (2013a). Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics, 34*(4), 457-479.

van Zeeland, H., & Schmitt, N. (2013b). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609-624.

Walker, C. M. (2011). What's the DIF? Why Differential Item Functioning Analyses Are an Important Part of Instrument Development and Validation. *Journal of Psychoeducational Assessment, 29*(4), 364-376.

Wang, R., Hempton, B., Dugan, J. P., & Komives, S. R. (2008). Cultural differences: Why do Asians avoid extreme responses. *Survey Practice, 1*(3), 1-7.

Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System, 25*(2), 261-274.

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.

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 287-307.

Watkins, H. S., Miller, C. A., & Poll, G. H. (2012). Lexical decay rate in adults with specific language impairment. *The Penn State McNair Journal, 18*, 143-154.

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. (2007). The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics, 28*(1), 46-65.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30*, 79-95.

Webb, S. (2020). *The Routledge Handbook of Vocabulary Studies* Routledge.

Webb, S., & Chang, A. C.-S. (2012). Vocabulary Learning through Assisted and Unassisted Repeated Reading. *The Canadian modern language review, 68*(3), 267-290.

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* (Vol. Oxford University Press). Oxford.

Webb, S., & Rodgers, M. (2009). Vocabulary Demands of Television Programmes. *Language Learning, 59*(2), 335-366.

Weltens, B., & Cohen, A. D. (1989). Language Attrition Research: An Introduction. *Studies in Second Language Acquisition, 11*, 127-133.

Weltens, B., Els, T. J. M. V., & Schils, E. (1989). The Long-Term Retention of French by Dutch Students. *Studies in Second Language Acquisition, 11*, 205-216.

Weltens, B., & Grendel, M. (1993). Attrition of Vocabulary Knowledge. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 135-157). John Benjamins.

Weltens, H. H. (1988). *The attrition of French as a foreign language*. Foris.

Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching, 50*(1), 1-31.

White, W. F. (1908). Reviews before and after vacation. *American Education, 10*(185-188).

Williams, M., Burden, R., & Lanvers, U. (2002). 'French is the Language of Love and Stuff': Student perceptions of issues related to motivation in learning a foreign language. *British Educational Research Journal, 28*(4), 503-528.

Williams, W., & Burden, R. (1997). *Psychology for Language Teachers*. Cambridge University Press.

Willis, D. (2003). *Rules, patterns, and words: Grammar and Lexis is English language teaching*. Cambridge University Press.

Willis, M., & Ohashi, Y. (2012). A model of L2 vocabulary learning and retention. *The Language Learning Journal, 40*(1), 125-137.

Xu, X. (2010). *English language attrition and retention in Chinese and Dutch university students* [PhD Thesis]. University of Groningen.

Yalçın, Ş., & Spada, N. (2016). Language Aptitude and Grammatical Difficulty: An EFL Classroom-Based Study. *Studies in Second Language Acquisition, 38*(2), 239-263.

Yoshii, M. (2006). L1 and L2 Glosses: Their Effects on Incidental Vocabulary Learning. *Language Learning & Teaching, 10*(3), 85-101.

Zhang, X. (2013). The I Don't Know Option in the Vocabulary Size Test. *TESOL Quarterly, 47*(4), 790-811.

APPENDICES

Appendix 1

A Description of the Participants (Pilot Study)

				Length				Self-report proficiency				
Gender	Age	L1	Level	L2 study	Residency	First contact	L3	R	W	L	S	Vocabulary
male	24	Urdu	PhD	9.0	3.0	5	NA	8	7	7	7	68
male	27	French	PhD	10.0	2.5	12	German, Japanese	7	6	9	9	67
male	27	Serbian	PEAP	8.0	0.5	15	Spanish, French	9	8	8	8	65
female	24	Chinese	MA	13.0	0.25	11	German	8	6	8	7	60
female	32	Spanish	MA	12.0	1.5	11	Dutch	8	5	7	6	63
male	29	Arabic	PhD	10.0	0.5	10	Turkish	8	5	7	5	41
male	28	Arabic	MSC	10.0	0.5	12	Turkish	8	8	8	8	42
female	23	Japanese	MSc	11.0	0.5	12	NA	4	3	2	2	27
male	29	Italian	PEAP	10.0	0.25	5	Spanish	5	4	6	4	40
male	35	Arabic	PEAP	0.5	0.5	34	NA	0	0	0	0	43
female	30	Arabic	PEAP	12.0	0.5	12	NA	6	4	5	6	25
female	32	Arabic	PEAP	1.0	1.25	12	NA	4	5	6	5	48
male	32	Thai	PhD	10.0	0.25	12	NA	5	4	4	5	35
male	29	Arabic	MSc	24.0	8.0	12	NA	8	5	8	8	32

Note. Length of L2 study and Length of Residency are presented as year.quarter year; First contact is presented as age; Self-reported proficiency was measured on a ten-point scale; maximum vocabulary score was 70 (Vocabulary Size Test levels 1-7 combined).

Appendix 2

Approved Ethics Form (Pilot Study)

Faculty of Arts Ethics Approval Form

Please submit this form to your School Ethics Officer **at least 2 months** before you plan to begin your research, along with:

- consent form
- written information sheet for participants
- signed declaration of ethical awareness
- questionnaire or focus group plan (if possible).

Please read the **Guidelines for Completing the Arts Ethics Form** (available on Workspace) before submitting the form to your School Ethics Officer.

Researcher name	Samuel Barclay
School/Department	School of English
Project Title	The effect of intralexical factors on the attrition of second language vocabulary knowledge
Date	22/01/2016
Email address	[redacted]

(1) Researcher Information – please tick as appropriate

☐ Member of Staff

☒ Postgraduate Researcher

Supervisors: Dr Ana Pellicer-Sanchez, Dr Michael Rodgers

☐ Member of staff obtaining approval for a module

Module Code:

Module Name:

Is the research funded by an external body or part of an external funding bid?

☐ Yes Funding Body: ☒ No

If yes, does the funding body require proof of ethics approval?

☐ Yes ☐ No

(2) Research aims/questions

Provide a brief summary of the research aims/questions [max 100 words]

The main aim of this study is to measure the effect of word length and part of speech on the attrition of second language vocabulary. A further aim is to better understand the effect of first language on second language lexical loss. In this study non-native speakers of English will independently study and be tested on a set of target items. A number of weeks later, they will be retested on the same set of target items. Following this, they will relearn the target items in addition to a set of previously unstudied items to test for a relearning advantage. This result would indicate latent knowledge of the target items.

(3) Methods

a) Please indicate which methods you will be using:

- ☒ Questionnaire
- ☐ Focus groups
- ☐ Interviews
- ☐ Observation
- ☐ Psychophysiological measures (e.g., response time, eyetracking, ERP etc.)
- ☐ Data found online
- ☐ Data produced by students (e.g., their essays)
- ☐ Other; please specify:

b) Please give brief details of how you will be employing these methods [max 200 words]

Materials will first be piloted with native and non-native participants. Native participants will complete a norm rating questionnaire to ascertain concreteness ratings for the target language. The materials will also be piloted on native-speaker participants.

The experiment will consist of a number of sessions. In the first session non-native participants will consider the aims and read the consent form. They will also complete a vocabulary size test and a language background questionnaire. After signing the consent form and sorting out questions, participants will be introduced to the learning instrument. After completing a practice session and being introduced to the learning log, participants will have the opportunity to ask any questions they have about any part of the study. This procedure will not take more than 60 minutes. They will receive a flash drive containing the learning software and a learning log which they will use for daily independent study of the target items over the next 7-10 days (each session should take no more than 50 minutes). After the period of study, participants will attend a testing session and complete a questionnaire. The latter will target prior knowledge of the target items. The administration of both should take 50 minutes. The subsequent testing session will consist of the same testing instrument and a questionnaire. This questionnaire will target any exposure to the target items since the learning procedure. Participants will take part in one final learning session using the learning software. This should take less than 60 minutes.

(4) Research Location

Please confirm where the research will take place:

☒ On Campus	☐ Outside the UK
☒ Elsewhere in the UK	☐ Online

If you are conducting your research outside of the UK, please state where:

Nottingham Trent University, Nottingham

(5) Research topics

a) Please confirm if your research involves any of the following:

Yes ☐	No ☒	Procedures likely to cause participants distress
Yes ☐	No ☒	Misleading participants about your research or withholding information
Yes ☐	No ☒	Investigation of sensitive issues (e.g., sexual, racial, religious or political attitudes, illegal activities etc.)
Yes ☐	No ☒	Investigation of personal topics (e.g., personal health, learning disabilities etc.)
Yes ☐	No ☒	Online data that requires a password to access

If you have ticked YES to any of the above, please provide more details below. Indicate any potential risk to participants, justify this risk and what steps will be taken to minimise it. For online data please provide details of the websites and how you will ensure consent is given.

(6) Participants, access and inducements

a) Please confirm if your sample will involve any of the following:

Yes ☐	No ☒	Participants under the age of 16
Yes ☐	No ☒	Adults of limited mental capacity
Yes ☒	No ☐	Participants recruited from special sources (e.g., educational institutions, prisons, hospitals etc.)

If you have ticked YES, please provide more detail information and justification:

Native English-speaking participants will be recruited from the teaching staff at Nottingham Trent University, where the researcher is a member of staff working as a lecturer and as part of the management team of the Pre-sessional English for Academic Purposes course. Participation will be voluntary. Non-native participants will be recruited from the undergraduate and postgraduate student community at Nottingham Trent University and the University of Nottingham. Participation will be voluntary. In the case of voluntary recruitment not resulting in the required number of participants, funding will be applied for and non-native participants might be given a £6 (or equivalent printer credits) compensation for their participation. Please note that the Nottingham Trent University Ethics Office has been made aware of this study. They have stated that I do not need to apply for ethics approval at NTU if ethics approval is granted by the University of Nottingham. I have also obtained informal permission to conduct the study from my line manager.

b) Please confirm if you will be offering inducements for taking part:

Yes ☒ No ☐

If YES, please provide more detailed information and justification:

As explained above, non-native participants may receive compensation for their participation. Native speaker participation will be voluntary.

c) Please confirm if there is a risk of participants being identified in any form of dissemination

Yes ☐ No ☒

If you have answered YES please provide more detail information and justification

If you have answered NO please confirm how you will protect participants' identities

(7) Data Storage & Dissemination

a) Please confirm that you will be storing your data in password-protected files

Yes ☒ No ☐

b) Please confirm if you will be destroying the data seven years after publication

Yes ☒ No ☐

If you have answered NO, please provide a justification and give details of where the data will be deposited

b) Please provide an indication of any intended dissemination or impact activities (if such activities are planned after the project is approved, please inform your School Ethics Officer of these changes and update consent procedures appropriately)

The dissemination activities will be the usual activities of a research project, i.e., conference presentations and research paper. Participants' information will not be disclosed in any of these activities.

(8) Declaration

Signed Samuel Barclay

Date 22/01/16

Office use only:

Approved Dominic Thompson Date 12/02/16

(School Ethics Officer)

Confirmed Jen Birks Date 15/02/16

(Second School Ethics Officer)

Send for full committee approval ☐

Approved ______________________ Date______________

(Faculty Ethics Officer)

Appendix 3
Consent Form (Pilot Study)

Study on Vocabulary Learning

Informed Consent

1. The main aim of this project is to understand what happens to knowledge of words once they have been learned. I hope that this study will help students develop stronger vocabulary knowledge in the future.

2. Sam Barclay is the leader of this project. He is at the School of English, University of Nottingham. His phone number is 0115 848 4405. His email is samuel.barclay@nottingham.ac.uk. Please feel free to contact him with any questions about the project.

3. In this study participants will be asked to take part in three stages. The first stage is the learning stage. Participants will study 72 words. There are daily tasks which last between 30 and 45 minutes each. Stage Two is the testing stage. One week after the beginning of Stage One, we will meet again and complete a vocabulary test to measure how many words you learned. Stage Three is a testing and learning stage. Participants take another test at the beginning of Stage Three. After this, they study 18 more words.

4. There are no known risks or hazards involved in participating in this study. I have provided my name and contact information should you have any questions about this research.

5. At any time you have a right to stop participating in the project. Also, after the data is collected you may request that your data not be used.

Please cross out as appropriate:

- I confirm that the purpose of the study has been explained and that I have understood it — Yes No
- I have had the opportunity to ask questions and they have been successfully answered — Yes No
- I understand that my participation in this study is voluntary and that I am free to withdraw from the study at any time, without giving a reason and without consequence — Yes No
- I understand that all data are anonymous and that there will not be any connection between the personal information provided and the data — Yes No
- I understand that there are no known risks or hazards associated with participating in this study — Yes No
- I confirm that I have read and understood the above information and that I agree to participate in this study — Yes No
- I confirm that I am over 16 years of age — Yes No

Participant's signature: ______________________________________ Date: __________

Participant's Name (in block capitals): ____

Researcher's signature: ______________________________________ Date: __________

Appendix 4

Screen Shots of the Form Recall and Form Recognition Instruments

Form Recall

1	**Look at the definition in the column entitled "definitions".**	
2	**Type the word that matches the definition in the column called "Your answer".**	
3	*Example*	
4	If you knew that the definition "*an educational institution at the highest level*" is	
5	the definition of "university" you would answer as follows.	
6		
7	**Your answer**	**Definition**
8	University	An educational institution at the highest level
9		
10	***The test***	
11	**Your answer**	**Definition**
12		
13		to catch or arrest someone who is doing something wrong
14		an image placed next to a word to make people notice it
15		to respect someone because they are old or important

Form Recognition

1	**Type "Y" next to the correct answer.**			
2	**If you think you know the answer but are not sure, please type "Y" next to your best guess.**			
3	**If you have no idea what the answer is, type an "Y" next to answer option 5.**			
4				
5	**1**		lop	***to catch or arrest someone who is doing something wrong***
6			conflate	
7			nab	
8			pucker	
9			I don't know	
10				
11	**2**		supposition	***an image placed next to a word to make people notice it***
12			bib	
13			asterisk	
14			camaderie	
15			I don't know	
16				
17	**3**		kip	***to respect someone because they are old or important***
18			decant	
19			dapple	
20			venerate	
21			I don't know	

Appendix 5
Approved Ethics Form (Study 1)

Faculty of Arts Ethics Approval Form

Please submit this form to your School Ethics Officer **at least 2 months** before you plan to begin your research, along with:

- consent form
- written information sheet for participants
- signed declaration of ethical awareness
- questionnaire or focus group plan (if possible).

Please read the **Guidelines for Completing the Arts Ethics Form** (available on Workspace) before submitting the form to your School Ethics Officer.

Researcher name	Samuel Barclay
School/Department	School of English
Project Title	The effect of intralexical factors on the attrition of second language vocabulary knowledge
Date	25/05/2016
Email address	[redacted]

(1) Researcher Information – please tick as appropriate

☐ Member of Staff

☒ Postgraduate Researcher

Supervisors: Dr Ana Pellicer-Sanchez, Dr Michael Rodgers

☐ Member of staff obtaining approval for a module

Module Code:

Module Name:

Is the research funded by an external body or part of an external funding bid?

☐ Yes Funding Body: ☒ No

If yes, does the funding body require proof of ethics approval?

☐ Yes ☐ No

(2) Research aims/questions

Provide a brief summary of the research aims/questions [max 100 words]

The main aim of this study is to measure the effect of word length and part of speech on the attrition of second language vocabulary. A secondary aim is to better understand the effect of first language on second language lexical loss. In this study non-native speakers of English will independently study and be tested on a set of target items. A number of weeks later, they will be retested on the same set of target items. Following this, they will relearn the target items in addition to a set of previously unstudied items to test for a relearning advantage. This result would indicate latent knowledge of the target items.

(3) Methods

a) Please indicate which methods you will be using:

- ☒ Questionnaire
- ☐ Focus groups
- ☐ Interviews
- ☐ Observation
- ☐ Psychophysiological measures (e.g., response time, eyetracking, ERP etc.)
- ☐ Data found online
- ☐ Data produced by students (e.g., their essays)
- ☐ Other; please specify:

b) Please give brief details of how you will be employing these methods [max 200 words]

All materials will first be piloted on native-speaker participants. The experiment will consist of a number of sessions. In the first session non-native participants will consider the aims and read the consent form. They will also complete a vocabulary test and a language background questionnaire. After signing the consent form and sorting out questions, participants will be introduced to the learning instrument. After completing a practice session and being introduced to the learning log, participants will have the opportunity to ask any questions they have about any part of the study. They will receive a flash drive containing the learning software and a learning log which they will use for studying the target items. The first learning session will take place on this first day. On day two, the students will again study target words. On day three they will review the target language and complete a test and a questionnaire. The latter will target prior knowledge of the target items. Some weeks later, a delayed test, consisting of the same testing instrument and a questionnaire, will be conducted. This questionnaire will target any exposure to the target items since the learning procedure. Participants will take part in one final learning session using the flashcard software to check for a relearning benefit.

(4) Research Location

Please confirm where the research will take place:

☒	On Campus	☐	Outside the UK
☒	Elsewhere in the UK	☐	Online

If you are conducting your research outside of the UK, please state where:

Nottingham Trent University, Nottingham

(5) Research topics

a) Please confirm if your research involves any of the following:

Yes ☐ No ☒ Procedures likely to cause participants distress

Yes ☐ No ☒ Misleading participants about your research or withholding information

Yes ☐ No ☒ Investigation of sensitive issues (e.g., sexual, racial, religious or political attitudes, illegal activities etc.)

Yes ☐ No ☒ Investigation of personal topics (e.g., personal health, learning disabilities etc.)

Yes ☐ No ☒ Online data that requires a password to access

If you have ticked YES to any of the above, please provide more details below. Indicate any potential risk to participants, justify this risk and what steps will be taken to minimise it. For online data please provide details of the websites and how you will ensure consent is given.

(6) Participants, access and inducements

a) Please confirm if your sample will involve any of the following:

Yes ☐ No ☒ Participants under the age of 16

Yes ☐ No ☒ Adults of limited mental capacity

Yes ☒ No ☐ Participants recruited from special sources (e.g., educational institutions, prisons, hospitals etc.)

If you have ticked YES, please provide more detail information and justification:

Native English-speaking participants will be recruited from the teaching staff at Nottingham Trent University, where the researcher is a member of staff working as a lecturer and as part of the management team of the Pre-sessional English for Academic Purposes course. Participation will be voluntary. Non-native participants will be recruited from the student community at Nottingham Trent University and the University of Nottingham. Participation will be voluntary. In the case of voluntary recruitment not resulting in the required number of participants, funding will be applied for and non-native participants might be given a £6 (or equivalent printer credits) compensation for their participation. Please note that the Nottingham Trent University Ethics Office has been made aware of this study. They have stated that I do not need to apply for ethics approval at NTU if ethics approval is granted by the University of Nottingham. I have also obtained permission to conduct the study from my line manager.

b) Please confirm if you will be offering inducements for taking part:

Yes ☒ No ☐

If YES, please provide more detailed information and justification:

As explained above, non-native participants may receive compensation for their participation. Native speaker participation will be voluntary.

c) Please confirm if there is a risk of participants being identified in any form of dissemination

Yes ☐ No ☒

If you have answered YES please provide more detail information and justification

If you have answered NO please confirm how you will protect participants' identities

(7) Data Storage & Dissemination

a) Please confirm that you will be storing your data in password-protected files

Yes ☒ No ☐

b) Please confirm if you will be destroying the data seven years after publication

Yes ☒ No ☐

If you have answered NO, please provide a justification and give details of where the data will be deposited

b) Please provide an indication of any intended dissemination or impact activities (if such activities are planned after the project is approved, please inform your School Ethics Officer of these changes and update consent procedures appropriately)

The dissemination activities will be the usual activities of a research project, i.e., conference presentations and research paper. Participants' information will not be disclosed in any of these activities.

(8) Declaration

Signed Samuel Barclay

Date 25/05/2016

Office use only:

Approved Dominic Thompson Date 19/07/16

(School Ethics Officer)

Confirmed Jen Birks Date 24/07/16

Send for full committee approval ☐

Approved ______________________ Date ______________

(Faculty Ethics Officer)

Appendix 6

Screen Shots of the Concreteness Norming Instrument

Instructions

Instructions

Nouns may refer to people, places and things that can be seen, heard, felt, smelled or tasted or to more abstract concepts that cannot be experienced by our senses. The purpose of this experiment is to rate a list of words with respect to "concreteness". Any word that refers to objects, materials or people should receive a high concreteness rating; any word that refers to an abstract concept that cannot be experienced by the senses should receive a low concreteness rating. Think of the words "chair" and "independence." "Chair" can be experienced by our senses and therefore should be rated as highly concrete. Using the scale below, you might type 7 next to this word; "independence" cannot be experienced by the senses as such and therefore should be rated as highly abstract. Using the scale below, you might type a 1. These words are given as an example below.

The scale

Highly Concrete						Highly Abstract
7	**6**	**5**	**4**	**3**	**2**	**1**

An example

WORD	RATING	DEFINITION
chair	7	A piece of furniture for one person to sit on, which has a back, a seat, and four legs
independence	1	The freedom and ability to make your own decisions in life, without having to ask other people for permission, help, or money

Instrument

TARGET ITEMS	RATING	DEFINITIONS
keg		a round wooden container with a flat top and bottom
orb		a bright ball-shaped object such as the sun or the moon
bib		a piece of cloth or plastic tied under a baby's face
tic		a sudden movement of a muscle in your face
wick		the piece of thread in a candle that burns when you light it
tusk		a long pointed tooth of animals such as elephants
sled		a small vehicle used for sliding over snow
scab		a layer of dried blood that forms over a cut
frond		the large long leaf
prong		a sharp point of something such as a fork
visor		the part of a hat that can be lowered to protect your face

Appendix 7

The Language Background Questionnaire (Studies 1 and 2)

Please provide the following personal details and language background information. If you prefer not to give an answer to a question, please leave it blank.

1. Are you male or female? ______________________
2. How old are you? ______________________
3. What is your native language? ______________________
4. What course do you study/plan to study? ______________________
5. How long have you studied English? ______________________
6. At what age did you first contact English? ______________________
7. How long (in total) have you lived in an English-speaking country? ____________
8. What other languages do you speak? ______________________

Please now provide self-ratings for your level of proficiency in English:

1 = extremely poor, almost no knowledge → 10 = extremely good, almost native like.

Reading:	1	2	3	4	5	6	7	8	9	10
Writing:	1	2	3	4	5	6	7	8	9	10
Listening:	1	2	3	4	5	6	7	8	9	10
Speaking:	1	2	3	4	5	6	7	8	9	10

If you have any other comments, please write them in the box below.

Thank you so much for your help

Appendix 8

Form Recall and Recognition Instrument (nb. actual version administered via MS Excel)

NAME: ______________________________

Instructions:

Thank you for your participation today. This test should take about 20 minutes. The results of the test will not be used on PEAP. This test is part of a research project I am conducting for my PhD. Please take your time and complete it seriously.

The purpose of this test is to measure how many of the words that you learned you are able to remember. This test is split into two parts. These are outlined below.

Part One:

In Part One, you will see a definition. Your task is to write the word that matches this definition. If you do not know the word, you can leave a blank space. However, there is no penalty for guessing, so please do try if you think you know the answer. Below, there are two examples.

Your answer	Definition
Table	a piece of furniture that has of a flat top supported by legs
umbrella	a round object used for protection from rain

When you have finished Part One, please begin Part Two. However, please note that you cannot return to Part One after starting Part Two.

Part Two

In part two, you will see a definition and four words. Please write a tick (✓) on the line next to the word that matches the definition. If you do not know the word, you can write a tick next to *I don't know*, however, there is no penalty for guessing, so please do try if you think you know the answer.

Ex.	***a piece of furniture that has of a flat top supported by legs***			**Ex.**	***a round object used for protection from rain***	
	✓	Table				Wallet
		Dog				Picture
		Car				Floor
		restaurant			✓	Umbrella
		I don't know				I don't know

PART ONE
Look at the definition in the column entitled "definitions". Type the word that matches the definition in the column called "Your answer".

Example
If you knew that the definition "*an educational institution at the highest level*" is the definition of *"university"* you would answer as follows.

	Your answer	Definition
ex	*University*	An educational institution at the highest level

The test

	Your answer	Definition
1		to catch or arrest someone who is doing something wrong
2		an image placed next to a word to make people notice it
3		to respect someone because they are old or important
4		an unmarried woman who is old
5		to make a process happen more quickly
6		a group of small islands
7		to hit someone gently
8		a round wooden container with a flat top and bottom
9		to bring food that you have already eaten back into your mouth
10		to walk with high steps
11		a child who acts badly
12		a feeling of friendship and trust among people
13		to give your opinion about something
14		a person who has very strong feelings
15		a machine that looks strange and is unlikely to work well
16		a sudden movement of a muscle in your face
17		to kill something suddenly with force
18		to officially begin studying at a university
19		to make a sound by hitting two metal objects together
20		to combine two or more things to form a single new thing
21		a sea fish that is blue and silver, and has a strong taste
22		to make something shorter

23		a very small piece of burnt wood
24		a small wild animal with a long body, short legs and sharp teeth
25		to make someone feel annoyed
26		a musical instrument like a piano
27		to pour wine from one bottle into another
28		a large round metal pot for boiling water over a fire
29		a quality of not giving up something easily
30		a small group of people who spend their time together
31		a piece of cloth or plastic tied under a baby's face
32		to turn on and off quickly
33		to sleep somewhere that is not your home
34		to interrupt and try to embarrass someone who is speaking in public
35		to play and move around in a happy way
36		to do something to make somebody angry with you
37		a bright ball-shaped object such as the sun or the moon
38		one apartment in a building with several apartments
39		to move your legs
40		a person who enjoys watching other people

This is the end of Part One. When you move onto Part Two, you cannot return to Part One, so please take a moment to check your answers.

PART 2:

Type "Y" next to the correct answer. If you think you know the answer but are not sure, please type "Y" next to your best guess. If you have no idea what the answer is, type a "Y" next to answer option 5.

1

to catch or arrest someone who is doing something wrong	
	lop
	conflate
	nab
	pucker
	I don't know

2

an image placed next to a word to make people notice it	
	supposition
	bib
	asterisk
	camaraderie
	I don't know

3

to respect someone because they are old or important	
	kip
	decant
	dapple
	venerate
	I don't know

4

an unmarried woman who is old	
	tenacity
	cauldron
	spinster
	zenith
	I don't know

5

to make a process happen more quickly	
	twinge
	expedite
	keg
	valour
	I don't know

7

to hit someone gently	
	emblazon
	suckle
	bop
	decant
	I don't know

8

a round wooden container with a flat top and bottom	
	jus
	mackerel
	marten
	keg
	I don't know

9

to bring food that you have already eaten back into your mouth	
	regurgitate
	heckle
	ambulate
	belittle
	I don't know

10

to walk with high steps	
	antagonise
	frolic
	prance
	chastise
	I don't know

11

a child who acts badly	
	antiquarian
	doe
	voyeur
	imp
	I don't know

6	*a group of small islands*
	harpsichord
	rosary
	archipelago
	orb
	I don't know

13	*to give your opinion about something*
	totter
	scintillate
	pontificate
	Kip
	I don't know

14	*a person who has very strong feelings*
	prophesy
	zealot
	condominium
	schism
	I don't know

15	*a machine that looks strange and is unlikely to work well*
	bib
	detritus
	contraption
	morgue
	I don't know

16	*a sudden movement of a muscle in your face*
	tic
	nab
	spinster
	executor
	I don't know

17	*to kill something suddenly with force*
	bungle
	venerate
	zap
	expedite
	I don't know

12	*a feeling of friendship and trust among people*
	inquisition
	asterisk
	archipelago
	camaraderie
	I don't know

19	*to make a sound by hitting two metal objects together*
	jangle
	ambulate
	aerate
	antagonise
	I don't know

20	*to combine two or more things to form a single new thing*
	prance
	conflate
	nab
	pacify
	I don't know

21	*a sea fish that is blue and silver, and has a strong taste*
	voyeur
	mackerel
	condominium
	pancreas
	I don't know

22	*to make something shorter*
	bemoan
	truncate
	frolic
	whoosh
	I don't know

23	*a very small piece of burnt wood*
	keg
	rubric
	modality
	cinder
	I don't know

18

to officially begin studying at a university	
	emboss
	bop
	sanctify
	matriculate
	I don't know

24

a small wild animal with a long body, short legs and sharp teeth	
	tenacity
	marten
	clique
	connotation
	I don't know

25

to make someone feel annoyed	
	fondle
	regurgitate
	irk
	siphon
	I don't know

26

a musical instrument like a piano	
	trepidation
	harpsichord
	asterisk
	vortex
	I don't know

27

to pour wine from one bottle into another	
	expedite
	loiter
	encumber
	decant
	I don't know

28

a large round metal pot for boiling water over a fire	
	imp
	zealot
	corset
	cauldron
	I don't know

29

a quality of not giving up something easily	
	canary
	tenacity
	condominium
	camaraderie
	I don't know

31

a piece of cloth or plastic tied under a baby's face	
	tic
	cinder
	bib
	audacity
	I don't know

32

to turn on and off quickly	
	pontificate
	renege
	zap
	scintillate
	I don't know

33

to sleep somewhere that is not your home	
	maraud
	matriculate
	kip
	truncate
	I don't know

34

to interrupt and try to embarrass someone who is speaking in public	
	heckle
	embezzle
	jangle
	blanch
	I don't know

35

to play and move around in a happy way	
	conflate
	recast
	frolic
	venerate
	I don't know

30	*a small group of people who spend their time together*
	contraption
	mirage
	clique
	parenthesis
	I don't know

36	*to do something to make somebody angry with you*
	antagonise
	bob
	regurgitate
	quench
	I don't know

37	*a bright ball-shaped object such as the sun or the moon*
	zealot
	orb
	effigy
	comptroller
	I don't know

38	*one apartment in a building with several apartments*
	emissary
	imp
	minnow
	condominium
	I don't know

39	*to move your legs*
	ambulate
	truncate
	tousle
	conflate
	I don't know

40	*a person who enjoys watching other people*
	slurry
	artifice
	voyeur
	cauldron
	I don't know

This is the end of Part Two. Thank you for your participation.

Appendix 9
Approved Ethics Form (Study 2)

Ethics Application Form: Student Research

Anyone conducting research under the auspices of the Institute (staff, students or visitors) where the research involves human participants or the use of data collected from human participants, is required to gain ethical approval before starting. This includes preliminary and pilot studies. Please answer all relevant questions in terms that can be understood by a lay person and note that your form may be returned if incomplete.
For further support and guidance please see accompanying guidelines and the Ethics Review Procedures for Student Research http://www.ucl.ac.uk/srs/research-ethics-committee/ioe or contact your supervisor or IOE.researchethics@ucl.ac.uk.

Before completing this form you will need to discuss your proposal fully with your supervisor(s).

Please attach all supporting documents and letters.

For all Psychology students, this form should be completed with reference to the British Psychological Society (BPS) Code of Human Research Ethics and Code of Ethics and Conduct.

Section 1 Project details

a.	Project title	The Effect of Intralexical Factors on the Decay of Foreign Language Vocabulary Knowledge			
b.	Student name	Samuel Barclay			
c.	Supervisor/Personal Tutor	Dr Ana Pellicer-Sanchez			
d.	Department	CCM			
e.	Course category (Tick one)	PhD/MPhil ☒		EdD	☐
		MRes ☐		DEdPsy	☐
		MTeach ☐		MA/MSc	☐
		ITE ☐			
		Diploma (state which) ☐			
		Other (state which) ☐			

f.	Course/module title	Research Degree: Culture, Communication, and Media
g.	**If applicable,** state who the funder is and if funding has been confirmed.	Na
h.	Intended research start date	01/05/2017
i.	Intended research end date	01/05/2018
j.	Country fieldwork will be conducted in *If research to be conducted abroad please ensure travel insurance is obtained through UCL http://www.ucl.ac.uk/finance/insurance/travel*	China

k.	Has this project been considered by another (external) Research Ethics Committee?	
	Yes ☐	External Committee Name:
	No ☒ ⇨ *go to Section 2*	Date of Approval:

If yes:

- Submit a copy of the approval letter with this application.
- Proceed to Section 10 Attachments**.**

Note: Ensure that you check the guidelines carefully as research with some participants will require ethical approval from a different ethics committee such as the National Research Ethics Service (NRES) or Social Care Research Ethics Committee (SCREC). In addition, if your research is based in another institution then you may be required to apply to their research ethics committee.

Section 2 Project summary

Research methods (tick all that apply)

Please attach questionnaires, visual methods and schedules for interviews (even in draft form).

☐ Interviews ☐ Focus groups ☐ Questionnaires ☐ Action research ☐ Observation ☒ Literature review	☒ Controlled trial/other intervention study ☐ Use of personal records ☐ Systematic review ⇨ ***if only method used go to Section 5.*** ☐ Secondary data analysis ⇨ ***if secondary analysis used go to Section 6.*** ☐ Advisory/consultation/collaborative groups ☐ Other, give details:

Aims and research questions

The main aim of this study is to measure the effect of word length and part of speech on the attrition of second language vocabulary. A secondary aim is to better understand the effect of task type on second language lexical loss. In this study, non-native speakers of English will independently study and be tested on a set of target items. Several weeks later, they will be retested on the same set of target items. Following this, they will relearn the target items in addition to a set of previously unstudied items to test for a relearning advantage.

The project aims to investigate the following research questions:

1. What is the effect of part of speech (PoS) and word length on the decay of written form recognition knowledge and written form recall knowledge?
2. What is the effect of frequency of exposure in this process?
3. What is the effect of meaning presentation method in this process?

Research Design and Methodology

Participants. The participants will be English as a foreign language learners at universities in China. To recruit participants, I will approach two universities in which I have contacts (Communication University of China, and University of Nottingham Ningbo). I will request their help in disseminating a call for participation in the project. This call will include the information sheet describing the study (please find attached). Potential participants will be asked to contact me if they are interested in taking part. Interested potential participants will then be offered an opportunity to ask further questions about the study either by email or during an induction before they decide whether to consent to take part.

Design. The project will consist of three sessions. In the first session, participants will consider the aims and read the consent form. They will also complete a vocabulary test and a language background questionnaire. After signing the consent form and sorting out questions, participants will be introduced to the learning instrument. After completing a practice session and being introduced to the learning log, participants will have the opportunity to ask any questions they have about any part of the study. They will receive a flash drive containing the learning software and a learning log which they will use for studying the target items. The first learning session will take place on this first day. On the second day, the students will again study the target words and complete a test. The latter will target prior knowledge of the target items. Sometime later, a delayed test, consisting of the same testing instrument and a questionnaire, will be conducted. This questionnaire will target any exposure to the target items since the learning procedure. At this time, participants will also take part in one final learning session using the flashcard software to check for a relearning benefit.

Data collection. Each session is expected to take 60-90 minutes. The sessions will take place at the Chinese universities.

Data analysis. Learning and test data will be extracted from the flash drive using a macro I have developed. It will then be analysed to look for any decay that has occurred. This data will then be analysed quantitatively, considering several predictor variables as well as meaning presentation type and frequency of exposure.

Section 3 Participants

Please answer the following questions giving full details where necessary. Text boxes will expand for your responses.

a.	Will your research involve human participants?	Yes ☒	No ☐ ⇨ *go to Section 4*

b. Who are the participants (i.e., what sorts of people will be involved)? Tick all that apply.

☐ Early years/pre-school	☐ Unknown – specify below
☐ Ages 5-11	☒ Adults *please specify below*
☐ Ages 12-16	☐ Other – specify below
☐ Young people aged 17-18	

NB: Ensure that you check the **guidelines** (Section 1) carefully as research with some participants will require ethical approval from a different ethics committee such as the National Research Ethics Service (NRES).

Participants will be university students

c. If participants are under the responsibility of others (such as parents, teachers or medical staff) how do you intend to obtain permission to approach the participants to take part in the study?

(Please attach approach letters or details of permission procedures – see Section 9 Attachments.)

na

d. How will participants be recruited (identified and approached)?

The participants will be English as a foreign language learners at universities in China. To recruit participants, I will approach two universities in which I have contacts (Communication University of China, and University of Nottingham Ningbo). I will request their help in disseminating a call for participation in the project. This call will include the information sheet describing the study (please find attached). Potential participants will be asked to contact me if they are interested in taking part.

e. Describe the process you will use to inform participants about what you are doing.

Potential participants will be emailed a written explanation of the study. Interested potential participants will be offered an opportunity to ask further questions about the study either by email or during an induction before they decide whether to consent to take part.

f. How will you obtain the consent of participants? Will this be written? How will it be made clear to participants that they may withdraw consent to participate at any time?

	See the guidelines for information on opt-in and opt-out procedures. Please note that the method of consent should be appropriate to the research and fully explained. Prior to participating in the study, potential participants will be asked to provide written consent to taking part in the research project. During the induction and on the consent form (which I will explain) it will clearly state that participants can withdraw from the study at any time without having to provide reason. It will be explained that this can be done by emailing me.
g.	**Studies involving questionnaires:** Will participants be given the option of omitting questions they do not wish to answer? Yes ☐ No ☐
	If **NO** please explain why below and ensure that you cover any ethical issues arising from this in section 8.
h.	**Studies involving observation:** Confirm whether participants will be asked for their informed consent to be observed. Yes ☐ No ☐
	If **NO** read the guidelines (Ethical Issues section) and explain why below and ensure that you cover any ethical issues arising from this in section 8.
i.	Might participants experience anxiety, discomfort or embarrassment as a result of your study? Yes ☐ No ☒
	If **yes** what steps will you take to explain and minimise this? If **not**, explain how you can be sure that no discomfort or embarrassment will arise? I have run a similar study previously in my course of study and found no discomfort or embarrassment to arise. One potential risk relates to the amount of time which is involved in participating in the study. To minimise this risk, the information sheet will provide full information about the time requirements of participation. However, other than imposition on participants' time, the research has no potential risk.
j.	Will your project involve deliberately misleading participants (deception) in any way? Yes ☐ No ☒
	If **YES** please provide further details below and ensure that you cover any ethical issues arising from this in section 8.

k.	Will you debrief participants at the end of their participation (i.e., give them a brief explanation of the study)? Yes ☒ No ☐
	If **NO** please explain why below and ensure that you cover any ethical issues arising from this in section 8.
l.	Will participants be given information about the findings of your study? (This could be a brief summary of your findings in general; it is not the same as an individual debriefing.) Yes ☒ No ☐
	If **no**, why not?

Section 4 Security-sensitive material

Only complete if applicable

Security sensitive research includes: commissioned by the military; commissioned under an EU security call; involves the acquisition of security clearances; concerns terrorist or extreme groups.

a.	Will your project consider or encounter security-sensitive material?	Yes ☐ *	No ☒
b.	Will you be visiting websites associated with extreme or terrorist organisations?	Yes ☐ *	No ☒
c.	Will you be storing or transmitting any materials that could be interpreted as promoting or endorsing terrorist acts?	Yes ☐ *	No ☒

* *Give further details in **Section 8 Ethical Issues***

5 Systematic review of research

Only complete if applicable

a.	Will you be collecting any new data from participants?	Yes ☒ *	No ☐
b.	Will you be analysing any secondary data?	Yes ☐ *	No ☒

** Give further details in **Section 8 Ethical Issues***

*If your methods do not involve engagement with participants (e.g., systematic review, literature review) **and** if you have answered **No** to both questions, please go to **Section 10 Attachments.***

Section 6 Secondary data analysis Complete for all secondary analysis

a.	Name of dataset/s		
b.	Owner of dataset/s		
c.	Are the data in the public domain?	Yes ☐ No ☐ ***If no,** do you have the owner's permission/license?* Yes ☐ No* ☐	
d.	Are the data anonymised?	Yes ☐ No ☐ *Do you plan to anonymise the data?* Yes ☐ No* ☐ *Do you plan to use individual level data?* Yes* ☐ No ☐ *Will you be linking data to individuals?* Yes* ☐ No ☐	
e.	Are the data sensitive (DPA 1998 definition)?	Yes* ☐	No ☐
f.	Will you be conducting analysis within the remit it was originally collected for?	Yes ☐	No* ☐
g.	**If no,** was consent gained from participants for subsequent/future analysis?	Yes ☐	No* ☐
h.	**If no,** was data collected prior to ethics approval process?	Yes ☐	No* ☐

** Give further details in **Section 8 Ethical Issues***

*If secondary analysis is only method used **and** no answers with asterisks are ticked, go to **Section 9 Attachments.***

Section 7 Data Storage and Security

Please ensure that you include all hard and electronic data when completing this section.

a.	Confirm that all personal data will be stored and processed in compliance with the Data Protection Act 1998 (DPA 1998). *(See the Guidelines and the Institute's Data Protection & Records Management Policy for more detail.)*		Yes ☒
b.	Will personal data be processed or be sent outside the European Economic Area?	Yes ☒ *	No ☐

*** If yes,** please confirm that there are adequate levels of protections in compliance with the DPA 1998 and state what these arrangements are below.

	The data will be collected and may be initially analysed in China. All research data will be stored in line with UCL's Research Data Policy. All data will be kept securely: digital data will be kept in a password protected system to which only I have access, and analogue data will be stored in a locked cabinet.
c.	Who will have access to the data and personal information, including advisory/consultation groups and during transcription? None but me.
During the research	
d.	Where will the data be stored? In a password protected folder on my computer, which is also password protected.
e.	Will mobile devices such as USB storage and laptops be used? Yes ☒ * No ☐ * **If yes,** state what mobile devices: Laptop * **If yes,** will they be encrypted?: Yes
After the research	
f.	Where will the data be stored? In a password protected folder on my computer.
g.	How long will the data and records be kept for and in what format? Data will only be kept in digital format. It will be deleted after a period of 10 years.
h.	Will data be archived for use by other researchers? Yes ☐ * No ☒ * **If yes,** please provide details.

Section 8 Ethical issues

Are there particular features of the proposed work which may raise ethical concerns or add to the complexity of ethical decision making? If so, please outline how you will deal with these.

It is important that you demonstrate your awareness of potential risks or harm that may arise as a result of your research. You should then demonstrate that you have considered ways to minimise the likelihood and impact of each potential harm that you have identified. Please be as specific as possible in describing the ethical issues you will have to address. Please consider / address ALL issues that may apply.

Ethical concerns may include, but not be limited to, the following areas:

- Methods
- Sampling
- Recruitment
- Gatekeepers
- Informed consent
- Potentially vulnerable participants
- Safeguarding/child protection
- Sensitive topics
- International research
- Risks to participants and/or researchers
- Confidentiality/Anonymity
- Disclosures/limits to confidentiality
- Data storage and security both during and after the research (including transfer, sharing, encryption, protection)
- Reporting
- Dissemination and use of findings

Vulnerable participants

No vulnerable groups will be targeted. All research participants will be adults.

International research

The proposed research project will be carried out at two universities in China. I am travelling to the research destinations as part of my work, and will conduct the data colletion in addition to my professional duties. Travel and accommodation is covered by my employer. Also, a thorough risk assessment has been performed by my employer (Nottingham Trent University).

Risks to participants and/or researchers

A potential risk relates to the amount of time which is involved in participating in the study. To minimise this risk, the information sheet will provide full information about the time requirements of participation. Other than imposition on participants' time, the research has no potential risk. No deception is involved in the study.

Confidentiality and anonymity

All data will honour assurances of confidentiality and anonymity. Participant confidentiality (named identity) will be maintained and remain with me. Using unique identifiers, all participants will be anonymised for data analysis and any release/ dissemination outside the project. To identify which participants have provided all levels of consent, the researcher will keep a name list with first names linked to the unique identifiers. This sheet will remain with me or locked away, and will be password-protected. At a later date, this sheet will be destroyed. In this way, all data can only be referenced by a unique identifier.

Sensitive data

No 'sensitive' data under the definition of the Data Protection Act 1998 will be collected.

Data storage/security

All research data will be stored in line with the IOE's Information Security Management Policy and UCL's Research Data Policy. All data will be kept securely: digital data will be kept in a password protected system to which only I have access, and analogue data will be stored in a locked cabinet.

Data sharing

All research participants will be informed about the data archiving and sharing process, and their written informed consent will be sought for the sharing of the data.

Informed consent

The participants' informed consent will be obtained by explaining the purpose and the process of the research in the induction. As outlined above, participants will be provided with an information sheet and will be given the opportunity to review the information sheet before indicating an interest in participating in the study. Interested potential participants will then be offered an opportunity to ask further questions about the study before they decide whether to consent to take part. The consent forms will also be reviewed with the participants during the induction, and they will be given ample time to read the forms before signing them (if they agree to participate). Participants will be assured that participation or non-participation will not impact on their evaluation in the course they take. The appropriate written consent forms will be obtained from each participant prior to the start of the first learning session.

Potential Benefits

Participants will have the opportunity to acquire vocabulary and learn their vocabulary size. I will also give the participants advice on good practice when learning vocabulary. Also, the insights gained will further our knowledge about the decay of explicitly acquired foreign language vocabulary knowledge which may help in designing tasks that lead to more robust knowledge.

Incentives

Participants will not be offered any incentives to participate in the research.

Dissemination

Participants will receive a summary about the overall results of the study when it is completed. No individual level information will be provided.

Section 9 Further information

Outline any other information you feel relevant to this submission, using a separate sheet or attachments if necessary.

Section 10 Attachments Please attach the following items to this form, or explain if not attached

a.	Information sheets and other materials to be used to inform potential participants about the research, including approach letters	Yes ☒	No ☐
b.	Consent form	Yes ☒	No ☐
	If applicable:		
c.	The proposal for the project	Yes ☐	No ☒
d.	Approval letter from external Research Ethics Committee	Yes ☐	No ☒

e.	Full risk assessment	Yes ☐	No ☒

Section 11 Declaration

	Yes	No
I have read, understood and will abide by the following set of guidelines.	☒	☐
BPS ☐ BERA ☒ BSA ☐ Other (please state) ☐		
I have discussed the ethical issues relating to my research with my supervisor.	☒	☐
I have attended the appropriate ethics training provided by my course.	☒	☐

I confirm that to the best of my knowledge:

The above information is correct and that this is a full description of the ethics issues that may arise in the course of this project.

Name	Samuel Barclay
Date	04/04/2017

Please submit your completed ethics forms to your supervisor

Departmental use

If a project raises particularly challenging ethics issues, or a more detailed review would be appropriate, you may refer the application to the Research Ethics and Governance Administrator (via IOE.researchethics@ucl.ac.uk) so that it can be submitted to the Research Ethics Committee for consideration. A Research Ethics Committee Chair, ethics representatives in your department and the research ethics coordinator can advise you, either to support your review process, or help decide whether an application should be referred to the Research Ethics Committee.

Reviewer 1	
Supervisor name	Ana Pellicer-Sánchez
Supervisor comments	The form outlines all the main ethical concerns of the study. I do not have any other concerns.
Supervisor signature	Signature removed
Reviewer 2	

Advisory committee/course team member name	Andrea Revesz	
Advisory committee/course team member comments	I see no ethical issues arising other than the ones addressed in the application.	
Advisory committee/course team member signature	Signature removed	
Decision		
Date decision was made	17/04/2017	
Decision	Approved	☒
	Referred back to applicant and supervisor	☐
	Referred to REC for review	☐
Recording	Recorded in the student information system	☐

Once completed and approved, please send this form and associated documents to the relevant programme administrator to record on the student information system and to securely store.

Further guidance on ethical issues can be found on the IOE website at http://www.ucl.ac.uk/srs/research-ethics-committee/ioe and www.ethicsguidebook.ac.uk

Appendix 10
The Checklist Test Used to Determine Prior Knowledge of the Target Items (Study 2)

Instructions:

If Known type "1". If unknown do not type anything. In the examples, "apple" is unknown and "banana" is known.

Example:	apple		banana	1

skirt		normal		cambule		pasture		ridout	
frost		empire		tax		lawn		ambulate	
mackerel		fort		pursue		independent		manufacture	
quote		generation		fortune		social		kip	
champion		angel		tooley		wine		naked	
entire		bib		baldock		difficult		conflate	
herd		spinster		attach		irk		respect	
develop		scream		museum		gown		import	
nab		quorant		stage		peer		roar	
tic		factor		grasp		lean		duffin	
prance		contraption		marten		quit		opportunity	
complain		popular		camaraderie		cauldron		adair	
matriculate		pocock		lovely		condominium		pride	
balfour		dignity		magic		hubbard		voyeur	
stretch		holy		galpin		acklon			
antagonise		frolic		dowrick		expedite			
threaten		berrow		mundy		frame			
bop		zap		wray		event			
debt		pring		pontificate		melt			
slender		regurgitate		hell		owe			
gift		motor		desperate		imp			
counsel		hen		withrow		endure			
jump		profit		scintillate		merit			
jangle		heckle		plot		venerate			
cinder		dwell		pity		register			
distinct		striking		contest		fix			
wage		slight		keg		nerve			
eckett		decant		disease		harpsichord			
orb		curious		merry		webbert			
theater		bance		prefer		introduce			
truncate		kiley		seize		knit			
oblige		tenacity		buttle		relief			
thread		tip		clique		elect			
justice		snell		zealot		asterisk			
hire		archipelago		blank		noise			

Appendix 11

The Target items, L1 Equivalents and L2 Definitions (Study 2)

Item	PoS	Length	L1 equivalent	L2 definition
bib	noun	3	(名词) 围嘴	(noun) a piece of cloth or plastic tied under a baby's face
keg	noun	3	(名词) 桶	(noun) a round wooden container with a flat top and bottom
tic	noun	3	(名词) 抽搐（面部）	(noun) a sudden movement of a muscle in your face
orb	noun	3	(名词) 球状物	(noun) a bright ball-shaped object such as the sun or the moon
voyeur	noun	6	(名词) 偷窥狂	(noun) a person who enjoys watching other people
cinder	noun	6	(名词) 煤渣	(noun) a very small piece of burnt wood
clique	noun	6	(名词) 小圈子	(noun) a small group of people who spend their time together
zealot	noun	6	(名词) 狂热者	(noun) a person who has very strong feelings
asterisk	noun	8	(名词)星号	(noun) an image placed next to a word to make people notice it
spinster	noun	8	(名词) 老处女	(noun) an unmarried woman who is old
cauldron	noun	8	(名词) 大锅	(noun) a large round metal pot for boiling water over a fire
mackerel	noun	8	(名词) 鲭鱼	(noun) a sea fish that is blue and silver, and has a strong taste
contraption	noun	11	(名词) 奇妙的新装置	(noun) a machine that looks strange and is unlikely to work well
archipelago	noun	11	(名词) 群岛	(verb) a group of small islands
harpsichord	noun	11	(名词) 竖琴	(noun) a musical instrument like a piano
condominium	noun	11	(名词) 公寓	(noun) an apartment in a building with several apartments
kip	verb	3	(动词) 睡觉（不在家里）	(verb) to sleep somewhere that is not your home
irk	verb	3	(动词) 激怒	(noun) to make someone feel annoyed
nab	verb	3	(动词) 逮捕	(verb) to catch or arrest someone who is doing something wrong
bop	verb	3	(动词)轻轻一击	(verb) to hit someone gently
decant	verb	6	(动词) 将（酒等）自瓶中倒入另一容器	(verb) to pour wine from one bottle into another
prance	verb	6	(动词) 昂首阔步	(verb) to walk with high steps

frolic	verb	6	(动词)嬉戏	(verb) to play and move around in a happy way
heckle	verb	6	(动词)起哄	(verb) to interrupt and try to embarrass someone who is speaking in public
conflate	verb	8	(动词) 合并	(verb) to combine two or more things to form a single new thing
truncate	verb	8	(动词) 缩短	(verb) to make something shorter
venerate	verb	8	(动词) 尊敬	(verb) to respect someone because they are old or important
expedite	verb	8	(动词) 加速	(verb) to make a process happen more quickly
matriculate	verb	11	(动词) 正式入学（大学）	(verb) to officially begin studying at a university
scintillate	verb	11	(动词) 闪烁	(verb) to turn on and off quickly
pontificate	verb	11	(动词) 夸夸其谈	(verb) to give your opinion about something
regurgitate	verb	11	(动词) 使反胃	(verb) to bring food that you have already eaten back into your mouth

Appendix 12

The L1 Survey Instrument Used to Measure Prior Knowledge, Intersessional Exposure, and Perceived Usefulness (Study 2)

"KNOWN" - Did you know any words before Session 1. Type a "Y" next to any word that you knew before studying

"SEEN" - Have you seen a word since the last session? Type Y next to words you have seen.

"USEFUL" - How useful is each word? 1 = completely useless; 2 = useless; 3 = not really useful; 4 = a little useful; 5 = useful; 6 = very useful

USB#

			Known	Seen	Useful
Example	English	英語	Y		6
	Vocabulary	詞彙		Y	3

			Known	Seen	Useful
1	nab	逮捕			
2	asterisk	星号			
3	venerate	尊敬			
4	spinster	老处女			
5	expedite	加速			
6	archipelago	群岛			
7	bop	轻轻一击			
8	keg	桶			
9	regurgitate	使反胃			
10	prance	昂首阔步			
11	pontificate	夸夸其谈			
12	zealot	狂热者			
13	contraption	奇妙的新装置			
14	tic	抽搐（面部）			
15	matriculate	正式入学（大学）			
16	frolic	嬉戏			
17	conflate	合并			
18	mackerel	鲭鱼			
19	truncate	缩短			
20	cinder	煤渣			
21	irk	激怒			
22	harpsichord	竖琴			
23	decant	将（酒等）自瓶中倒入另一容器			

24	cauldron	大锅			
25	clique	小圈子			
26	bib	围嘴			
27	scintillate	闪烁			
28	kip	睡觉（不在家里）			
29	heckle	起哄			
30	orb	球状物			
31	condominium	公寓			
32	voyeur	偷窥狂			

Appendix 13

The L2 Survey Instrument Used to Measure Prior Knowledge, Intersessional Exposure, and Perceived Usefulness (Study 2)

"KNOWN" - Did you know any words before Session 1. Type a "Y" next to any word that you knew before studying

"SEEN" - Have you seen a word since the last session? Type Y next to words you have seen.

"USEFUL" - How useful is each word? 1 = completely useless; 2 = useless; 3 = not really useful; 4 = a little useful; 5 = useful; 6 = very useful

USB#

			Known	Seen	Useful
Examples	English	A language spoken in the UK and the USA	**Y**		**6**
	Vocabulary	words		**Y**	**3**

			Known	Seen	Useful
1	nab	to catch or arrest someone who is doing something wrong			
2	asterisk	an image placed next to a word to make people notice it			
3	venerate	to respect someone because they are old or important			
4	spinster	an unmarried woman who is old			
5	expedite	to make a process happen more quickly			
6	archipelago	a group of small islands			
7	bop	to hit someone gently			
8	keg	a round wooden container with a flat top and bottom			
9	regurgitate	to bring food that you have already eaten back into your mouth			
10	prance	to walk with high steps			
11	pontificate	to give your opinion about something			
12	zealot	a person who has very strong feelings			
13	contraption	a machine that looks strange and is unlikely to work well			
14	tic	a sudden movement of a muscle in your face			

15	matriculate	to officially begin studying at a university			
16	frolic	to play and move around in a happy way			
17	conflate	to combine two or more things to form a single new thing			
18	mackerel	a sea fish that is blue and silver, and has a strong taste			
19	truncate	to make something shorter			
20	cinder	a very small piece of burnt wood			
21	irk	to make someone feel annoyed			
22	harpsichord	a musical instrument like a piano			
23	decant	to pour wine from one bottle into another			
24	cauldron	a large round metal pot for boiling water over a fire			
25	clique	a small group of people who spend their time together			
26	bib	a piece of cloth or plastic tied under a baby's face			
27	scintillate	to turn on and off quickly			
28	kip	to sleep somewhere that is not your home			
29	heckle	to interrupt and try to embarrass someone who is speaking in public			
30	orb	a bright ball-shaped object such as the sun or the moon			
31	condominium	one apartment in a building with several apartments			
32	voyeur	a person who enjoys watching other people			

Appendix 14

Model Selection Process for Learning Burden (Frequency and Time of Exposure) Models (Study 2)

Summary of Mixed -Effect Model Comparisons for Frequency of Exposure, Strict condition (Study 2)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Code of meaning presentation; Usefulness; code * usefulness; Vocabulary; Additional language; Age; Word length; PoS; length * PoS; length *code; PoS * code	By-participant + by-item random intercepts.	360.67	NA	NA	NA
2	Model 1 – Age	Same as Model 1	360.46	-0.21	X^2 (1) = 1.79	*P* = .18
3	Model 2 – Additional language	Same as Model 1	358.50	-1.96	X^2 (1) = 0.04	*P* = .85
4	Model 3 – Vocabulary	Same as Model 1	361.09	2.59	X^2 (1) = 4.60	*P* = .03*
5	Model 3 – Length * PoS	Same as Model 1	357.60	-0.90	X^2 (1) = 1.11	*P* = .29
6	Model 5 - PoS*Code	Same as Model 1	357.76	0.16	X^2 (1) = 2.16	*P*=< .14
7	Model 6 - Length * Code	Same as Model 1	355.91	-1.85	X^2 (1) = 0.14	*P* = .71
8	Model 7 - Class	Same as Model 1	359.38	3.47	X^2 (1) = 5.47	*P* = .02*
9	Model 7 - Length	Same as Model 1	380.60	24.69	X^2 (1) = 26.69	*P* < 0.001***

Summary of Mixed -Effect Model Comparisons for Frequency of Exposure (Lenient condition) (Study 2)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Code of meaning presentation; Usefulness; code * usefulness; Vocabulary; Additional language; Age; Word length; PoS; length * PoS; length *code; PoS * code	By-participant + by-item random intercepts.	450.23	NA	NA	NA
2	Model 1 – Age	Same as Model 1	450.01	-0.22	X^2 (1) = 1.78	*P* = .18
3	Model 2 – Additional language	Same as Model 1	448.02	-1.99	X^2 (1) = 0.01	*P* = .93
4	Model 3 – Vocabulary	Same as Model 1	451.33	3.31	X^2 (1) = 5.31	*P* = .02*
5	Model 3 – Length * PoS	Same as Model 1	447.10	-0.92	X^2 (1) = 1.08	*P* = .29
6	Model 5 - PoS*Code	Same as Model 1	447.05	-0.05	X^2 (1) = 1.95	*P* = .16
7	Model 6 - PoS	Same as Model 1	449.69	2.64	X^2 (1) = 4.64	*P* = .03*
8	Model 6 - Length * Code	Same as Model 1	445.68	-1.37	X^2 (1) = 0.63	*P* = .42
9	Model 8 - Length	Same as Model 1	470.33	24.65	X^2 (1) = 26.66	*P* < 0.001***

Summary of Mixed -Effect Model Comparisons for Frequency of Exposure (form recognition condition) (Study 2)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Code of meaning presentation; Usefulness; code * usefulness; Vocabulary; Additional language; Age; Word length; PoS; length * PoS; length *code; PoS * code	By-participant + by-item random intercepts.	660.82	NA	NA	NA
2	Model 1 – Age	Same as Model 1	659.89	-0.93	X^2 (1) = 1.07	*P* = .30
3	Model 2 – Additional language	Same as Model 1	657.89	-2	X^2 (1) <0.01	*P* = .95
4	Model 3 – Vocabulary	Same as Model 1	661.23	3.34	X^2 (1) = 5.34	*P* = .02*
5	Model 3 – Length * PoS	Same as Model 1	657.1	-0.79	X^2 (1) = 1.20	*P* = .27
6	Model 5 - PoS * Code	Same as Model 1	656.1	-1	X^2 (1) = 1.34	*P* = .24
7	Model 6 - PoS	Same as Model 1	659.43	3.33	X^2 (1) = 4.99	*P* = .02*
8	Model 6 - Length * Code	Same as Model 1	657.33	1.23	X^2 (1) = 2.89	*P* = .08
9	Model 8 - Length	Same as Model 1	683.42	26.09	X^2 (1) = 28.09	*P* < .001***

Summary of Mixed -Effect Model Comparisons for Time of Exposure (Strict condition) (Study 2)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Code of meaning presentation; Usefulness; Vocabulary; Additional language; Age; Word length; PoS; length * PoS; PoS * Code; Length * Code	By-participant + by-item random intercepts.	1547.5	NA	NA	NA
2	Model 1 – Age	Same as Model 1	1557.8	10.3	X^2 (1) = 14.30	*P* < 0.001***
3	Model 1 - Additional language		1545.5	-2	X^2 (1) = 0.74	*P* = 0.39
4	Model 3 – Vocabulary	Same as Model 1	1571.5	26	X^2 (1) = 29.29	*P* < 0.001***
5	Model 3 – Length * PoS	Same as Model 1	1542.8	-2.7	X^2 (1) = 0.61	*P* = 0.43
6	Model 5 – PoS * Code	Same as Model 1	1548.4	5.6	X^2 (1) = 7.59	*P* = 0.005**
7	Model 5 - Length * Code	Same as Model 1	1583.4	40.6	X^2 (1) = 42.59	*P* < 0.001***

Summary of Mixed -Effect Model Comparisons for Time of Exposure (lenient condition) (Study 2)

Model description					Test against prior model	
Model	Fixed Effects	Random Effects	AIC	ΔAIC	Statistic	Significance
1	Code of meaning presentation; Usefulness; Vocabulary; Additional language; Age; Word length; PoS; length * PoS; PoS * Code; Length * Code	By-participant + by-item random intercepts.	1713.7	NA	NA	NA
2	Model 1 – Age	Same as Model 1	1726	12.3	X^2 (1) = 14.24	*P* < .001***
3	Model 1 - Additional language		1712.8	-0.9	X^2 (1) = 1.05	*P* = .30
4	Model 3 – Vocabulary	Same as Model 1	1742.5	29.7	X^2 (1) = 31.72	*P* < .001***
5	Model 3 – Length * PoS	Same as Model 1	1711.3	-1.5	X^2 (1) = 0.50	*P* = .47
6	Model 5 – PoS * Code	Same as Model 1	1716.7	5.4	X^2 (1) = 7.40	*P* = 0.007**
7	Model 5 - Length * Code	Same as Model 1	1759.2	47.9	X^2 (1) = 49.90	*P* < 0.001***

Summary of Mixed -Effect Model Comparisons for Time of Exposure (recognition condition) (Study 2)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Code of meaning presentation; Usefulness; Vocabulary; Additional language; Age; Word length; PoS; length * PoS; PoS * Code; Length * Code	By-participant + by-item random intercepts.	2081.2	NA	NA	NA
2	Model 1 – Age	Same as Model 1	2092	10.8	X^2 (1) = 12.76	*P* < .001***
3	Model 1 - Additional language		2080.8	-0.4	X^2 (1) = 1.56	*P* = .21
4	Model 3 – Vocabulary	Same as Model 1	2109.9	29.1	X^2 (1) = 31.10	*P* < .001***
5	Model 3 – Length * PoS	Same as Model 1	2079.4	-1.4	X^2 (1) = 0.57	*P* = .45
6	Model 5 – PoS * Code	Same as Model 1	2084.2	4.8	X^2 (1) = 6.86	*P* = .008**
7	Model 5 - Length * Code	Same as Model 1	2135.2	55.8	X^2 (1) = 57.82	*P* < .001***

Appendix 15

Model Selection Procedure for Decay Analysis (Study 2)

Summary of Generalised Linear Mixed -Effect Model Comparisons for Retention of Learned Items (strict condition) (Study 2)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Code of meaning presentation; Usefulness; Code*Usefulness; Vocabulary; Additional language; Time of exposure; Frequency of exposure; Time*Code; Frequency*code; Word length; PoS; length * PoS; PoS * Code; Length * Code	By-participant + by-item random intercepts.	2003.9	NA	NA	NA
2	Model 1 – Additional language	Same as Model 1	2002.7	1.2	X^2 (1) = 0.86	*P* =.35
3	Model 2 - Vocabulary	Same as Model 1	2003.2	0.5	X^2 (1) = 2.48	*P* = .11
4	Model 3 – Length*PoS	Same as Model 1	2001.9	-1.3	X^2 (1) = 0.70	*P* = .40
5	Model 4 – PoS*Code	Same as Model 1	2000.7	-1.2	X^2 (1) = 0.79	*P* = .37
6	Model 5 – PoS	Same as Model 1	2005.6	4.9	X^2 (1) = 6.89	*P* = .009**
7	Model 5 - Length * Code	Same as Model 1	1998.7	-2	X^2 (1) <0.001	*P* = .92
8	Model 7 - Word length	Same as Model 1	1999.9	1.2	X^2 (1) = 3.19	*P* =.07
9	Model 8 – Time*Code	Same as Model 1	1998.5	-1.4	X^2 (1) = 0.56	*P* =.45
10	Model 9 - Frequency*Code	Same as Model 1	1999.4	0.9	X^2 (1) = 2.91	*P* = .08
11	Model 10 – Time of exposure	Same as Model 1	2017.3	17.9	X^2 (1) = 19.93	*P* <.001***
12	Model 10 – Frequency of exposure	Same as Model 1	2016	16.6	X^2 (1) = 18.65	*P* <.001***

Summary of Generalised Linear Mixed -Effect Model Comparisons for Retention of Learned Items (lenient condition) (Study 2)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Code of meaning presentation; Usefulness; Code*Usefulness; Vocabulary; Additional language; Time of exposure; Frequency of exposure; Time*Code; Frequency*code; Word length; PoS; length * PoS; PoS * Code; Length * Code	By-participant + by-item random intercepts.	2198.7	NA	NA	NA
2	Model 1 – Additional language	Same as Model 1	2198.6	-0.1	X^2 (1) = 1.90	*P* =.16
3	Model 2 - Vocabulary	Same as Model 1	2198	-0.6	X^2 (1) = 1.44	*P* = .23
4	Model 3 – Length*PoS	Same as Model 1	2196	-2	X^2 (1) < 0.01	*P* = .98
5	Model 4 – PoS*Code	Same as Model 1	2195.7	-0.3	X^2 (1) = 1.70	*P* = .19
6	Model 5 – PoS	Same as Model 1	2201.7	6	X^2 (1) = 7.99	*P* = .004**
7	Model 5 - Length * Code	Same as Model 1	2193.8	1.9	X^2 (1) = 0.03	*P* = .86
8	Model 7 - Word length	Same as Model 1	2192.7	-1.1	X^2 (1) = 0.95	*P* =.33
9	Model 8 – Time*Code	Same as Model 1	2193.4	0.7	X^2 (1) = 2.68	*P* =.10
10	Model 9 - Frequency*Code	Same as Model 1	2194.8	1.4	X^2 (1) = 3.43	*P* = .06
11	Model 10 – Time of exposure	Same as Model 1	2204.3	9.5	X^2 (1) = 11.50	*P* < .001***
12	Model 10 – Frequency of exposure	Same as Model 1	2215.3	20.5	X^2 (1) = 22.51	*P* < .001***

Summary of Generalised Linear Mixed -Effect Model Comparisons for Retention of Learned Items (recognition condition) (Study 2)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Code of meaning presentation; Usefulness; Code*Usefulness; Vocabulary; Additional language; Time of exposure; Frequency of exposure; Time*Code; Frequency*code; Word length; PoS; length * PoS; PoS * Code; Length * Code	By-participant + by-item random intercepts.	823.1	NA	NA	NA
2	Model 1 – Additional language	Same as Model 1	821.2	1.9	X^2 (1) = 0.07	*P* =.78
3	Model 2 - Vocabulary	Same as Model 1	821.1	-0.1	X^2 (1) = 1.87	*P* = .17
4	Model 3 – Length*PoS	Same as Model 1	819.4	-1.7	X^2 (1) = 0.28	*P* = .59
5	Model 4 – PoS*Code	Same as Model 1	821	1.6	X^2 (1) = 3.66	*P* = .05
6	Model 4 – Length * Code	Same as Model 1	817.4	-2	X^2 (1) < 0.01	*P* = .95
7	Model 6 - Word length	Same as Model 1	816.4	-1	X^2 (1) = 1.04	*P* = .30
8	Model 7 - Time*Code	Same as Model 1	814.8	-1.6	X^2 (1) = 0.43	*P* =.51
9	Model 8 – Frequency*Code	Same as Model 1	813.5	-1.3	X^2 (1) = 0.69	*P* =.40

Appendix 16
Approved Ethics Form (Study 3)

Doctoral Student Ethics Application Form

Anyone conducting research under the auspices of the Institute of Education (staff, students or visitors) where the research involves human participants or the use of data collected from human participants, is required to gain ethical approval before starting. This includes preliminary and pilot studies. Please answer all relevant questions in simple terms that can be understood by a lay person and note that your form may be returned if incomplete.

Registering your study with the UCL Data Protection Officer as part of the UCL Research Ethics Review Process

If you are proposing to collect personal data i.e., data from which a living individual can be identified **you must be registered with the UCL Data Protection Office before you submit your ethics application for review**. To do this, email the complete ethics form to data-protection@ucl.ac.uk. Once your registration number is received, add it to the form* and submit it to your supervisor for approval.

If the Data Protection Office advises you to make changes to the way in which you propose to collect and store the data this should be reflected in your ethics application form.

Section 1 Project details				
a.	Project title			The Effect of Intralexical Factors on the Decay of Foreign Language Vocabulary Knowledge
b.	Student name and ID number (e.g., ABC12345678)			Samuel Barclay 17075213
c.	***UCL Data Protection Registration Number**			Z6364106/2018/11/62 Date issued 20/11/2018
c.	Supervisor/Personal Tutor			Dr Ana Pellicer-Sanchez
d.	Department			CCM
e.	Course category (Tick one)	PhD ☒	EdD ☐	
		DEdPsy ☐		
f.	**If applicable,** state who the funder is and if funding has been confirmed.			
g.	Intended research start date			20/10/2018
h.	Intended research end date			20/10/2019
i.	Country fieldwork will be conducted in *If research to be conducted abroad please check www.fco.gov.uk and submit a completed travel risk assessment form (see guidelines). If the FCO advice is against travel this will be* ***required*** *before ethical*			UK

	approval can be granted: *http://ioe-net.inst.ioe.ac.uk/about/profservices/international/Pages/default.aspx*	

j. Has this project been considered by another (external) Research Ethics Committee?

Yes ☐	External Committee Name:
No ☒ ⇨ *go to Section 2*	Date of Approval:

If yes:

- Submit a copy of the approval letter with this application.
- Proceed to Section 10 Attachments**.**

Note: Ensure that you check the guidelines carefully as research with some participants will require ethical approval from a different ethics committee such as the National Research Ethics Service (NRES) or Social Care Research Ethics Committee (SCREC). In addition, if your research is based in another institution then you may be required to apply to their research ethics committee.

Section 2 Research methods summary (tick all that apply)

☒ Interviews
☐ Focus groups
☒ Questionnaires
☐ Action research
☐ Observation
☒ Literature review

☒ Controlled trial/other intervention study
☐ Use of personal records
☐ Systematic review ⇨ ***if only method used go to Section 5.***
☐ Secondary data analysis ⇨ ***if secondary analysis used go to Section 6.***
☐ Advisory/consultation/collaborative groups
☐ Other, give details:

Aims and research questions

The main aim of this study is to measure the effect of manner of meaning and form presentation on the learning and subsequent attrition of second language vocabulary. Further aims of this study are to better understand the effect of participant variables (e.g., language learning aptitude) and the speed of lexical production prior to a retention interval on second language lexical loss. In this study, non-native speakers of English will complete a measure of language learning aptitude. They will then independently study and be tested on a set of target items. Several weeks later, they will be retested on the same set of target items. Following this, they will relearn the target items in addition to a set of previously unstudied items to test for a relearning advantage.

The project aims to investigate the following research questions:

1. What is the effect of the mode of meaning and form presentation on the decay of written form recognition knowledge and written form recall knowledge?
2. What is the effect of language learning aptitude in this process?
3. To what extent does frequency of exposure interact with language learning aptitude?
4. To what extent does the speed with which responses are provided prior to the retention interval predict target item retention?

Research Design and Methodology

Participants

The participants will be English language learners at Nottingham Trent University (NTU). To recruit participants, I will disseminate a call for participation among pre-sessional EAP students. This will be possible as I work full-time on the pre-sessional EAP programme at NTU. This call will include the information sheet describing the study (attached). Potential participants will be asked to contact me if they are interested in taking part. Interested potential participants will then be offered an opportunity to ask further questions about the study either by email or during an induction before they decide whether to consent to participate.

Design

A subsample of participants will first take part in a pilot study to measure the effectiveness of two language learning aptitude instruments. This pilot will take place over two sessions, held a week apart. In the first session, the goals of the pilot will be explained, and the learners will have the opportunity to ask questions. Learners will then sign a consent form and complete an instrument commonly used to measure language learning aptitude. One week later, they will complete another instrument commonly used to measure language learning aptitude. Based on a comparison of the results, an instrument will be chosen for the main study.

The main project will consist of three sessions. In the first session, participants will consider the aims and read the consent form. They will also complete a vocabulary test, a language background questionnaire, and various components of a language learning aptitude test. After signing the consent form and sorting out questions, participants will be introduced to the learning instrument. After completing a practice session, participants will have the opportunity to ask questions about any part of the study. They will receive a flash drive containing the learning software which they will use for studying the target items, the test of attainment, the test of retention, a test of vocabulary size, and a measure of language learning aptitude. In addition, a keystroke logging programme (Inputlog) will be loaded onto the USB. This will be used to measure the speed and automaticity with which participants produce the responses, and will only be used during the study, and thus will not record any personal information. Participants will be told about the use of this software during the induction. The first learning session will take place on this first day.

On the second day, the students will again study the target words and complete a test of attainment. The latter will target knowledge of the target items. Sometime later, a delayed test, consisting of the same testing instrument and a questionnaire, will be conducted. This questionnaire will target any exposure to the target items since the learning procedure. At this time, participants will also take part in one final learning session using the flashcard software to check for a relearning benefit. The keystroke logging software will be used for each learning and testing session. Participants will be repeatedly informed about its use.

After the final session, participants who displayed typical patterns of decay may be invited to participate in an interview with the researcher. In this interview, I will ask them why they feel they were able to retain some words better than others. These interviews will be recorded (prior permission for this will be sought), and transcribed. Additionally, the data from the keystroke logging software might be used in a stimulated recall protocol. The purpose of this would be to validate the construct of the various keystroke logging indices (e.g., total learning time, etc.). During these sessions, learners would be presented with their learning data and asked to recall what they were doing at key points in the learning process. This data would then be used to justify the inclusion of such indices in later analysis.

Data collection

Each session is expected to take 60-90 minutes. The sessions will take place on the university campus.

Data analysis

Learning and test data will be extracted from the flash drive using a macro I have developed. In addition, the keystroke logging data will be concatenated with this learning data to show the student production for each learning trial (i.e., each interaction with a flashcard), and their learning behaviour (i.e., pauses prior to production, revisions, etc.). These data will be analysed to look for any decay that has occurred. This analysis will involve the development of a series of statistical models including several predictor variables, meaning (L1 equivalent or L2 definition) and form (written only, or written + speaking) presentation mode, frequency of exposure, and several indices from the keystroke logging data (e.g., pause duration prior to production, in-word pauses, number of revisions, total learning time).

Section 3 Research Participants (tick all that apply)

☐ Early years/pre-school	☒ Adults *please specify below*
☐ Ages 5-11	☐ Unknown – specify below
☐ Ages 12-16	☐ No participants
☐ Young people aged 17-18	

NB: Ensure that you check the guidelines carefully as research with some participants will require ethical approval from a different ethics committee such as the National Research Ethics Service (NRES) or Social Care Research Ethics Committee (SCREC).

Section 4 Security-sensitive material (only complete if applicable)

Security sensitive research includes: commissioned by the military; commissioned under an EU security call; involves the acquisition of security clearances; concerns terrorist or extreme groups.

a.	Will your project consider or encounter security-sensitive material?	Yes ☐ *	No ☒
b.	Will you be visiting websites associated with extreme or terrorist organisations?	Yes ☐ *	No ☒
c.	Will you be storing or transmitting any materials that could be interpreted as promoting or endorsing terrorist acts?	Yes ☐ *	No ☒

* *Give further details in **Section 8 Ethical Issues***

Section 5 Systematic reviews of research (only complete if applicable)

a.	Will you be collecting any new data from participants?	Yes ☒ *	No ☐
b.	Will you be analysing any secondary data?	Yes ☐ *	No ☒

* *Give further details in **Section 8 Ethical Issues***

*If your methods do not involve engagement with participants (e.g., systematic review, literature review) **and** if you have answered **No** to both questions, please go to **Section 8 Attachments.***

Section 6 Secondary data analysis (only complete if applicable)

a.	Name of dataset/s		
b.	Owner of dataset/s		
		Yes ☐	No ☐

c.	Are the data in the public domain?		*If no, do you have the owner's permission/license?* Yes ☐ No* ☐	
d.	Are the data special category personal data (i.e., personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation)?		Yes* ☐	No ☐
e.	Will you be conducting analysis within the remit it was originally collected for?		Yes ☐	No* ☐
f.	**If no,** was consent gained from participants for subsequent/future analysis?		Yes ☐	No* ☐
g.	**If no,** was data collected prior to ethics approval process?		Yes ☐	No* ☐

* *Give further details in **Section 8 Ethical Issues***

*If secondary analysis is only method used **and** no answers with asterisks are ticked, go to **Section 9 Attachments.***

Section 7 Data Storage and Security

Please ensure that you include all hard and electronic data when completing this section.

a.	**Data subjects -** Who will the data be collected from? Participants will be university students enrolled on a pre-sessional English for Academic Purposes course at Nottingham Trent University				
b.	**What data will be collected?** Please provide details of the type of personal data to be collected I will collect biographical details including age, first language, and English language proficiency. I will also ask for participants' names; however, this will be later anonymised to maintain participant confidentiality (see below).				
c.	**Is the data anonymised?**	Yes	☒	No*	☐
	Do you plan to anonymise the data?	Yes*	☒	No	☐
	Do you plan to use individual level data?	Yes*	☒	No	☐
	Do you plan to pseudonymise the data?	Yes*	☒	No	☐
	* *Give further details in **Section 8 Ethical Issues***				

e.	**i. Disclosure** – Who will the results of your project be disclosed to? Participants will receive a summary about the overall results of the study when it is completed. Additionally, results will be shared with the supervisory team and disseminated through typical academic activities (i.e., presentation, publication) **ii. Disclosure** – Will personal data be disclosed as part of your project? No
f.	**Data storage** – Please provide details on how and where the data will be stored i.e., UCL network, encrypted USB stick**, encrypted laptop** etc. All data will be stored in line with the IOE's Information Security Management Policy and UCL's Research Data Policy. Data will be kept securely: digital data in a password protected system (only I have access); analogue data stored in locked cabinet. ** Advanced Encryption Standard 256 bit encryption which has been made a security standard within the NHS
g..	**Data Safe Haven (Identifiable Data Handling Solution)** – Will the personal identifiable data collected and processed as part of this research be stored in the UCL Data Safe Haven (mainly used by SLMS divisions, institutes and departments)? Yes ☐ No ☒
h.	How long will the data and records be kept for and in what format? Data will only be kept in digital format. It will be deleted after a period of 10 years. Will personal data be processed or be sent outside the European Economic Area? (If yes, please confirm that there are adequate levels of protections in compliance with GDPR and state what these arrangements are) No Will data be archived for use by other researchers? (If yes, please provide details.) No
i.	If personal data is used as part of your project, describe what measures you have in place to ensure that the data is only used for the research purpose e.g., pseudonymisation and short retention period of data' All data will honour assurances of confidentiality and anonymity. Participant confidentiality (named identity) will be maintained and remain with me. Using unique identifiers, all participants will be anonymised for data analysis and any release/ dissemination outside the project. To identify which participants have provided all levels of consent, the researcher will keep a name list with first names linked to the unique identifiers. This sheet will remain with me or locked away, and will be password-protected. At a later date, this sheet will be destroyed. In this way, all data can only be referenced by a unique identifier.
	* *Give further details in **Section 8 Ethical Issues***

Section 8 Ethical issues

Please state clearly the ethical issues which may arise in the course of this research and how will they be addressed.

All issues that may apply should be addressed. Some examples are given below, further information can be found in the guidelines. *Minimum 150 words required.*

− Methods − Sampling − Recruitment − Gatekeepers − Informed consent − Potentially vulnerable participants − Safeguarding/child protection − Sensitive topics	− International research − Risks to participants and/or researchers − Confidentiality/Anonymity − Disclosures/limits to confidentiality − Data storage and security both during and after the research (including transfer, sharing, encryption, protection) − Reporting − Dissemination and use of findings

Vulnerable participants
No vulnerable groups will be targeted. All research participants will be adults.

Risks to participants and/or researchers
A potential risk relates to the amount of time which is involved in participating in the study. To minimise this risk, the information sheet will provide full disclosure about the time requirements of participation. Another potential risk relates to the use of keystroke logging software as this records all participant interaction with the computers used for the study. To avoid inadvertently recording personal information, this software will be accessed via USB that will only be inserted into the computers (university computers) directly prior to commencement of the study. Furthermore, participants will be told that this software is being used, and instructed not to engage with documents, etc. outside the confines of the experimental design. Thus, no personal information (usernames, passwords, etc.) will be recorded. Therefore, other than imposition on participants' time, the research has no potential risk. No deception is involved in the study.

Confidentiality and anonymity
Please refer to Section 7 (i)

Sensitive data
No 'sensitive' data under the definition of the Data Protection Act 1998 will be collected.

Data storage/security
All research data will be stored in line with the IOE’s Information Security Management Policy and UCL’s Research Data Policy. All data will be kept securely: digital data will be kept in a password protected system to which only I have access, and analogue data will be stored in a locked cabinet.

Data sharing
All research participants will be informed about the data archiving and sharing process, and their written informed consent will be sought for the sharing of the data.

Informed consent
The participants’ informed consent will be obtained by explaining the purpose and the process of the research during the induction. As outlined above, participants will be provided with an information sheet and will be given the opportunity to review the information sheet before indicating an interest in participating in the study. Interested potential participants will then be offered an opportunity to ask further questions about the study before they decide whether to consent to take part. The consent forms will also be reviewed with the participants during the induction, and they will be given ample time to read the forms before signing them (if they agree to participate). Participants will be assured that participation or non-participation will not impact

on their evaluation in the course they take. The appropriate written consent forms will be obtained from each participant prior to the start of the first learning session.

Potential Benefits
Participants will have the opportunity to acquire vocabulary and learn their vocabulary size. I will also give the participants advice on good practice when learning vocabulary. Additionally, the insights gained will further our knowledge about the decay of explicitly acquired foreign language vocabulary knowledge, which may help in designing tasks that lead to the formation of more robust knowledge.

Incentives
Participants will not be offered any incentives to participate in the research.

Dissemination
Participants will receive a summary about the overall results of the study when it is completed. No individual level information will be provided.

Please confirm that the processing of the data is not likely to cause substantial damage or distress to an individual Yes ☒

Section 9 Attachments Please attach the following items to this form, or explain if not attached

a.	Information sheets, consent forms and other materials to be used to inform potential participants about the research *(List attachments below)*	Yes ☒	No ☐
	If applicable/appropriate:		
b.	Approval letter from external Research Ethics Committee		Yes ☐
c.	The proposal ('case for support') for the project		Yes ☐
d.	Full risk assessment		Yes ☐

Section 10 Declaration

I confirm that to the best of my knowledge the information in this form is correct and that this is a full description of the ethical issues that may arise in the course of this project.

I have discussed the ethical issues relating to my research with my supervisor. ☒ ☐

I have attended the appropriate ethics training provided by my course. ☒ ☐

I confirm that to the best of my knowledge:

The above information is correct and that this is a full description of the ethics issues that may arise in the course of this project.

Name	Samuel Barclay
Date	12/10/2018

Departmental use

If a project raises particularly challenging ethics issues, or a more detailed review would be appropriate, the supervisor **must** refer the application to the Research Development Administrator (via ioe.researchethics@ucl.ac.uk so that it can be submitted to the IOE Research Ethics Committee for consideration. A departmental research ethics coordinator or representative can advise you, either to support your review process, or help decide whether an application should be referred to the REC. If unsure please refer to the guidelines explaining when to refer the ethics application to the IOE Research Ethics Committee, posted on the committee's website.

Student name	Samuel Barclay	
Student department	Culture, Communication, and Media	
Course	PhD	
Project title	The Effect of Intralexical Factors on the Decay of Foreign Language Vocabulary Knowledge	
Reviewer 1		
Supervisor/first reviewer name	Ana Pellicer-Sánchez	
Do you foresee any ethical difficulties with this research?	I believe all the ethical issues have been appropriately addressed in this form.	
Supervisor/first reviewer signature	Signature removed	
Date	19/11/2018	
Reviewer 2		
Second reviewer name	Andrea Revesz	
Do you foresee any ethical difficulties with this research?	I cannot foresee any ethical issues other than the ones addressed in the application.	
Supervisor/second reviewer signature	Signature removed	
Date	23 Nov 2018	
Decision on behalf of reviews		
Decision	Approved	☒
	Approved subject to the following additional measures	☐

	Not approved for the reasons given below	☐
	Referred to REC for review	☐
Points to be noted by other reviewers and in report to REC		
Comments from reviewers for the applicant		

Once it is approved by both reviewers, students should submit their ethics application form to the Centre for Doctoral Education team: IOE.CDE@ucl.ac.uk.

Appendix 17

Language Background Questionnaire (Study 3)

Background information

Please provide the following personal details and language background information. If you prefer not to give an answer to a question, please leave it blank.

1. Are you male or female? ______________________
2. How old are you? ______________________
3. What is your native language? ______________________
4. What course do you study/plan to study? ______________________
5. How long have you studied English? __________
6. At what age did you first contact English? ______________________
7. How long (in total) have you lived in an English-speaking country? __________
8. What other languages do you speak?
9. How often do you used electronic flashcards to learn English vocabulary

   never - rarely - sometimes - quite often - often - always

Please now provide self-ratings for your level of proficiency in English:

1 = extremely poor, almost no knowledge → 10 = extremely good, almost native like.

Reading:	1	2	3	4	5	6	7	8	9	10
Writing:	1	2	3	4	5	6	7	8	9	10
Listening:	1	2	3	4	5	6	7	8	9	10
Speaking:	1	2	3	4	5	6	7	8	9	10

If you have any other comments, please write them in the box below.

Thank you so much for your help

Appendix 18

Llama Instruction Document (Study 3)

LLAMA B

What?

This test measures your memory and how quickly you can connect words with meaning.

How?

You will see a number of shapes. You need to learn the names of these shapes in 2 minutes. There is a test after you learn the names.

1. Write your full name in the blue boxes at the top left.
2. Do not change the numbers – 2003 and 120.
3. When you are ready click the arrow. This starts the timer and activates the test. You can now click on an image and see the name in the centre of the screen. **YOUR TASK**: learn the names of as many of the twenty objects as you can in two minutes. Do not take notes.
4. After the timer in the centre runs out, the testing phase will begin. Press the ⇨ button to see the name of an image. You need to click on the image that matches the name.
5. After clicking on an image press ⇨ to see the next name.
6. When you finish, your score is displayed.
7. To close the programme click

Interpreting your score

0-20 – poor score
25-45 – average score
50-70 – above average score
75-100 – outstanding score.

LLAMA D

What?

This measures how well you can remember sounds

How?

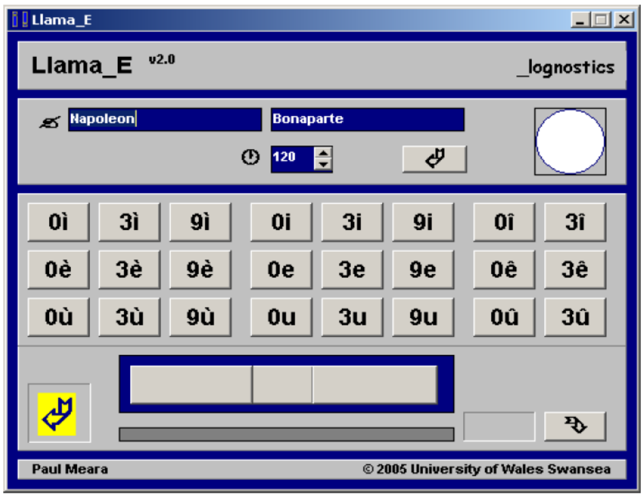
You hear a number of sounds from a language you will not know. You have to decide whether some of these are repeated.

1. Write your full name in the blue boxes at the top left. Click ⏎ to begin.
2. You will hear 10 words in an unfamiliar language. Listen carefully.
3. When the ⇨ is activated, click it to hear the next word. If you have heard a word before, click ☺, if you have not heard a word before click 😐.
4. When you have clicked on a face, click ⇨ to hear the next word.
5. When the test is finished, you can see your results.
6. To close the programme click ⇲

Interpreting your score

0-10 – poor score
25-35 – average score
40-60 – above average score
75-100 – outstanding score.

LLAMA E



What?

This tests how well you can connect sounds to letters.

How?

You will see letters and hear sound that are unfamiliar. You need to learn the pronunciation of each letter.

1. Write your full name in the blue boxes at the top left.
2. Do not change the numbers – 2003 and 120.
3. When you are ready click the arrow . This starts the timer and activates the test.
4. Click on one of the buttons to hear the pronunciation of that text. Your task is to learn the pronunciation of each button. You have two minutes to learn.
5. After two minutes, click ⇨ to start the test. Each time you click ⇨ you will hear a word. At the same time, two possible spellings for this word are presented. Choose the correct spelling.
6. There are twenty questions on the test.
7. When you have finished, your score will be displayed.
8. To close the programme click

Interpreting your score

0-15 – poor score
20-45 – average score
45-70 – above average score
75-100 – outstanding score.

Appendix 19

Model Selection Procedure for Learning Burden Analysis (Frequency and Length of Exposure) (Study 3)

Summary of Mixed -Effect Model Comparisons for Frequency of Exposure Study 3 (Strict condition)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Additional language; Familiarity with flash cards; Group; L1; LLAMA B; LLAMA D; LLAMA E; Mode of form presentation; PoS; vocabulary size; Word length; Mode*LLAMA B; Mode*LLAMA D; Mode*LLAMA E	By-participant + by-item random intercepts.	423	NA	NA	NA
2	Model 1 – Flash card familiarity	Same as Model 1	427	4	X^2 (1) = 5.82	*P* = .016*
3	Model 1 – Additional language	Same as Model 1	421	-2	X^2 (1) = .20	*P* = .66
4	Model 3 – Group	Same as Model 1	419	-2	X^2 (1) <.001	*P* > .99
5	Model 4 – PoS	Same as Model 1	425	6	X^2 (1) = 7.66	*P* = .006**
6	Model 4 - Length	Same as Model 1	446	27	X^2 (1) = 28.70	*P*< .001***
7	Model 4 - L1	Same as Model 1	418	-1	X^2 (1) = 5.10	*P* = .07
8	Model 7 - Vocabulary Size	Same as Model 1	430	10	X^2 (1) = 12.10	*P* <.001***

Summary of Mixed -Effect Model Comparisons for Time of Exposure Study 3 (Strict condition)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Additional language; Familiarity with flash cards; Group; L1; Llama B; Llama D; Llama E; Mode of form presentation; PoS; vocabulary size; Word length; Mode*Llama B; Mode*Llama D; Mode*Llama E	By-participant + by-item random intercepts.	1255	NA	NA	NA
2	Model 1 – Flash card familiarity	Same as Model 1	1254	-1	X^2 (1) = .92	*P* = .34
3	Model 2 – Additional language	Same as Model 1	1252	-2	X^2 (1) = .02	*P* = .88
4	Model 3 – Group	Same as Model 1	1250	-2	X^2 (1) =.22	*P* = .64
5	Model 4 – PoS	Same as Model 1	1249	-1	X^2 (1) = 1.35	*P* = .24
6	Model 5 - Length	Same as Model 1	1284	35	X^2 (1) = 36.70	*P*< .001***
7	Model 5 - L1	Same as Model 1	1253	4	X^2 (1) = 7.40	*P* = .02*
8	Model 5 - Vocabulary Size	Same as Model 1	1267	18	X^2 (1) = 19.80	*P* <.001***

Summary of Mixed -Effect Model Comparisons for Frequency of Exposure Study 3 (Lenient condition)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Additional language; Familiarity with flash cards; Group; L1; Llama B; Llama D; Llama E; Mode of form presentation; PoS; vocabulary size; Word length; Mode*Llama B; Mode*Llama D; Mode*Llama E	By-participant + by-item random intercepts.	500	NA	NA	NA
2	Model 1 – Flash card familiarity	Same as Model 1	503	3	X^2 (1) = 5.24	*P* = .02*
3	Model 1 – Additional language	Same as Model 1	498	-2	X^2 (1) = .12	*P* = .73
4	Model 3 – Group	Same as Model 1	496	-2	X^2 (1) =.01	*P* = .94
5	Model 4 – PoS	Same as Model 1	500	4	X^2 (1) = 6.45	*P* = .01*
6	Model 4 - Length	Same as Model 1	523	27	X^2 (1) = 29.50	*P*< .001***
7	Model 4 - L1	Same as Model 1	497	1	X^2 (1) = 5.23	*P* = .07
8	Model 7 - Vocabulary Size	Same as Model 1	508	11	X^2 (1) = 13.30	*P* <.001***

Summary of Mixed -Effect Model Comparisons for Time of Exposure Study 3 (Lenient condition)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Additional language; Familiarity with flash cards; Group; L1; Llama B; Llama D; Llama E; Mode of form presentation; PoS; vocabulary size; Word length; Mode*Llama B; Mode*Llama D; Mode*Llama E	By-participant + by-item random intercepts.	1365	NA	NA	NA
2	Model 1 – Flash card familiarity	Same as Model 1	1365	0	X^2 (1) = 1.11	*P* = .29
3	Model 2 – Additional language	Same as Model 1	1363	-2	X^2 (1) = .01	*P* = .91
4	Model 3 – Group	Same as Model 1	1361	-2	X^2 (1) =.16	*P* = .69
5	Model 4 – PoS	Same as Model 1	1360	-1	X^2 (1) = 1.08	*P* = .30
6	Model 5 - Length	Same as Model 1	1396	36	X^2 (1) = 37.70	*P*< .001***
7	Model 5 - L1	Same as Model 1	1363	3	X^2 (1) = 7.66	*P* = .02*
8	Model 5 - Vocabulary Size	Same as Model 1	1378	18	X^2 (1) = 19.80	*P* <.001***

Summary of Mixed -Effect Model Comparisons for Frequency of Exposure Study 3 (Recognition condition)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Additional language; Familiarity with flash cards; Group; L1; Llama B; Llama D; Llama E; Mode of form presentation; PoS; vocabulary size; Word length; Mode*Llama B; Mode*Llama D; Mode*Llama E	By-participant + by-item random intercepts.	734	NA	NA	NA
2	Model 1 – Flash card familiarity	Same as Model 1	734	0	X^2 (1) = 2.13	*P* = .14
3	Model 2 – Additional language	Same as Model 1	733	-1	X^2 (1) = .47	*P* = .49
4	Model 3 – Group	Same as Model 1	731	-2	X^2 (1) =.07	*P* = .79
5	Model 4 – PoS	Same as Model 1	735	4	X^2 (1) = 5.89	*P* = .02*
6	Model 4 - Length	Same as Model 1	756	25	X^2 (1) = 27.20	*P*< .001***
7	Model 4 - L1	Same as Model 1	731	0	X^2 (1) = 3.70	*P* = .16
8	Model 7 - Vocabulary Size	Same as Model 1	738	7	X^2 (1) = 19.34	*P* <.002**

Summary of Mixed -Effect Model Comparisons for Time of Exposure Study 3 (Recognition condition)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Additional language; Familiarity with flash cards; Group; L1; Llama B; Llama D; Llama E; Mode of form presentation; PoS; vocabulary size; Word length; Mode*Llama B; Mode*Llama D; Mode*Llama E	By-participant + by-item random intercepts.	1769	NA	NA	NA
2	Model 1 – Flash card familiarity	Same as Model 1	1767	-2	X^2 (1) = .11	*P* = .74
3	Model 2 – Additional language	Same as Model 1	1766	-1	X^2 (1) = .41	*P* = .52
4	Model 3 – Group	Same as Model 1	1765	-1	X^2 (1) =.87	*P* = .35
5	Model 4 – PoS	Same as Model 1	1764	-1	X^2 (1) = .97	*P* = .32
6	Model 5 - Length	Same as Model 1	1797	33	X^2 (1) = 35.50	*P*< .001***
7	Model 5 - L1	Same as Model 1	1768	4	X^2 (1) = 8.41	*P* = .02*
8	Model 5 - Vocabulary Size	Same as Model 1	1781	27	X^2 (1) = 18.90	*P* <.001***

Appendix 20

Model Selection Procedure for Decay Analysis (Study 3)

Summary of Generalised Linear Mixed -Effect Model Comparisons for Decay of Learned Items Study 3 (strict condition)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Additional language; Familiarity with flash cards; Group; L1; Llama B; Llama D; Llama E; Mode of form presentation; PoS; vocabulary size; Word length; Mode*Llama B; Mode*Llama D; Mode*Llama E; Frequency of exposure; Time of exposure	By-participant + by-item random intercepts.	1339	NA	NA	NA
2	Model 1 – Flash card familiarity	Same as Model 1	1337	-2	X^2 (1) = .31	*P* = .58
3	Model 2 – Additional language	Same as Model 1	1338	1	X^2 (1) = 2.43	*P* = .12
4	Model 3 – Group	Same as Model 1	1336	-2	X^2 (1) =.28	*P* = .60
5	Model 4 – PoS	Same as Model 1	1340	4	X^2 (1) = 6.43	*P* = .01*
6	Model 4 - Length	Same as Model 1	1334	-2	X^2 (1) = .02	*P*= .90
7	Model 6 - L1	Same as Model 1	1354	21	X^2 (1) = 24.4	*P* < .001***
8	Model 6 - Vocabulary Size	Same as Model 1	1335	1	X^2 (1) = 3.46	*P* =.06

Summary of Generalised Linear Mixed -Effect Model Comparisons for Decay of Learned Items Study 3 (lenient condition)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Additional language; Familiarity with flash cards; Group; L1; Llama B; Llama D; Llama E; Mode of form presentation; PoS; vocabulary size; Word length; Mode*Llama B; Mode*Llama D; Mode*Llama E; Frequency of exposure; Time of exposure	By-participant + by-item random intercepts.	1394	NA	NA	NA
2	Model 1 – Flash card familiarity	Same as Model 1	1392	-2	X^2 (1) = .36	*P* = .55
3	Model 2 – Additional language	Same as Model 1	1392	0	X^2 (1) = 1.87	*P* = .17
4	Model 3 – Group	Same as Model 1	1390	-2	X^2 (1) =.46	*P* = .50
5	Model 4 – PoS	Same as Model 1	1394	4	X^2 (1) = 6.10	*P* = .01*
6	Model 4 - Length	Same as Model 1	1389	-1	X^2 (1) = .13	*P*= .72
7	Model 6 - L1	Same as Model 1	1411	22	X^2 (1) = 26.2	*P* < .001***
8	Model 6 - Vocabulary Size	Same as Model 1	1389	0	X^2 (1) = 2.91	*P* =.08

Summary of Generalised Linear Mixed -Effect Model Comparisons for Decay of Learned Items Study 3 (recognition condition)

Model description					Test against prior model	
Model	Fixed effects	Random effects	AIC	ΔAIC	Statistic	Significance
1	Additional language; Familiarity with flash cards; Group; L1; Llama B; Llama D; Llama E; Mode of form presentation; PoS; vocabulary size; Word length; Mode*Llama B; Mode*Llama D; Mode*Llama E; Frequency of exposure; Time of exposure	By-participant + by-item random intercepts.	1507	NA	NA	NA
2	Model 1 – Flash card familiarity	Same as Model 1	1506	-1	$X^2 (1) = .62$	*P* = .43
3	Model 2 – Additional language	Same as Model 1	1505	-1	$X^2 (1) = .95$	*P* = .33
4	Model 3 – Group	Same as Model 1	1505	0	$X^2 (1) = 1.82$	*P* = .18
5	Model 4 – PoS	Same as Model 1	1510	5	$X^2 (1) = 7.18$	*P* = .007**
6	Model 4 - Length	Same as Model 1	1503	-2	$X^2 (1) = .72$	*P*= .40
7	Model 6 - L1	Same as Model 1	1520	17	$X^2 (1) = 24.00$	*P* < .001***
8	Model 6 - Vocabulary Size	Same as Model 1	1505	2	$X^2 (1) = 3.24$	*P* =.07