RETRIEVAL INDUCED FORGETTING AND ADULT SECOND LANGUAGE VOCABULARY ACQUISITION: PRELIMINARY INSIGHTS FROM A WELSH LANGUAGE TRAINING STUDY

by

Lyam Bailey

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
July 2017

TABLE OF CONTENTS

LIST OF FIGURES iv

ABSTRACT v

LIST OF ABBREVIATIONS USED vi

ACKNOWLEDGEMENTS vii

CHAPTER 1: INTRODUCTION 1

1.1 ADULT SECOND LANGUAGE LEARNING 1

1.1.1 Preface 1

1.1.2 Second Language Vocabulary Acquisition In Adults: A Challenging Endeavour 1

1.1.3 Native Language Influences In Second Language Vocabulary Acquisition 2

1.1.4 Vocabulary Acquisition And Native Language Influence: Moving Forward 5

1.2 RETRIEVAL INDUCED FORGETTING 5

1.2.1 Introduction To Retrieval Induced Forgetting 5

1.2.2 Modulating Factors In Retrieval Induced Forgetting 7

1.2.3 Competing Theoretical Accounts Of Retrieval Induced Forgetting 9

1.2.4 ERP Correlates Of Retrieval Induced Forgetting 11

1.2.5 Retrieval Induced Forgetting And Language Learning 14

1.3 THE CURRENT INVESTIGATION 15

1.3.1 The Research Question 15

1.3.2 Our Approach 16

1.3.3 Behavioural And Electrophysiological Hypotheses 17

CHAPTER 2: METHODOLOGY 19

2.1 SUBJECTS 19

2.2 STIMULI AND APPARATUS 19

2.3 PROCEDURE 20

2.4 BEHAVIOURAL DATA COLLECTION 22

2.5 BEHAVIOURAL DATA ANALYSIS 22

2.6 ELECTROPHYSIOLOGICAL DATA COLLECTION 23

2.7 ELECTROPHYSIOLOGICAL DATA ANALYSIS 24

CHAPTER 3: RESULTS......25
3.1 BEHAVIOURAL DATA......25
3.2 ERP DATA......29
3.3 BEHAVIOR X ERP INTERACTION......32
CHAPTER 4: DISCUSSION......34
4.1 BEHAVIOURAL FINDINGS......34
4.2 ERP FINDINGS......38
4.3 CONCLUSIONS AND FUTURE RECOMMENDATIONS......41
REFERENCES......44
APPENDIX A......50
LIST OF WELSH WORDS USED IN THE CURRENT INVESTIGATION WITH THEIR RESPECTIVE ENGLISH TTRANSLATIONS
APPENDIX B......51
EXAMPLES OF PICTURE STIMULI USED IN THE CURRENT INVESTIGATION, FOR PICTURE SET 1 (USED IN WELSH LEARNING PHASE), SET 2 (USED IN RETRIEVAL-PRACTICE PHASE) AND SET 3 (USED IN WELSH TEST PHASE)

LIST OF FIGURES

Figure 1: ERP waveforms at an anterior recording site, time-locked to the onset of category-exepmplar word stems or words pairs, in the retrieval-practice or relearning condition respectively. Taken from Johansson et al. (2007) 13

Figure 2: Frequency distributions for raw RT scores (left panel), and log-transformed RT scores (right panel). 25

Figure 3: Mean error rates for the retrieval-practice X phase interaction. Error bars represent 95% CIs. ** indicates $p \leq .001$, following Bonferroni correction. 26

Figure 4: Mean RT values for the retrieval-practice X phase interaction. Error bars represent 95% CIs. All p values were derived from post-hoc Bonferroni comparisons. * indicates $p \leq .01$, ** indicates $p \leq .001$ 27

Figure 5: Regression slope for main effect of L1 frequency on RT, collapsing across phase and retrieval-practice variables. Error margins represent 95% CIs. 28

Figure 6: Regression slope for the retrieval-practice X L1 frequency interaction, plotted separately for Baseline and Test conditions. Values represent RTs to RpE+ items minus RTs to RpE- items, computed for each individual word (i.e. the magnitude of the retrieval-practice effect). The *p* values (derived from Bonferroni-corrected post-hoc comparisons) pertain to Test condition and Baseline condition, respectively. Error margins represent 95% CIs. 28

Figure 7: Scalp topographic plots for the ERP difference between RpE+ and control items during retrieval-practice at 200 ms intervals. Dots represent electrode locations 29

Figure 8: ERP waveforms time-locked to the onset of RpE+ and control items during retrieval-practice, across representative electrodes from each of the 9 ROIs used in the current investigation. **A:** Left anterior (AF7), **B:** Medial anterior (Pz), **C:** Right anterior (F6), **D:** Left central (C3), **E:** Medial central (Cz), **F:** Right central (C4), **G:** Left posterior (P7), **H:** Medial posterior (Pz), **I:** Right posterior (P6). Note that by convention, electrical potential on the y-axis is plotted negative-up . .31

Figure 9: Mean response amplitudes for RpE+ minus control items during retrieval-practice for Anterior (Ant), Central (Cent) and Posterior (Post) scalp surfaces. Results are presented for the early time window (left panel), and the late time window (right panel). Error bars represent 95% CIs. 32

Figure 10: Regression slopes for the relationship between ERP difference wave (computed by subtracting EEG responses to RpE+ items from those of control items, during retrieval practice) and RT difference at test (computed by subtracting RTs to RpE+ items from RTs to RpE- items), across multiple ROIs. Lines represent left (L), medial (M) and right (R) divisions of anterior (Ant), medial (Med) and posterior (Post) scalp areas. Error bars represent 95% confidence intervals. Data is presented for both the early time window (top panel) and the late time window (lower panel). 33

ABSTRACT

The current investigation explored the role of retrieval-induced forgetting (RIF), a process whereby retrieval of information from memory perturbs access to related material, in second language (L2) vocabulary acquisition. Prior research has indicated that retrieving words in one language may cause difficulties remembering equivalent words in another language. Here, subjects learned novel Welsh words, half of which were then retrieved in English. EEG was recorded throughout the experiment with a view to assess whether ERP components previously associated with RIF were present. Behavioural results showed that, at a final test of Welsh knowledge, subjects exhibited significantly longer RTs to items that had been retrieved in English, compared to those which had not. This is taken as preliminary evidence for a role of RIF during L2 acquisition. ERP components previously associated with RIF were not elicited, likely due to differences in experimental design between previous research and the current investigation.

LIST OF ABBREVIATIONS USED

Abbreviation	Definition
ERP	Event related potential
RpE-	Items not retrieved in English during the current investigation
RpE+	Items retrieved in English during the current investigation
L1	Native language
RT	Reaction time
RIF	Retrieval induced forgetting
L2	Second language
Rp-	Unpracticed items belonging to a practiced category
Nrp	Unpracticed items belonging to an unpracticed category

ACKNOWLEDGEMENTS

My thanks to Aaron Newman for his advice and guidance throughout the course of this research project. Additional thanks to Mike Lawrence for his assistance with data analysis, and to Tracy Taylor-Helmick and Raymond Klein for their advice on data interpretation and further developments on the work presented in this thesis. To Laura, I cannot thank you enough for your support throughout the past two years – for your encouragement, for listening to me babble endlessly about variable interactions, and for enduring seemingly countless iterations of pilot experiments. I couldn't have done it all without you.

CHAPTER 1: INTRODUCTION

1.1 ADULT SECOND LANGUAGE LEARNING

1.1.1 Preface

Adult second language learning is an enormous field - with dozens of books, hundreds of research articles, and even a number of Doctoral dissertations written on the topic to date. While this topic is indeed central to the current investigation, it would not be possible to provide a comprehensive overview of the *entire* field within the relatively small scope of this manuscript. As such, the following section provides an overview of those aspects of L2 acquisition which relate directly to the research question addressed by the current investigation.

1.1.2 Second Language Vocabulary Acquisition In Adults: A Challenging Endeavour

Second language (L2) learning is an extremely challenging task for most adults. While a great deal of controversy surrounds the relative abilities of children versus adults to acquire new language, it is generally agreed that adult learners are at a severe disadvantage. This is perhaps most clearly demonstrated by the fact that, despite investing huge amounts of time, effort and (very often) financial resources, most adult L2 learners still fail to achieve native-like proficiency.

Vocabulary acquisition represents an early and crucial foundation for successful language learning, and even in this adults struggle to achieve native-like levels of proficiency. This is perhaps best illustrated by Hellman (2008, 2011) who reports that only around 75% of adults immersed in an L2 for a *minimum of ten years* achieved scores on tests of vocabulary size equivalent to those of native speakers. A large review of research on formal L2 tuition by Schmitt (2008) estimates that a vocabulary size of at least 8,000 word families (equal to approximately 34,000 individual words; Nation; 2006) is necessary to achieve functionality in English written discourse. It is easy to see how such a goal would be daunting for L2 learners – indeed, Schmitt cites estimates of between 400 and 1,000 hours of formal English tuition necessary to attain an English vocabulary of just 1,000 word families. In order to attain 4,000 word families – only half of Schmitt's estimated 'necessary minimum' – some estimates call for over 2,000 hours of formal tuition (see Schmitt, 2008; Table 1). It seems that even this basic component of language acquisition poses a daunting challenge for many adults.

While relatively little effort has been made to understand factors that may *impair* vocabulary acquisition - one of which is the focus of the current investigation - facilitative factors such as age of acquisition, amount of exposure, and motivation (amongst others) have received an enormous amount of empirical attention in recent decades. Given the focus of the current investigation on native language (L1) use as a barrier to L2 vocabulary acquisition, the issue of L1 influence warrants special attention.

1.1.3 Native Language Influences In Second Language Vocabulary Acquisition

The question of how one's native language may affect the L2 vocabulary learning process has been the focus of numerous investigations in recent decades. The results of such widespread enquiry, it appears, remain decidedly mixed. While L1 knowledge appears to be beneficial to L2 learning in some contexts, it acts as a source of interference in others (Schmitt, 2008). For the most part, however, L1 knowledge seems to be largely facilitative during the *early* stages of L2 vocabulary learning.

A number of authors have suggested that L1 knowledge is in fact *essential* for initial L2 vocabulary learning. Hall (2006) describes a "parasitic strategy" model of vocabulary learning, whereby initial form-meaning mapping of novel words is entirely dependent on L1 knowledge. The hypothesis is, at its core, very intuitive - in plain terms the learner must utilize their L1 representation of a given word in order to learn the meaning of the new L2 form. To provide a functional example, if one were presented with the Welsh word for ball – *pel* – along with its English translation, one would immediately associate *pel* with the English word *ball*, and all semantic content that *ball* conferred. Similarly, the Revised Hierarchical Model (RHM; Kroll, Hell, Tokowicz, &

Green; 2010) proposes that while L2 learners / bilingual speakers hold independent lexical representations of a word in each language, the semantic content of both representations is based on a shared conceptual system. Moreover, Kroll et al. (2010) emphasize that during the early stages of L2 vocabulary acquisition, L1 form-meaning associations are stronger than those of L2. Because of this, newly acquired L2 vocabulary must be associated with L1 representations in order to be accessed - in simple terms, accessing the semantic meaning of an L2 word is necessarily mediated by direct L1-L2 translation.

Despite the facilitative effects of L1 use on L2 form-meaning mapping described above, it seems that such 'linguistic parasitism' may also impose certain constraints on one's ability to acquire or access L2 vocabulary. Such constraints may be loosely characterized by the colloquial expression "old habits die hard". In plain terms, it appears that over-dependence on L1 knowledge may cause difficulties in assimilating new lexical rules, or errors resulting from assumptions of equivalent translation between L1 and L2. For example, both de Groot, (2006) and Hamada & Koda (2008) report that low-proficiency learners were readily able to learn novel L2 words which followed phonological and orthographical patterns similar to that of their L1. By contrast, learning words with L1-atypical patterns presented a significant challenge, evidenced by reduced scores on tests of L2 word retrieval. In a different vein, Hemchua and Schmitt (2006) analyzed the content of English essays written by L1 Thai students, and report that the majority of errors made were attributable to L1 influence. For example, a common error concerned semantic content of a sentence – whereby a literal translation of an appropriate Thai sentence resulted in an inappropriately constructed English sentence. For example: *one advantage of living in a city is many <u>comfortable things</u> such as buses, trains, stores, etc.* (to provide some context, the authors suggest that '*comfortable things'* was translated from Thai words equivalent to '*facilities'*, hence the miscommunication). It is important to emphasize here that these findings are still consistent with the parasitic strategy hypothesis. Clearly, learners in Hemchua and Schmitt's (2006) study were depending on (albeit false) assumptions of equivalent translation between L1 representations and L2 forms, resulting in L2 semantic errors that would be perfectly acceptable in their L1.

It is worth mentioning that some work has also investigated modulating effects of L1 word features on corresponding L2 word learning success (note the focus here on *features of L1 words*, as opposed to relationships between L1 and L2 forms discussed above). A handful of studies carried out by de Groot and colleagues (Lotto & de Groot, 1998; De Groot & Keijzer, 2000; de Groot, 2006) investigated the influence of L1 word frequency on acquisition of corresponding L2 words. These studies report a *marginal* advantage for L2 forms of high-frequency, compared to low-frequency, L1 items. However, these effects were consistently small in magnitude and often lacked statistical significance. By contrast, L1 concreteness represents a much more influential modulating factor. A number of studies have consistently reported significantly higher rates of vocabulary acquisition success when learning novel words with concrete L1 referents, compared to those with abstract L1 referents (Lotto & de Groot, 1998 (De Groot & Keijzer, 2000; more recently Kaushanskaya & Rechtzigel, 2012; Palmer, Macgregor, & Havelka, 2013).

Overall, it seems that L1 knowledge can exert a variety of effects on L2 acquisition success. Having said this, one consideration which has received surprisingly little empirical attention is exposure to and/or explicit use of L1 as a source of interference during vocabulary acquisition. In this vein, Guekes and Zwisterlood (2016; Experiment 2) recently investigated the influence of L1 exposure on participants' ability to recall newly acquired pseudowords associated with common objects. Following an initial learning period, participants were required to name pictures in their newly-learned 'pseudolanguage', while naming latencies were recorded. In one condition, each trial began with an auditory distractor word which was identical to the to-be-named picture. Crucially, distractors were presented either in participants' native language, or in the pseudolangauge. The authors report that while identical pseudoword distractors facilitated pseudoword naming of pictures, presentation of identical L1 words showed no significant effect on pseudoword naming latency (relative to a baseline condition in which distractors were unrelated L1 words). These results indicate that L1 *exposure* (note the distinction from L1 *use*) has little impact on one's ability to access newly learned words for common objects.

1.1.4 Vocabulary Acquisition And Native Language Influence: Moving Forward

To reiterate the discussion above, it appears that L1 influences exert a variety of effects on new vocabulary acquisition – many of which perturb the learner's ability to form strong L2 representations. Despite the presence of many such deleterious influences, mere exposure to L1 referents does not appear to damage one's memory for corresponding L2 forms. However, the question of whether explicit L1 *use* may impair access to recently acquired L2 vocabulary remains to be investigated. To explore this issue, we turn our attention to an exciting and widely documented empirical phenomenon from studies of long-term memory: a phenomenon somewhat quizzically termed *retrieval induced forgetting*.

1.2 RETRIEVAL INDUCED FORGETTING

1.2.1 Introduction To Retrieval Induced Forgetting

Retrieval induced forgetting (RIF) refers to a phenomenon whereby retrieval of items from memory perturbs access to semantically associated information. In the traditional RIF paradigm (typically referred to as the *retrieval-practice* paradigm) participants are required to study lists of category-exemplar pairs, such as *fruits-orange; fruits-apple*; *fruits-kiwi, drinks-wine*; *drinks-beer* and so on. Following this, participants are guided to retrieve a subset of exemplars from one of the studied categories (for example, *fruits-orange*, but not *fruits-apple* or *fruits-kiwi*). In the traditional version of this paradigm retrieval is typically guided by means of a word-stem completion task, containing a category cue paired with a word-stem representing the intended exemplar - for example *fruits-or*___ for *orange*. Finally, participants' ability to recall all of the learned exemplars is tested.

The retrieval-practice paradigm generates three types of stimuli at test: typically denoted as Nrp, Rp-, and Rp+. Nrp refers to unpractised items belonging to an unpractised category (in our example above, *drinks-beer* and *drinks-wine*). Rp- refers to unpractised items belonging to a practised category (*fruits-apple* and *fruits-kiwi*); while Rp+ refers to practiced items (*fruits-orange*). Unsurprisingly Rp+ items are usually better recalled than Nrp or Rp-, and this is attributed to facilitation of these items due to retrieval practice. Curiously (and indeed, crucially), a consistent finding in such experiments is that participants show significantly worse memory for Rp- items compared to all others. It is critical to emphasize that Rp- items are recalled less well than Nrp, despite both types of item having not been practiced. In simple terms, it appears that retrieving items from a given category impairs participants' memory for other items in that category to a greater degree than 'passive' memory decay can account for. This effect, first coined 'retrieval induced forgetting' by Anderson and colleagues (Anderson, Bjork, & Bjork, 1994) has attracted a great deal of attention in cognitive science since its inception - with close to 200 related articles having been published by 2014 (Murayama, Miyatsu, Buchli, & Storm, 2014).

An important feature of RIF is that it is not confined to memory for written category-exemplar pairs. In the two decades since Anderson and colleagues' seminal paper (Anderson et al., 1994), a huge amount of research has explored the generalisability of RIF to a growing range of contexts. In brief, variants on the traditional retrieval-practice paradigm have elicited RIF effects using pictures (Ciranni & Shimamura, 1999; Ford, Keating, & Patel, 2004), visual scenes (Shaw, Bjork, & Handal, 1995), event narratives (MacLeod, 2002; Migueles & Garcia-Bajos, 2006), factual propositions (M. C. Anderson & Bell, 2001), mathematical problems (Phenix & Campbell, 2004), and even a logical problem-solving task (Storm et al., 2011). RIF has also been implicated in aspects of social cognition such as trait assignment to fictional characters (Macrae & Macleod, 1999; Dunn & Spellman, 2003) and social decision making (Iglesias-Parro & Gómez-Ariza, 2006). Moreover, RIF effects have been observed in a range of experimental contexts, such as socially shared retrieval (Coman, Manier, & Hirst, 2009; Stone et al., 2010), eyewitness testimony scenarios (MacLeod, 2002; Pica, Pierro, Bélanger, & Kruglanski, 2014) and language processing (Levy, McVeigh, Marful, & Anderson, 2007; but see section 1.2.5).

1.2.2 Modulating Factors In Retrieval Induced Forgetting

A great deal of research has been dedicated to understanding factors that modulate the presence and/or magnitude of RIF. While few such factors can be applied to *all* RIF research – that is, given the high degree of variation in stimulus type used and the type of memory being investigated – the literature does contain some broadly generalizable principles.

One such principle (which is somewhat obvious but nevertheless must be emphasised) is the necessity of *active retrieval-practice* between study and test. A number of studies employing 'mere exposure' conditions – whereby participants either passively view (Guekes & Zwisterlood, 2016; Verde, 2013) or 're-learn' (Johansson, Aslan, Bäuml, Gäbel, & Mecklinger, 2007; Spitzer, Hanslmayr, Opitz, Mecklinger, & Bäuml, 2008) a subset of previously studied items – have demonstrated that such conditions are *not* sufficient to elicit forgetting of Rp- items relative to Nrp baseline. Instead, it appears that RIF (as the name suggests) is a process specific to active retrieval alone. As an aside, there is evidence that *successful* retrieval-practice is not actually necessary to cause forgetting. Storm et al. (2006) developed a variant on the traditional retrieval-practice paradigm, whereby in one condition participants were posed with an 'impossible' retrieval task - specifically, during retrieval-practice they were provided with cues for non-existent exemplars of studied categories, such as *Fruit-Te*___. The authors report that participants exhibited near-identical rates of perturbation of memory for *all* non-practiced items within practiced categories (i.e. Rp-) relative to non-practiced item (Nrp) baseline, regardless of whether participants had engaged in 'possible' or 'impossible' retrieval-practice. In plain terms, it seems that RIF is not dependent on whether participants are actually able to recall Rp+ items during retrieval-practice – only that retrieval-practice of the relevant category occurs at all.

A second factor that has received considerable attention in the RIF literature is exemplar (or item) strength. It is worth mentioning at this point that virtually all research on this particular characteristic of RIF has used written category-exemplar pairs, whereby word frequency is generally taken as a proxy for strength of internal representation (this has also been phrased as "ease of accessibility"; (Anderson et al., 1994)). Based on this principle, some of the earliest research on RIF (Anderson et al., 1994; Experiments 2 & 3) distinguished 'strong' category exemplars (for example, *Fruits-Orange* and *Fruits-Apple*) from 'weak' category exemplars (for example, *Fruits-Kiwi* and *Fruits-Papaya*). This early work showed that strong exemplars were more susceptible to RIF than their weak counterparts, evidenced by significantly higher rates of forgetting amongst strong Rp- items compared to weak Rp- items. This finding has since been replicated in a number of studies (Bäuml, 1998; Hellerstedt & Johansson, 2014; Storm, Bjork, Bjork, & Nestojko, 2006; Williams & Zacks, 2001), although the theoretical implications remain vigorously debated (see section 1.2.3). Nevertheless, these findings have informed a great deal of subsequent research, most of which now strives to confine stimulus sets to strong exemplars, thus ensuring strong RIF effects at test (Murayama et al., 2014).

Contrary to the discussion above, not *all* published data has confirmed an effect of exemplar strength. In a deviation from traditional dependence on word frequency as a proxy for internal representation, Jakab & Raaijmakers, 2009 attempted to manipulate item strength by varying the frequency at which items appeared during the experiment's study phase. Contrary to previous work the authors report no difference in memory perturbation for strong vs. weak Rp- items, relative to Nrp baseline. The exact meaning of this finding is debatable – given its conflict with previous research, and the question of whether increased frequency of an item during a study phase truly represents a manipulation of its 'strength'. Regardless, the majority of published data seems to support the hypothesis that strong exemplars are more susceptible to RIF, and so for the purposes of the current investigation we will consider this to be accurate.

1.2.3 Competing Theoretical Accounts Of Retrieval Induced Forgetting

While RIF is a robust and widely replicated empirical phenomenon, there is some controversy surrounding its underlying cognitive mechanism. Something which is generally agreed upon is that when one attempts to retrieve an item belonging to a particular category from long-term memory (i.e. a target item), semantically associated items within the same category (i.e. non-target items) also become activated. This 'step' is referred to as competitor activation. According to most theoretical perspectives, this 'unwanted' competitor activation generates competition for selection of the appropriate target item. It logically follows that some cognitive mechanism must enable selection of the desired target item over competing non-targets, and it is generally agreed that RIF represents its end result. The precise nature of this mechanism, however, is where theoretical controversy arises.

Theoretical accounts of RIF are generally polarised to one of two camps: inhibition-based accounts and interference-based accounts. Inhibition-based accounts posit that active suppression (or *inhibition*) of non-target items is necessary to resolve competition, and thus enable selection of the appropriate target item. The result is facilitated retrieval of the target item (i.e. Rp+) in subsequent tests of recall, and long-lasting perturbation of memory for suppressed non-target items (Rp-). By contrast, interference-based accounts suggest that inhibition is not necessary to explain RIF, instead emphasising the importance of competition between activated items in long-term memory. According to this perspective, retrieval-practice strengthens the association between the retrieved target item and the cue associated with that items (i.e. the category it belongs to). This action subsequently occludes, interferes with, or 'steals' activation from competing non-target items associated with the same category (Storm & Levy, 2012). Crucially, this perspective permits perturbed access to non-target items in long-term memory without invoking inhibitory processes.

Each perspective has garnered various lines of support. Such evidence constitutes a wealth of published literature (see Anderson, 2003; Storm & Levy, 2012; for comprehensive discussions). However, as the theoretical nuances of RIF are actually of little *direct relevance* to the current investigation, we will only review a few such lines of evidence here.

One oft-cited line of evidence in support of the inhibitory-based account is the aforementioned effect of exemplar strength. Inhibitory theorists argue that strong competitor items generate more competition than weak ones, requiring a greater degree of inhibition necessary to overcome competition amongst the former items. This is supposedly evidenced by increased RIF susceptibility amongst strong exemplars compared to weak, which presumably signals increased inhibition of the former items (Anderson et al., 1994; Anderson, 2003).

A second line of evidence for the inhibitory account comes from so-called 'independent cue' variants of final test. To provide some context, much classic research on RIF used tests of final recall which involved category cues that were present in the study and/or retrieval practice phases of the experiment. For example, retrieval-practice of the word *orange* may be elicited using a word-stem completion task such as *fruit-or*____, and then at test participants would be prompted to recall items in the category *fruit* with a cue such as *fruit-*_____. If inhibition acts on memory representations (as opposed to associations between cue and item), then RIF effects ought to be present regardless of the type of cue used at test. This assumption has led to widespread use of independent cues - these are stimuli which were not present during study or retrieval-practice, but still convey information about studied items. For example, in the case of a category-exemplar pair such as *Fruit-Banana*, one may guide participants to retrieve the item during retrieval-practice using the cue *Fruit-B*_____, but at test provide an unseen yet informative cue such as *Yellow-*____. A number of studies using such 'conceptual' independent cues report RIF effects using a variety of stimulus types (Michael C Anderson & Spellman, 1995; Gómez-Ariza, Fernandez, & Bajo, 2012; Huddleston & Anderson, 2012; Saunders & MacLeod, 2006; Veling & van Knippenberg, 2004). An alternative form of independent cue is a perceptually driven one. For example, in a fragment completion task in using category-exemplar pairs, one may provide a cue such as *Fruits-Or*___ during retrieval-practice, followed by a previously unseen segment such as ___*ange* at test (Bajo, Gomez-Ariza, Fernandez, & Marful, 2006). That RIF effects are still present under such experimental manipulations is widely cited as evidence for an inhibitory mechanism which is used to resolve competition amongst memory representations.

By contrast, evidence for interference-based accounts (which assert that non-inhibitory processes can account for RIF) is generally less direct, and often takes the form of evidence *against* tenets of the inhibitory account. For example, Raaijmakers & Jakab (2013) have questioned the aforementioned effect of item strength (generally held as strong evidence for inhibition) - citing cases in which strong RIF effect sizes have been elicited using stimuli of low word frequency. These authors have also published findings which suggest *no* modulating effect of item strength on RIF (Jakab & Raaijmakers, 2009), but see discussion of these findings earlier. In a similar vein, some studies have failed to elicit RIF effects using independent cues (Camp, Pecher, & Schmidt, 2007; Perfect, Moulin, Conway, & Perry, 2002; Verde & Perfect, 2011). Again, such inconsistencies call supporting evidence for an inhibitory mechanism into doubt.

The above discussion provides only a brief and partial insight into the theoretical disputes surrounding RIF. The various arguments and lines of evidence for and against the inhibitory-based account are too numerous to address in this manuscript. However, it is worth mentioning that this account does not necessarily preclude the involvement of non-inhibitory mechanisms. Indeed, many of its proponents (including Anderson and colleagues; Anderson et al., 1994; Anderson, 2003; Anderson & Spellman, 1995) maintain that 'purely' competition-based factors also play a significant role in RIF. However, with criticisms, counter-criticisms and caveats abound, the role of inhibition in RIF remains hotly contested - and will likely remain so for some time.

1.2.4 ERP Correlates Of Retrieval Induced Forgetting

Given the high degree of empirical attention that RIF has received over the past two decades, naturally some effort has been made to identify neural correlates of this phenomenon. The first such investigation, by Johansson et al. (2007), used electroencephalography (EEG) to record neural activity during a retrieval-practice paradigm with category-exemplar pairs. The authors report that retrieval-practice of Rp+ items (compared to to a mere-exposure condition, in which participants 're-learned' category-exemplar pairs studied at the beginning of the experiment) was associated with two distinct event-related potentials (ERP) components. The first of these comprised a positive-going deflection over frontal electrodes during an early (200-1000 ms) time window, which onset around 200-300 ms following presentation of retrieval-guiding cues. The second component involved late-onsetting and prolonged positivity, again over frontal electrodes, during a late (1000-2000 ms) time window. Moreover, the authors report this latter ERP component was associated with behavioural performance at the final test stage of the experiment. Specifically, participants were split post-hoc into groups of those who showed a high degree of RIF at test (calculated by accuracy for re-learned items minus accuracy for retrieved items) versus those who showed a low degree of RIF. As illustrated in Figure 1, the authors report that the late ERP component was present (and statistically significant) in the 'high' group, but not in the 'low' group - providing strong support for this late-onset, prolonged positivity as a neural correlate of RIF.

Similar results were also obtained by Staudigl, Hanslmayr, and Bäuml (2010; supplementary analysis), who report a prolonged and positive-going ERP component over frontal electrode sites in a 750-1500 ms time window for retrieval of Rp+ items compared to re-exposure. While this effect was present in a slightly earlier time window to that examined by Johansson et al. (2007), it does share many of the characteristics of their reported results. Unfortunately Staudigl et al. (2010) do not report whether this ERP was predictive of later behavioural performance. Both Johansson et al. (2007) and Staudigl et al (2010) attribute this prolonged, positive-going frontal activity to competitor inhibition during target-retrieval, consistent with dominant theoretical accounts of RIF.

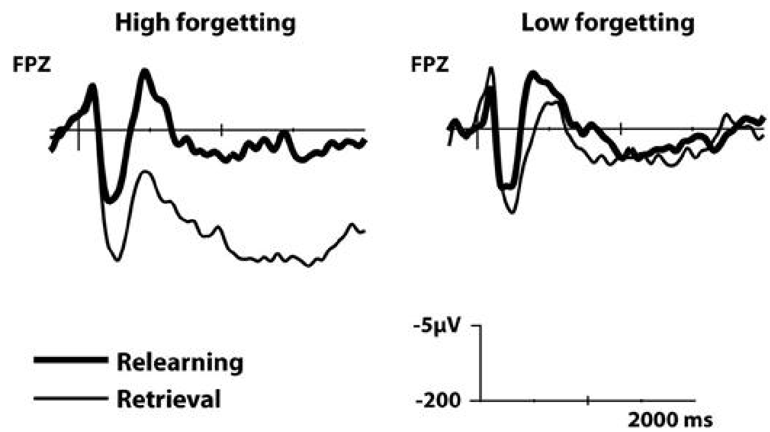

Figure 1: ERP waveforms at an anterior recording site, time-locked to the onset of category-exepmplar word stems or words pairs, in the retrieval-practice or relearning condition respectively. Taken from Johansson et al. (2007).

A more recent study by Hellerstedt and Johansson (2014) developed on the findings of Johansson et al. (2007) by distinguishing neural activity related to competitor activation - supposedly a vital precursor to RIF - from target retrieval. This was achieved by presenting participants with a word-stem completion task (such as 'Fruit-A___') during retrieval-practice, whereby the category cue ('Fruit') was presented 600ms before the exemplar-specific target cue ('A___'), allowing the authors to isolate neural activity associated with competitor activation elicited by the category cue. Using this paradigm, the authors contrasted strong competitor (i.e. high-frequency words) with weak competitor (i.e. low-frequency words) activation, by priming participants with lists of either strong or weak category exemplars prior to retrieval-practice (and assuming that the primed words would become activated upon presentation of the category cue).

The authors report that activation of strong competitors (compared to weak) following presentation of a category cue resulted in a positive-going deflection over frontal and fronto-central electrodes, onsetting around 300ms post-stimulus presentation. This ERP component was similar (in terms of onset latency, polarity and topography) to that of the 'early' ERP identified in Johansson et al.'s experiment, tentatively providing further support for its candidacy as a neural marker of competitor activation (which would have occurred in both experiments, following presentation of a category cue).

1.2.5 Retrieval Induced Forgetting And Language Learning

Despite efforts to push the boundaries of RIF to an increasingly wide range of contexts and cognitive domains, very little research has applied this phenomenon to language learning. To our knowledge, only one study by Levy et al. (2007) stands out. Using an adaptation of the retrieval-practice paradigm, Levy and colleagues investigated whether L2 (Spanish) vocabulary retrieval could temporarily perturb access to equivalent L1 (English) in moderately fluent Spanish speakers. The experiment was relatively simple, beginning with a study period in which participants' memory for L2 vocabulary was 'refreshed' by exposing them to Spanish words paired with matching pictures. Following this, participants completed what was essentially a cross-language equivalent of a retrieval-practice phase. Participants named a series of pictures of objects - each picture was named in either L1 or L2, depending on a cue appearing alongside the picture on each trial. Each picture was named multiple times, but always named in the same language. In a final test of recall, participants recalled *all* practised items in L1 with an independent-cue word-stem completion task.

The main finding of this investigation was that participants recalled significantly fewer items that had been named in L2 at final test, compared to items named in L1. In simple terms, it appears that retrieval-practice of L2 vocabulary perturbs access to equivalent L1 forms - demonstrating a clear RIF effect in the context of language learning. Indeed, Levy and colleagues attributed their findings to inhibition of memory representations for L1 forms during L2 retrieval-practice - consistent with theoretical accounts of 'classic' RIF. Before continuing, it is necessary to point out that these results are subject to a nontrivial caveat. By manipulating the number of times different items were named, the authors showed that their main effect was only significant for items named in L2 ten times. Items named five times showed a reduced and nonsignificant effect in the same direction, while items named once actually showed significantly *facilitated* L1 access at final test of recall. This finding provides an important methodological implication for further research - clearly multiple repetitions of an item in one language are necessary to induce forgetting in the other.

It appears, then, that RIF may have interesting implications for L2 acquisition in the context of L1 use. Surprisingly no work that we are aware of has investigated the 'inverse' of these results. That is to say, does using L1 impair access to memory representations for newly-learned L2 words? Such a finding may partially account for L2 learning advantages in immersion settings (in which learners experience minimal exposure to L1), and would hold interesting implications for the field of language learning more generally.

1.3 THE CURRENT INVESTIGATION

1.3.1 The Research Question

To reiterate the above discussion, it appears that RIF may extend to the context of language learning - specifically, retrieval of L2 terms amongst moderately fluent speakers perturbs access to L1. No study to date, however, has explored L1 use as a source of interference during the L2 vocabulary acquisition process. As such, the current investigation aims to explore whether retrieving L1 forms of newly learned L2 vocabulary may impact the learner's access to memory representations for those L2 words – i.e., whether L1 use may induce RIF during L2 learning. Of course, this line of inquiry necessarily depends on the prediction that newly learned L2 words will represent competing items (and thus, generate competition) during L1 retrieval. Given that newly learned words will necessarily be weakly represented, and given the positive correlations between item strength and magnitude of RIF typically reported in prior literature, it is likely that any effects observed in the current investigation will be substanially smaller in magnitude compared to 'classic' RIF effects.

1.3.2 Our Approach

To investigate potential deleterious effects on the process of vocabulary acquisition, one must first provide participants with representations of novel vocabulary. To this end, the Welsh language presents a useful tool. Welsh is somewhat unique in that, as a Gaelic-based language, it is rarely encountered in North America and highly distinct from widely used 'Romantic' languages such as English, French, or Spanish. However, despite its phonological and morphosyntactic distinctiveness, Welsh conveniently uses a Latin-based alphabet - permitting presentation of written forms to North American participants without much difficulty. As such, it provides a truly novel language which can be easily conveyed in the context of a laboratory experiment.

The current investigation employed an adapted version of Levy et al. (2007)'s paradigm, in which native English participants learned Welsh nouns for common objects. After an initial learning phase, participants retrieved the English forms for half of the objects with a simple picture-naming task, similarly to Levy et al. Finally, participants' memory for *all* of the recently learned Welsh words was tested, and accuracy and response latencies measured. Throughout the experiment neural activity was recorded with EEG, with a view to assess whether any potential behavioural effect could be attributed to the same underlying inhibitory activity seen in studies using a classic retrieval-practice paradigm. Adopting a similar notation style to that of previous RIF literature, from hereon in we will refer to items named in English during retrieval-practice as RpE+, and items *not* named in English during retrieval practice as RpE-.

Before continuing, it is necessary to emphasize that while the current investigation is based heavily on previous RIF literature, our experiment is fundamentally different from that of the classic retrieval-practice paradigm in a number of ways. It may be beneficial to highlight these differences to avoid later confusion. First, as outlined earlier, the classic retrieval-practice paradigm generates three types of stimulus - practiced items belonging to practiced categories (Rp+), unpracticed items belonging to practiced categories (Rp-) and unpracticed items belonging to unpractised categories (Nrp). By contrast, our paradigm entails only two types of item - those practiced in

English (RpE+) and those not practiced in English (RpE-). This leads to the second fundamental difference - retrieval-practice of Rp+ items usually leads to facilitated memory for those items at test. Due to the 'trans-language' nature of our paradigm, no stimulus/word retrieved during retrieval-practice is actually present at test (instead, English words are retrieved, while knowledge of Welsh words is assessed at test). As such, no set of items are afforded facilitation, as is the case with Rp+ items in a classic retrieval-practice paradigm. Finally, while Rp+ items in prior literature and RpE+ items in the current investigation are functionally comparable, in that these items are retrieved during retrieval-practice, it is crucial to emphasize that retrieval-practice is expected to have a *negative* effect on memory for L2 forms of these items, rather than a facilitative effect.

1.3.3 Behavioural And Electrophysiological Hypotheses

We predicted that English retrieval-practice of recently learned items would significantly perturb participants' memory for the Welsh forms of those items, evidenced by significantly higher error rates and response latencies to RpE+ items, relative to RpE- items, at a final test of recall.

To our knowledge, no previous study has used EEG to investigate RIF using picture stimuli. However, prior studies have associated 'classic' RIF with (1) an early-onset positive-going deflection over frontal electrodes, and (2) late-onset prolonged positivity, again over frontal electrodes, during retrieval-practice of Rp+ items. These processes are thought to signal competitor activation and subsequent competitor inhibition, respectively. As we predicted a RIF effect for newly learned vocabulary, we expected that similar neural processes would be present in the current investigation. As such, we predicted two distinct ERPs - an early frontal positive deflection (onsetting around 200-300 ms) followed by late and prolonged frontal positivity (onsetting around 1000 ms) during retrieval-practice of RpE+ items, compared to previously unseen control items. Replicating the work of Johansson et al. (2007), we focussed our analyses on an early (200-1000 ms post-stimulus) and a late (1000-2000 ms post-stimulus) time window of interest. Finally, if the ERP effects described above were truly reflective of neural activity giving rise to RIF, they ought to predict behavioural performance at final test of recall. As such, we predicted that ERP effects present during retrieval-practice would correlate positively with RTs and error rates to RpE+ items at test.

CHAPTER 2: METHODOLOGY

2.1 SUBJECTS

Twenty-two subjects (five males) aged between 18 and 25 participated in the current investigation. All subjects were native English speakers with no prior Welsh language experience; reported no neurological/psychological illness; were not taking any medications known to interfere with EEG recording; and had normal hearing and normal or corrected-to-normal vision. All but one subject was right-handed, as determined by a modified version of the Edinburgh Handedness Inventory (Oldfield, 1971). The EEG dataset from one subject was corrupted and had to be excluded, resulting in 22 behavioural data sets and 21 EEG data sets for analysis. All subjects were recruited via Dalhousie University's online subject pool, and were awarded with course credits following their participation. All procedures were approved by the Social Sciences and Humanities Research Ethics Board at Dalhousie University.

2.2 STIMULI AND APPARATUS

Participants were trained and tested on a set of 40 concrete Welsh nouns (see Appendix A). We intentionally selected phonologically and orthographically distinct Welsh nouns (i.e. they did not resemble any Latin-based language words) to ensure a high degree of novelty. Word frequency for the English form of each item was collected from the Corpus of Contemporary American English (COCA; Davies, 2008) online database. Three sets of pictorial stimuli (see Appendix B) were used throughout the experiment - one set of black-and-white line drawings, and two (different) sets of coloured line drawings. Each phase of the experiment was assigned one set of pictures (i.e., subjects viewed different pictures of the same objects between phases), and this was held constant across subjects. The motivation for this was primarily to reduce potential old-new ERP effects; but also to make the experiment more interesting and enjoyable for participants.

All visual stimuli were presented on a 24" LCD monitor (144 Hz refresh rate) against a plain white background. Auditory stimuli were recorded by a male speaker (with over 20 years of exposure to Welsh) to individual Waveform Audio File (WAV) files using Audacity(R) recording software (Audacity team, 2014). These WAV files were then normalised to a common amplitude, again using Audacity(R). During the experiment auditory stimuli were presented through a pair of *Mackie MR5* speakers. The experimental task, including stimulus presentation and behavioural response collection (see Procedure), was administered using PsychoPy 2.0 (Peirce, 2007).

2.3 PROCEDURE

Prior to beginning the experiment, subjects completed a short questionnaire with items pertaining to demographic information (age and sex), language history, and handedness. Each subject was tested individually over a single two-hour testing session. After being fitted with EEG recording equipment (see EEG data collection), subjects were seated in front of a computer monitor for the duration of the experiment. The experiment consisted of three phases: a Welsh learning phase, a retrieval-practice phase, and a Welsh testing phase.

During the Welsh learning phase, subjects were taught a set of Welsh nouns using a computerized two-alternative forced choice (2AFC) paradigm. Each trial began with a single printed Welsh noun presented on-screen for 2000 ms and accompanied by an audio clip of a male speaker pronouncing the word. This was followed by a blank screen for 1000 ms, after which subjects were presented with two line drawings of objects side-by-side onscreen— one of which matched the presented word— and required to indicate the matching picture with a left or right directional keyboard response. Subjects received feedback at the end of each trial indicating a correct or incorrect response. This procedure was used to train subjects on 40 Welsh nouns. Each noun was presented four times throughout the Welsh learning phase, and in each instance the matching picture was psuedorandomly paired with a different foil picture. Foil pictures did not represent items learned in Welsh, therefore an additional set of 40 pictures was incorporated into this phase. The first time a subject encountered a previously unseen noun they were forced to guess the matching picture, however the feedback combined with multiple repetitions of this procedure ensured that subjects learned correct picture-noun associations. Early piloting of this procedure demonstrated that subjects were able to learn 40 nouns in this manner, with > 90% accuracy at final test, following four repetitions of each.

During the retrieval-practice phase, subjects were required to retrieve the English forms of a subset of items seen in the previous phase, by naming them out loud in English. Each trial began with a black fixation cross presented for 500 ms. Following this, a single line-drawing of an object was presented at the center of the screen for 2000 ms (plus a variable duration of 0-200 ms, to avoid phase-locking of responses). Following this period, a cue appeared above the picture, prompting the subject to name the picture out-loud in English. A microphone situated near the participant and connected to the stimulus presentation computer detected each vocal response and triggered onset of the next trial. Subjects were instructed to withhold vocalizations until they were ready to respond, in order to avoid triggering the next trial prematurely. In total, 40 pictures were presented during the retrieval-practice phase; 20 of these were present in the Welsh learning phase, and 20 were previously unseen items. Previous work has shown that ≥10 repetitions of an item in one language is needed to cause forgetting of that item in another language (Levy et al., 2007), therefore each of the 40 items was presented 10 times during retrieval-practice, resulting in a total of 400 trials. The subset of items retrieved in English was counterbalanced across participants.

In the third and final phase of the experiment, subjects were tested on their knowledge of Welsh words using a near-identical 2AFC procedure to the Welsh learning phase. In this phase, however, each Welsh word was presented only once (as opposed to four times), and subjects did not receive any feedback following each trial. Finally, foil pictures in this phase were other learned-in-Welsh items, as opposed to the set used in the Welsh learning phase.

2.4 BEHAVIOURAL DATA COLLECTION

During the Welsh learning and testing phases, participant's error rate (whereby errors are defined as an incorrect response) and response time (RT) for each Welsh word was recorded based on their keyboard responses to the two-alternative forced choice task. For analysis purposes, error and RT data from the final presentation of each word during the Welsh learning phase was treated as a baseline measure of performance. This baseline data was contrasted with performance during the Welsh testing phase - generating a repeated-measures 'phase' variable with two levels: baseline and test. In addition, each Welsh word presented throughout the experiment was treated according to whether the English form was retrieved during retrieval-practice, thus generating a 'retrieval-practice' variable with two levels: retrieved in English (RpE+) and not retrieved in English (RpE-).

2.5 BEHAVIOURAL DATA ANALYSIS

All behavioural data analyses were conducted using R software (R Core Team, 2017). Two independent Linear Mixed Effects Regression analyses (LMER; implemented using the *mgcv* package in the R environment) were used to model variance in subjects' error rates and RT to each of the learned Welsh words. Both LMERs used a mixed-effects design which treated retrieval-practice, phase, counterbalancing group, and English word frequency (from hereon in, L1 frequency) as fixed-effect variables. Subject and individual word were treated as random-effect variables (the motivation for including the latter random effect was that there are likely some characteristics of individual Welsh words that affected performance on the behavioural task, which are not accounted for by L1 frequency). Because error rate represented dichotomous response data (i.e. one's response could either be correct or incorrect), the model treating error rate as the dependent variable used a logit-link binomial contrast. Where applicable, post-hoc multiple comparisons were conducted on main effects and interactions, using a Bonferroni adjustment.

2.6 ELECTROPHYSIOLOGICAL DATA COLLECTION

Electroencephalography (EEG) data was recorded from 64 electrode sites covering the scalp using an ANT Refa8 72 channel amplifier (Advanced NeuroTechnologies, Enschende, The Netherlands) connected to a Brain Products 64 ActiCap electrode system (Brain Products, Munich, Germany). The electrodes were mounted in a fitted cap with electroconductive gel applied to the subject's scalp at each site. The EEG was recorded with an average reference, and re-referenced offline to the average of the left and right mastoids. Horizontal and vertical elecro-oculograms (HEOG and VEOG) were recorded from bipolar electrodes positioned above and below the right eye, and on the outer cathi of the left and right eyes. EEG data was sampled at 500 Hz with a 167 Hz low-pass filter, referenced on-line to the average of all channels, and digitised using *Asalab* recording software. To avoid the disruption to the experiment that would have been caused by setting up the EEG recording half-way through the experiment, subjects' EEG responses were recorded throughout all three phases of the experiment.

All offline EEG preprocessing was conducted using MNE - a Python-based software package specialized for analyzing electrophysiological data (Gramfort et al., 2014). Due to the presence of some extremely noisy channels (with amplitudes to the order of ~1V, due to malfunctioning electrodes) in a number of datasets, bad channels were immediately identified and manually removed, and the data re-referenced to the average of all remaining channels. Trials containing movement/muscle artifact were removed based on visual inspection, and independent component analysis (ICA) used to correct regular sources of artifact such as blinks and saccades, using the fastICA algorithm (Hyvärinen & Oja, 2000). Subsequent to ICA, any channels that had been previously removed were interpolated using spherical splines, and then data were re-referenced to the average of the two mastoid electrodes (TP9 and TP10). Data from the retrieval-practice phase of the experiment was segmented into 2200 ms epochs which were time-locked to the presentation of each to-be-named picture (200 ms prior to each event marker and 2000 ms post).

2.7 ELECTROPHYSIOLOGICAL DATA ANALYSIS

Statistical analyses of ERP data were conducted on data pooled to 8 topographic regions of interest (ROIs): medial frontal (Fp1, Fp2, AF3, F1, F2, FZ), left frontal (F3, F5, F7, AF7), right frontal (AF8, F4, F6, F8), medial central (FC1, CP1, C1, Cz, CPz, FC2, C2, CP2), left central (T7, FC5, FC3, C5, C3, TP7, CP5, CP3), right central (FC4, FC6, FT8, C4, C6, T8, CP4, CP6, TP8), medial posterior (P1, Pz, P2, PO3, POz, PO4, O1, Oz, O2) left posterior (P7, P5, P3, PO7), and right posterior (P8, P6, P4, PO8). Data from every trial and electrode were entered into the analysis (i.e., no averaging was performed). Similarly to the behavioural data analysis, ERP data was analysed with two independent LMERs, for an early (200-1000 ms) and a late (1000-2000 ms) time window. Picture type (i.e., RpE+ or control) and ROI were treated as fixed-effect variables, while subject, word and subject X ROI interaction were treated as random effects variables. Where applicable, post-hoc multiple comparisons were conducted using a Bonferroni adjustment.

CHAPTER 3: RESULTS

3.1 BEHAVIOURAL DATA

Preliminary examination of subjects' RT data showed a positively skewed distribution, containing a number of extremely low RT outliers (see Figure 2, left panel). First, all trials with an RT of < 300 ms were removed. Regression analyses are more suited to normally distributed data, therefore a log transformation was performed on RT data (from hereon in, log-RT), which resulted in more normal distribution. Finally, log-RT data for trials on which participants made an incorrect behavioural response was removed (see Figure 2). Two independent Linear Mixed Effect Regressions (LMER) were used to analyze the effects of phase, retrieval-practice, counterbalancing group and L1 frequency on subjects' error rate and log-RT scores. Unless otherwise stated, statistical significance is assumed at $p \leq 0.01$.

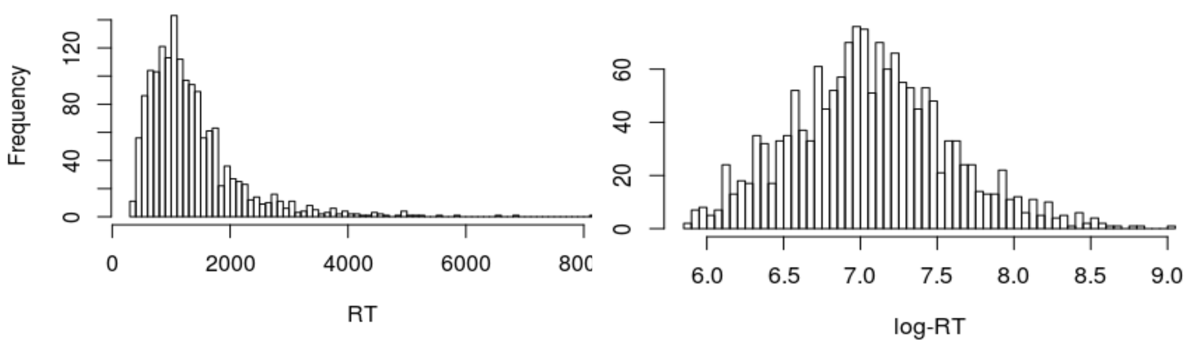

Figure 2. Frequency distributions for raw RT scores (left panel), and log-transformed RT scores (right panel).

For the model treating error rate as the dependent variable, as illustrated in Figure 3, subjects' error rates at baseline were significantly lower than error rates at test: $F(1, 1699) = 16.79$, $p < .001$. No other significant main effects, or significant interactions, were present ($p > 0.10$ in all cases). By contrast, the model treating log-RT as the dependent measure showed a wide range of effects. There was a significant main effect of L1 frequency, $F(1, 1525) = 4.25$, $p = .03$, whereby log-RTs correlated negatively with this variable (see Figure 5). As illustrated in Figure 4, RTs were significantly higher at test compared to baseline: $F(1, 1525) = 12.92$, $p < .001$. There was also a significant two-way interaction between phase and retrieval-practice, $F(1, 1525) = 7.21$, $p = .007$, again illustrated in Figure 4. Post-hoc comparisons revealed significantly higher log-RTs to RpE+ items at test compared to baseline ($p < .001$), while no such effect was present for RpE- items ($p = .519$). Moreover, log-RTs to RpE+ items were significantly higher than RpE- items during the test phase (p=.004), while this difference was not present at baseline (p=.461).

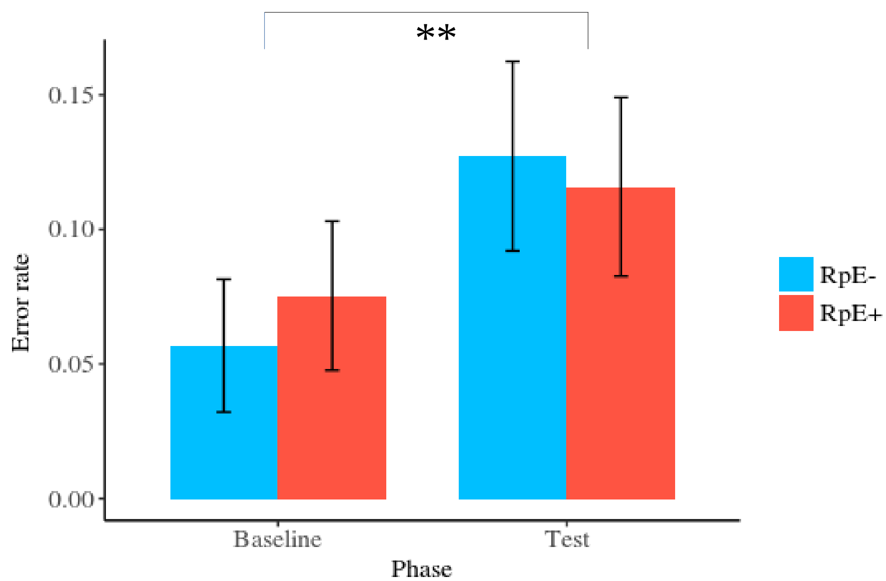

Figure 3. Mean error rates for the retrieval-practice X phase interaction. Error bars represent 95% CIs. ** indicates $p \leq .001$, following Bonferroni correction.

In addition to the effects described above, there was a significant three-way interaction between phase, retrieval-practice and L1 frequency: $F(1, 1525) = 6.00$, $p = .01$, whereby the log-RT difference between RpE+ and RpE- items was modulated by L1 frequency during the test phase, but not baseline. As illustrated in Figure 6, during the test phase RTs were higher for RpE+ items for more low-frequency words, and this difference decreased linearly with L1 frequency. Finally, there was a significant three-way interaction between phase, retrieval-practice and counterbalancing group: $F(1, 1525) = 6.81$, $p = .009$. No other significant main effects or interactions were present.

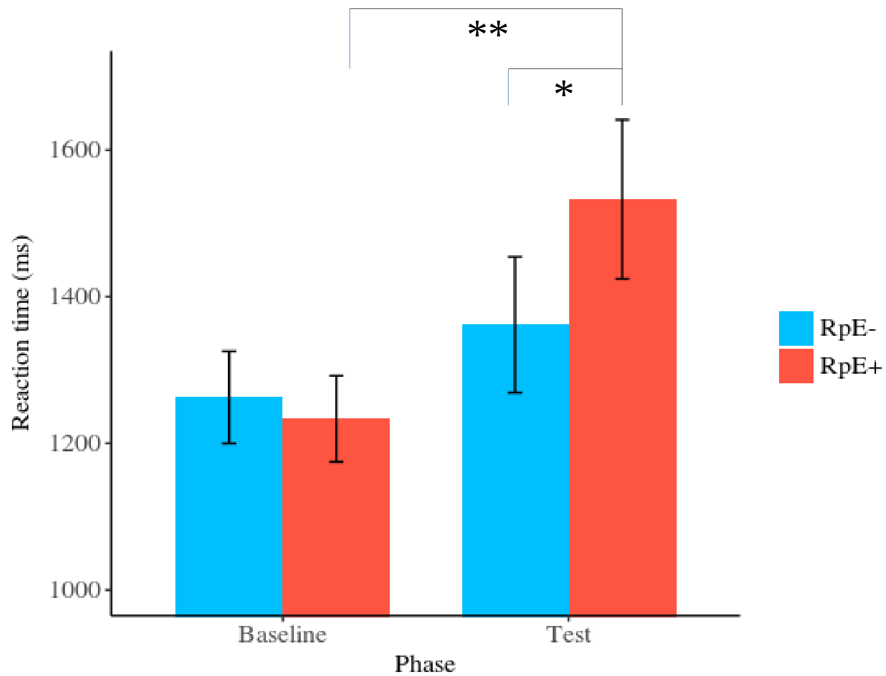

Figure 4. Mean RT values for the retrieval-practice X phase interaction. Error bars represent 95% CIs. All p values were derived from post-hoc Bonferroni comparisons. * indicates p ≤ .01, ** indicates p ≤ .001

For the purposes of comparison with prior research, we calculated an effect size which represented the magnitude of RIF in subjects' RT scores. Consistent with a recent meta-analysis of RIF research by (Murayama et al., 2014) we computed a standardized mean difference, *g*, between RpE+ and RpE- values at test. This yielded an effect size of $g = 0.19$.

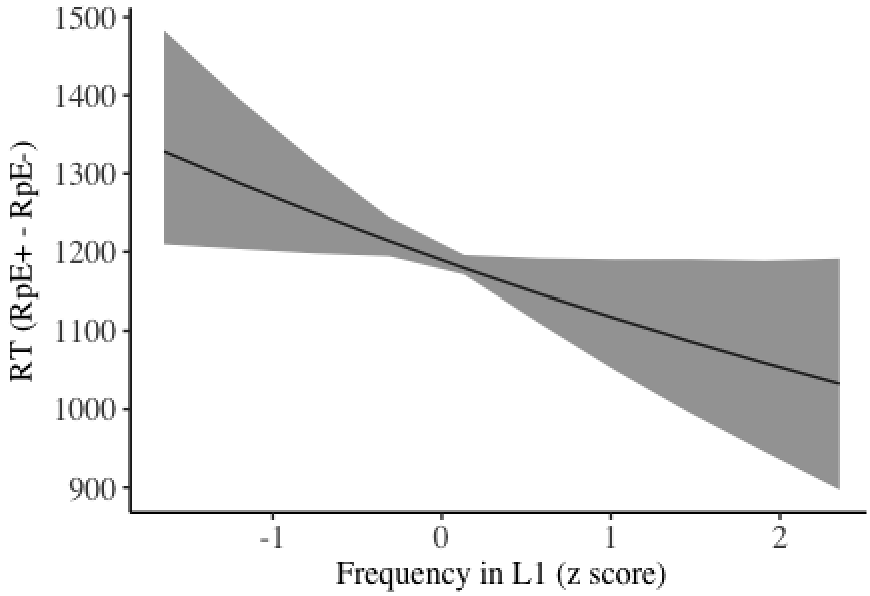

Figure 5. Regression slope for main effect of L1 frequency on RT, collapsing across phase and retrieval-practice variables. Error margins represent 95% CIs.

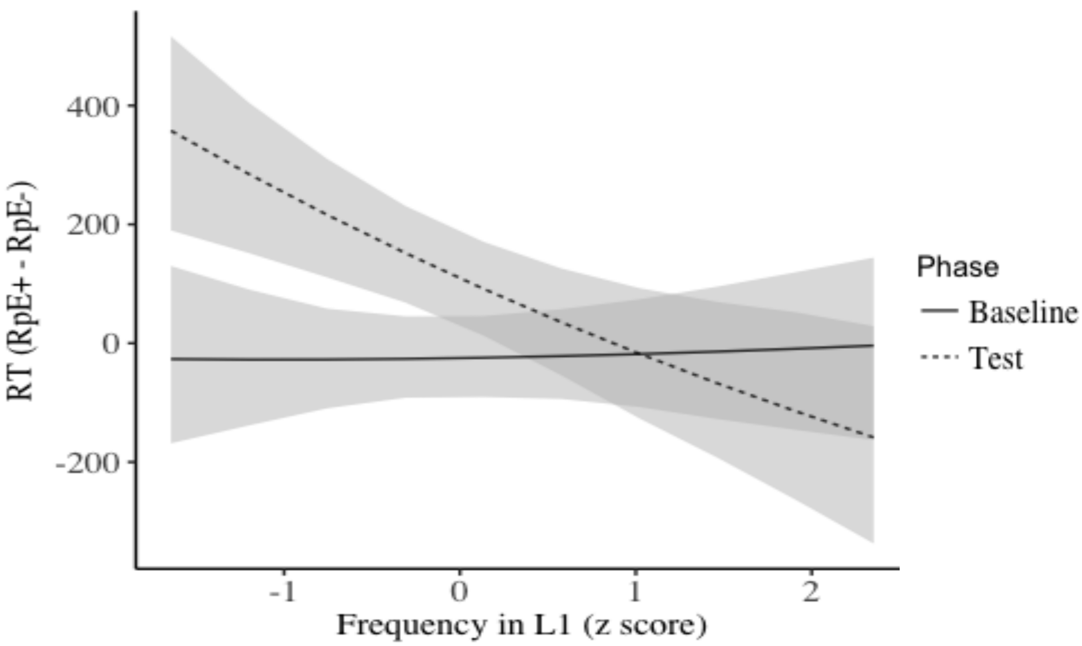

Figure 6. Regression slope for the retrieval-practice X L1 frequency interaction, plotted separately for Baseline and Test conditions. Values represent RTs to RpE+ items minus RTs to RpE- items, computed for each individual word (i.e. the magnitude of the retrieval-practice effect). The *p* values (derived from Bonferroni-corrected post-hoc comparisons) pertain to Test condition and Baseline condition, respectively. Error margins represent 95% CIs.

3.2 ERP DATA

Two independent LMERs were used to analyze the effects of picture type and ROI on subjects' mean response amplitude, during an early (200-1000 ms) and a late (1000-2000 ms) time window. Figure 8 contrasts ERP waveforms time-locked to the presentation of pictures during retrieval practice for RpE+ items and control (previously unseen) pictures. The topographical scalp distribution of the difference between these two conditions is illustrated in Figure 7, and mean ERP amplitudes for each time window are plotted in Figure 9.

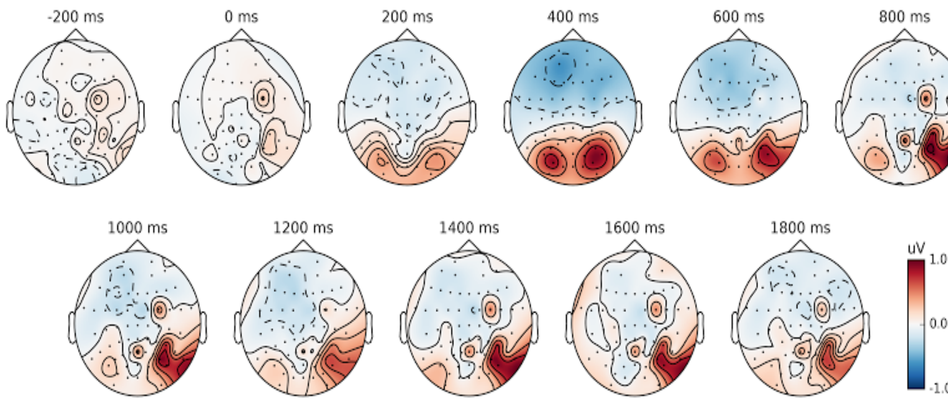

Figure 7. Scalp topographic plots for the ERP difference between RpE+ and control items during retrieval-practice at 200 ms intervals. Dots represent electrode locations.

Results from the early time window show a significant two-way interaction between picture type and ROI: $F(1, 8) = 134.18$, $p < .001$. Subjects showed a negative-going deflection (for RpE+ items compared to control) which was visible from around 300 ms post-stimulus presentation, persisted until around 600 ms, and was maximal over left, medial, and right anterior ROIs ($p < .001$ in all cases, according to Bonferroni post-hoc comparisons). Within the same time window, subjects also exhibited enhanced positivity (again, for RpE+ compared to control items) maximal over left, medial, and right posterior ROIs ($p < .001$ in all cases) which was visible from around 400 ms, and persisted until the end of the time window (1000 ms). The magnitude of signal change during the early time window is illustrated in Figure 9 (left panel).

The results of analyses on the late (1000-2000 ms) time window again show a significant two-way interaction between picture type and ROI: $F(1, 8) = 21.69$, $p < .001$. Subjects exhibited enhanced positivity (again, for RpE+ compared to control items) throughout the entire late time window, which was maximal over the right posterior ROI ($p < .001$). This positivity also bordered on significance, at the $p = .05$ level, over the medial central ROI ($p = .056$). No other significant effects were present. The magnitude of signal change during the late time window is illustrated in Figure 9 (right panel).

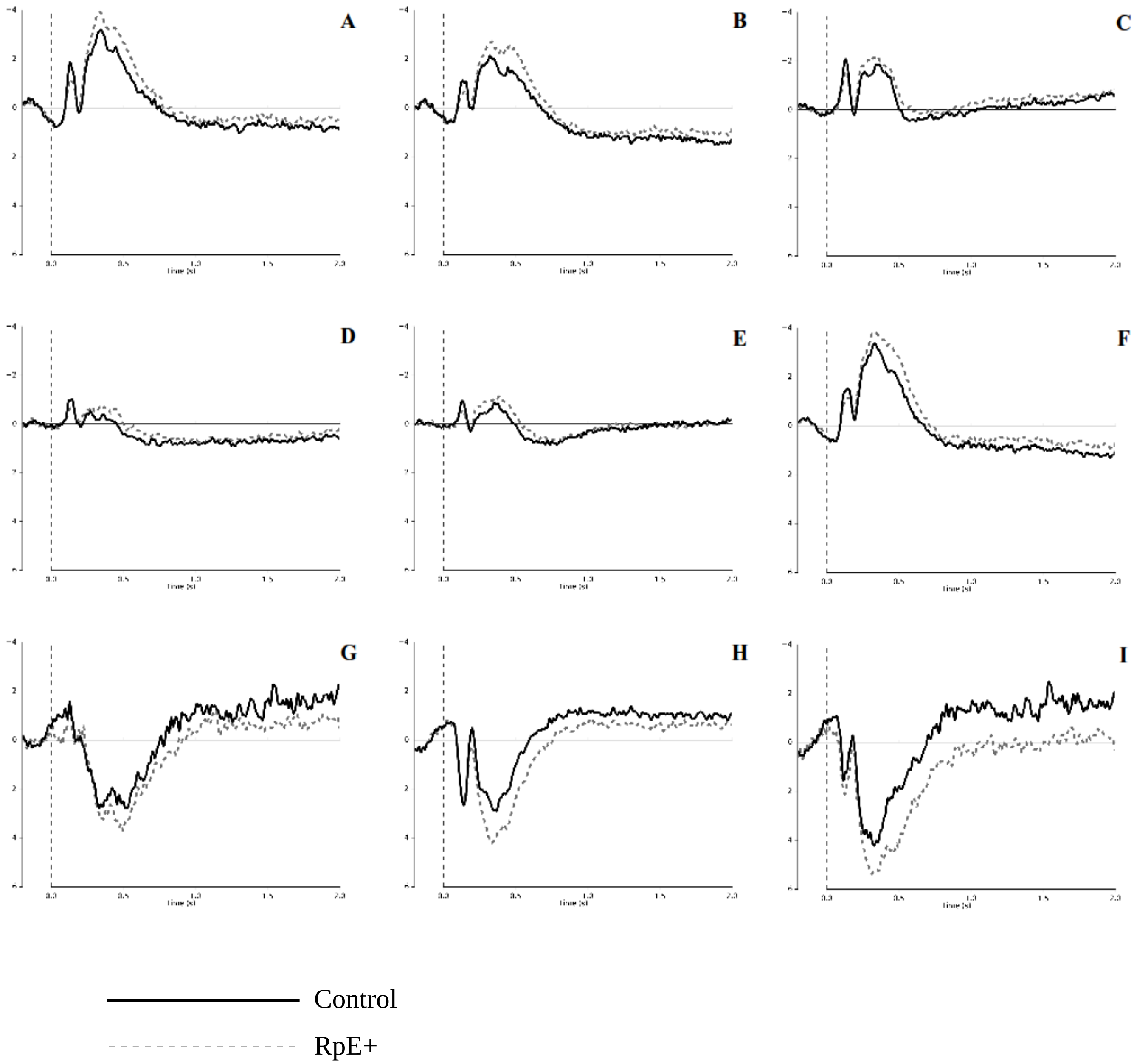

Figure 8. ERP waveforms time-locked to the onset of RpE+ and control items during retrieval-practice, across representative electrodes from each of the 9 ROIs used in the current investigation. **A:** Left anterior (AF7), **B:** Medial anterior (Pz), **C:** Right anterior (F6), **D:** Left central (C3), **E:** Medial central (Cz), **F:** Right central (C4), **G:** Left posterior (P7), **H:** Medial posterior (Pz), **I:** Right posterior (P6). Note that by convention, electrical potential on the y-axis is plotted negative-up.

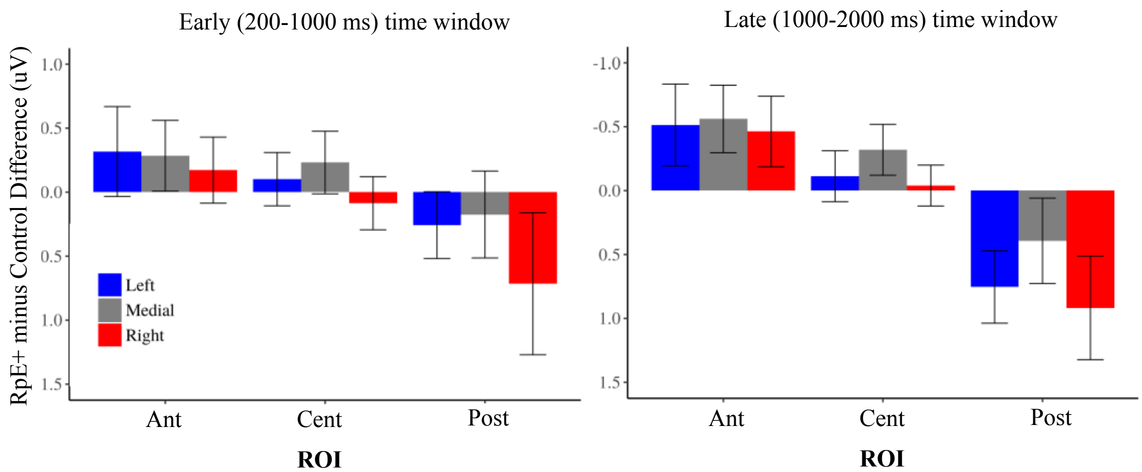

Figure 9. Mean response amplitudes for RpE+ minus control items during retrieval-practice for Anterior (Ant), Central (Cent) and Posterior (Post) scalp surfaces. Results are presented for the early (200-1000 ms) time window (left panel), and the late (1000-2000 ms) time window (right panel). Error bars represent 95% CIs.

3.3 BEHAVIOR X ERP INTERACTION

Two independent LMERs were used to investigate potential correlations between behavioural RT data at test and mean response amplitudes for each time window. Both models treated mean amplitude difference (for RpE+ - control items) as the dependent variable, and ROI and RT difference (for RpE+ items at test - RpE- items at test) as fixed-effect variables. Results from both time windows show no significant relationship between RT difference and mean amplitude difference, and no significant interaction between RT difference and ROI ($p > .50$ in all cases). In simple terms, mean amplitude difference between RpE+ and control items (in either time window) during retrieval-practice was not significantly predictive of RT scores at test. These results are illustrated in Figure 10.

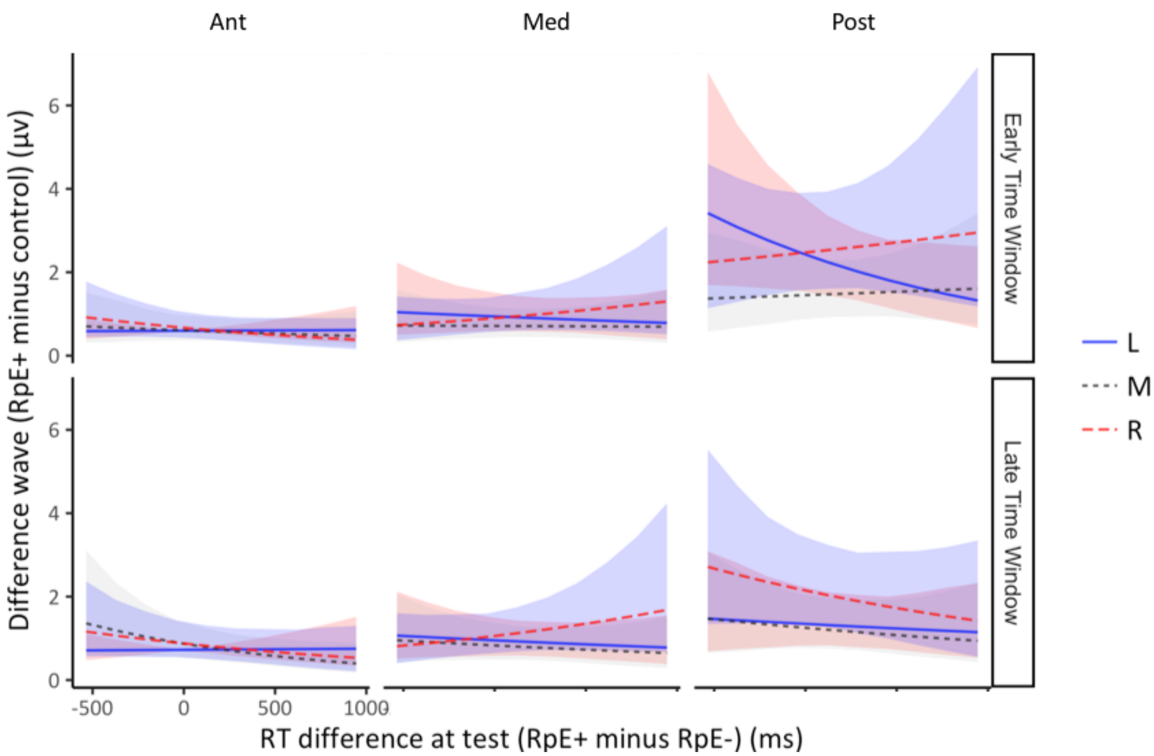

Figure 10. Regression slopes for the relationship between ERP difference wave (computed by subtracting EEG responses to RpE+ items from those of control items, during retrieval practice) and RT difference at test (computed by subtracting RTs to RpE+ items from RTs to RpE- items), across multiple ROIs. Lines represent left (L), medial (M) and right (R) divisions of anterior (Ant), medial (Med) and posterior (Post) scalp areas. Error bars represent 95% confidence intervals. Data is presented for both the early time window (top panel) and the late time window (lower panel).

CHAPTER 4: DISCUSSION

4.1 BEHAVIOURAL FINDINGS

The current investigation aimed to test the hypothesis that using one's native language to refer to items recently learned in a second language would impair one's memory for the newly learned L2 words. This hypothesis was partially supported - while there was no significant effect of retrieval-practice (and no phase X retrieval-practice interaction) on subjects' error rate to L2 items at test, L1 retrieval-practice did significantly perturb subjects' RTs to L2 items at test. The results illustrated in Figure 4 show significantly longer RTs to items retrieved in L1 (RpE+) compared to items not retrieved in L1 (RpE-) at test, while this difference was not present at baseline – indicateing a clear retrieval induced effect. Moreover, the results also show significantly longer RTs to RpE+ items at test compared to baseline, whereas effect was absent for RpE- items - again suggesting an effect that was exclusive to RpE+ items. Overall, this pattern of results indicates 'targeted' slowing of responses that appears to be entirely attributable to retrieval-related processes, providing support for our hypothesis with respect to RT data.

This finding is consistent with previous work – while most prior RIF research has focused on word retrieval at final test (by means of word-stem completion tasks, for example), a number of studies have investigated RIF effects on participant's response latencies to test material. The first such study that we are aware of measured participants' RTs on a lexical decision task, whereby they decided if category-exemplar pairs presented at test were congruent or incongruent (Perfect et al., 2002; Experiment 5). The authors report that participants' responses were significantly longer for Rp- compared to Rp+ items. Similar patterns of results have been reported by studies employing variants on Perfect et al.'s lexical decision task (Veling & van Knippenberg, 2004), written exemplar recognition (Lev-Ari & Keysar, 2014; Racsmany, Conwav, Garab, & Nagymate, 2008; Verde & Perfect, 2011), and written sentence recognition (Gómez-Ariza, Lechuga, Pelegrina, & Bajo, 2005).

The RT results observed in the current investigation showed an estimated effect size of g = 0.19. By comparison, Murayama (2014) places estimated effect sizes from prior RIF research substantially higher, at g = 0.35, with a 95% confidence interval (CI) = [0.32,0.38], when averaging across their whole sample of 512 studies. However, when Murayama's analysis was confined to studies using RT as a test measure, the estimated effect size was g = 0.18, CI = [0.07,0.28]. In light of this, it appears that the RT effect observed in the current investigation was not only supportive of our hypothesis, but of a similar magnitude to those of comparable prior RIF research.

Given that our hypothesis was supported by the RT data, it is curious that we did not observe similar effects (or, at least, a trend) in the error rate data. It is likely that this incongruity may be attributed to a ceiling effect in subjects' performance on the behavioural task. Error rates were extremely low throughout the entire experiment - around 5-7% at baseline, and 10-12% at test. By contrast, error rates in previous RIF research using written material (summarized here as the percentage of items that participants failed to recall, for the sake of comparison with our data) tend to be substantially higher - for example Anderson et al. (1994) reported error rates of 20-70% (collapsing across all experimental conditions). Of course, it may not be fully appropriate to compare error rates from such studies with those of our own – given that word-stem completion tasks (typically employed at final test when using written material) are fundamentally different from our 2AFC task on a number of levels. However, one must consider that experiments testing participants' recognition for previously studied items (which are arguably more similar to our 2AFC task) have still yielded error rates in a higher range to those of the current investigation. For example, Hicks and Starns (2004) had participants judge whether items at test had previously appeared during the study phase, and report the percentage of items not correctly recognized around 35-38% and 26-29% for Rp- and Nrp items respectively. Subsequent replications using similar old-new recognition tasks have reported error rates around 40-50% for Rp- items, and 30-40% for Nrp items (Spitzer & Bäuml, 2007; Verde & Perfect, 2011). Finally, the only RIF study reporting error rates on a forced choice task (with three potential responses on each trial) that we are aware of reported the percentage of items not correctly identified at 63% and 50% for Rp- and Nrp items respectively (Saunders & MacLeod, 2002; Experiment 2). In light of these error rate ranges observed in previous research, it seems reasonable to surmise that the 2AFC task was not a sufficiently challenging measure to effectively tap RIF effects in terms of subjects' accuracy on a final test of memory for newly learned L2 words.

One unexpected result which is worth addressing is that subjects' error rates were significantly higher at test compared to baseline. This was most likely a product of the current investigation's experimental design. Subjects were trained on L2 vocabulary during the Welsh learning phase, immediately after which their baseline performance was measured. The test phase, however, did not occur until approximately 20-25 minutes later, after the retrieval-practice phase. It seems likely that the increase in error rates between baseline and test (which were separated by a long and fatiguing task) was merely due to passive memory decay of the L2 words.

A number of other significant effects warrant attention. First, the main effect of L1 frequency is not entirely surprising. Previous research has shown that L2 words with high-frequency L1 forms are generally better remembered than those with low-frequency L1 forms (de Groot, 2006; De Groot & Keijzer, 2000; Lotto & de Groot, 1998). As such, this result merely reflects facilitated memory of Welsh words with high-frequency English forms when collapsing across all other experimental conditions. Far more interesting is the significant three-way interaction between phase, retrieval-practice and L1 frequency - suggestive of a modulating effect of L1 frequency in our main results. One might assume that (as mentioned above) L2 words with high-frequency L1 forms are better encoded, and therefore more strongly represented in memory. One may expect, then (according to inhibitory-based accounts of RIF), greater inhibition of the more strongly represented items, reflected by a positive correlation between L1 frequency and magnitude of the RIF effect. Such a result would be consistent with previous studies repeatedly showing a positive correlation between item frequency and magnitude of RIF.

Instead, however, the results of the current investigation show a highly significant *negative* correlation (see Figure 6).

In order to interpret this unanticipated result, one must consider the basic theoretical assumptions surrounding RIF. As previously discussed, it is generally held that when one attempts to retrieve a target item from memory, related non-target (i.e., competitor) items compete for selection. Selection of the appropriate target item will then depend (to a greater or lesser extent, depending on one's theoretical perspective) on the associative strength between the retrieval cue and the target item, relative to that of competing items. In the case of the current investigation, L1 forms represent target items during retrieval-practice, and L2 words (in theory) represent competing items. One could argue from an inhibitory-based perspective that, because high-frequency L1 words are strongly represented, little or no inhibition of the competing L2 form is needed to achieve selection of the appropriate L1 target. By contrast, low-frequency L1 words are less well represented, increasing the likelihood that an L2 competitor will present "sufficient" competition to necessitate its inhibition. Such an account would explain not only why a modulating effect of L1 frequency was present, but also why subjects' responses to L2 words at the high end of the L1 frequency range were seemingly unaffected by retrieval practice (see Figure 6).

Finally, there is the significant three-way interaction between phase, retrieval-practice and counterbalancing group. Subjects were counterbalanced according to which items were assigned to RpE+ and RpE- items, therefore one may be tempted to attribute this result to a stimulus effect. However, the cause of this result likely lies in an oversight during data collection. Instead of adopting the standard practice of *interleaving* counterbalancing conditions throughout the data collection period, all subjects in one counterbalancing condition were tested concurrently, after which subjects in the other counterbalancing condition were tested. While this methodological detail may seem trivial, there is evidence that undergraduate participants' performance on tasks which measure RT may be significantly affected by the time-point during the academic term at which they are tested (Fawcett et al., unpublished data; obtained via personal communication with J Fawcett, July 2017). Indeed, our sample consisted entirely of undergraduate subjects who were nearing exam periods towards the latter part of our data collection period. As such, this three-way interaction likely does not hold any theoretical significance.

4.2 ERP FINDINGS

In addition to our behavioural hypotheses, the current investigation aimed to explore whether neural processes similar to those underlying 'classic' RIF would be present in this experimental context. In brief, we expected to observe two distinct ERPs following presentation of RpE+ items (compared to control items) during retrieval-practice: an early positive-going deflection over frontal electrodes during a 200-1000 ms time window (theoretically representative of competitor activation), followed by sustained frontal positivity during a 1000-2000 ms time window (representative of competitor inhibition). Neither of these predictions were met. During the early time window, there was a significant *negative*-going deflection, onsetting around 300 ms post-stimulus presentation over left, right and medial frontal ROIs, accompanied by posterior positivity which onset around 400 ms and persisted until the end of the time window. In the late time window there was no significant ERP difference between RpE+ and control items over frontal ROIs, while right-sided posterior positivity (continuing from the early time window) persisted until the end of the epoch (2000 ms).

First, to address the unexpected findings during the early time window, one must consider the premise of our hypothesis - that is, that the current investigation would elicit competitor activation during retrieval practice, a process which has been previously associated with the early frontal positive-going deflection (Hellerstedt & Johansson, 2014). As alluded to above, it is likely that newly learned L2 vocabulary constitutes very weak memory representations. One must consider that, by contrast, previous studies reporting the predicted ERP effect have used RIF paradigms employing well-known (i.e., strongly represented) L1 items. As demonstrated by Hellerstedt and Johansson, the magnitude of the predicted ERP effect is dependent on the strength of an item's representation. In simple terms, activation of weak competitors results in a diminished ERP effect relative to strong competitors. As such, it may be the case that newly learned vocabulary are too weakly represented to provide competitor activation of sufficient magnitude to be detected using our EEG paradigm.

While the the above explanation may account for an *absence* of predicted effects, it does not explain the presence of the negative-going waveform over frontal ROIs. Here, one must consider that our hypotheses were dependent on the assumption that any potential RIF effects present in the current investigation would be underlied by identical neural processes/mechanisms as in prior studies. In plain terms, while it is reasonable to expect competitor activation during the early time window, one must consider that there are still profound methodological differences between the current investigation and prior research in this area - for example, the introduction of an L2 into the RIF paradigm, and the use of pictures as opposed to printed word forms. It logically follows that different cognitive processes would be engaged during the current investigation - such as parallel activation of L1 and L2 items, as opposed to 'within-L1' competitor activation. As such, the unexpected ERPs that were present during the early time window may simply reflect engagement of different cognitive processes that were elicited by our experiment and not those of Johansson et al. (2007), Staudigl et al (2010), and Hellerstedt and Johansson (2014).

The complete absence of significant ERP differences over frontal ROIs during the late time window may be attributed to one of two possibilities. The first is inappropriate experimental design—while previous studies on the neural correlates of RIF have used relearning/re-exposure of previously studies items as a baseline for comparison with retrieval-practice, the current investigation involved retrieval-practice of *both* control and RpE+ items. In retrospect, this baseline condition may not have been fully appropriate. It may be the case that retrieval of *any* L1 word for a given picture entails some inhibition of competitor items. For example, when shown a picture of a shirt (all of our control items were articles of clothing), participants may be inhibiting competing memory representations for other types of clothes. If this is indeed the case, it would logically follow that the predicted experimental effect may be masked by our control condition. An alternative to this explanation once again returns to the issue of weak memory representation of newly learned L2 vocabulary. It is possible that, L2 competitors are so weakly represented in memory, very little inhibition of is required for successful L1 retrieval. One could argue that due to this this level of inhibition being quite minimal, associated neural activity would be difficult to detect with EEG.

Overall, the current investigation failed to elicit ERP effects which, based on prior work in this area, presumably reflect neural correlates of competitor activation and subsequent inhibition giving rise to RIF. However, the ERPs that were elicited were remarkably similar in terms of onset latency, time course and (in the case of the early time window) topography to those we predicted. One may argue that these could reflect neural correlates of RIF in the unique context of the current investigation. As discussed above, the experimental design of the current investigation was profoundly different from prior work in this area. Could it be the case, then, that the introduction of an L2, or the use of pictures instead of printed words, modulated the spatial and polar properties of the predicted ERPs? This is possible, yet unlikely for a number of reasons.

First, none of the observed ERP effects were predictive of behavioural performance at test. As is evident in Figure 9, there was no significant correlation between magnitude of amplitude change in posterior regions (or, indeed, any region) and RT performance at test. As stated earlier, any neural correlate of RIF ought to predict later behavioural performance - therefore it seems unlikely that the ERP effects we observed are not related to our primary research question.

The second reason is that the observed ERP effects strongly resemble old/new effects found in studies of recognition memory. In brief, a number of studies have shown that exposure to ‘old’ stimuli that have been previously seen during the course of an experiment (compared to ‘new’, unseen stimuli) tends to elicit a distinct pattern of neural activity. This pattern comprises an early negative-going deflection which onsets around 300 ms over frontal electrodes, accompanied by a prolonged positive going deflection which onsets around 500-800 ms, and can persist for up to 1000 ms before returning to baseline (Curran, 2000; Curran & Hancock, 2007; Wilding, 2000; and others). These old/new ERP effects bear remarkable similarity to those observed in the current investigation in terms of onset latency, polarity and topographical distribution. One must also consider that during retrieval-practice, participants in the current investigation viewed pictures of RpE+ items, which had been previously seen in the previous learning phase (albeit different pictures of the same items), along with previously unseen control items. With this in mind, an 'old/new'-based interpretation of the observed ERP effects seems appropriate, given their remarkable similarity to ERPs generated by studies of recognition, and given the design of the current investigation.

With all of the above considered, the ERP results of the current investigation remain unclear. On the one hand, it is possible that they reflect some context-specific neural correlate of RIF - given the uniqueness of our RIF paradigm. However, the available evidence suggests that these unexpected results are more likely an artifact of our experimental design.

4.3 CONCLUSIONS AND FUTURE RECOMMENDATIONS

The current investigation has provided some encouraging preliminary insights into the potential role of RIF during the L2 vocabulary acquisition process. Our behavioural data suggest that retrieving words in L1 causes some slowing of access to recently formed memory representations of their L2 forms (although it seems that this effect is greater for less frequent L1 words). This central finding opens numerous avenues of inquiry. For example, is this impairment of newly-formed L2 word representations temporary, or does it have a lasting impact on one's memory for those words? Must L1 retrieval occur immediately following L2 vocabulary acquisition in order for RIF effects to take place? Does RIF affect only vocabulary acquisition, or can it impact more complex components of language learning such as morphosyntax? Of course, the first issue to address in future research is the lack of a behavioural effect in our error data. As discussed earlier, this is likely attributable to the 2AFC task simply being too easily to effectively tap RIF effects. In terms of a more challenging probe of L2 knowledge, a great deal of previous RIF studies have used word-stem completion tasks a test. As such, this approach may be beneficial for further research in this area. Alternatively, if word-stem completion tasks are still too easy, tests of 'free recall' in which participants are simply asked to produce all the L2 words from the experiment that they can recall may provide a useful tool.

The current investigation failed to elicit ERP correlates of the observed behavioural effect traditionally associated with RIF. Precisely why this is the case is unclear, but likely relates to weak representation of L2 competitor items at test, potentially inappropriate experimental design, masking of predicted ERPs due to old/new effects, or perhaps some combination of all the above. Regarding the issue of inappropriate experimental design, prior research has used relearning/re-exposure conditions as a baseline for comparison with retrieval-practice. This approach was avoided in the current investigation because initial piloting showed that mere exposure to L1 words for common objects was a somewhat onerous for participants. Pilot participants often became fatigued and/or failed to concentrate on the stimuli for the necessary length of time (approximately 20 minutes). That being said, retrieval-practice of unseen items clearly does not present a viable baseline condition for comparison with retrieval-practice of newly learned items. Future research should endeavour to develop a baseline condition which does not entail retrieval-practice, but is still engaging enough to hold participants' attention.

Regarding deeper theoretical implications of our results, it is unclear whether newly learned L2 words are subject to inhibition or interference during L1 retrieval. One may argue that, because different sets of pictures were used in each phase of the experiment, the behavioural results from the test phase were obtained using a so-called 'independent cue' form of final test. This line of reasoning could be used to argue for an inhibition-based interpretation, however one must consider that merely using a different picture of the same item may not represent a *truly* independent cue. A more robust means of testing interference vs. inhibition in the context of this paradigm would be to employ a wholly different probe of participants' L2 knowledge at test, such as L2 word production or word-stem complettion.

Overall, the findings of the current investigation are promising. They indicate that L1 retrieval, a previously neglected factor in the field of second language acquisition, has a deleterious effect on access to memory representations for newly learned L2 vocabulary. Subsequent developments on these initial findings will likely hold valuable insights for the field of L2 learning in general and, ultimately, may inform optimum strategies for adult L2 learners.

REFERENCES

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445. https://doi.org/10.1016/j.jml.2003.08.006

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: the role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology. General*, *130*(3), 544–570.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *20*(5), 1063–1087.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: memory retrieval as a model case. *Psychological Review*, *102*(1), 68–100. https://doi.org/10.1037/0033-295X.102.1.68

Audacity Team (2014). Audacity(R): Free Audio Editor and Recorder [Computer program]. Version 2.0.0 retrieved April 20th 2014 from http://audacity.sourceforge.net/. [Audacity(R) software is copyright (c) 1999-2014 Audacity Team. [Web site: http://audacity.sourceforge.net/. It is free software distributed under the terms of the GNU General Public License.] The name Audacity(R) is a registered trademark of Dominic Mazzoni.]

Bajo, M. T., Gomez-Ariza, C. J., Fernandez, A., & Marful, A. (2006). Retrieval-Induced Forgetting in Perceptually Driven Memory Tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1185–1194.

Bäuml, K. (1998). Strong items get suppressed, weak items do not: The role of item strength in output interference. *Psychonomic Bulletin & Review*, *5*(3), 459–463. https://doi.org/10.3758/BF03208822

Camp, G., Pecher, D., & Schmidt, H. G. (2007). No retrieval-induced forgetting using item-specific independent cues: evidence against a general inhibitory account. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(5), 950–958. https://doi.org/10.1037/0278-7393.33.5.950

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology*, 1403–1414.

Coman, A., Manier, D., & Hirst, W. (2009). Forgetting the Unforgettable Through Conversation: Socially Shared Retrieval-Induced Forgetting of September 11 Memories. *Psychological Science*, *20*(5), 627–633. https://doi.org/10.1111/j.1467-9280.2009.02343.x

Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition*, *28*(6), 923–938. https://doi.org/10.3758/BF03209340

Curran, T., & Hancock, J. (2007). The FN400 indexes familiarity-based recognition of faces. *NeuroImage*, *2*(36), 464–471. https://doi.org/10.1016/j.neuroimage.2006.12.016

Davies, M. (2008) The Corpus of Contemporary American English (COCA): 520 million words, 1990-present. Available online at http://corpus.byu.edu/coca/

de Groot, A. M. B. (2006). Effects of Stimulus Characteristics and Background Music on Foreign Language Vocabulary Learning and Forgetting. *Language Learning*, *56*(3), 463–506. https://doi.org/10.1111/j.1467-9922.2006.00374.x

De Groot, A. M. B., & Keijzer, R. (2000). What Is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting. *Language Learning*, *50*(1), 1–56. https://doi.org/10.1111/0023-8333.00110

Dunn, E. W., & Spellman, B. A. (2003). Forgetting by remembering: Stereotype inhibition through rehearsal of alternative aspects of identity. *Journal of Experimental Social Psychology*, *5*(39), 420–433. https://doi.org/10.1016/S0022-1031(03)00032-5

Ford, R. M., Keating, S., & Patel, R. (2004). Retrieval-induced forgetting: A developmental study. *British Journal of Developmental Psychology*, *22*(4), 585–603. https://doi.org/10.1348/0261510042378272

Gómez-Ariza, C. J., Fernandez, A., & Bajo, M. T. (2012). Incidental retrieval-induced forgetting of location information. *Psychonomic Bulletin & Review*, *19*(3), 483–489. https://doi.org/10.3758/s13423-012-0228-8

Gómez-Ariza, C. J., Lechuga, M. T., Pelegrina, S., & Bajo, M. T. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition*, *33*(8), 1431–1441. https://doi.org/10.3758/BF03193376

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Guekes, S., & Zwisterlood, P. (2016). Novel L2 words do not facilitate but interfere with their L1 translations during picture naming – behavioural and event-related potential evidence. *Language, Cognition and Neuroscience*, *31*(8). Retrieved from http://www.tandfonline.com/doi/abs/10.1080/23273798.2016.1209530

Hall, C. J. (2006). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *IRAL - International Review of Applied Linguistics in Language Teaching*, *40*(2), 69–87. https://doi.org/10.1515/iral.2002.008

Hamada, M., & Koda, K. (2008). Influence of First Language Orthographic Experience on Second Language Decoding and Word Learning. *Language Learning*, *58*(1), 1–31. https://doi.org/10.1111/j.1467-9922.2007.00433.x

Hellerstedt, R., & Johansson, M. (2014). Electrophysiological Correlates of Competitor Activation Predict Retrieval-Induced Forgetting. *Cerebral Cortex*, *24*(6), 1619–1629. https://doi.org/10.1093/cercor/bht019

Hemchua, S., & Schmitt, N. (2006). An analysis of lexical errors in the English compositions of Thai learners. *Prospect*, 3–25.

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review*, *11*(1), 125–130. https://doi.org/10.3758/BF03206471

Huddleston, E., & Anderson, M. C. (2012). Reassessing critiques of the independent probe method for studying inhibition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(5), 1408–1418. https://doi.org/10.1037/a0027092

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, *13*(4), 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5

Iglesias-Parro, S., & Gómez-Ariza, C. J. (2006). Biasing decision making by means of retrieval practice. *European Journal of Cognitive Psychology*, *18*(6), 899–908. https://doi.org/10.1080/09541440500501189

Jakab, E., & Raaijmakers, J. G. W. (2009). The Role of Item Strength in Retrieval-Induced Forgetting. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(3), 607–617.

Johansson, M., Aslan, A., Bäuml, K.-H., Gäbel, A., & Mecklinger, A. (2007). When Remembering Causes Forgetting: Electrophysiological Correlates of Retrieval-Induced Forgetting. *Cerebral Cortex*, *17*(6), 1335–1341. https://doi.org/10.1093/cercor/bhl044

Kaushanskaya, M., & Rechtzigel, K. (2012). Concreteness effects in bilingual and monolingual word learning. *Psychonomic Bulletin & Review*, *19*(5), 935–941. https://doi.org/10.3758/s13423-012-0271-5

Kroll, J. F., Hell, J. G. V., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment*. *Bilingualism: Language and Cognition*, *13*(3), 373–381. https://doi.org/10.1017/S136672891000009X

Lev-Ari, S., & Keysar, B. (2014). Executive control influences linguistic representations. *Memory & Cognition*, *42*(2), 247–263. https://doi.org/10.3758/s13421-013-0352-3

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: the role of retrieval-induced forgetting during second-language acquisition. *Psychological Science*, *18*(1), 29–34. https://doi.org/10.1111/j.1467-9280.2007.01844.x

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. https://doi.org/10.1111/1467-9922.00032

MacLeod, M. (2002). Retrieval-induced forgetting in eyewitness memory: forgetting as a consequence of remembering. *Applied Cognitive Psychology*, *16*(2), 135–149. https://doi.org/10.1002/acp.782

Macrae, C. N., & Macleod, M. D. (1999). On recollections lost : When practice makes imperfect. *Journal of Personality and Social Psychology*, *77*(3), 463–473.

Migueles, M., & Garcia-Bajos, E. (2006). Influence of the Typicality of the Actions in a Mugging Script on Retrieval-Induced Forgetting. *Psicologica: International Journal of Methodology and Experimental Psychology*, *27*(1), 119–135.

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: a meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin*, *140*(5), 1383–1409. https://doi.org/10.1037/a0037505

Nation, I. (2006). How Large a Vocabulary is Needed For Reading and Listening? *Canadian Modern Language Review*. https://doi.org/10.3138/cmlr.63.1.59

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Palmer, S. D., Macgregor, L. J., & Havelka, J. (2013). Concreteness effects in single-meaning, multi-meaning and newly acquired words. *Brain Research*, *1538*, 135–150. https://doi.org/10.1016/j.brainres.2013.09.015

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Perfect, T. J., Moulin, C. J. A., Conway, M. A., & Perry, E. (2002). Assessing the inhibitory account of retrieval-induced forgetting with implicit-memory tests. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*(6), 1111–1119.

Phenix, T. L., & Campbell, J. I. (2004). Effects of multiplication practice on product verification: integrated structures model or retrieval-induced forgetting? *Memory & Cognition*, *32*(2), 324–335.

Pica, G., Pierro, A., Bélanger, J. J., & Kruglanski, A. W. (2014). The Role of Need for Cognitive Closure in Retrieval-Induced Forgetting and Misinformation Effects in Eyewitness Memory. *Social Cognition*, *32*(4), 337–359. https://doi.org/10.1521/soco.2014.32.4.337

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raaijmakers, J. G. W., & Jakab, E. (2013). Is Forgetting Caused by Inhibition? *Current Directions in Psychological Science*, *22*(3), 205–209. https://doi.org/10.1177/0963721412473472

Racsmany, M., Conwav, M. A., Garab, E. A., & Nagymate, G. (2008). Memory awareness following episodic inhibition. *The Quarterly Journal of Experimental Psychology*, *61*(4), 525–534.

Saunders, J., & MacLeod, M. D. (2002). New evidence on the suggestibility of memory: the role of retrieval-induced forgetting in misinformation effects. *Journal of Experimental Psychology. Applied*, *8*(2), 127–142. https://doi.org/10.1037/1076-898X.8.2.127

Saunders, J., & MacLeod, M. D. (2006). Can inhibition resolve retrieval competition through the control of spreading activation? *Memory & Cognition*, *34*(2), 307–322. https://doi.org/10.3758/BF03193409

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review*, *2*(2), 249–253. https://doi.org/10.3758/BF03210965

Spitzer, B., & Bäuml, K.-H. (2007). Retrieval-induced forgetting in item recognition : Evidence for a reduction in general memory strength. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(5), 863–875.

Spitzer, B., Hanslmayr, S., Opitz, B., Mecklinger, A., & Bäuml, K.-H. (2008). Oscillatory Correlates of Retrieval-induced Forgetting in Recognition Memory. *Journal of Cognitive Neuroscience*, *21*(5), 976–990. https://doi.org/10.1162/jocn.2009.21072

Staudigl, T., Hanslmayr, S., & Bäuml, K.-H. T. (2010). Theta Oscillations Reflect the Dynamics of Interference in Episodic Memory Retrieval. *Journal of Neuroscience*, *30*(34), 11356–11362. https://doi.org/10.1523/JNEUROSCI.0637-10.2010

Storm, B. C., Bjork, E. L., Bjork, R. A., & Nestojko, J. F. (2006). Is retrieval success a necessary condition for retrieval-induced forgetting? *Psychonomic Bulletin & Review*, *13*(6), 1023–1027. https://doi.org/10.3758/BF03213919

Storm, B. C., & Levy, B. J. (2012). A progress report on the inhibitory account of retrieval-induced forgetting. *Memory & Cognition*, *40*(6), 827–843. https://doi.org/10.3758/s13421-012-0211-7

Veling, H., & van Knippenberg, A. (2004). Remembering Can Cause Inhibition: Retrieval-Induced Inhibition as Cue Independent Process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 315–318.

Verde, M. F. (2013). Retrieval-induced Forgetting in Recall: Competitor Interference Revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1433–1448. https://doi.org/10.1037/a0032975

Verde, M. F., & Perfect, T. J. (2011). Retrieval-induced forgetting in recognition is absent under time pressure. *Psychonomic Bulletin & Review*, *18*(6), 1166–1171. https://doi.org/10.3758/s13423-011-0143-4

Wilding, E. L. (2000). In what way does the parietal ERP old/new effect index recollection? *International Journal of Psychophysiology*, *35*(1), 81–87. https://doi.org/10.1016/S0167-8760(99)00095-1

Williams, C. C., & Zacks, R. T. (2001). Is retrieval-induced forgetting an inhibitory process? *The American Journal of Psychology; Urbana*, *114*(3), 329–54.

APPENDICES

Appendix A: list of Welsh words used in the current investigation with their respective English translations

Welsh noun	English equivalent	Welsh noun	English equivalent
olwyn	wheel	granc	crab
caddair	chair	llywnog	fox
fwrdd	table	aderyn	bird
llwy	spoon	mochyn	pig
ty	house	ci	dog
pel	ball	madfall	lizard
trol	cart	gwydd	goose
gwely	bed	llyfant	frog
rhaw	shovel	pysgod	fish
faddon	bathtub	blaidd	wolf
nodwydd	needle	llew	lion
oergell	fridge	gloyn byw	butterfly
tegell	kettle	buwch	cow
cloi	lock	gwiddon	witch
haenell	plate	lleian	nun
morthwyl	hammer	dewin	wizard
telyn	harp	dyn tan	fireman
tatws	potato	athro	teacher
cwych	boat	plymiwr	diver

Appendix B: Examples of picture stimuli used in the current investigation, for picture set 1 (used in Welsh learning phase), set 2 (used in retrieval-practice phase) and set 3 (used in Welsh test phase)