Alliteration and assonance as mnemonic devices in second language word-pair learning

Michael Green

Centre for Language and Communication Research

School of English, Communication and Philosophy

Cardiff University

Thesis submitted for the degree of PhD

Cardiff University

December 2022

Abstract

The central question addressed in this thesis is to what extent phonological patterns, in particular alliteration and assonance, aid the recall and retention of word-pairs for Japanese L1 learners of English.

The research builds on previous findings from a series of classroom-based quasi-experimental work, principally from the team of Boers, Lindstromberg and Eyckmans, which shows a mnemonic advantage for collocations and compounds that have phonological patterns, compared to equivalent word strings with no phonological overlap. This advantage appears in both free- and cued-recall tests, and across a variety of temporal intervals (up to two weeks). Much of the prior research has drawn participant samples from a Dutch L1 speaking population. Furthermore, these studies have mainly used target items deemed to be familiar to the participants.

This thesis is motivated by the need to question if the previous empirical findings generalise to a population whose L1 phonological constructs are different from those of Dutch L1 speakers. The purpose is to test if Japanese L1 speakers have a different perception of alliteration and assonance, and if so, whether this impacts on their learning behaviour. A further aim is to investigate whether the mnemonic effect applies to unfamiliar target items. In addition, the thesis considers the extent to which the cognitive process of form-based priming underpins the mnemonic effect.

A series of four experiments are conducted which progressively examine the processing advantage conferred by alliterating and assonating patterns. Different sets of experimental stimuli are used, including high-frequency, low-frequency, and pseudoword items. Treatment phases often incorporate a dictation activity when using familiar word stimuli, or a study phase when using unfamiliar stimuli. A variety of testing instruments are adopted to measure recall of the written forms of the stimuli, or the forms plus meanings of novel stimuli, over differing periods of time. One study uses a Lexical Decision Task to ascertain if phonological patterns aid lexical processing.

Overall, the findings indicate that phonological patterns do confer a small mnemonic advantage for known stimuli, though the effect dissipates with time. However, the extent to which orthographic similarity plays a facilitatory role remains unclear. When participants are asked to learn novel word-pairs the results are more ambiguous; alliteration seems to have a greater mnemonic effect than assonance, but the cognitive challenge of learning new material appears to mitigate any robust mnemonic effects. The data from the Lexical Decision Task do not support any strong claims that perceptual priming is the determining factor for the processing advantage.

In answer to the central question, it can be inferred from the findings that both phonological and orthographic patterns are a useful pedagogical tool for helping language learners recall and retain multi-word strings.

Acknowledgements

This thesis would not have been completed without the unfailing support and encouragement of my supervisor Professor Alison Wray. I would also like to thank my second supervisor Dr. Andreas Buerki for his guidance and suggestions on the statistical analyses.

Special thanks to Kimi for coping so well with a cantankerous curmudgeon. I would also like to express my gratitude to my mum and sister for their moral support. Finally, I appreciate the students who generously participated in this research.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Focus and contextualisation

The start of my TEFL (Teaching English as a Foreign Language) career coincided with the publication of what turned out to be an influential book for my professional development, *The Lexical Approach* (Lewis, 1993), which still remains on the recommended reading lists of many EFL teacher-training syllabi, for example, the Cambridge Diploma in Teaching English to Speakers of Other Languages (DELTA). The core premise of the approach is to encourage students to tackle language learning at a larger scale than the individual word, to look at vocabulary in terms of multi-word 'chunks'.

This seminal book introduced me to the importance of collocational knowledge and also challenged my assumption that a language consisted essentially of its productive rules of grammar, with words just slotted in. Indeed, many of my learners, across disparate cohorts, appear to still have consistent difficulty with noticing, remembering, and producing the lexical combinations that are a mark of second language (henceforth L2[1]) proficiency. This inability to forge appropriate links does not always affect intelligibility, but it can distract from the message, and may even "irritate the recipient" (Nesselhauf, 2005, p. 3). Some examples generated by my students are given here:

[1] In the fields of EFL and applied linguistics, the terms 'native speaker' and 'non-native speaker' are common reference points. They do, however, lack precise definitions, and to some may have inherent assumptions about the superiority of the former and the inferiority of the latter. For these reasons, the terms L1 and L2 will be used; for example, L1 English speakers refers to those who have English as a 'mother tongue', and L2 English speakers refers to those who are learning English as a second (third etc.) or foreign language.

Example 1.1 Learner Errors.

There was a girl **driving a bike*. (riding a bike)
We went to the countryside and **made tents*. (pitched tents)
I **made a trip* to Mount Fuji. (went on a trip)

But how are students who, perhaps, are not able to notice chunks easily, supposed to learn to identify and remember them? This thesis addresses one aspect of the possible answer.

Paul Nation (Nation, 2001, p. 502; 2014, p. 31), arguably one of the most influential vocabulary researchers among current EFL practitioners, notes that chunks often have characteristic phonological patterns, such as alliteration and assonance (defined in section 1.4), and advises L2 learners to pay deliberate attention to them, advice that has more recently been echoed by Szudarski (2017).

As a first step to the systematic exploration of this suggestion, Chapter 1 establishes some basic information relating to it. Firstly, are multi-word chunks sufficiently widespread in the target language to merit study? Secondly, are the difficulties I observed in my own students common to many or most L2 learners? Thirdly, are phonological patterns themselves common within these chunks? Fourthly, do such patterns facilitate acquisition for L2 learners?

The first two questions are addressed below (1.2 and 1.3, respectively). The third will be considered in section 1.5, and the fourth question is the basis for the studies reported in this thesis. Thus, in brief, this thesis explores the extent to which Nation's (2013, 2014) and Szudarski's (2017) advice is useful. Specifically, the research reported here explores ways of supporting Japanese learners of English into increasing their retention of multi-word strings, by testing their capacity to attend to phonological features within them.

1.2 The extent of multi-word chunks in English

Nation (2001, 2014) implies that multi-word chunks are pervasive enough to deserve targeting by teachers and learners. But what evidence is there for that assumption? One of the challenges with answering that question is pinning down what usefully counts as a multi-word chunk. A great many different types have been identified, and an inexhaustive list could include the following: binomials (for instance, *safe and sound*), collocations (for example, *deep sea*), compounds (such as *wrapping paper*), discourse markers (*on the other hand*), idioms (*have bigger fish to fry)*, phrasal verbs (*come across*), proverbs (*the early bird catches the worm*), sentence stems (*The most likely reason for … was* …), and similes (*as busy as a bee*). Before we can assess how pervasive such multi-word chunks are, we need a definition, to establish what should be counted. Terminology for the phenomenon differs according to the purpose of the research and theoretical orientation. For example, Wray (2002), with a focus on modes of processing, offers the following definition of *formulaic sequence* (FS):

> A sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (p. 9)

One advantage of this definition is that it is maximally broad - anything that looks like it might be prefabricated is included, making *formulaic sequence* a useful umbrella term. Thus, the term covers chunks such as idioms that are semantically opaque, that is, their meaning cannot be gleaned from the individual components (for example, *beat around the bush*), chunks that are syntactically irregular and cannot be generated from a grammatical rule (for example, *by and large*), and chunks that are semantically transparent and syntactically regular (for example, *It was lovely to see you*) (examples from Wray, 2000). Although there is little in this thesis that requires, or assumes, that the vocabulary items investigated are holistically stored and processed, the term formulaic sequences (FSs) will be used as and when the phenomenon is mentioned. It should be noted that many of the types of FS are not of primary interest, as the focus on phonological patterns invites a relatively narrow subset, as will be outlined later.

Although it may seem intuitive that FSs are ubiquitous in English, the full extent of formulaic language is hard to calculate, partly because the boundaries of what constitutes a formulaic expression are fuzzy. Some published estimates have been made based on intuition, for example, Pawley & Syder's (1983, p. 213) figure of 'several hundreds of thousand', while other estimates have been based on the analysis of frequency of occurrence in large corpora[2] (see Table 1.1).

Table 1.1 Examples of estimates of the extent of Formulaic Sequences (FSs) in English.

Authors	Estimate	Basis for Estimate
Sorhus (1977)	one FS every five words	130,000-word corpus of spoken Canadian English
Jackendoff (1995)	approximately 80,000 FSs	Quiz-show corpus
Altenberg (1998)	80% of the corpus	London-Lund Corpus of Spoken English (approximately half a million running words)
Erman & Warren (2000)	52% of the written corpus, 58% of the spoken corpus	London-Lund Corpus of Spoken English, Lancaster-Oslo-Bergen Corpus of written English, one novel

Figures vary according to the nature of the texts in the corpus, with written and spoken discourse having different forms and distributions of formulaic language, and because of differences in what is being counted as a FS. For example, Altenberg's

[2] Corpus estimates based on frequency of occurrence do not take into account other defining characteristics of FSs such as familiarity within an L1 speaking community, predictability, pragmatic function, and fixedness (Siyanova-Chanturia & Pellicer-Sánchez, 2019, p. 3).

(1998) corpus analysis focused on 'recurrent word combinations', which are defined as continuous strings of 3 - 5 words occurring at least ten times in identical form. Such a definition yields a large estimate for the pervasiveness of formulaic language in spoken discourse, though this number includes repetitions and fragments of larger chunks (for example, *if you see* and *out of the*), which would seem unusual vocabulary items for a learner to focus on. In contrast, Erman & Warren (2000) calculated the number of 'prefabs', the defining criterion for which is 'restricted exchangeability' in which one member of the chunk cannot be replaced by a synonym without changing the meaning, function or idiomaticity: for example, *not bad* cannot be changed into *not lousy*, or *I'm afraid*, used to soften bad news, cannot be changed to *I'm scared* (Erman & Warren, 2000, p. 32). Interestingly, their working definition of a prefab excludes compositional collocations - for example*, dark night,* where neither word determines the choice of the other - on the grounds that such expressions fail to meet the determining criterion of 'restricted exchangeability'.

Estimates such as those in Table 1.1 have led to the generally accepted notion that FSs permeate adult L1 discourse. Furthermore, some types of FS, namely collocations and compounds, are added to the English lexicon at a faster rate than new words (Hanks, 2013). Such ubiquity has led to a consensus that FSs should be included in L2 pedagogy. However, in addition to the sheer scale of the learning task, it is also commonly accepted that L2 learners struggle to master this aspect of linguistic competence, as we shall see in the following section.

1.3 Why might formulaic sequences be difficult for L2 language learners?

1.3.1 Persistent challenges in L2 learning and teaching

Every EFL teacher is familiar with being met with a learner's utterance that is comprehensible, grammatically sound, pragmatically appropriate, yet sounds 'odd'.

'*That's not quite how we would say it'* may be a common response, perhaps followed, if the situation demands, by a suggestion of a more idiomatic reformulation.[3] Pawley & Syder (1983) note that few L2 learners ever acquire the full repertoire of FSs; instead they over-generalise and produce 'unnatural or highly marked' speech, for example, **You are pulling my legs* (ibid. 1983, p. 215). It seems that learning FSs poses "an immense problem for even the most proficient" of L2 learners (Wray, 2000, p. 468), and there are claims that a lack of competence in producing the right word combinations is evident in any analysis of a learner's L2 output (Hill, 2000, p. 49). Several explanations have been proposed as to why these chunks are "very difficult to learn" (Irujo, 1986, p. 236). Three broad themes can be discerned in these explanations: the nature of the chunks themselves, factors related to the learning environment, and factors related to teaching and testing, and these will now be considered in turn.

Possibly one of the most challenging aspects of formulaic language is easy to overlook – much of it is composed of multi-word strings, and it is very unlikely a learner will be able to combine words if they do not know one or more of them. Moreover, the multi-faceted nature of formulaic language means much of it can be non-literal, and / or syntactically irregular. As a result, even if a student knows the meaning of the noun *bean* and the verb *spill*, the meaning of the phrase *spill the beans* will still remain opaque. Likewise, changing the grammatical structure of a FS can produce deviant combinations – the modal verb cannot be removed from *it'll do* (**it does*), and *I guess* cannot be negated (**I don't guess*) without a change in meaning or function. Lack of saliency is another factor; learners may not notice the bonds between words, particularly short words that look familiar and may seem semantically transparent (for example, *put up with)*. These bonds traverse the blank spaces between the individual orthographic units, spaces which predispose our literate selves to conceive of language as composed only of word-sized building

---

[3] As Taylor (2012, p. 282) notes, the term 'idiomatic' has dual meanings. Here it is used in the sense of conforming to usage norms, as opposed to referring to a semantically opaque or non-compositional phrase, like *kick the bucket*.

blocks. Furthermore, learners often cannot fall back on a literal translation equivalent due to a lack of congruency between the L1 and L2, hence the common errors in V+N collocations, for example, **do a mistake* or **say the truth*. An additional potential impediment is the deceptive cognate, or false friend, for example, the Dutch equivalent of *find out* means invent, rather than discover.

Perhaps the most effective way to pick up FSs is by encountering them regularly in the output of reliable speakers and writers (Pawley & Syder, 1983; Sinclair, 1991). However, the learning environment, particularly in an EFL setting, means there is a dearth of exposure to the target language or practice opportunities beyond the limited class-time, especially compared to the massive amounts of input involved in L1 acquisition, so students cannot pick up FSs in the ambient dominant language. Any incidental learning opportunities, such as encountering FSs in a book, film, or song, for example, do not allow for any negotiation of meaning, or clarification, or feedback on production. Learning in the L2 environment, of course, will alleviate this problem to some extent.

The third explanation as to why there are persistent challenges in acquiring FSs relates to teaching materials and syllabi. There seems to be little systematic treatment of FSs – even if Irujo's (1986, p. 237) comment that "we do not teach them very well" is not quite so valid as it was 35 years ago. This may partly be to do with the complexity of the phenomenon, or partly the fact that the single word reigns supreme in many EFL textbooks. If there is a focus on formulaic language, textbook exercises often split up the combinations and rely on a trial-and-error approach to matching the components, followed by corrective feedback, with little to show for in terms of learning gains (see the case made by Boers, Demecheleer, Coxhead, & Webb, 2014; Strong & Boers, 2019).

Finally, the impact of testing cannot be overlooked. In my personal teaching context, many of my Japanese L1 undergraduates seem to equate 'vocabulary learning' with poring over bilingual vocabulary books, often containing decontextualised, low frequency, individual words – a practice that has been noted as a demotivating factor for Japanese learners of English (Kikuchi, 2013, p. 215). This habit is, perhaps, a result of negative washback from Japanese university entrance exams, which seem to pose a considerable lexical burden on candidates.

For example, Kaneko's (2013) analysis of the Tokyo University entrance exam found 98% text coverage, deemed necessary for adequate comprehension (Schmitt, Jiang, & Grabe, 2011), required knowledge of about 5,000 word families, much higher than the 3,700 word families that the average Japanese student is estimated to leave high school with (McLean, Hogg, & Kramer, 2014).

In sum, learning L2 FSs can be a slow and challenging process. The intervention proposed by Nation (2001, p. 502; 2014, p. 31) to accelerate acquisition involves one specific feature of many formulaic word strings - phonological patterning. Such patterning does not account for all FSs, but there is a notable tendency (see section 1.5), especially for those intended to be memorised. The next section will look at the connections between phonological patterns and memory.

1.3.2 Phonological patterns and memory

In Peters' (1983) seminal account of first language acquisition, phonological patterns of similarity, such as rhyme, alliteration, and assonance, are posited as heuristic, or learning, devices that young children employ to segment or compare 'units' of meaning in the intermittent stream of speech sounds. Research by Jusczyk, Goodman & Baumann (1999) shows that even nine-month-old infants are already sensitive to alliteration in CVC syllables, which suggests that patterns of similarity are pertinent to building the L1 vocabulary. Phonological patterns also emerge in early childhood wordplay, reflecting the ludic, or playful, function of language (Crystal, 1997, p. 198), as can be seen in the rhyme and alliteration in the unsolicited spontaneous poem from a 3-year-old child (Example 1.2, from Dowker, 1989, p. 192):

Example 1.2 Child's Poem.

> There was a lill-lull, lill and a lell.
>
> Little light flashed on it.
>
> Little lollipop sat on his bum.

There is a longstanding and widespread use of phonological patterns, sometimes referred to as rhetorical devices, in advertising, slogans and brand names (such as *Coca Cola*, *Kit Kat*, *Tim Tams, Dunkin' Donuts*), fictional characters (*Donald Duck, Wonder Woman, Scooby Doo*), nursery rhymes, and of course, poetry. Such patterns also appear in meticulously crafted works of prose, see Example 1.3.

Example 1.3 from Dylan Thomas, Portrait of the Artist as a Young Dog (1940).

> Mr. Farr trod delicately and disgustedly down the dark, narrow stairs like a man on ice.

In short, phonological patterns seem omnipresent in the English language, which begs the question: why? The fact that two consecutive words will share a sound is a statistical probability, given a finite phonemic and orthographic inventory, though this does not explain why, for example, the alliterating phrase *belly button* is far more common than the non-alliterating counterpart *tummy button*,[4] or why we say *time will tell*, rather than **time will say.* Some researchers have hypothesised that phonological patterns play a key role in the incremental conventionalisation of word strings. For example, the central claim of Benczes (2013) is that phonological patterns motivate the selection of components in novel compounds, such as *cuddle puddle*, *snail mail,* and *street spam*. In such compounds, phonological patterns serve several functions, such as enhancing salience, helping the listener decipher the meaning of a novel expression,[5] and, of particular relevance to this thesis, aiding long-term retention (Benczes, 2013, p. 168) – for a similar suggestion, see Boers & Stengers (2008) and Gries (2011).

---

[4] A Google search shows approximately 120 times more instances for *belly button*.

[5] For example, if a reader / listener comes across the creative compound *brain gain*, the phonological similarity with *brain drain* can help to decode the meaning (example from Benczes, 2013, p. 170).

It seems almost axiomatic that phonological patterns can act as an aid to memory; for instance, rhyme is used to learn the fates of the six wives of Henry VIII (‘*Divorced, beheaded, died, divorced, beheaded, survived.*’), the number of days in the month (*Thirty days hath September, April, June and November …*), and historical events (*In fourteen hundred and ninety-two, Columbus sailed the ocean blue*). The literature on Oral Traditions and the genres of epic and ballads (for a comprehensive survey see Rubin, 1995) proposes that phonological patterns, together with meaning and imagery, are constraints that cue memories and restrict choices, enabling thousands of lines of songs, stories and poems to be kept in memory and transmitted for centuries.

Yet, as we shall see in Chapter 2, it can be surprisingly hard to pin down compelling empirical evidence that phonological patterns actually do have a mnemonic effect.

1.3.3 Overview of the Thesis

The central question of the thesis is to what extent phonological patterns might help L2 learners internalise new word strings. Existing studies, reviewed in Chapter 2, suggest they might, albeit only under certain circumstances. Inspired by those studies, this thesis will progressively explore what phonological factors might contribute to making new word pairs more memorable. The remainder of this chapter will define the key terms and then address the central background question of how common phonological patterning is in particular types of English formulaic language. Chapter 2 is a critical engagement with the current research literature that contextualises and justifies the studies that follow, culminating in specific research questions. Chapters 3 to 6 report four quantitative investigations that empirically explore the series of research questions, prompted by gaps and questions arising from the current research. Chapter 7 pulls together conclusions from those studies and relates them back to the research literature, so as to evaluate the extent to which the research questions were answered and where it leaves us. In response to these conclusions, the discussion then broadens to consider the determinants of L2 learners' internalisations of multi-word input. Chapter 8 presents a short conclusion with pointers to opportunities for future research.

1.4 Defining terms

1.4.1 The importance of defining key terms in research

According to Pierrehumbert, Beckman and Ladd (2000, p. 278), one of the hallmarks of a successful scientific community is the 'maintenance of a common vocabulary'. Without an established nomenclature there could be no systematic enquiry; after all, if you are not measuring the same thing then you cannot replicate and verify it, and little scientific progress can be made. Such a common vocabulary could be considered especially important in the broad field of Applied Linguistics, dealing as it does with the messy, highly complex natural phenomenon of language, where words may be used with senses that differ among sub-disciplines as well as from colloquial usage. Establishing a terminology is not always a straightforward task (c.f. the difficulties even in determining what constitutes a 'word' in Wray, 2015).[6]

Therefore, defining terms is critical - to ensure central concepts and terminology are shared, to reduce ambiguity when evaluating the strengths and generalisability of claims, and to avoid misunderstanding. The terms most central to the ensuing discussion are defined here. Other terms will be defined later as they become more relevant, for example 'priming' (section 2.6) and 'mora' (Chapter 3).

---

[6] Notwithstanding the many subtle and technical issues associated with the exact definition of 'word', as explored by Wray (2015), for the present purposes a 'word' will be a spoken form that, when written, would be represented by a string of letters with a space either side.

1.4.2 'Phonological patterns'

In contrast to a phonetic analysis or description, which is at the level of fine-grained and non-discrete concrete sounds, a phonological analysis or description is more cognition-centric, involving a "reduction to the essential information, to what speakers and hearers *think* [emphasis added] they are saying and hearing" (McMahon, 2002, p. 3). English L1 speakers, on hearing the political soundbite '*bung a bob for a Big Ben bong'*,[7] would probably not deny that the words start with the 'same sound', and this is the entry point for conceptualising a 'phonological pattern' in the context of this research. In this thesis, the term 'phonological pattern' will have the very narrow meaning of 'linguistic sounds recurring'. This, of course, opens up the question of what exactly counts as a recurrence, and how soon that recurrence would need to take place to constitute a 'phonological pattern', an issue that will be revisited at various points during the thesis.

At the more general level, in terms of determining what constitutes a 'pattern', perhaps the simplest one would be reduplication, that is, full repetition of the whole word (for example, *bye-bye*, *boo-boo*). Reduplication may also be partial, involving rhyme (for example, *razzle-dazzle, easy-peasy*), or it may occur as ablaut reduplication,[8] where the first word usually has a high vowel sound and the second a low vowel sound (*tip-top*, *wishy-washy*). Reduplication receives a lot of attention in the linguistics literature, perhaps as it often involves the interaction of phonology and morphology. Reduplication appears to be a common pattern across all languages, though some languages may employ it more than others: Crystal (1997, p. 176) notes, for example, that in the context of onomatopoeic expressions, Japanese

[7] Alexander Boris de Pfeffel Johnson in a BBC Breakfast interview, quoted in *The Guardian* Jan 14th 2020.

[8] Ablaut reduplication in English is characterised by identical vowel *quantity* in the stressed syllables and distinct vowel *qualities* in the two halves (Minkova, 2002).

favours reduplicative forms (equivalent to English *pitter-patter* and *ding-dong*), while English more often uses one-word forms (e.g. *bang*, *splash* and *plop*).[9]

The two phonological patterns at the heart of this thesis are alliteration and assonance, which both fall within the extensive taxonomy of rhyme. Even though literary definitions are the obvious starting point, in fact, they are of only limited help. *The Princeton Encyclopedia of Poetry and Poetics* (Greene et al., 2012, p. 1189) reports that definitions of even 'rhyme' itself are "unsystematic and inconsistent". Many of the debates central to the appreciation or technical construction of poetry will shed little light on the cognitively oriented questions arising from the focus on learnability in this thesis. Nevertheless, it turns out that neither 'alliteration' nor 'assonance' even has an entry in linguistics dictionaries such as *A Dictionary of Linguistics and Phonetics* (Crystal, 2011), so recourse must be made to other works of reference. The brief overviews that now follow therefore navigate the boundary between literary and cognitive perspectives.

1.4.3 'Alliteration'

Alliteration is sometimes labelled 'head rhyme' or 'initial rhyme' and broadly refers to "the repetition of the same sounds – usually initial consonants of words or of stressed syllables – in any sequence of neighbouring words" (Concise Oxford Dictionary of Literary Terms, Baldick, 1996, p. 96). However, the precise parameters of what counts as alliteration are underspecified in that definition. If we compare it with the definition in *The Princeton Encyclopedia of Poetry and Poetics (Greene et al., 2012, p. 40)*, "the repetition of the sound of the initial consonant or consonant cluster in stressed syllables close enough to each other for the ear to be affected",

---

[9] Interestingly, Minkova (2002, p. 139) notes that novel word formation in English using ablaut reduplication has come to a 'virtual standstill': there were only three attested word-formations in the twentieth-century using this form of patterning: *ping-pong* (1900), *ticky-tacky* (1960) and *hip-hop* (1985) in the Random House compact unabridged dictionary.

we see that alliteration is dependent on what affects the hearer, which in turn delimits the distance between the syllables.

Before considering some questions arising from these definitions, it is worth noting where they agree. Firstly, both sources assert that alliteration is concerned with sounds,[10] not orthography, so *kitchen cupboard* presumably alliterates for English L1 speakers, whereas *cellar cupboard* does not. Secondly, both definitions refer to initial consonants and / or initial consonant clusters. It should be noted that initial vowels are excluded. This is because, in this sort of formal literary definition, all vowel-based patterns constitute 'assonance' (see section 1.4.4). Thirdly, alliteration is only deemed a feature of stressed syllables. These two exclusions mean that a phrase such as *important information* would not be viewed as alliterative, since it features initial vowels in unstressed syllables. Fourthly, notwithstanding the mention in the Oxford definition that alliteration is "usually" on the initial syllable, the definitions do not limit alliteration to word-initial position. Interestingly, this means that a word-pair like *in**te**lligent substi**tu**tion* alliterates according to these sources.

A rather vague aspect of the definitions is their handling of proximity – "close enough" (Princeton) and "any sequence" (Oxford). The lack of parameters sails close to the '*I'll know it when I see it*' approach. For instance, in Example 1.4, a reader is likely to be able to identify the alliteration:

Example 1.4 from Dylan Thomas, Under Milk Wood (1954, p. 4).

> Come closer now.
> Only you can hear the houses sleeping in the streets in the slow deep salt and silent black, bandaged night.

---

[10] The term 'sound' is used here, rather than a more technical term such as 'phoneme', because of questions regarding whether alliteration does indeed require phonemic repetition, or whether similar phonemes can also alliterate by virtue of sharing sub-phonemic features. This issue will be addressed further in Chapter 7.

Possible instances are the initial /k/ in line 1, the repeated /s/ patterns and the repetition of /b/. But should we view *now* as alliterating with *night* or does this violate the constraint of what constitutes ‘close enough’? Presumably the alliterating words must fall within the range of an average working memory span, and the cognitive orientation of this thesis will favour considering this aspect. But it cannot be assumed that all memory spans are the same, nor that the poet aims to anticipate what they are for each potential reader or listener. Furthermore, we cannot ignore the potential for the written form to suggest alliteration that the ear would not pick up due to distance: the written form enables material to be revisited, whereas in audition the message is heard and then quickly fades.[11]

In the context of a linguistic study, the key question is how alliteration should be operationalised in designing stimuli. This turns out to be quite intricate, but those discussions will be postponed until after the existing research on alliteration and assonance in language learning has been reviewed in Chapter 2 and the details of methodological decisions are explored. For now, it will suffice to stipulate that, in the context of this thesis, ‘alliteration’ is the repetition, in two consecutive words, of the same consonant sound in an initial stressed syllable. This definition precludes repetition of initial vowels and instances of consonance (intra-word consonant repetition, for example *cu**rr**ent t**r**end*) for practical reasons relating to the research designs used later. The definition also includes cases where one word starts with a consonant cluster, see *black* and *bandaged* in Example 1.4, the justification being that this seems to be standard in poetry, as can be seen in the f-fl-fr and br-bl-b patterns found in Example 1.5:

Example 1.5 from Samuel Taylor Coleridge, The Rime of the Ancient Mariner (1834).

The fair breeze blew, the
white foam flew,
The furrow followed free:

[11] In a similar vein, ‘eye rhyme’ only applies to the written form, for example, *plough, cough, through, dough.*

We were the first that ever
burst
Into that silent sea.

1.4.4 ‘Assonance’

The other phonological pattern central to this thesis is assonance, otherwise known as ‘vowel rhyme’ or ‘vocalic rhyme’. The *Concise Oxford Dictionary of Literary Terms* (Baldick, 1996, p. 26) defines it as “the repetition of identical or similar vowel sounds in the stressed syllables (and sometimes in the following unstressed syllables) of neighbouring words”, and in *The Princeton Encyclopedia of Poetry and Poetics* (Greene et al., 2012, p. 94) as “the repetition of a vowel or diphthong in nonrhyming stressed syllables near enough to each other for the echo to be discernible”. As with alliteration, both sources agree that sounds are of paramount concern rather than spelling, so English L1 speakers would probably agree that *thin* and *night* do not assonate in Example 1.6 below, whereas *breeze*, *creased* and *streets* do.

Example 1.6 from Dylan Thomas, Under Milk Wood (1954, p. 62).

The thin night darkens.

A breeze from the creased water sighs the streets close under Milk waking Wood.

Is there, however, a ‘discernible echo’ between *thin* and *Milk*? This imprecise aspect of proximity, in both assonance and alliteration, is not critical to this thesis: the focus will be on the effect of phonological patterns in consecutive words, but it does to some extent highlight the fuzziness of the concepts involved. In this thesis, assonance is operationalised as the repetition of a vowel or diphthong sound in the prominent syllable of two consecutive words, for example, *main gate*.

1.4.5 ‘Word-pair’

The term ‘word-pair’ is used here to mean real or artificial words (non-words or pseudo-words) adjacent to each other in natural and / or artificial settings, intended to be learnt as a single unit. Essentially, they are an experimental tool, and the definition includes those aspects of vocabulary that would conventionally be called collocations and compounds. Additional features of the ‘word-pair’ will be mentioned in the context of individual experiments, as they become relevant.

1.5 How extensive are phonological patterns in English?

An alliterative or assonating pattern may be of little mnemonic value to a second language learner if the pattern rarely occurs in the speech and writing of the target language community. However, Nation (2014, p. 31) states that “about 20% of phrases in English”[12] display some sort of phonological patterning. This figure appears to be based on the publications from a group of researchers whose experimental findings, reviewed in detail in Chapter 2, are briefly summarised below.

By doing hand counts, mostly from established printed dictionary sources,[13] these researchers, most prominently a team led by Frank Boers, have been able to arrive at an estimate for the extent of certain phonological patterns in a variety of FSs (see Table 1.2, in which the examples are selected from those given in the original

---

[12] Nation (2014) uses the word ‘phrases’ as a broad term for a variety of FSs, including collocations and idioms.

[13] Specifically, the *Oxford Dictionary of Idioms*, the *Collins Cobuild Dictionary of Idioms*, and the *Macmillan English Dictionary for Advanced Learners.* As these hand counts were done over a considerable span of time, different editions of the dictionaries were consulted.

sources). Variation in the totals is to be expected when doing hand counts from different published sources because, for instance, publishers and their lexicographical teams may split phraseological categories into idioms or proverbs in different ways, or one dictionary may classify an item as a 'phrase' rather than as an 'idiom' (Moon, 2015, p. 319).

Table 1.2 Hand-counts of figurative FSs displaying alliteration and assonance.

Phonological Pattern	Type of FS	% of FSs having the pattern	Example
Alliteration	idioms	12.7 [A] – 15 [C]	*through thick and thin*
	binomials	28 [B]	*black and blue*
	similes	42 [B]	*as right as rain*
Assonance	idioms	7 [D] – 13.5 [E]	*a dead end*
	binomials	12.7 [E]	*high and mighty*
	proverbs	24 [F]	*time flies*

[A] Boers & Lindstromberg (2005, p. 227) [B] Lindstromberg & Boers (2008a, p. 202).

[C] Boers & Stengers (2008). [D] Boers & Lindstromberg (2009, p. 114).

[E] Boers, Lindstromberg & Eyckmans (2014b, p. 94). [F] Lindstromberg (2012).

To overcome other methodological issues, researchers have established protocols for framing how they search through the reference material: for example, counting main entries only, and limiting their focus to verb / noun content words only (and so excluding, for example, items like *drop off* or *log on* from the count, even though they assonate).

A further source of discrepancy in the estimates in Table 1.2 lies in the fact that some items have alternatives, for example *beat one's breast / chest,* or *go to ground / earth*; how such phrases are treated in the counting process may affect the percentages. It should also be noted that the estimates in Table 1.2 do not include multi-word units displaying reduplication (for example, *shoulder to shoulder*), rhyme (for example, *fat cat*), consonance (for example, *above board*), or mixed phonological patterns such as alliteration + assonance (for example, *carry the can*).

Thus, the overall amount of phonological patterning in these types of FS is probably higher.

The above estimates regarding the extent of alliteration and assonance are based on dictionary sources. This is perhaps due to the fact that idioms and proverbs have proved difficult to detect in real-world contexts, despite them forming part of the “stock-in-trade” of vocabulary teaching (McCarthy, 1998, p. 4). The ability of idioms and proverbs to evade comprehensive corpora searches (c.f. Moon, 1998, 1999) lies in their relative scarcity and their potential for variability in terms of “inflection, open slots, adjectival or adverbial modification and passivisation, … derivation, compounding, negation, [and] distribution over multiple clauses” (Hughes, Filimonov, Wray, & Spasić, 2021, p. 281). Thus, though the canonical form of figurative phrases may occur infrequently in a corpus (Moon, 2015), there is evidence that they occur more often in guises that deviate from their dictionary entry-type form (Charteris-Black, 1999; Philip, 2005).

Since idioms and proverbs are typically described as ‘marginal’ items in the English language, attempts have been made to quantify the extent of phonological patterns in other, more ubiquitous, types of unit, namely collocations and compounds – see Table 1.3 (examples are taken from the sources where available; underlined examples have been inferred from the published descriptions).

As Table 1.3 indicates, the detected proportion of collocations and compounds that have a phonological pattern is much lower than that for idioms and the like (Table 1.2). However, this might not be entirely attributable to that distinction of types.

Table 1.3 Estimates for collocations and compounds displaying alliteration and assonance.

Phonological Pattern	Type of FS	%	Example
Alliteration	academic collocations [A]	2.8 [B]	*capitalist country*
	N + N compounds [C]	7.8 [B]	<u>*baby boom*</u>
Assonance	academic collocations [A]	5.5 [B]	*mental health*
	Adj + N collocations [D]	9.7 [D]	*rich history*
	N + N compounds [C]	7 [B]	<u>*town house*</u>

[A] Based on the Academic Collocation List (Ackermann & Chen, 2013).

[B] Lindstromberg (2017).

[C] Based on the 5000 most frequent monosyllabic word lemmas in the COCA corpus (Davies, 2008 to present).

[D] Boers, Lindstromberg & Eyckmans (2014b, p. 95) looked at the 10 most frequent noun collocates of the 100 most frequent monosyllabic adjectives in the COCA corpus (Davies, 2008 to present).

The figures for the Academic Collocations List (ACL) (Ackermann & Chen, 2013) are consistently lower than those derived directly from corpora. Lindstromberg (2017, p. 6) suggests that the low incidence of alliteration in the Academic Collocation List (ACL) seen in Table 1.3 may partly be due to the fact that 21% of the content words start with a vowel sound (for example, *alternative approach*), and thus fall outside the scope of alliteration (and also assonance unless, they have an initial stressed syllable). In comparison, only 9% of the content words in the high-frequency Adj + N collocations used in Boers, Lindstromberg & Eyckmans (2014b)[14] start with a vowel

[14] My own counts of the 2469 collocations in the Academic Collocations List show initial vowel sounds in 19.9% of the left-ward components and 29.9% of the right-ward collocates. ‘Vowel sound’ was operationalised to exclude those sounds which are produced like vowels, but which pattern as consonants in English, such as vocalic glides /w/ and /j/ and the voiceless fricative /h/, on the grounds that an English L1 speaker would probably say that collocations such as *hot head* or *wet wood* alliterate rather than assonate. However, there is a lot of repetition of both the initial word and the second word across multiple entries in the ACL, for example, *annual conference* and *annual meeting,* and 42 entries have *information*

sound. To some extent, this reflects the fact that, as Lindstromberg (2017, p. 14) notes, most high-frequency words in English tend to be monosyllabic and start with a consonant, whereas many academic words may have vowel-initial prefixes (for example, *interact, interpersonal, international, unintended, unskilled, unlimited*).

Again, the estimates in Table 1.3 do not include phonological patterns such as rhyme, consonance, or even mixed alliteration / assonance. Nor do the estimates for the collocations include alliteration and assonance displayed in words beyond the scope of their search terms, such as prepositions or adverbs (for example, *well educated*). Taking such exclusions into consideration, it is likely that the figures in Tables 1.2 and 1.3 are conservative projections of the total extent of phonological patterns within certain types of FS.

What Tables 1.2 and 1.3 do not tell us is whether we should be surprised by the proportion of FSs that have a phonological pattern. What might we expect by chance? In order to establish an answer to that, Boers et al. have provided evidence supporting the idea that phonological patterns occur at above-chance levels in English. For example, Boers, Lindstromberg & Eyckmans (2014b, p. 95) took 197 binomials from the *Oxford Idioms Dictionary for Learners of English* (2006) and split each binomial after the conjunction to create a randomised list of 197 novel binomial combinations: for instance, *cut and run, win or lose* > *cut and lose, win or run*. By comparing the distribution of assonance in the arbitrary binomials (2.8%) with the distribution of assonance in the original binomials (12.7%), the authors concluded that the observed distribution of assonance in that particular repertoire was significantly higher than chance (Pearson's *chi*-squared test = 22.5, $p < .0001$).

Further evidence is supplied by Gries (2011), a study partly motivated by a previous corpora-driven analysis of the polysemy of the verb *to run* (Gries, 2006), in which it was found that many of the *run* phrases alliterated, for example, *to run rampant, to run riot, to run roughshod, to run the risk*. Gries (2011) investigated

---

as the second word, and this could skew the results. Further analysis on the rightward components was done, in which all instances of repetition were removed (but leaving morphological differences, so that *effect, effects, effective, effectively* count as four entries), leaving 738 unique entries. Of these, 28.7% (212) start with a vowel sound.

phonological similarity in 211 high-frequency[15] Verb + Noun Phrase idioms (for example, *turn the tables*) from the *Collins Cobuild Dictionary of Idioms,* and phonological similarity in *way* constructions (for example, *it wound its way*) from the British National Corpus (BNC).

Starting with the idioms, using word-initial phoneme type and token frequencies from the CELEX database[16] (Baayen, Piepenbrock, & Gulikers, 1996) on a random sample of 667 non-idiomatic Verb + Noun Phrases from the International Corpus of English (ICE), Gries calculated a baseline percentage of alliteration of 4.7% - 6% (the estimate depends on whether token or type frequencies of word-initial phonemes were used). The differences between these baseline percentages and those occurring in the 211 dictionary idioms (11.3% alliteration) were statistically significant.

A similar method was adopted to measure the amount of alliteration in 5,831 *way*-constructions in the BNC, with a baseline estimate based on a control group of transitive verb phrases with *way* as the direct object. Gries found levels of alliteration 2.44 times higher than the baseline percentage with regard to type frequencies, and 2.59 times higher with regard to token frequencies, and both differences were statistically significant according to exact binomial tests (2011, p. 502).

These findings indicate that particular types of English FSs have above-chance levels of certain phonological patterns. But is this true of all languages? For if it is not, it might impact on the capacity of learners to exploit the opportunities in English for using phonological cues to assist with learning and retention. If there are differences between languages, what might cause them? One possibility is differences in the proportion of words with initial stress: despite the fact that word stress could, in theory, be assigned to any syllable in a long word, stress is not

---

[15] Using the ordinal frequency labelling in the *Collins Cobuild Dictionary of Idioms*, which highlights those idioms with a frequency of once per 2 million words or higher, each target idiom occurred at least 105 times in the *Cobuild* corpus (Gries, 2011, p. 493).

[16] Although a comparatively small database with approximately 18 million tokens, the CELEX database has full phonological annotation.

evenly distributed across syllable-stress patterns in English, and there are strong tendencies for stress to occur in some positions more than others (Clopper, 2002). Cutler & Carter (1987) report that in their 200,000-word corpus of spontaneous British English conversation, 90% of the content words had an initial stressed syllable.

In contrast, as noted by Boers & Lindstromberg (2005), many Romance languages stress the final syllable in multi-syllabic words. Furthermore, the morpho-syntactic patterning of a language may obscure any phonological pattern – in Dutch, for example, the formation of a past participle often involves the insertion of a *ge-* prefix, and this consonantal addition could obscure an alliterative pattern. Boers & Lindstromberg (2005, p. 234) report hand-counts from a Dutch dictionary of idioms (de Groot, 1999) which suggest prototypical alliteration occurs in 7.3% of idioms, and their hand-counts from a French idiom dictionary (Rey & Chantreau, 1979) found only 7% of the target phrases alliterated (c.f. the estimated range of 12.7 – 15% for English idioms in Table 1.2). If, then, it is the case that speakers of some languages are more accustomed to encountering, and thus see the potential value of encountering, these patterns compared to speakers of other languages, then research into how well English L2 learners pick up on phonological patterns might do well to take into account what is going on in their L1, a point that will be returned to in Chapter 3.

While there are estimates that might, for reasons outlined above, be rather conservative, for the extent of alliteration and assonance in some varieties of FSs in English, there seem to be no current evaluations of the extent of phonological patterns for other types of units such as compounds (for example, *broadband, pickpocket*), discourse markers (for example, *It is safe to say that, first and foremost*), trinomials (for example, *cool, calm and collected*), or exclamations (*Good grief!*). There is more work to be done on calculating the range of phonological similarity in English, but it is beyond the scope of the present study.

Having said that, a full understanding of the distribution of phonological patterns in FSs must entail asking why they might be preferred by speakers over non-alliterating and non-assonating alternatives. The next chapter explores this question by asking to what extent these phonological patterns have a mnemonic advantage for speakers.

Chapter 2 What evidence is there for memory advantages due to phonological patterning in word-pairs?

2.1 Introduction

This chapter takes the initial steps in answering the central question articulated in the title by examining existing research. It will become evident that one research team in particular has dominated the work in this area, either by virtue of their own studies or others influenced by them. There is always a risk with such domination, and so part of what will be covered is to evaluate the potential for what is known in this domain to be skewed in some way, not so much by poor quality research as, possibly, assumptions made about the underlying nature of the phenomena, and / or how experiments are best designed to investigate them. For this reason, the chapter will ask various, more fundamental, questions relating to phenomena and assumptions.

The review questions that will be addressed are:

- Review Question 1: What evidence is there that phonological patterning aids recall?
- Review Question 2: What evidence is there that L2 learners have better recall with phonological patterning?
- Review Question 3: How has the research of Boers and colleagues contributed to our understanding of phonological patterns in L2 word-pair learning?
- Review Question 4: What are the purported mechanisms that underpin the mnemonic effects of phonological patterns?

2.2 What evidence is there that phonological patterning aids recall?

As mentioned in Chapter 1, sound patterns such as alliteration and rhyme are frequently found in advertising, brand names and fictional characters, and it may feel intuitive that there is 'something' about sound patterns that makes them so ubiquitous. In this chapter evidence for a mnemonic effect of phonological patterns will be assessed from a broad range of disciplines: marketing, cognitive and experimental psychology, and second language acquisition.

2.2.1 Phonological Patterns in Marketing

Advertisers exploit all the linguistic components of a language to increase brand awareness: phonological devices such as alliteration (for example, *Coca-Cola*), orthography (unusual or non-standard spelling, such as *Kool-Aid*), morphology (using affixation, for example *Jell-O*), semantics (such as the use of metaphor, as in *DieHard* batteries), and syntax (diverging from the norm, *Got Milk?*). However, much of the empirical evidence concerns how potential customers evaluate the product being advertised, with little compelling evidence for any mnemonic effect of phonological patterning.

One much-cited example from the literature is McQuarrie & Mick (1996), in which the authors aimed to develop a taxonomy of rhetorical devices most commonly used in advertisements. Rhetorical devices were defined to include rhyme, alliteration[1] and assonance, amongst other 'artful deviations' from expectations, which were all assumed to perform a function that makes a difference in how an advertisement is received. The researchers assembled a random sample of advertising headlines from the six magazines with the highest advertising revenue in 1990 - 1991, in the

[1] 'Alliteration' was defined as three or more repetitions of a consonant. A double repetition of a consonant was referred to as 'chime'. See Table 2.1 for examples.

hope of reflecting a broad range of editorial content and audience. They found that over two-thirds of the advertisements in the sample contained a rhetorical device listed in their taxonomy, and they hypothesised that the presence of such devices would have consequences for how the advertisement is processed.

To test this claim, they ran an experiment to gain insights into if there was a difference in how consumers rated advertising headlines containing rhetorical devices, versus headlines which were literal or nonfigurative. Data was collected from 131 English L1 undergraduates. A list of literal and rhetorical headlines (n = 52) was compiled, with a boldface label indicating the product category. The literal and rhetorical headlines were interspersed, and three different orders of presentation were used. Table 2.1 shows three sample test items, though it should be noted that participants were not made aware of the taxonomy. Participants were asked to evaluate each headline on a 10-point rating scale, from 'clever / artful' to 'plain / matter of fact' – it is unclear from the report if such descriptive anchors are used only at the extremes of the scale.

Table 2.1 Sample test items from McQuarrie & Mick (1996).

Category	Headline	Product & Source	Taxonomy
AUTO	*No one knows the land like a Navajo.*	Mazda cars in *Car & Driver* magazine	Assonance & Alliteration
COMMUNICATIONS	*The best in the business.*	AT&T telecommunications in *Business Week* magazine	Chime
TOBACCO	*Introducing the new Virginia Slims 10-pack.*	Virginia Slims cigarettes in *Cosmopolitan* magazine	Literal

The results showed that headlines with rhetorical devices were judged more artful and cleverer than the literal headlines, and that this difference was statistically significant. The authors concluded that consumers are sensitive to the presence of rhetorical devices in advertising.

This is a widely cited paper, though in relation to the present thesis, it is unfortunate that the authors did not test productive recall of the headlines, or test recognition at a later point in time, to establish if the 'clever' ones were more memorable or not. Furthermore, it could be argued that asking participants if there is something 'clever' in a text is somewhat of a leading question.

A further example from the field of marketing is Argo, Popa & Smith (2010), in which the authors demonstrated over the course of six experiments that participants experience positive 'affect' (that is, positive feelings and emotions) when exposed to novel brand names displaying sound repetition. Although this set of experiments focuses on affective attitudes towards words, and not recall of those words, the suggestion that the presence of a phonological pattern can cognitively affect how a participant reacts to a word is pertinent to this thesis, and thus warrants closer attention.

In Study 1 (Argo et al., 2010, p. 98), 47 participants were randomly assigned to a condition in which they were exposed to a novel brand of ice-cream, the name of which either did or did not contain sound repetition, in this case consonance (word-internal alliteration), for example *sepsop* vs. *sepfut*. The authors wanted to see if the participants would evaluate the ice cream differently according to (a) whether it had sound repetition in the name, and/or (b) whether they said the name aloud or read it silently. Each participant tasted the same ice cream sample twice, though they were told it was a different ice cream. A series of Likert scales were used to assess the participants' views on, for example, how the brand names made them feel, and the extent to which they enjoyed the ice cream. The names and taste were evaluated significantly higher when the word contained sound repetition and the participant said the word aloud, and there was also a significant interaction between the two variables. Thus, the authors concluded that the sound pattern cognitively affects how a brand name is received.[2]

One paper that does directly address the issue of advertising and memorability is McQuarrie & Mick (1992). The authors conducted two experiments that suggest

---

[2] An example of a less than compelling interpretation of research data from the field of marketing is Davis et al. (2016), in which the authors concluded that alliterative promotions were more desirable than non-alliterative promotions after participants in one experiment overwhelmingly chose the offer of "2 Twix" over "2 Snickers". This overlooks the equally plausible explanation of an aversion to peanuts rather than a rhetorical device influencing the choice of snack.

manipulating the puns or wordplay used in advertising headlines can improve the unaided free recall of the headlines as well as greater self-reported liking for the advertisement and more positive attitudes towards the advertised product. However, since their operationalisation of wordplay did not incorporate phonological overlap, they offer little insight into the current question.

The mnemonic effect of phonological patterns in television commercials is a central feature of Lowrey, Shrum & Dubitsky (2003). In this study, 480 brand names were taken from a pre-existing data set created by a market research company. This data set included ‘copytesting’ reports, that is, consumer responses to real television commercials featuring the brands. The data gathering for the copytesting was done by recruiting participants[3] over the telephone and inviting them to preview new programmes (with commercial breaks) on unused cable television channels in the participants’ own homes. The participants were contacted the following day and asked if they had watched the programmes. If they had watched the material, the market research company interviewed them further and collected data on the programmes and the commercials. This interview data included: whether the participants could recall seeing commercials for particular types of products, whether they could remember the brand names featured in the commercials, and what content they could recall from the commercials. Eliciting information about the content was designed to eliminate ‘false positives’ – that is, participants could offer a well-known brand name for a product category cue, without having seen the commercial. Each commercial had between 180 - 200 consumer responses, though it is unclear if there were 480 commercials, or some brand names appeared in multiple commercials.

Lowrey et al. coded the brand names according to 23 ‘linguistic features’, two of which were ‘alliteration’ and ‘assonance’, then cross-referenced the coded names with the recall data from the market research company. A multiple regression analysis tested for the association between the linguistic properties and recall of the brand names. The authors found that some of the linguistic features were a

[3] Participants from a variety of U.S cities were tested, and the age distribution was quota-controlled to approximate the U.S. population.

statistically significant predictor of brand-name memorability, though the effect size was small (p = .001, R^2 = .07). However, it is not clear how much of a role alliteration and assonance play in these findings; of the 23 linguistic features, only 11 appeared with sufficient frequency in the brand names data set to warrant retention in the analysis, but exactly which ones is not elucidated. A further impediment to evaluating the evidence is the lack of definitions for 'alliteration' and 'assonance'.[4]

The evidence from the field of marketing outlined above suggests that phonological patterns affect how language is perceived yet they can shed little light on whether such patterns impact on memorability. Considering the number of resources spent on advertising, and the prevalence of phonological patterns in advertising, this is rather surprising. To find convincing evidence of the memorability of alliterating and assonating lexical items, one must look at other fields of enquiry.

2.2.2 Phonological Patterns in Cognitive and Experimental Psychology

Evidence from the fields of cognitive and experimental psychology suggests that, in addition to semantic features, phonological ones can influence lexical access, and there is a long history of laboratory experiments investigating how the structural organisation of material influences the way it is learned and recalled. As so much has been published in this area, it might be helpful to focus this review on oft-cited studies which purport to directly test for a mnemonic effect of phonological patterns or furnish evidence to support the notion that such patterns can affect lexical processing. Some of this work was carried out decades ago, but it forms the basis for our understandings and underpins later research. In the following section, therefore, the claims from the following three papers will be evaluated:

---

[4] The only two examples given in the paper suggest the terms refer to word-internal similarity: "does *Ajax* exhibit alliteration?" and "does *Ajax* exhibit assonance?" (2003, p. 11).

- Rapp & Samuel (2002): phonological patterns affect the choice of words a speaker produces.
- Bower & Bolton (1969): word-pairs with phonological patterns are more easily learned.
- Lea, Rapp, Elfenbein, Mitchel & Romine (2008): alliteration in poetry can affect recall of the text.

Rapp & Samuel (2002)

In the first experiment reported in Rapp & Samuel (2002), 88 English L1 undergraduates were given written sentence pairs with a missing word, and asked to fill in the blank with a reasonable one-word completion, for example:

1) *The man walked into the bank and slipped on some ice. He'd gone to deposit his payment and nearly broke his _________.*

2) *Joe opened the present and hoped he wouldn't get any clothes. He looked inside the container and found some* _________.

Semantically plausible completions for 1) above could be *leg*, *arm*, *foot*, *ankle* and so on. The central research question was whether changing the word *payment* (which the authors termed a 'prime') for *check*, would make the participants more likely to answer with the 'target' *neck*. Similarly, would switching the prime *container* with *box* in 2), elicit the rhyming target *socks*?

The authors also varied the design so that any phonological priming would have to cross a sentence boundary, for example:

3) *Joe opened the birthday present and looked inside the box. He found some _________ even though he hoped he wouldn't get any clothes.*

Participants completed 100 sentence pairs: eight within-sentence rhyme primes, eight within-sentence non-rhyme primes, eight across-sentence rhyme primes, and eight across-sentence non-rhyme primes. There were an additional 68 filler sentences with a blank space in a random position to ensure participants remained unaware of the experiment's purpose, and all the sentences were randomised and counterbalanced.

Using a two-factor ANOVA (Rhyme / Non-Rhyme Prime x Within-Sentence / Between-Sentence), Rapp & Samuel found that the presence of a word with phonological overlap with the potential target resulted in an almost 50% increase in the production of a rhyming word over the baseline - rhymes following a rhyme-prime (M = 29.8%), rhymes following a non-rhyme prime (M = 20.1%) (Rapp & Samuel, 2002, p. 567). Furthermore, participants produced more rhymes in the within-sentence rhyme condition (M = 32%) than the between-sentences rhyme condition (M = 27.5%), and the difference was statistically significant ($p < .05$). The authors suggest this is evidence that the strength of a rhyme prime is modulated by the presence of an intervening syntactic boundary, and that it complements the findings of Hudson & Tanenhaus (1985) who found that the influence of a prime may decline across intervening words. The authors conclude that people are sensitive to both the meaning and the surface structure of words when accessing lexical representations and that their results support interactive models of word production (for example, Dell, 1986) wherein activation of phonological information can feed back to influence lemma selection.

These findings support the notion that sound patterns (in this case, rhyme) seem to affect the choice of words. How confident can we be in these findings? One strength of the design is that only 16 out of the 100 sentence pairs contained a rhyme prime, so it is difficult to argue that the participants themselves were likely to develop a bias or a strategy for producing rhymes over the course of the experimental session. Were the experimental items free from bias? The authors only give five examples of the sentence pairs used in the experiment, yet closer inspection raises one issue in particular that might make interpreting the results less clear-cut. In the appendix we find this example:

The store was difficult to find when I went the wrong way. It was night and I was supposed to turn _____.

Here, the range of possible one-word semantically meaningful lexical items is very restricted (either *left* or *right*), so there is a very high probability that a participant would produce a rhyming target, potentially biasing the results. If we look at the other examples of sentence pairs used in the experiment, we can see the range of potential items is wider; items of clothing, bones that can be broken, murder weapons and insects, but again, there are subtle restrictions:

Joe opened the present and hoped he wouldn't get any clothes. He looked inside the container and found some (<u>item of clothing</u>).

The presence of the determiner means some items seem to be less semantically plausible (for example, some *suits*?) and may favour plural nouns like socks, which themselves seem, to this reader, to be the archetypal gift of clothing. Without seeing all the sentence pairs, it is difficult to fully assess the validity of the claims.

Could lexical access have been more influenced by frequency and/or familiarity than by the presence of a preceding phonologically overlapping prime? Again, it is hard to say based on only the five reported samples of test items, but if we look at test item 1) again and then consider the possible answers, this explanation cannot be discounted:

The man walked into the bank and slipped on some ice. He'd gone to deposit his check and nearly broke his _________.

Table 2.2 Corpus frequency of potential targets.

Potential responses	COCA[1] raw frequency	Frequency per million
broke his leg	185	0.17
broke his neck	146	0.13
broke his arm	112	0.10
broke his ankle	63	0.06
broke his foot	36	0.03

[1] Corpus of Contemporary American English.

Table 2.2 shows the outcome of checking phrase frequency in the COCA corpus. The data in Table 2.2 suggests that participants would choose a rhyming word (*neck*) or assonating word (*leg*) because it was more frequent or familiar to them in the sentential context.

It would have been interesting if rhyme had been operationalised to encompass other phonological patterns such as alliteration. As it was, the adopted marking scheme involved coding the written responses only for rhyme and non-rhyme, presumably based on the researchers' subjective judgements. This means that any alliterating or assonating responses would have been disallowed across all the conditions, for example:

He'd gone to deposit his payment and nearly broke his <u>pelvis</u>.

Finally, the authors found that the effect seemed to dissipate when the priming word and the response were on either side of a syntactic division. This alludes to the fuzzy notion of proximity touched on in Chapter 1, where phonological patterns were defined as occurring within a span "close enough … for the ear to be affected" (Greene et al., 2012, p. 40). Although this thesis focuses on adjacent words in a word-pair, it does raise the question of how a mnemonic effect can be impacted by distance: for example, consider *time will tell* versus *time will most definitely tell*, or *time, she said, will tell*.

Bower & Bolton (1969)

A common finding in experiments on paired-associate learning[5] is that pairs with phonological similarity are learned faster and recalled better than ones without that similarity. This has proved to be a robust finding, and one of the earliest papers to try to explain this phenomenon is Bower & Bolton (1969). The authors hypothesised that

[5] An experimental task in which participants are asked to learn a pair of words together, usually pairs of unrelated short words such as *cat* and *pen*, or a pseudoword and a meaning.

rhymes are mnemonic because awareness of the rhyming relationship restricts the range of response alternatives that need to be considered for a given stimulus.

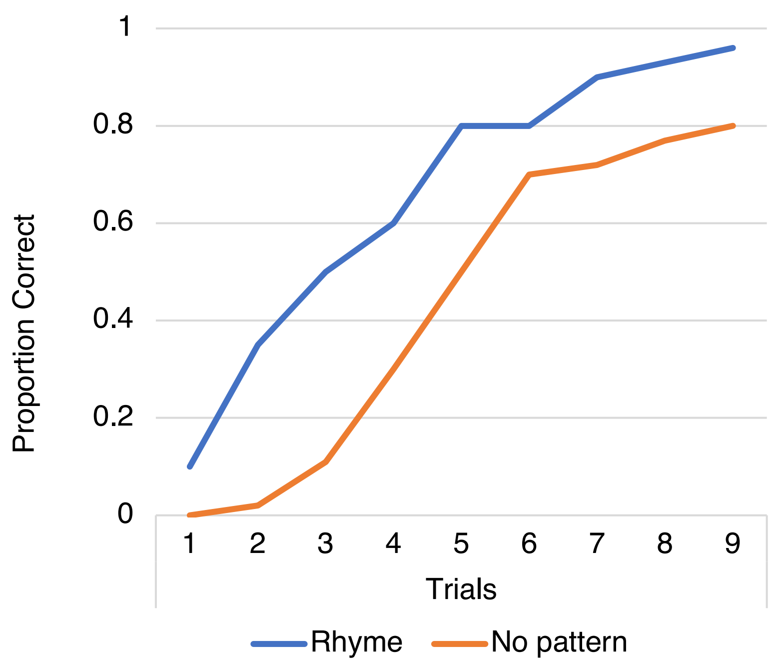
To test their 'response-restriction' hypothesis, the authors conducted an experiment with English L1 undergraduates, divided into two experimental groups. Each group were shown a list of 36 paired associates, all consonant-vowel-consonant (CVC) words, and almost all nouns, matched for frequency. The list for Group 1 (9 participants) consisted of 18 rhyming pairs written in red ink, for example, *can – man*, *hat – mat*, mixed with 18 non-rhyming pairs in black, for example, *lad – buy*, *box - lip*. The list for Group 2 (9 participants) consisted of the same 18 non-rhyming pairs as in Group 1, plus 18 pairs in which the second word in each pair was related in form to the first word by altering the last letter (in red ink), such as *hat – ham*, *bin – bit*. That is, the paired-associate items alliterated and assonated.

The participants were told about the nature of their lists prior to the study phase, so those in Group 1 were told that the red cue word had a rhyming response, and for those in Group 2, the red cue word had an alliterating and assonating response. The study phase employed an anticipation learning methodology: on each trial the participant is shown the cue word, attempts to guess the response word, and is then shown the correct response. The technique involved the use of a memory drum[6], and the lists were shown nine times, each time in a random order. The authors found that for both groups, the phonologically similar pairs were anticipated correctly more often, on every trial, than the unrelated pairs. In Group 1, the mean correct responses per item over the nine trials was 5.95 for rhyme pairs and 4.11 for unrelated pairs. A t-test showed this difference to be statistically significant ($t(8) = 4.26$, $p < .001$). Likewise, for Group 2, the alliterating and assonating pairs had a large advantage (mean correct responses = 5.38) over the unrelated pairs (mean correct responses = 3.61) and the difference was statistically significant ($t(8) = 6.19$, $p < .001$). The authors suggested that alliteration + assonance restricts response alternatives as much as rhyme does and produces an equal amount of facilitation in paired-associate learning.

[6] Preceding the use of desktop computers, a memory drum was a rotating cylinder to which a list of words was attached, visible for a fixed time period through a hole in a screen.

This is an interesting study and seems to pertain to the question posed at the beginning of the chapter. However, there are some caveats which should be borne in mind when assessing this evidence. Participants are effectively asked to guess the associate word on the first round, so it would be expected that the scores would be zero for the non-matching pairs and slightly higher for the ones with phonological patterns due to the restricted choice. The authors do not comment on whether the incorrect responses were at least candidates conforming to the rules for that list, for example *can – pan* rather than the 'correct' answer of *can – man* in Group 1. The sets are repeated nine times, and presumably the participants are told the lists are stable, because the scores do increase, as can be seen from Figure 2.1. It can be surmised that the mean scores are low on account of the random guesses in the initial trials. However, it seems that this experiment is confounding two different activities and their effects: guessing and memory.

Figure 2.1 Learning Curves for Rhyming List (adapted from Bower & Bolton, 1969).



Bower & Bolton's notion of 'response-restriction' has parallels with Rubin's (1995) account of memory in Oral Traditions in which sound patterns form constraints that help the speaker recall a ballad or epic poem.

Lea, Rapp, Elfenbein, Mitchel & Romine (2008)

This study addresses the issue of recall from memory and reports three computer-based experiments on the mnemonic effect of alliteration in poetry and prose. Because of its relevance to the ensuing discussion, this study merits a more detailed description.

In the first experiment, 40 English L1 speakers read aloud poems on a computer screen at their own pace, advancing line by line by pressing a key on a response box. At a certain point in each poem, a 'Get Ready' message appeared for 500ms, followed by a target word. Participants responded 'yes' if the word had appeared previously in the poem, and 'no' if it had not, and their response times were recorded. The poem resumed after each response, and each poem finished with a yes / no comprehension question. The authors had edited the poems by inserting a 'target' line near the beginning of the poem, and a 'cue' line towards the end. Well-known poems were avoided to ensure the participants were relying on episodic memory, and that they would not notice the revisions, and free-verse poems were chosen to avoid any potential mnemonic effects of rhyme or metre. All other instances of alliteration were removed from the poems.

For 21 poems, three versions were made, and the alliterative overlap was manipulated by modifying the target and cue lines in one of three ways (see Table 2.3, where the tested item is *barn*). Either the target line had no alliteration, to provide a baseline response, or the target line had different alliteration than the cue line, or the target line had the same alliteration as the cue line. The cue line was immediately followed by a probe word which was taken from the target line but did not alliterate.

Table 2.3 Example of inserted lines in modified poem extract.

Target line (inserted near the beginning of the poem)	**Condition**
all along the creek-winding road, past Stuart's barn	no alliteration
all along the raw and rutted road the reddish barn	different alliteration
all along the way-winding road, wary whispers of the old barn	same alliteration
Cue Line (inserted approximately ten lines later in the poem) and Probe Insertion Point ^	
the wooden willowy warp of wildcarrot ^ leaf	(recognition probe: *BARN*)

The participants had to hit the 'Yes' key on the response box if they thought the probe word (*barn*) had occurred approximately ten lines earlier in the poem and their response time was recorded on the computer. In all three conditions *barn* appeared at the end of the target line so the correct answer to the probe in this case was always affirmative. Therefore, a further twenty-four 'filler' poems were included to provide negative responses for the recognition task and to mask the alliterative parts of the experimental stimuli from the participants. The authors hypothesised that if the phonological pattern did not 'reactivate' previous text information, the response times would be similar across all three conditions. On the other hand, faster response times would be recorded if the alliterative sound reactivated information related to similar sounds earlier in the text. If the presence of alliteration alone is sufficient to reactivate previously encountered information, the different-alliteration condition would also show faster response times.

The mean reaction time to the recognition probes was 85ms faster when the probes appeared in the same-alliteration versions than when they appeared in different-alliteration or no-alliteration poems. A Repeated Measures ANOVA was conducted with participants (F_1) and items (F_2) as random variables. The results showed the overall differences between related means was statistically significant at $p < .05$. Post-hoc LSD tests (Fisher's least significant difference test) showed the same-alliteration mean was different from both the no-alliteration mean and the different-alliteration mean, and the latter two means did not differ. An analysis of the accuracy data showed no differences for probes in the three different conditions and the authors held that this did not support the interpretation that the results were due to a speed-accuracy trade-off.

Experiment 1 was then replicated (Experiment 2, Lea et al., 2008, p. 712), with a new cohort of participants (32 English L1 undergraduates), using the same materials and procedures, except this time the poems were read silently. The results matched those of Experiment 1; response times were, on average, more than 108ms faster in the same-alliteration condition than in the other two conditions and the difference was found to be statistically significant.

In their third experiment, the authors used 24 revised passages of narrative prose (and 36 filler passages) rather than poetry, on the basis that the previous results may have been obtained due to the use of different reading strategies specific to poetry. Experiment 3 followed the same procedure as Experiment 2: a new group of 24 participants read the passages silently on the computer screen. Again, the mean response times were significantly faster in the same-alliteration condition than the other two conditions. The authors summarised that "repeating alliterative sounds in both poetry and prose resulted in measurable and reliable reactivation of backgrounded portions of the texts" (p. 713).

This is an interesting set of results and directly addresses the issue of the mnemonic effect of a phonological pattern. Considering the amount of text that the participants read (over 7,000 words in Experiment 1 if the examples are indicative of the test material, and 11,000 in Experiment 3), it is striking that the authors managed to get these results and that participants did not suffer from task-fatigue[7].

The participants in these three experiments were unaware of the focus on alliteration, yet the sound pattern was found to aid processing, as evidenced by faster response times. This is an important finding, because, as we shall see in the later discussion of findings from second language acquisition, Boers and his colleagues (section 2.3) usually found a mnemonic effect only *after* drawing the participants' attention to the phonological patterns. Is it perhaps a question of the *amount* of alliteration? In the work of Boers et al., participants only saw or heard the sound patterns across two components of a target phrase, for example*, private*

[7] In each experiment, response times more than three standard deviations from the participant's mean were taken out of the analysis, resulting in a loss of less than 5% of the experimental data in total.

property, *slippery slope*, *tall tree* (from Boers, Lindstromberg, & Eyckmans, 2012). On the other hand, in Lea et al. (2008), according to the examples given, participants encountered eight alliterating words in two lines of poetry before they were given the recognition probe (*way, winding, wary, whispers, wooden, willowy, warp, wildcarrot*) and six alliterating words in the two lines of prose (*rusty, raced, recklessly, relived, returned, rendezvous*). Would Lea et al. have found similar results if there had been fewer alliterating words in the target and cues lines? In fact, there is one potential way to approach an answer to this question from within Lea et al.'s report. The narrative prose passages had fewer alliterating words in the target and cue lines than the poems, to make them as inconspicuous as possible. Did this reduction in alliteration make a difference? Although it is only indicative of a difference, it is worth considering the reported effect sizes which can be compared as they are not totally dissimilar treatments: the effect sizes were greater in Experiment 1, poems read aloud, compared to Experiment 3, prose read silently (the percentage variance effect sizes are .219 and .143 respectively), so this could be taken as a sign that the amount of alliteration that the participant encounters affects processing[8].

What is the theoretical model which underpins the effect found in Lea et al. (2008)? The authors claim that Resonance Theory (see for example, Gerrig & O'Brien, 2005; Myers & O'Brien, 1998) can account for the effects of alliteration that they found in the three experiments. The validity of this claim cannot be assessed without a brief explanation of Resonance Theory and an overview of some of the evidence supporting it. The central tenet of the theory is that text processing, that is, how a reader creates mental representations from the discourse, is grounded in memory-based cognitive processes. The theory attempts to explain how information presented earlier in a text can become 'reactivated' and become readily available to a reader, even when a text is coherent and there is no need for a conscious search for information related to the current input. Gerrig & O'Brien (2005) summarise the theory thus:

---

[8] It should also be noted that there was no statistically significant difference in the error rates for each condition, only faster on-line processing as indicated by response times.

> Concepts from earlier portions of the discourse and general world knowledge resonate as a function of the degree of match to the input. The match depends on the overlap of both semantic and contextual features among concepts. Memory elements that are contacted by the initial signal in turn signal to other elements. During this resonance process, activation builds and when the process stabilizes, the most active elements enter working memory and become part of the active portion of the discourse model (p. 229).

Resonance thus seems to be an automatic product of similarity between immediate semantics and wider textual context in working memory, and as such, it is not immediately obvious that it would extend to phonological form. Evidence offered in another study, by Myers & O'Brien (1998), hinges mostly on how readers resolve anaphoric references when reading; for example, slower reading times on target sentences are recorded in the face of processing difficulties which arise when antecedent text information contradicts a current reference. The ability to resolve anaphors depends on several factors such as the extent to which the text has elaborated on the antecedent, the presence of other potential antecedents or distractors, the distance between the anaphor and the antecedent, and general world knowledge (ibid. p. 134). An additional factor is 'featural overlap' between the antecedent and the anaphor, though featural in this context appears limited to semantic features, for example, evidence is cited that a background antecedent (such as *burglar*) is reactivated more quickly when the anaphor is a direct repetition, rather than a synonym (*criminal*). Extending the scope of Resonance Theory to explain the influence of non-semantic features of a text, namely the phonological overlap in Lea *et al.* (2008), seems valid only if a broader definition of resonance is adopted. Myers & O'Brien (1998) suggest this is possible: "The basic process in the resonance model is pattern matching" (p. 134) which, it could be argued, could extend to phonological pattern matching. This issue of what exactly is underpinning the mnemonic effect will be returned to later in this chapter.

So far, reference has been made to empirical work using English L1 participants. The next section assesses the evidence for a mnemonic effect of phonological patterns with L2 learners.

2.3 What evidence is there that L2 learners have better recall with phonological patterning?

As mentioned previously, the investigation into the mnemonic effect of phonological patterns in word-pairs with L2 learners has been dominated by a single team of senior academics, originally working out of Belgium[9] and the UK.[10] An overview of the team's prolific work can be seen in Table 2.4. This series of small-scale quasi-experiments[11] began when Boers and Lindstromberg (2005) undertook retrospective analyses of classroom experiments originally designed to evaluate the facilitative effects of etymological explanation.[12] They found that higher levels of recollection had been recorded for English idioms that displayed alliteration, as in *bite the bullet* and *get someone's goat*. This superior level of recall was calculated, post hoc, to be at greater than chance likelihood, and these results provided the impetus for developing a series of further experiments.

The experiments in Table 2.4 all explore whether English L2 participants identify or recall word-pairs with phonological patterns better than equivalent sequences with no such pattern. The designs are generally iterative, involving a partial replication of a previous experiment and adding a new element or adjusting the procedure. Collectively, these classroom-based studies build the case that phonological patterns

---

[9] Frank Boers was based at Erasmus University College of Brussels and the University of Antwerp, both in Belgium, then at Victoria University of Wellington, New Zealand. At the time of writing, he is based at Western University, Ontario, Canada. June Eyckmans works from Ghent University in Belgium.

[10] Seth Lindstromberg is based at Hilderstone College, UK.

[11] Study designs are non-randomised: participants are drawn from pre-existing convenience samples of intact classes.

[12] An approach to teaching idiomatic expressions via their original, literary meaning (Boers & Lindstromberg, 2005).

Table 2.4 Quasi-experiments by Boers, Lindstromberg and Eyckmans on the mnemonic effect of phonological patterns on recalling word-pairs with L2 English learners.

Study *(phonological pattern)*			L1 [1]	N [2]	Proficiency [3]	Familiarity of Target Items	Awareness Raising
1	Lindstromberg & Boers (2008a) *(alliteration)*	Exp 1	Dutch	25	B2 CEFR	Known	Yes
		Exp 2	Dutch	31	B2 – C1 CEFR	Known	-
		Exp 3	Dutch	29	> B2 CEFR	?	Yes
2	Lindstromberg & Boers (2008b) *(assonance)*		Dutch	35	Not stated	Known	Yes
3	Boers, Lindstromberg & Eyckmans (2012) *(alliteration)*		Malay, Chinese	27	"Upper Int. / Advanced"	Known	No
4	Boers, Lindstromberg & Eyckmans (2014a) *(alliteration)*	Exp 1	Dutch[4]	36	B2 CEFR	Known	No
		Exp 2	Dutch	47	B2 CEFR	Known	No
5	Boers, Lindstromberg & Eyckmans (2014b) *(assonance)*	Exp 1	Malay, Chinese	55	"Advanced"	Known	No
		Exp 2	Dutch	44	B2 CEFR	Known	Yes
6	Boers, Lindstromberg & Eyckmans (2014c) *(consonance)*		Dutch, Dutch bilinguals[1]	36	B2 CEFR	Known	No
7	Boers, Eyckmans & Lindstromberg (2014) *(alliteration & consonance)*		Dutch, Dutch bilinguals[1], Latvian	38	B2 CEFR	Known	Yes
8	Lindstromberg & Eyckmans (2014) *(assonance)*	Exp 1	7 Indo-European L1s	25	"Advanced Bilingual"	Known	Yes
		Exp 2	6 Indo-European L1s	17	B2 CEFR	Known	Yes
9	Boers, Lindstromberg & Webb (2014) *(alliteration)*		Japanese	54	"High-Intermediate"	Known	Yes
10	Eyckmans, Boers & Lindstromberg (2016) *(alliteration)*		Dutch	65	A2 CEFR	Unknown	Yes
11	Eyckmans & Lindstromberg (2017)*(alliteration & assonance)*		Dutch	50	B2 CEFR	Unknown	Yes
12	Lindstromberg & Eyckmans (2017) (*assonance*)		Dutch	81	B2 CEFR	Known	Yes

[1] L1 refers to the participants' first language. The bilingual participants in Studies 6 and 7 were reported to have Dutch as an L1 plus another (unstated) language (not English).
[2] N refers to the total number of participants in the sample.

[3] CEFR = Common European Framework of Reference for Languages.
[4] Of the 36 participants, 18 had Dutch as their L1, 8 were bilinguals (Dutch + another L1), 7 had Indo-European L1s.

in word-pairs have a mnemonic effect for L2 learners to some degree.

Given the large number of studies, it will be useful to examine them as a set, looking for their key similarities in terms of test items and method. In the same way, the findings can be compared and jointly evaluated. This approach will also help pinpoint the most appropriate options for extending their work, as reported in Chapters 3 to 6. These similarities and the experimental results are discussed in the account below; the publications will be referred to by their number from the first column of Table 2.4 and individual experiments are indicated in brackets, so for example, Study 5(2) refers to the second experiment reported in Study 5 (Boers et al., 2014b).

2.3.1 How are 'word-pairs' operationalised?

In examining these studies, the first thing to establish is what exactly they claim to be testing. The authors refer to the stimulus items used in the experiments in Table 2.4 in various ways: as 'collocations', 'idioms', 'MWUs' (Multi-Word Units) or simply as 'chunks' or 'phrases'. To some extent, these terms seem to be interchangeable for them, though not all test stimuli are of entirely the same type. Due to the gradient nature of such phenomena, the boundary between what constitutes a collocation and what constitutes an idiom is not always clear; as Liu (2010, p. 5) notes, the term *collocation* is far from unambiguous and can vary depending on the research purpose and theoretical orientation. Generally speaking, the experiments operate on three kinds of word-pairs consisting of two open-class words, sometimes with an intervening grammatical word:

- verb + noun (phrase), for example *play a part* in Study 1(3).
- adjective + noun, for example *deep sea* in Study 5(1).

- noun + noun, for example *time frame* in Study 7.

While most studies use pairs with relatively transparent meanings, some deploy more opaque, idiomatic items, in which the meaning cannot be determined from their individual constituent words, for example *bite the bullet* and *jump the gun* in Study 11. Such items are sourced from published idiom dictionaries, for example the *Oxford Idioms Dictionary for Learners of English* (Parkinson, 2006).

To ensure that the constituent parts of the stimulus word-pairs are not just a random association of words, the authors measure their collocational strength. This is frequently done with reference to mutual information (MI) scores or *t*-scores, statistical measures commonly used in corpus linguistics to ensure that there is a statistically significant level of co-occurrence between the words in the corpus. Thus, the authors are able to create a pool of target items and control items that are evenly balanced in this regard. Although the number of target items varies from 3 to 16, most of the studies use about 12, matched with 12 controls and supplemented with a number of filler items and extra items used to avoid primacy and recency effects.[13] As can be seen in Table 2.4, the most common phonological patterns investigated are alliteration and assonance, though Study 6 and 7 also incorporate some items displaying consonance, sometimes referred to as intra-word alliteration, for example *important point* (Study 6).

As with all tests of lexical recall, the list of potential confounding variables is considerable, as words can vary along several dimensions: semantic, structural, and affective. To control for affective factors, Boers and colleagues avoid using word-pairs with 'striking associations' which may render one word-pair more memorable than its control counterpart, for example, *same sex* (Lindstromberg & Boers, 2008a, p. 210). Aside from omitting test stimuli with distinctive connotations, it is hard to see how else this feature can be controlled for in a classroom-based experiment – in Study 12 Lindstromberg and Eyckmans claim to have balanced test items in terms of

[13] It is a common procedure in tasks of list recall to use 'dummy' items at the beginning and end of the list, as there is ample evidence showing that participants tend to have better recall of items at the beginnings and ends of lists. These dummy items are not used in the subsequent data analysis.

‘emotiveness and personal relevance’ but without stating how this was achieved (Lindstromberg & Eyckmans, 2017, p. 131).

Overall, target items and control items have also been matched for other variables that are known to affect memory performance. It is recognised, for instance, that more frequent words are generally easier to recall (for example Gregg, 1976), so the authors have balanced whole-phrase frequency of occurrence, which is usually assessed by reference to large corpora such as COCA (Corpus of Contemporary American English, Davies, 2008 to present) or the BNC (British National Corpus). As there is robust evidence that shorter words are easier to recall than longer ones (for example Baddeley, Thomas, & Buchanan, 1975),[14] word-pairs are matched in terms of orthographic and syllabic length, as well as other features such as their grammatical relationship, word class, and concreteness of meaning.

Concreteness[15] relates to the extent to which the concept denoted by a word refers to a perceptible entity, action or property that can be experienced through the physical senses, for example *kick*, or *sugar*. The referents of abstract words, like *infinity*, cannot be seen or touched. In general, concrete words are more easily remembered than abstract words. One prevalent theoretical explanation for this can be found in dual-coding theory (see for example Clark & Paivio, 1991), which proposes that concrete words are better remembered than abstract words because they activate perceptual or sensory memory codes, as well as verbal codes. An alternative explanation, context-availability theory (for example, Schwanenflugel, Akin, & Luh, 1992), accounts for this phenomenon in terms of a more accessible semantic network for concrete words than for abstract words. Evidence from the

---

[14] This so-called ‘word-length effect’ seems to apply to free- and serial-recall of uniform lists, that is, lists composed of either long or short words. When lists are compiled of stimuli of different orthographic length, the opposite is sometimes observed, that is, the longer items have a mnemonic advantage (see Katkov, Romani, & Tsodyks (2014) for a proposed retrieval mechanism).

[15] The authors in Table 2.4 sometimes also refer to *imageability*, a sub-feature of concreteness, which is the ability of a word to elicit an internal image. Because word ratings of concreteness and imageability are highly correlated (Brysbaert, Warriner, & Kuperman, 2014), the two terms are often used synonymously.

neurosciences, using techniques such as event-related functional magnetic resonance imaging (fMRI), suggests different neural processes for concrete and abstract words (for example, Fliessbach, Weis, Klaver, Elger, & Weber, 2006). Concreteness is such a powerful predictor of word recall in memory research that there are large databases of concreteness ratings, for example the MRC Psycholinguistic Database (M. Coltheart, 1981), but this raises a question. Is it valid to assume this mnemonic advantage applies to word-pairs?

There are empirical answers to that question, which will be addressed presently, but there is an underlying theoretical one that is worth mentioning. It relates to whether we consider word-pairs to be processed as item + item or as a single two-part item. This is core to discussions of formulaic language (Wray, 2002), and potentially important for the central question addressed here. Is the learning of collocational pairs actually about learning a single item, or is it about learning some mnemonic method for recalling item 2 after recalling item 1? This might make a difference to how long any alliterative or assonating feature needs to remain active as a mnemonic: forever, if it is a matter of retrieving item + item, but only during the initial learning phase if the collocation becomes a lexical entry in its own right.

The authors in Table 2.4 certainly seem to think concreteness ratings can be applied to word-pairs, and one database, outlined in Brysbaert et al. (2014), does include ratings for some compound nouns and phrasal verbs. However, in light of the fact that most published norms are for single words, how can the concreteness of a word-pair be gauged? In those studies in Table 2.4 where concreteness has been controlled, the authors have asked English L1 speakers to rate the items on a scale. For instance, in Study 6, the authors asked 11 speakers to rate the second word in each word-pair on a scale of 1 (least concrete) to 10 (most concrete). The set of instructions read: “How concrete / imageable is each word in the 2nd column when it follows the word in the 1st column? Could you give each word in the 2nd column a concreteness / imageability rating from 10 (the most) to 1 (the least)?”. The speakers then scored each item on a handout, an example of which can be seen below:

Leftward collocate	Rightward collocate	Concreteness – Imageability Score (1-10)
Private	PROPERTY	
Private	COLLECTION	

In other studies, for example Boers, Lindstromberg & Eyckmans (2012), the authors have asked the English speakers to rate the complete word-pair. Such rating systems enable the authors to calculate the means for each item and ensure the test items and the control items are evenly matched along this dimension. This seems like a reasonable method since the MRC Psycholinguistics Database is itself derived from merging three sets of norms which are based on subjective ratings for printed words.

2.3.2 What are the similarities in terms of experimental procedures?

This section addresses the question of what exactly the participants *do* with the target items in the Boers et al. experiments. Answering this question will hopefully clarify the optimal way to explore the issue further in new experiments and draw attention to any methodological issues.

The majority of the experiments have a treatment phase, sometimes referred to as a study phase, followed by a test phase. As the treatment phase requires more extensive explanation, the simpler test element will be outlined first. Tests are usually unannounced and take the form of free-recall and / or cued-recall. In a free-recall test, participants are asked to write down, in any order, all the test items they can remember from the study phase. In a cued-recall test, the participants are given the first component of a word-pair from the treatment, and are asked to recall the second word; for example, the experimenter says *tall* and the participants are expected to write down *tree*. As free-recall tests are challenging and may lead to a floor-effect, they are sometimes followed by a cued-recall test. Studies 1(1) and 2 involve recognition tests in which the participants are asked to tick the items to which they were previously exposed. Testing is usually done in stages; immediate tests,[16] then delayed tests, which range from one hour after the initial treatment, to up to two weeks.

The treatment phase in earlier publications (Studies 1 and 2) involved asking participants to manually sort slips of paper into piles of phonologically patterned word-pairs and piles of items with no pattern, ahead of a testing phase. Later studies, however, focus on two main procedural themes: dictations and study handouts, though there is often a combination of both.[17]

Since the change in methods cuts across other features of similarity and difference, summaries of the task components will be used as an aid to the discussion. For instance, Study 3 is illustrative of an experiment with minimal engagement on the part of the participants; the experimenter reads aloud each item twice from a randomised list, and the participants are asked to repeat aloud the item and then write it on a sheet of paper. The papers are collected and there immediately follows an unannounced free-recall test: participants are given a clean sheet of paper on which to write down as many of the items as they can remember from the dictation. This sequence of actions can be represented by (A) below, showing what each participant does with the word-pair, prior to the testing phase:

(A) HEAR (x 2) > REPEAT > WRITE > **TEST PHASE**

In contrast, the procedure used in Study 4(2) has a more involved experimental design: five different versions of a list of 28 items are distributed to the participants, so that no participant has the same order of items as the participant they are sitting

[16] Clearly, the logistics of collecting and distributing test papers in a classroom involves a short amount of time between a study phase and a testing phase. In some experiments, participants are asked to change seats in this interval, in the hope of disrupting any attempt to rehearse test items.

[17] There is no record of *why* the experimental procedure changed from a sorting task to a dictation / handout activity. It could be speculated that the latter is more expedient to mark, compared to the former which would involve unclipping each participant's pack of cards and checking if the slips of paper are grouped correctly.

next to. The experimenter reads aloud each item twice from another jumbled list and the participants are asked to repeat it sub-vocally. Then, they are given time to find the item on their list and rate it on a scale according to whether they have never / sometimes / often heard, read or used the word-pair before.[18] After this stage, participants are paired up and dictate their lists to each other, presumably taking it in turns to read out their list or listen and write down their partner's list, though it is not clear if they can refer to their own list as they are listening. The papers are collected, and there is an immediate free-recall test followed by a cued-recall test. After one week, there is another free-recall test followed by another cued-recall test. This study can be represented as (B):

(B) HEAR (x 2) > REPEAT > SCAN > RATE > DICTATE < > HEAR > WRITE > **TEST PHASE**

If we can accept (A) and (B) as points on a continuum of participant engagement, then nearly all the experiments in Table 2.4 fall somewhere between. Such variation in the study phase leads to the obvious question: which one leads to the most compelling results? Before appraising the experimental findings, however, one key factor must be addressed. Consider (C):

(C) HEAR (x 2) > WRITE > CIRCLE > **TEST PHASE**

This is the procedure used in Study 12 (it is not explicitly stated if the participants in this study vocalised the items after hearing them in the dictation). The procedure in (C) is almost identical to (A) above, save for the stage when the teacher draws the participants' attention to the phonological pattern, in this case assonance, and asks them to circle the items that do not display the pattern. Raising awareness of the phonological pattern during the treatment phase is an important variable and is used in the majority of the experiments (see the last column in Table 2.4). It might be

---

[18] This rating activity, designed to promote deeper engagement with each word-pair, is not sensitive to whether or not the participants have never / sometimes / often encountered the words as individual entities.

reasonable to predict that drawing the participants' attention to the phonological pattern would have an effect. That is, raising awareness results in participants recalling more alliterating / assonating items than items with no pattern, to a greater than chance likelihood. As we shall see in the next section, this is not always the case.

2.3.3 What are the experimental findings?

As will become apparent, the experiments from this team of researchers have produced conflicting results (see Table 2.5). Of particular note is that in the experiments depicted in design (A) and (B) (Studies 3 and 4(2) respectively), with no awareness-raising component, there was a statistically significant greater recall of test items with a phonological pattern, compared to recall of items without a pattern. In the experiment *with* awareness-raising, depicted in (C) (Study 12), there was no such finding. This clearly requires some discussion, but first, in reviewing these experimental outcomes, it might be helpful to look at those studies which failed to yield statistically significant results in the direction of a phonological effect, to see if there are any discernible issues with the research designs or procedures that may account for the lack of findings.

Study 1(3)

The methodology used in this study was quite different from that of the other studies, as it involved a study phase in which two groups of participants were exposed to approximately ten hours of authentic listening materials over a three-month period. Twenty test items were taken from the listening materials, eight with a phonological pattern (seven displaying alliteration, one with rhyme), and 12 control items with no phonological pattern. These items were targeted in transcript-based gap-filling exercises over the three-month period. The classroom teacher pointed out the phonological patterns in the test items in the audio transcripts to one group

Table 2.5 An overview of the findings in the experiments of Boers et al. on the mnemonic effect of phonological patterns in word-pairs.

Study	Awareness-Raising?	Statistically Significant Finding?	Phonological Pattern
1(1)	✓	✓	alliteration
1(2)		not applicable	alliteration, assonance, rhyme
1(3)	✓		alliteration
2	✓	✓	assonance
3		✓	alliteration
4(1)		✓	alliteration
4(2)		✓	alliteration
5(1)			assonance
5(2)	✓	✓	assonance
6		✓[1]	consonance
7	✓		alliteration, consonance
8(1)	✓	✓	assonance
8(2)	✓	✓	assonance
9		✓	alliteration
10	✓	✓	alliteration
11	✓	✓	alliteration, assonance
12	✓		assonance

[1] A statistically significant finding, but not in the predicted direction.

(the experimental group), but not to the other (the control group). After the study phase, there were two unannounced tests which focused on the form of the test items. The participants in the experimental group recalled more alliterating / rhyming items than the control group, but the difference was not large enough to be of statistical significance. This result seems to lend support to the idea of a mnemonic effect for word-pairs displaying phonological patterns. However, the results must be treated with caution as the test items and control items were *not* matched for important variables such as length, frequency of occurrence, syntax or concreteness.

Study 5(1)

This study used a typical procedure of a teacher-led dictation (twenty high-frequency word-pairs, half of which showed assonance), followed by an immediate free recall test. There was no attention-raising. Participants recalled more assonating items than control items, but the difference was not large enough to be statistically significant. The participants were described as having Malay or Chinese as their L1, making this is one of the few studies in Table 2.4 in which the L1 is not Dutch.

Study 6

This study is the only one in which the phonological pattern under investigation is exclusively consonance, defined as 'across-word consonant repetition' (Boers et al., 2014c, p. 51), for example, *time frame*. There was no attention-raising prior to the dictation, nor is it stated if the participants were instructed to repeat / sub-vocalise each item during the dictation phase. Participants recalled more control items than target items in both the immediate cued recall test and the one-week delayed cued recall test, and the difference was statistically significant.

Study 7

In Study 7, there *was* an intervention to bring attention to the alliteration and consonance in the target word-pairs; after the dictation activity, and after having heard and seen examples of alliteration / consonance on the classroom blackboard, the participants were asked to tick the items on their lists that they thought displayed a 'form of consonant repetition across the two words' (Boers, Eyckmans, et al., 2014, p. 363). On analysing the results of this task, the authors found that none of the 38 participants had identified all 12 instances of phonological repetition, and that 21% of the control items had been ticked, for example, *popular demand* and *time span*. The authors suggested that one possible explanation for the large number of errors was what they termed 'orthographic interference'.[19] That is, the influence of the orthographic form could obscure the phonological pattern (for example, the repetition of /k/ in *a **q**uestion **c**oncerning*), or wrongly suggest one (for example, inter-word /t/ repetition *impor**t**ant **t**hing*). It could be argued that this problem was due to task instructions which did not emphasise enough the auditory nature of the phenomenon. The results of the immediate cued recall test showed that alliterating items were recalled more than items with consonance and the control items, but the results were not statistically significant. In the delayed tests, the control items were recalled more, but not to a statistically significant degree.

Study 12

After a teacher-led dictation of 28 collocations, half of which manifested assonance, the participants were given a worksheet with the list of items and were asked to circle those items which did not display assonance. After the papers were collected, there was an immediate free recall test followed by a one-week delayed free recall test. In both tests, the control items were recalled more than the assonating word-pairs, but the difference was not statistically significant.

In those experiments with an awareness-raising intervention, the researchers have sometimes asked the participants to identify the items *with* the phonological pattern under investigation (Studies 1(1), 5(2), 7, 10, 11). This is most commonly operationalised by asking the participants to circle or tick word-pairs in their dictated lists (as opposed to sorting slips of paper into phonologically patterned / non-patterned piles). In two other experiments, however, the researchers have asked the

[19] This finding is similar to the one reported in Study 1(2), which looked solely at the ability of the participants to identify and sort phonological patterns in word-pairs on slips of paper. Although the Dutch L1 participants were language majors at a B2-C1 CEFR level, only 13 out of the 31 participants were able to sort the word-pairs into alliterating, assonating and rhyming piles; the majority resorted instead to semantic groupings, such as *fine wine, fish dish, fast food, bean soup*. The authors concluded that the ability to autonomously notice phonological patterns in test items cannot be taken for granted.

participants to mark those items which *lack* any phonological pattern (Studies 8 and 12). In study 12, the rationale given for this step in the procedure is that a mnemonic advantage for a word-pair may arise due to an effect of the identifying task itself, rather than a phonological pattern having intrinsic memorability. That is, ticking or circling a word-pair renders that item more memorable compared to a word-pair that has not been ticked. Following this line of argument, asking participants to circle, say, all the consonant clusters in a word list would give a word such as *strength* a recall advantage over *academic*, all other things being equal. On the other hand, it could be argued that deciding if a word-pair has a certain feature, such as a phonological pattern, is a binary yes-no decision, a decision which would apply equally to all the items on a list, with the same 'cognitive effort', as the authors note in Study 1 (Lindstromberg & Boers, 2008a, p. 211) and Study 2 (Lindstromberg & Boers, 2008b, p. 431). In the two experiments reported in Study 8, per-item analyses found negative correlations between recall of an item and the number of times it had been circled, either correctly or incorrectly (Lindstromberg & Eyckmans, 2014, p. 29). The authors interpreted this as a lack of evidence of task-induced bias from the awareness-raising task.

In sum, the studies described above have not produced a statistically significant result in the predicted direction. However, with the exception of Studies 6 and 12, there is a tendency for participants to recall more items with a phonological pattern than control items, with or without an awareness-raising step.

Turning now to the studies with mixed results in Table 2.4, it is typically the case that any mnemonic advantage observed in the immediate testing phase has dissipated in the interval before the delayed testing phase. For example, in Studies 4(1), 4(2) and 8(1) a significant mnemonic effect can be seen in the immediate cued recall tests, but not in the delayed tests – one day later in Study 8(1) and one week later in the two experiments in Study 4.

Study 9 is a further example of a study with mixed results. This study warrants further attention because the experimental design is quite different, and it is the only experiment that used Japanese L1 participants, a point that will be returned to later (see section 2.4.4). Study 9 investigated the incidental[20] learning of alliterative phrases during an extensive reading study phase. The target items were twelve idiomatic V+N (or V+N phrase) word-pairs, three of which alliterated (*cut corners, face the fact, run the risk*). The authors created additional sentences which included the twelve word-pairs, and inserted these sentences into a Graded Reader, so that each word-pair appeared five[21] times in the text. Word-pairs were matched in terms of length, frequency, strength of co-occurrence (with reference to *t*-scores in the BNC), concreteness and semantic transparency (both rated by English L1 speakers) and L1 congruency (as rated by Japanese L1 speakers). The 54 participants were pre-tested on their knowledge of the word-pairs one week before the treatment, first with a test of receptive knowledge of form - a pen-and-paper multiple-choice test, for example:

lose _____

a) cigarette	*b) touch*	*c) demand*	*d) meat*	*e) church*	*f) I don't know*

Participants were then tested on their knowledge of meaning; they were given the twelve target word-pairs and asked to write a definition in Japanese. One week later, in the study phase, the participants read the modified graded reader and at the same time listened to an audio recording of the story. When the audio was finished, participants were given the same tests, unannounced, although with a different order of test items. One week later, the same tests, in random order, were done again.

In analysing the test results, the authors looked at gains made from the pre-test to the delayed post-test and found that the participants scored more highly on the multiple-choice tests for the three alliterative word-pairs than the non-alliterating word pairs, and that this difference was statistically significant. However, there was

[20] Perhaps 'semi-incidental' would be more accurate because the participants' attention is explicitly drawn to the target word-pairs.

[21] Five appearances was deemed suitable as previous research on vocabulary gains in incidental reading had found this sufficient to produce a measurable effect (Webb, Newton, & Chang, 2013).

no significant difference between alliterating word-pairs and non-alliterating word pairs on the translation tests. It could be asked what reliable and valid inferences can be drawn from a target sample of only three items. The authors justify this low number by arguing that three alliterative phrases out of twelve is a ratio with ecological validity as it reflects the amount of alliteration in many types of English multi-word units (see section 1.4). The authors acknowledge that one potential drawback with the design is a test-taking effect – the participants had already seen the intact word-pairs seven times (excluding the multiple-choice tests); once in the pre-test translation test, five times in the text, and once in the immediate post-test translation, before the one-week delayed round of testing. Nor is it clear how participants could be prevented from reading / hearing the items outside of the experimental setting. However, the authors point out that this exposure to the intact forms applied equally to the alliterating and non-alliterating word-pairs.

Of the remaining studies that have found a mnemonic effect, a few merit additional remarks. The procedures used in Studies 1(1) and 2 were not atypical, involving sorting and dictation tasks, followed by immediate free-recall tests and delayed recognition tests (a two-week delay in Study 1(1) and a one-week delay in Study 2). However, what is noteworthy about these two experiments is the effect sizes[22] that the researchers report. In reviewing the studies in Table 2.4, comment has so far been made only on the results of null hypothesis significance tests (NHSTs), and whether a result is 'statistically significant', with reference to the conventional alpha value of $p = .05$ (meaning a 5% probability of a Type I error). These NHST results can be supplemented with effect sizes to show substantive significance. In Study 1(1), participants recalled / recognised more alliterative word-

[22] The term 'effect size' is used inconsistently in the literature as it can refer to a number of things:

- a statistic which estimates the magnitude of an effect, for example, *Cohen's d*
- the actual value calculated from a statistic, for example $d = .5$
- an interpretation of an estimated magnitude of an effect, for example '*medium'*

Here, 'effect size' is defined as "a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest." (Kelley & Preacher, 2012, p. 140).

pairs than control items in both tests, the differences were statistically significant ($p = .004$ in the immediate free recall test, $p = .0018$ in the delayed recognition test) and the effect sizes were calculated[23] as Cohen's $d = .88$, and $d = .93$ respectively. In Study 2, which looked at assonance, the researchers report effect sizes of $d = .87$ and $d = 1.03$. These values have been interpreted as 'large', according to Cohen's 'one size fits all' descriptors. To put them into perspective, it may be helpful to compare them to Plonsky and Oswald's (2014, p. 890) rough estimate for a mean effect size of $d = .69$ for the domain of L2 research as a whole.[24] However, one caveat which concerns the validity of the results in Study 1(1) is that the variable of concreteness was not a factor in the material design.

No other experiments in Table 2.4 have managed to replicate such substantial effect sizes, though two of the authors have conducted two small-scale meta-analyses of the mnemonic effect of assonance (Lindstromberg & Eyckmans, 2014, 2017). The later meta-analysis, using results from four studies, though not the positive results reported in Study 2 due to a loss of data, concluded that the odds of a participant recalling a word-pair with assonance are approximately 14% higher than recalling a word-pair with no assonance, after an awareness-raising component. The authors equate this to a very small effect size of $d = .075$ (Lindstromberg & Eyckmans, 2017, p. 133).

2.3.4 What issues arise from this suite of experiments?

Several questions emerge from the findings from the experiments described above. Firstly, how generalisable are these results to a wider population? The participants in these experiments were mostly undergraduate students, though Study 8(1) used in-service TEFL teachers and Study 10 used high-school students (13-14

[23] Effect sizes were calculated retrospectively and reported in Study 5 (Boers et al., 2014b, p. 97).

[24] Publishing bias against nonsignificant findings necessitates a broad interpretation of this estimate.

years old). Furthermore, most of the university students were English language majors. Although language proficiency is not a variable that their research explicitly explores, it is worth noting that the vast majority of participants are "Upper-Intermediate"[25] or "Advanced", with reference sometimes made to the comparable B2 and C1 level descriptors of the Common European Framework of Reference for Languages (CEFR). In justifying which label participants are given, recourse is most often made to the classroom teachers' estimates of their student's proficiency, although reference is also made to course requirements (Study 4 and 7), or results from the Vocabulary Size Test (Nation & Beglar, 2007) (Study 3 and 5). Given that most of the participants are quite proficient and study English language at university level, a valid line of enquiry would be to investigate if the same results can be found with participants who may be less attuned to the phonology of their L2.

The prevalent L1 in this body of research is Dutch, which is unsurprising given that two of the researchers were based in Belgium, where most of the data was collected. The fact that the majority of the participants in Table 2.4 spoke Dutch as their L1 will be explored in more detail in Chapter 3. The only indication that the observed mnemonic effect is applicable to other populations is given in Studies 9, 5(1) and 3. As noted above, Study 9 used Japanese L1 speakers and an extensive reading study phase. The results are slightly indicative of a mnemonic advantage for alliterating word-pairs, but it remains to be seen if we can expect the same kind of results from a study phase involving deliberate attention to a phonological pattern with Japanese L1 speakers. Study 5(1) found no recall advantage for word pairs displaying assonance with Malay and Chinese speakers, but Study 3, also with Malay and Chinese speakers, did find an advantage for word pairs displaying alliteration. Both Studies 5(1) and 3 lacked an awareness-raising component.

---

[25] One anomaly crops up in Study 9, in which the learners are described as "high intermediate" (Boers, Lindstromberg, et al., 2014, p. 88), yet their average score on the 2000 level of the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001) was 20.4 out of a maximum 30, which seems to me rather low for a post-intermediate learner.

2.3.5 Do the effects extend to the recall of previously unknown word-pairs?

The status of the target items as 'known' or 'new' raises further questions. In the experiments in Table 2.4, the status of a word-pair as 'known' is most often gauged indirectly, with reference to the corpus frequency of the items and the high proficiency level of the participants. For example, there is an assumption that an upper-intermediate language learner would be expected to know a word-pair such as *tall tree*. A lack of spelling errors in the dictation exercises is sometimes taken as additional evidence that the items are familiar. The fifth column in Table 2.4 shows that only two of the studies, 10 and 11, involve participants learning 'new' word-pairs.[26] The authors of these studies suggest that where the test items are 'new' to the learners, as opposed to already 'known', a different memory subsystem is being called upon. Before looking further into the memory architecture that is purportedly being tested throughout the experiments in Table 2.4, it may be helpful to briefly review Studies 10 and 11, to establish the role played by this factor in their experimental design and results.

In Study 10, the Dutch L1 participants were tested on the target items one week prior to the study phase, in order to establish that the items were indeed unfamiliar. The target items were 32 idiomatic word-pairs (V+N), half of which alliterated. The pre-test took the form of a handout on which the items were embedded into sentences and the verb omitted. The participants were asked to supply a suitable verb, for example:

---

[26] Eyckmans and Lindstromberg (2017, p. 4) state that two other experiments, 1(3) and 9, used 'unknown' word-pairs, but in Study 1(3) some of the target items in the gap-fill exercises were deemed "sufficiently predictable for students to fill in before checking their responses" (Lindstromberg & Boers, 2008a, p. 216), which to me suggests an expected degree of prior familiarity. As noted previously, Study 9 only had three target items with a phonological pattern, a small sample size from which to draw firm conclusions.

When he was first elected mayor he believed he was going to be able to _________ mountains and make life better for everyone in the city. He soon realised there were limits to what he could achieve.

In the study phase, the participants were given the intact items on a handout with the Dutch translations, for example, *move mountains: bergen verzetten*. The treatment group received an additional instruction: to identify and tick the items which had a phonological pattern. Participants were asked to study the items for an upcoming fill-in-the-blank test. There was an immediate test after the study phase, but the authors report their greater interest in longer-term retention, so the results of the delayed test received greater scrutiny. This unannounced test had the same format as the pre-test (though in a different order) and occurred ten days after the study phase. The analysis focused on the gains made between the pre-test and the delayed post-test: the intervention of identifying the phonological pattern was found to be more strongly associated with gains than the no-intervention condition.

One issue that arises from this study concerns the way that vocabulary knowledge is assessed in the sentence completion pre-test (see Appendix B in Eyckmans et al., 2016, p. 136). Given that the participants are schoolchildren estimated to be at the A2 CEFR level, the pre-test appears to have quite a heavy vocabulary load. A vocabulary profile (Cobb, n.d.) shows the pre-test includes low frequency vocabulary, for example *ardent* (7K), *ancestry* (8K), *solace* (9K), *sweatshop* (off-list), with 98% coverage (the level deemed suitable for independent reading) reached at 6K. It is not unreasonable to think that if a participant struggles to understand the sentence, they will have difficulty knowing which verb fits semantically. Consider the following example, where ‘blah’ is used in place of unknown vocabulary: if an A2 learner does not know *mayor* or *achieve* (both 3K lemmas), or even *elect* and *limit* (2K), then *climb* seems a valid response:

When he was first blah blah he believed he was going to be able to _________ mountains and make life better for everyone in the city. He soon realised there were blah to what he could blah.

In the treatment phase, the authors also make the assumption that supplying the Dutch translation for the idiom is sufficient for the participants to understand the target item, without questioning if they are familiar with the idiom in their L1.

In addition to the lexical load, a further question arises regarding the status of the items as 'unknown'. Failure to provide the verb *move* in the example above, may constitute evidence of a lack of productive knowledge of the idiom, but this does not preclude receptive knowledge of it. Word knowledge is multi-faceted and complex; Nation's (2001, p. 27) taxonomy, for instance, has eighteen different components. Clearly, it would not have been feasible for the authors in Study 10 to assess multiple aspects of word knowledge for 32 target items in a classroom-based experiment with 65 teenagers. However, assigning a target item the status of 'unknown' or 'known' belies the complexity of the phenomenon, and this is an issue which will be revisited in Chapter 7.

The methodology in Study 11 differed only in that the pre-test came immediately before the study phase. This study used 26 idiomatic items: six alliterated, six assonated and one showed both (*live off the fat of the land*). For the pre-test, the authors report that the Dutch L1 participants were asked to supply one suitable noun and were given the first letter and the definition.[27]

MISS THE M __________. Fail to achieve the result that was intended.

In the study phase, the participants were asked to study the complete items with definitions and example sentences taken from an idiom dictionary. As in Study 10, the treatment group was asked to mark those items which had a phonological pattern. The analysis focused on the gains made between the pre-test and the

[27] The inclusion of the definition is implied, as it is not described in the published procedure (Eyckmans & Lindstromberg, 2017, p. 9). Without a definition, it is not clear how knowledge of the target idiom (e.g., *miss the mark)* can be tested, as there could be numerous plausible noun candidates, for example, *man*, *meeting*, *message, moment*, *money*, *movie* and many others are all attested in the COCA corpus as collocates of *miss the*, though they vary in degree of idiomaticity.

unannounced one-week delayed test, and the authors found greater recall for the sound repeating items than the control items.

Both of these experiments tested participants on their recall of the "formal lexical makeup" (Eyckmans et al., 2016, p. 134) of the target items, that is, participants were assessed on productive aspects of orthographical form and collocational usage. The meaning of the items was not tested. The sentence completion delayed-test used in Study 10 may appear to involve semantics, but the sentential co-text is not strictly necessary, that is, a participant just needs to recall the verb they saw with *mountain* in the study phase, and, if they were in the treatment group, they know that there is a chance of the verb starting with an 'm' sound. Had the task been to supply the word-pair to the cue *"bergen verzetten*: ________", then this would be testing recall of both form and meaning. Interestingly, the immediate test in Study 10 did test semantic knowledge; participants had to supply a V+N word-pair that fitted the meaning of the sentence, without any cues. However, apart from descriptive statistics, no inferential analysis was done on these test results. The immediate test in Study 11 is a little more ambiguous, as participants were asked to fill in two gaps (the pre-test had one gap) but were given initial letter cues and a definition, for example:

M __________ THE M __________. Fail to achieve the result that was intended.

Again, it could be argued that the definition is surplus to requirements, in that the participants only need to recall the composition of the missing words rather than any semantic features.

In sum, Studies 10 and 11 look at the role of phonological patterns in the recall of the written form of a new target word-pair. It remains to be seen if alliteration and assonance can also help with the recall of the *meaning* of a new item, a question that will be considered in Chapters 6 and 7. The authors suggest that the tests they use tap "into" semantic memory (Eyckmans & Lindstromberg, 2017, p. 4), whereas the remainder of the studies in Table 2.4 test the recall of known items held "in" episodic memory (Lindstromberg & Eyckmans, 2017, p. 130). The use of the prepositions *into* and *in* is telling; linguistic models are partly constituted by our ruling metaphors, and the key metaphor for memory is that of storage. The next section will look at these memory storage systems and the underlying mechanisms that might contribute to the effects observed in the studies in Table 2.4.

2.4 What are the underlying mechanisms that might contribute to the effects observed?

The experimental tests used in the studies in Table 2.4 are assumed to measure the recollection of language from the episodic and semantic memory ‘stores’. There have been numerous analogies for the phenomenon of human memory, most often involving a spatial metaphor. Since antiquity, memories have been thought of as discrete objects kept in particular locations within a conceptual receptacle; memories are ‘things’ that can be stored, searched for, and may or may not be re-collected. Various receptacles have been postulated: an aviary, a library, a house, a storeroom, a computer, even a cow’s stomach (see Roediger, 1980 for an overview) and a ‘good’ memory implies a large capacity to hold information. Advances in the neurosciences have led to a reformulation of this idea of a memory as a ‘thing’, to that of memory as a process, namely, the reactivation of specific groups of neurons distributed across interconnected neural regions, though the literature still retains the inherent metaphors of memories as ‘traces’ or ‘representations’. Conceptualising memory without conflating storage and processing could be due to the distinction being partly erased in levels-of-processing theory (Craik & Lockhart, 1972), which holds that storage is a by-product of processing.

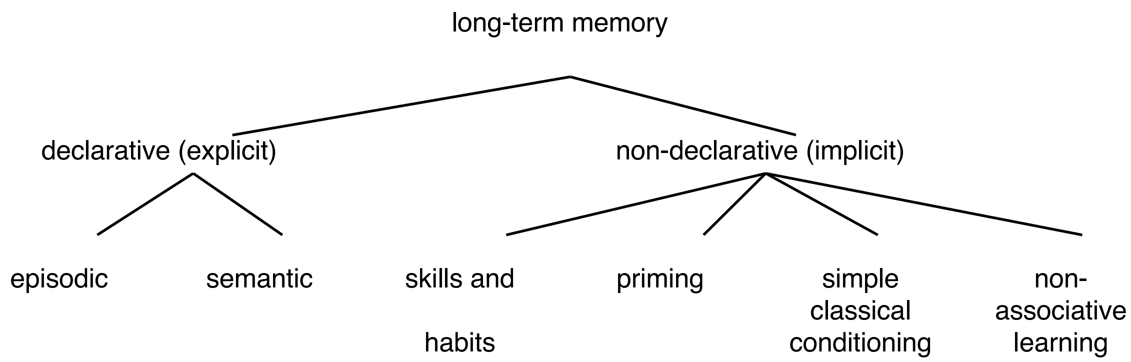
Formerly envisioned as a single unitary space, the receptacle has been reconceptualised and undergone a process of fractionation into separate ‘stores’, and this fractionation is supported by evidence from neuropsychological experiments on dissociations, in which one form of memory is impaired while another remains intact. The differences between episodic and semantic memory are often illustrated in the literature with examples such as the following: stating that *Caerdydd* is the capital of Wales demonstrates a retrieval of generic information from the semantic store, whereas my recollection of the time spent walking in a downpour when first visiting the city is recalling autobiographical information from the episodic store. In terms of the experiments outlined in Table 2.4, Eyckmans and Lindstromberg (2017) make the distinction as follows:

> Episodic memory would be drawn on, for instance, when following an instruction like this, ‘Please write down, from memory, as many as you can of the two-word phrases that you sorted and then dictated to each other at the beginning of class last Tuesday.’ Semantic memory, which tends to be more durable and which underpins most everyday language use, does not preserve the link between something that is known … and the occasion when it was learned (p.4).

In the context of learning an L2, however, the interaction of these two memory systems poses some questions about testing methodologies. On the face of it, a *free* recall test would draw on the episodic system, though in a *cued*-recall test (in which, for example, the researcher says *cut* to elicit the target *corners*) it seems plausible to suggest that a participant could recruit semantic knowledge to supply the missing word if recall from episodic memory fails. Eyckmans and Lindstromberg’s (2017) account of tapping into an episodic store is also consistent with a particular research finding listed in Table 2.4: the mnemonic effect dissipates in the time interval between the treatments and the test phases. Semantic memories, in contrast, should not decay in this fashion. However, it does raise a number of issues. Firstly, if successful language acquisition and use depend on the semantic memory system, then why test the episodic memory system? The majority of the studies in Table 2.4 tested recognition or recall of ‘known’ word-pairs recently presented and thus probably drawn from episodic memory. If a phonological pattern is a mnemonic for a word-pair temporarily held in episodic memory, how does this benefit language learners who need to draw on long-term semantic representations? To answer this, we need to determine what exactly is the relationship between the episodic and the semantic memory systems.

The literature on recall and recognition from episodic and semantic memory provides no definitive answers to the question of what this association is; unsurprisingly, the issue is rather complex. To paraphrase Ellis (2003, p. 46), this is a subject area where every content word can seem like a research discipline in itself. On the relationship between episodic and semantic memory, there seems to be broad, but not universal, agreement that they are both part of secondary or long-term memory;[28] Baddeley (2001), building on a conceptualisation by Squire a decade earlier, identifies episodic and semantic memory as components of declarative memory, as seen in Figure 2.2.

Figure 2.2 Classification of long-term memory (from Baddeley, 2001, p. 1346).



While semantic memory seems to be a fairly settled concept, episodic memory has been further refined by some theorists. For example, Conway and Pleydell-Pearce (2000) make the durative distinction between an episodic memory system with a short temporal span for recent recollective experiences, and an autobiographical memory for longer term accumulation of personal knowledge; whereas for Tulving (for example, Tulving & Schacter, 1990) episodic memory has three components: self, ‘autonetic’ awareness of one’s experiences and subjectively sensed time, though these two approaches do not seem incompatible.

Although, as already noted, there is no universally accepted answer regarding the relationship between the episodic and semantic memory systems, there is general consensus that they are interdependent and can affect each other during encoding and retrieval (Greenberg & Verfaellie, 2010). Baddeley (2001, p. 1346) suggests that semantic memory represents the accumulated ‘residue’ of many learning episodes; it may seem intuitive that information has to ‘go through’ episodic memory first before

[28] As distinct from primary or short-term memory.

being encoded in the semantic system – that is, we must experience the *what*, *where*, and *when* before it becomes part of the semantic knowledge store.[29] This idea of an accumulation of episodes also has parallels with usage-based exemplar models of linguistic knowledge. For example, Taylor's (2012, p. 286) conception of the mental lexicon as a "repository of memories of encountered language", and Ambridge's (2020, p. 514) model of language acquisition in which every speech event is stored as an episodic memory containing both fine-grained phonetic detail and extra-linguistic factors such as the context and the speaker's identity (these ideas will be returned to later in Section 7 on phonological similarity).

What are the conditions that need to be met for an event to leave a durable trace that can withstand the passage of time? According to Tulving (2001), whether or not information is encoded in long-term memory depends on various aspects such as the novelty of the incoming information, and levels-of-processing (c.f. Craik & Lockhart, 1972), with 'deeper' semantically based processing leading to stronger encoding than 'shallower' visual or phonological processing. Other factors such as repetition / rehearsal, saliency, recency, personal relevancy, and emotional content (which can enhance attention and perception) are also known to enhance encoding (for example Adelman & Estes, 2013). Post-encoding processes such as reactivation, during both sleep and wake states, are also thought to help consolidate and strengthen memories (Sneve et al., 2015).

The hippocampus is thought to play a pivotal role in episodic memory encoding and retrieval, whereas semantic memory is most often associated with activity in neural networks across widespread cortical regions (Shallice & Cooper, 2011, p. 51). However, there is experimental evidence that, over time, hippocampus-dependent

[29] In contrast, Tulving's SPI model (which stands for serial encoding, parallel storage, independent retrieval) is the "reverse of common sense" (2001, p. 1508): information must first pass from a perceptual memory system to the semantic system before it can be encoded into the episodic memory system. The parallel storage of the model means that a learning event does not leave a single memory trace but rather a bundle of features organised hierarchically, and parallel retrieval allows for retrieval from only one system, though frequently information from different systems is used jointly in any given act of retrieval.

memories “acquire a more cortically based representation” (Sweegers & Talamini, 2014, p. 53), suggesting that episodic memories can become decontextualised as they age: a subjective memory of an event is retrieved and re-encoded, in due course, to become part of our “mental thesaurus” (Tulving, 1972, p. 386). Perhaps this helps shed light on the question posed near the beginning of this section - why test episodic memory? Under the right conditions, something is learned:

> It seems plausible … that any learner able to access a context-rich episodic memory of a recently encountered lexeme is well-placed to adopt it for use – in a still on-going conversation, for example – with the result that it becomes more integrated into their productive vocabulary (Lindstromberg & Eyckmans, 2014, p. 22).

To recap, the studies in Table 2.4 have built a body of evidence that suggests a word-pair with a phonological pattern has a mnemonic advantage in recall from episodic memory compared to an equivalent word-pair with no such pattern. The advantage mostly pertains to better recall of the written form of the word-pair. As it appears this recall can be bridged to semantic memory with time, this may prove to be a useful learning tool if the findings can be extended to improved recall of both the form and meaning of a novel word-pair.

However, the question remains: *why* would a phonological pattern confer a mnemonic advantage? What is the intrinsic operation that might supply the results seen in this body of research? One possible explanation given by the authors in Table 2.4 is that recall of one part of the word-pair ‘primes’ recall of the other part, due to orthographic and / or phonological similarity (Boers et al., 2014a, p. 295; Eyckmans et al., 2016, p. 129; Eyckmans & Lindstromberg, 2017, p. 3; Lindstromberg & Boers, 2008a, p. 206; Lindstromberg & Eyckmans, 2017, p. 3). The next section will define the term ‘priming’ and evaluate this claim in light of evidence from relevant priming studies.

2.5 Priming

This section assesses the plausibility of ‘priming’ as an explanatory factor for the mnemonic advantage of alliterating or assonating word-pairs, as seen in the experiments outlined in Table 2.4. In order to accomplish this, the term ‘priming’ will be defined, followed by a brief delineation of how priming is measured and how it is thought to operate, before turning to the extensive literature for the most salient and robust support.

2.5.1 Defining ‘priming’

More than a century has passed since it was first observed that people can identify a word faster if they have recently been exposed to another word with some characteristic that brings it to mind. In the earliest work, the priming characteristic identified was semantic, not phonological, though a wider range of triggers was soon recognised. As Jiang (2012, pp. 13-21) indicates in his overview, Cattell’s extensive work on the chronometry of ‘mental operations’ in the late nineteenth century (for example, Cattell, 1887) helped pave the way for the subliminal perception experiments popular in the 1960s and what would later become known as ‘priming’ experiments. The term ‘priming’ appears to do a lot of heavy lifting in the literature: as seen previously in Figure 2.2, it can refer to a *component* of implicit long-term memory (Baddeley, 2001). The term can also be used to apply to a research *paradigm* or *tool* for examining lexical processing (for example McDonough & Trofimovich, 2009). Moreover, Shallice & Cooper (2011) variously employ the term ‘priming’ to cover a functional imaging *procedure* (p. 182), a short-term memory *process* (p. 245), a general *characteristic* of cognitive subsystems (p. 249), and a *property* of processing (p. 249). Distinctions between these disparate usages are not always clearly articulated. Such terminological diffusion may have arisen partly because priming has been the focus of research across several disciplines, notably neuropsychology, cognitive psychology, psycholinguistics and, increasingly, SLA studies.

Additional complexity arises due to the identification of different types of priming, for example, auditory, acoustic, phonological, and phonetic priming. The lack of a precise nomenclature is perhaps the result of the putative belief that priming, as a cognitive process, occurs at many different levels of lexical representation, which in turn map roughly onto linguistic levels of description: phonological, semantic, syntactic, and morphological. For the time being, priming can be stated in its broadest terms as referring to "a change in the speed, bias or accuracy of the processing of a stimulus, following prior experience with the same, or a related, stimulus" (Henson, 2003, p. 54). To differentiate from the Word Association Task (WAT) paradigm in which participants are often asked to *produce* a word in relation to a given stimulus, priming experiments most often focus on how the relationship between prior information and a given stimulus affects the participants' *response* to that stimulus, for example, deciding whether or not it is a real word. Going forward, as we gain a more explicit understanding of priming, a finer-grained definition will be proposed, pertinent to the findings in Table 2.4.

One characterisation that *is* most often clearly distinguished in the literature is a conceptual / perceptual dichotomy. Broadly speaking, conceptual priming is dependent on the degree of semantic processing required in the generation of a response (McDonough & Trofimovich, 2009, p. 2). Reference to different sub-types of conceptual priming can be found in the literature, dependent on the relationship between the primes and the targets: associative priming (in which words are related semantically but are not part of the same semantic category e.g. *sugar - sweet*); categorical priming (in which words are members of the same semantic category e.g. *furniture - table*); and mediated priming (in which words are not directly linked but are mediated by another concept e.g. *stripes - lion*, where the relationship is mediated by *tiger*), examples from McDonough & Trofimovich (2009, p. 62).

Insofar as Boers and his colleagues used target word-pairs and control word-pairs that were equally matched for semantic variables, conceptual priming can be ruled out as a rationale for the mnemonic effect in Table 2.4: the assumption being that there is no inherent mnemonic advantage in the meaning of a phonologically-patterned word-pair, compared to a word-pair with no pattern, for example *deep sea* (matched vowel) vs. *deep hole* (used in Study 5, Boers et al., 2014b). However, (as noted previously) compiling lists of lexical items that are balanced on multiple dimensions is notoriously difficult; how should one gauge the level of semantic association of pairs like *deep sea* and *deep hole* without knowing the personal experiences of the participants? The word-pairs may have very different subjective, internal representations for a keen diver versus a keen potholer; emotive factors such as valence[30] and arousal could result in different relative levels of priming for them. Perhaps the only way to adequately avoid this issue would be to use non-words (see Chapter 6).

In contrast to conceptual priming, perceptual priming is said to occur when the stimulus and the response are either identical, or there is some repetition of the form, be it the auditory, syntactic, morphological, phonological or orthographical structure (Marsden, 2009, p. 12). For example, research suggests that following the letter string *greet*, language users will process the word *STREET*[31] more quickly and more accurately than a word with no phonological and / or orthographic overlap (example from Zwitserlood, 1996). In relation to the body of experimental research reviewed in Table 2.4, the researchers have hypothesised that the word-pair *deep sea* has a mnemonic advantage over the comparable word-pair *deep hole* due to a perceptual / form-priming process, specifically, phonological priming.

Thus, in consideration of the present context, the term *priming* can be used to refer to how the processing of one word (the prime) automatically facilitates the processing, recognition and / or retrieval of a subsequent word (the target) due to repetition, or similarity, of the phonological form. By extension, the term can also be used to describe the cognitive mechanism underlying this facilitation, which will be explored below. The term 'priming effect' will be used to describe the measurable extent of this facilitation.

Priming effects are usually determined by measuring the reaction times (RTs) of keyboard presses, in milliseconds (ms), and the accuracy of such responses, in a

---

[30] A concept borrowed from Social Psychology, referring to the positivity or negativity of emotions (Brendl, Miguel, & Higgins, 1996, p. 95).

[31] Published priming studies often follow the convention of reporting specific examples of primes in lower-case and targets in uppercase.

laboratory setting. The priming research paradigm encompasses a wide range of 'indirect' experimental tasks for observing and quantifying cognitive behaviour; for example, eye-tracking recording, Lexical Decision Tasks, and reading aloud protocols. It should be noted that such research is founded on an inferential framework for relating observable behaviour to the cognitive system. Thus, interpreting RT findings entails a commitment to an implicit assumption that RTs and accuracy, dubbed 'the meat and two veg' of experimental psychology (Henson, 2005, p. 195), reflect an 'online' cognitive process.

The distinction between conceptual and perceptual priming clearly bears some resemblance to how responses are commonly categorised in the Word Association Task (WAT) research paradigm, in which, as noted above, participants are given a cue and are asked to produce a written or spoken response – 'the first word that comes to mind' - as opposed to react to a target that may or may not have been primed. Influenced by Saussure's (1916) syntagmatic / paradigmatic distinction, in the WAT methodology associations generated by a cue are often grouped into one of three types of relationship (Fitzpatrick & Thwaites, 2020):

- meaning - based associations (paradigmatic responses), wholly or partially synonymous with the cue and sharing the same syntactic features;
- position - based associations (syntagmatic responses), often collocative or contiguous with the cue;
- form - based associations (clang responses), based on the phonology or orthography of the cue.

However, finer-grained taxonomies of response have also been developed (for example, Fitzpatrick, 2006), as it has been recognised that word associations can be complex and idiosyncratic (Fitzpatrick & Thwaites, 2020).

As WAT and priming studies both draw inferences about the architecture of the mental lexicon and how words are perceived, stored, organised and retrieved, it is perhaps unsurprising that they both utilise designs that appear to involve priming. Since one of the key differences between the two research paradigms is that in WAT experiments, the participants consciously generate a candidate target to a cue, WAT responses could be viewed as a kind of *strategic* priming effect. This might seem at odds with the aforementioned description of a methodology that asks participants to say ‘the first word that comes to mind’, but some researchers (for example, Wilks & Meara, 2007) have questioned the assumption of WAT responses reflecting ‘direct access’ to the lexicon, and suggest that there may be a continuum of automaticity of response, dependent on respondent strategies and task demands.

In contrast, priming in its ‘purer’ form is held to be an implicit memory process and is thus often described as unintentional, unconscious or automatic (Tulving & Schacter, 1990) and is not associated with any awareness of the underlying process – indeed, any such awareness is held to imply that something other than priming is occurring, such as decision-making or strategy use (Shallice & Cooper, 2011, p. 247). Such explicit strategies include attempting to guess upcoming targets, trying to find relationships between the primes and the targets, or tuning in to experimental ‘bias’ such as word-length for primes vs. targets. Insofar as participants do become aware of the processes, results may thus reflect ad-hoc products of the experimental task rather than a priming effect (Lucas, 2000, p. 619). To avoid this, in many priming studies the presentation of the stimulus is manipulated. For instance, in a ‘masked’ priming experiment the prime stimulus will be presented so quickly (for example, 50ms) that the participants are not aware of it and so cannot employ conscious strategies when reacting to the target stimulus.

Before looking more closely at how the cognitive process of priming may function in relation to the methodologies adopted in the experiments outlined in Table 2.4, it might be beneficial to distinguish between phonological priming and other types of perceptual priming that may have occurred during the course of the experiments in Table 2.4, but which may not be the principal explanatory factor for the mnemonic effects of alliteration and assonance.

Auditory priming

According to McDonough & Trofimovich (2009, p. 20), auditory priming refers to the unintentional facilitated processing of language due to prior experience of the spoken form: or as Church & Fisher (1998, p. 525) put it, "each time a word is heard, a lasting representation of the sound pattern is encoded that then facilitates or biases subsequent identification of the same word." In effect, the very fact that a word is repeated may serve to entrench or consolidate it in the mental lexicon of the speaker, illustrating implicit learning on a micro-level (J. Taylor, 2012, p. 212). This processing advantage for repeated words has been demonstrated with L2 learners (Trofimovich & Gatbonton, 2006), so it seems reasonable to assume that as the words were repeated orally in many of the experiments in Table 2.4, by the experimenter and by the participants themselves in the dictation activities, auditory priming would have occurred. But this kind of repetition priming would have applied to all the words in an experiment,[32] irrespective of the condition, and for this type of priming to occur, there does not need to exist a perceptual similarity between the words. As we shall see, Boers et al. have argued that a word-pair such as *big band* has a mnemonic advantage over *hot meal* due to a different type of priming occurring.

[32] Syntactic priming can also be ruled out on a similar basis: target word-pairs and control word-pairs were of the same grammatical class. Acoustic priming is a phenomenon not relevant to the current investigation. As a research paradigm, acoustic priming investigates the involvement of auditory neural systems during conceptual-oriented tasks. For example, evidence has been found that acoustic conceptual features recruit auditory brain regions even when implicitly presented through visual words, for instance, the written word *bell* activates auditory neural systems, as opposed to a word like *pillow* (Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008).

Articulatory / Phonetic priming

In essence, this type of priming occurs when the processing of a spoken word is facilitated by previous exposure to language that overlaps in terms of phonetic / articulatory features. For instance, a listener will be faster in identifying *bat* after having heard *peel* due to the shared feature of a word-initial bilabial stop, compared to a word with no phonetic overlap, such as *win* (Luce, Goldinger, Auer, & Vitevitch, 2000). In the experiments of Boers et al. the word-pairs in the phonologically patterned sets would benefit from this effect: *big* primes *band*, for example. But what are the parameters within which this effect occurs? How much time can pass before this effect dissipates? Do intervening words disrupt the effect, as in *a big ugly bag*? Would a previous occurrence not of *bag* but of *bin*, *tag, lamb* or *bab* suffice? In other words, how different could some previous event be, phonologically, and still act as a prime? Could a sound with the same manner of articulation but not place, or vice versa, be a prime (c.f. Luce et al., 2000)? Could a sound that is the same phoneme in the language, but that is a different allophone on account of the phonological environment, be a prime?

As we shall see, it is not easy to draw a cogent line between phonetic and phonological priming as the relationship between phonetic substance and the content of phonological units is far from clear (Burton-Roberts, Carr, & Docherty, 2000, p. 2). In Section 1.3.2, where 'phonological patterns' were operationalised as 'linguistic sounds recurring', a distinction was drawn between a phonetic representation of a word-pair and its psychological, listener-centred phonological representation. The definition of priming herein, based on the repetition or similarity of phonological segments in a word-pair, very much depends on a cognition-centric stance, focused on what the brain hears, as opposed to the ear. When an alliterating word-pair is spoken, a phonetic analysis of the speech stream may show great variation in the characteristics of the two words, even though a listener interprets the auditory code as two words having 'the same sound': Port (2007), for example, shows how sound spectrograms for the syllables [di] and [du] show large differences in terms of the frequency and the resonance of the stop consonant, yet listeners perceive similarity.

The situation becomes more complicated when we try to gauge similarity across languages, as no two languages share exactly the same phonological inventories. Take, for example, the word-pair *kitchen cupboard* – an English L1 user would probably say it alliterates as both words start with the same /k/ sound, despite the differences in articulation: for the initial consonant in *kitchen* the hump of the tongue is raised to the point of being palatal and the lips are more widely spread, while for the initial consonant in *cupboard* the hump of the tongue is further back and the mouth is more open. Using the International Phonetic Alphabet (IPA), the initial consonant sounds could be transcribed as a [c] (the *kitchen* sound) and [k] (the *cupboard* sound), and in English [k] and [c] do not signal contrasting meanings, that is, they are variations (or allophones) of the phoneme /k/. However, a Hungarian speaker would most probably object to the claim that the two parts of *kitchen cupboard* start with the same sound, because in Hungarian, [k] and [c] do signal a contrast, hence we find minimal pairs like *kuka* [kuka] 'dustbin' and *kutya* [kuca] 'dog'. As McMahon (2002, p. 19) notes, "The phoneme system of a speaker's native language, and specifically the difference between pairs of sounds which contrast and pairs which do not, strongly condition her perceptions." Viewed from this perspective, whether or not a word-pair assonates or alliterates might depend on who is listening. This distinction between phonetic and phonological levels of representation is important because patterns involve repetition or similarity, and the similarity of two sounds (or the perception of similarity) is a "somewhat tricky notion" (Giegerich, 1992, p. 32): an understatement that will be looked at in more detail later (Chapter 3).

The notions of alliteration and assonance are predicated on sounds being matched across different words. As just indicated, they must at some level be considered identical or in some other way equivalent. The standard theoretical frameworks for such a view recognise the phoneme as a stable, delineated entity in the speaker's phonological system. However, many researchers have cast doubt on the plausibility of such a stark distinction between phonological categories and phonetic realisations (for further discussion, see Janet Pierrehumbert et al., 2000; Port, 2007). Of particular relevance here are two questions arising from that work. First, is it necessary or reliable to assume that alliteration and assonance are constrained to the phonemic level? For example, could manner of articulation be sufficient to trigger the observed beneficial effects of alliteration? If so, two stop consonants with different places of articulation (e.g. *top cat*) could be considered to alliterate and found to enhance memorability. Second, is it safe to assume that a speaker of a second language will perceive two sounds as alliterative in the same manner as a first language speaker, even if both the L1s have the 'same' phoneme? Pierrehumbert et al's (2000) account suggests that the gross category of the phoneme might disguise many differences in not only the phonetic experience but also the phonological inventory of the two speakers. These issues need not be developed further at this point, since the empirical model of Boers et al. does not take them into account. However, some of their implications will be gradually unpacked in later chapters, culminating in a fuller discussion in Chapter 7.

It also worth remembering that some of the treatment phases in the design and procedures of the experiments listed in Table 2.4 involved giving the participants the target word-pairs in written form, either on slips of paper to be grouped according to pattern, or on study handouts. According to the *Universal Phonological Principle* (see for example Perfetti & Liu, 2005, p. 194; Verhoeven & Perfetti, 2017, p. 6), such visual input, as well as the act of spelling the word-pairs in the dictation phases, would automatically activate a phonological representation at some level, which may be sufficient to build upon a priming effect.

Orthographic priming

Lastly, we must consider to what degree orthographic priming, a further type of perceptual / repetition priming, plays a role in the mnemonic effect. The role of orthography and phonology is much debated in visual word recognition, both in terms of the extent to which they are involved, and their order along a temporal dimension. Ferrand & Granger (1994) give the following as an example:

> In the primed perceptual identification paradigm where both primes and targets are masked and briefly presented in different case, the percentage of correct target

> identification has been shown to vary as a function of prime-target orthographic overlap. Thus, word targets preceded by orthographically similar primes (e.g*., couch – TOUCH*) are easier to recognize than when preceded by unrelated controls (e.g., *flown – TOUCH*) (p. 366).

However, it is not clear from these examples how a prime and a target can overlap orthographically without also overlapping phonologically. In many languages with direct phoneme-grapheme correspondence it would be challenging to separate the two. In English, it is sometimes possible to tease them apart due to the substantial degree of spelling-sound irregularity, consider *plough, cough, through, dough*. In the work of Boers et al. there is most probably a considerable overlap between orthographic and phonological priming, and this is likely even when a word has to be read holistically to get the pronunciation, for example, *yacht*, according to the Universal Phonological Principle mentioned previously. This is an issue that will be discussed more fully in Chapter 7.

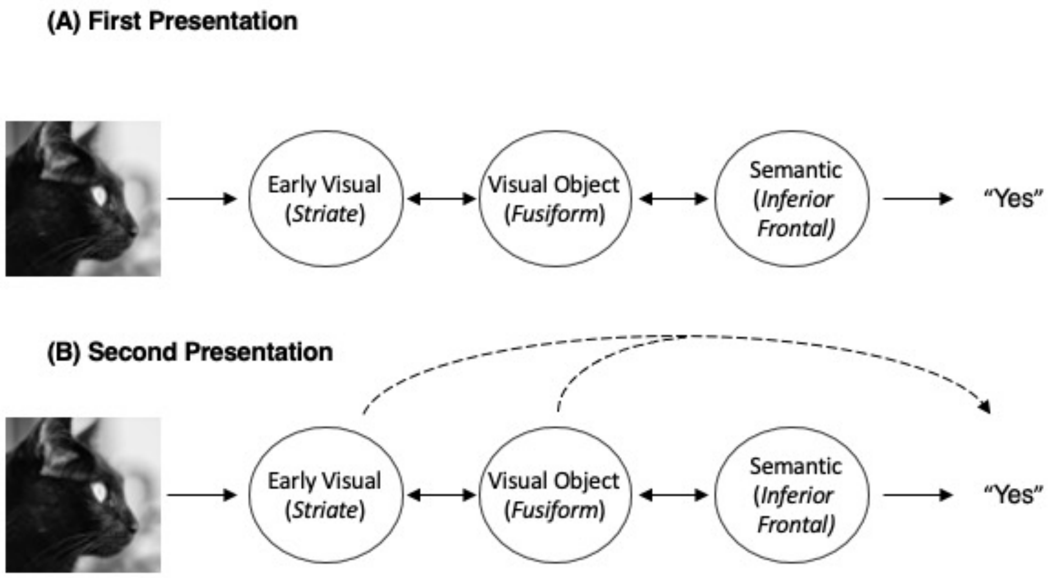
2.5.2 How does priming work?

Considerable intellectual effort has gone into understanding the priming phenomenon, as evidenced by an extensive literature spanning decades. This section briefly outlines two different approaches, one based on a neuroanatomical locus of memory, and one centred on a network model of the mental lexicon.

A "mechanistically plausible" (Shallice & Cooper, 2011, p. 251) theory of how priming functions is that it involves building and strengthening specific stimulus-response links (Logan, 1990), resulting in automaticity, an implicit measure of long-term procedural memory (Figure 2.2). Logan's (1990) account holds that re-exposure to a stimulus leads to a bypassing, or curtailment, of the neural processes that were activated in the initial presentation, resulting in facilitated subsequent behaviour. This formation of a more direct mapping[33] between a repeated stimulus and its response results in a decrease in neural activity within the specific cortical regions activated in the initial presentation (see Figure 2.3).

It should be noted that the schematic in Figure 2.3 belies the complexity involved, even in a seemingly 'simple' Yes / No decision task, and many composite processes are omitted: obeying task instructions,[34] preparing and executing eye movements, multifaceted decision processes, and sensory-motor gestures. The biological marker

Figure 2.3 Schematic of hypothetical component processes in a visual judgement task. Participants are asked to respond to a picture by pressing either the 'Yes' or the 'No' key, according to whether the depicted object is a living thing or not. Several stages are involved in the initial presentation (A), shown with presumed associated cortical regions. (B) A repeated stimulus can effectively bypass some of the component processes, resulting in faster reaction times and decreased neuronal activity in the bypassed regions. Adapted from Horner & Henson (2008).



[33] Made possible due to synaptic plasticity, the ability of the brain to change and adapt to new information, thought to underlie associative or Hebbian learning: "neurons wire together if they fire together" (Lowel & Singer, 1992, p. 211).

[34] Following instructions has been labelled "a process we cannot currently scientifically characterise" (Shallice & Cooper, 2011, p. 57).

of the decrease in neural activity, which can be measured in fMRI and PET studies,[35] has been termed Repetition Suppression (RS) (Horner & Henson, 2008, p. 1979). Although RS has been measured not only in visual repetition priming but in frontal brain areas related to semantic retrieval (for example, Demb et al., 1995), the exact localisation of the architecture of semantic knowledge is much debated; Patterson et al. (2007) suggest that semantic networks are widely distributed, and partly organised to conform to the neuroanatomy of sensory, motor and linguistic systems. Despite the physiological evidence, it is not clear how the stimulus-response account would apply to the phonological priming that is purported to happen in the word-pair experiments. How could there be a more direct mapping between *tall* to *tree*, compared to *tall* to *man,* in the absence of extra repetition or exposure, and which neural processes would be bypassed? Moreover, if priming is held to be an implicit, 'ballistic' process, why do participants often need their attention drawn to the phonological patterns for a mnemonic effect to be observed? Perhaps a clue to this puzzle can be found in a 'spreading activation' account of priming.

> One does not think of the typical priming paradigms, such as naming, [or] lexical decision … as memory tasks of the same character as paired-associate recall or sentence recognition. Nonetheless, the claim made here is that the same spreading activation mechanism is involved in memory retrieval (Anderson, 1983, p. 264).

Modelling lexical access in terms of spreading activation takes as its starting point the assumption that lexical knowledge is arranged in a network of connections. This comparison invites us to visualise a dense "multidimensional cobweb" of words (Aitchison, 1987, p. 72), with multiple intra-lexical connections reflecting semantic, phonological, orthographic, syntactic, and encyclopaedic knowledge (Wilks & Meara, 2002, p. 303). The basic mechanism in spreading activation accounts is that the activation of one 'unit' of representation, for example, of a word, may flow along pathways to other units in the network, which themselves become activated and

---

[35] Two types of neuroimaging techniques, functional Magnetic Resonance Imaging (fMRI) and Positron Emission Topography (PET).

more readily available for processing and production, and this ‘excitatory’ impetus spreads out along various connections in the network.

Mechanisms of this sort can be found in a wide range of theories covering word recognition, memory, and speech production; hence reference to ‘activation’ of some sort can be found in interactive activation models (for example, McClelland & Rumelhart, 1981), connectionist models (for example, Grossberg & Stone, 1987), cohort models (for example, Marslen-Wilson, 1984), and other spreading activation theories (for example, Anderson, 1983; Dell, 1986). Although such accounts may share core similarities in terms of the postulated encoding, storage and retrieval processes, few have solved the intractable “units problem” (Dell, 1986, p. 294) by stipulating what exactly constitutes a cognitive unit of representation - perhaps this reflects a drawback of the inherent relationship between network models and an underlying metaphor of ‘memory-as-storage’; it necessitates a delineation of the thing being stored.

One conception of the ‘unit’ that still has current currency, and which will be adopted going forward, is that of the network ‘node’, which for Dell (1986) represents “concepts, words, morphemes, phonemes, and phonemic features, and … syllables and syllabic constituents” (Dell, 1986, p. 286). The term ‘syllabic constituents’ refers to consonant clusters and syllabic rimes, so presumably the characteristics and granularity of the node must vary across languages (see Chapter 3 for a related discussion regarding the Japanese language).

The amount of activation that spreads between nodes is determined by the strength of the links between them, which in turn depends on the user’s language experience, the amount of repetition or, in experiments, the number of relevant trials. The consequence of spreading activation is that the presentation of a word can lead to an increased likelihood of producing a related word soon afterwards, and the amount of activation of a unit determines the cognitive capacity available to process information, and thus the rate and probability of recall.[36] Hence, *doctor* primes NURSE because when *doctor* is presented, NURSE is activated due to repeated

[36] A verbal, qualitative description such as this can, however, obscure the mathematical summative nature of many of the early models (c.f. Anderson, 1983; Dell, 1986).

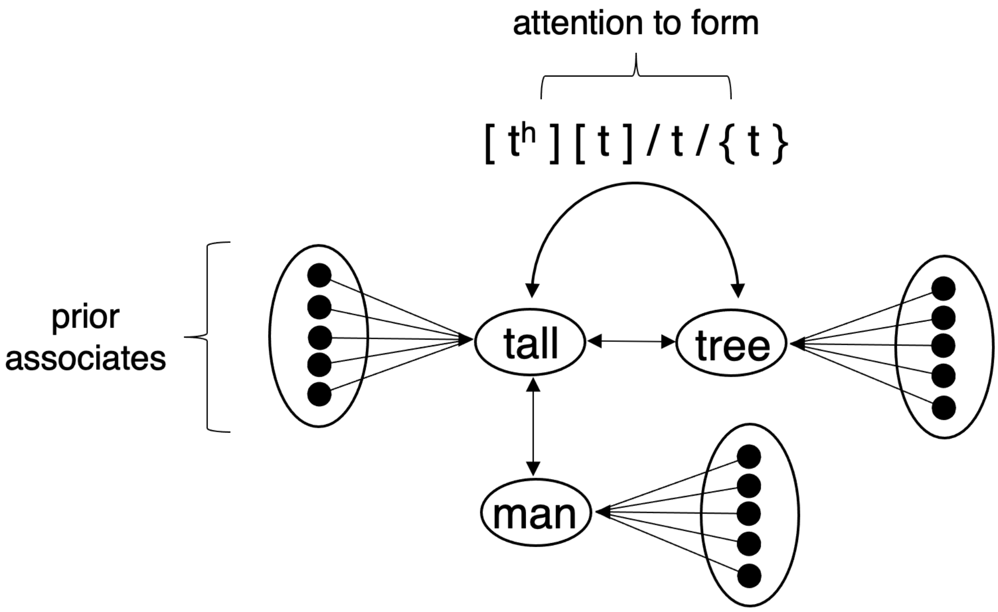
experiences of association; when NURSE is subsequently presented it takes less additional effort to reach a ‘recognition threshold’ because it was already partially activated (example from Meyer & Schvaneveldt, 1971).

Two other features of the spreading activation theory deserve brief mention: links can atrophy or become dormant, which helps to account for memories fading; and links can be ‘weighted’, which helps with modelling a more sensitive account of activation that can cover not only relative exposure but also potentially qualitative features like salience.

How would a spreading activation theory account for the mnemonic effect of a phonological pattern in a word-pair? One possible way to provide additional pathways, or enhance existing pathways, between two network nodes is to perform an *elaboration*: “one of the most potent manipulations that can be performed in terms of increasing a subject’s memory for material” (Anderson, 1983, p. 285). An example of an elaboration given by Anderson is that of a research participant, trying to learn the paired-associate *dog – chair*, constructing a mini-narrative which supposedly improves recall: “Elaborations can influence recall by redirecting activation towards the to-be-recalled material, providing additional sources of activation, and providing a means for reconstructing what has been studied” (p. 292). The narrative elaboration that leads from *dog* to *chair* is semantic in nature, but the idea has been extended to elaborations of form or structure in Barcroft’s TOPRA model (c.f. Barcroft, 2002): copying a word, or crossing out vowels, are given as examples of structural elaboration which can aid recall of a lexical form.

Likewise, in the experiments outlined in section 2.3, the act of ticking or circling the word-pairs which displayed assonance or alliteration could be construed as constituting a structural elaboration, and thus creating an additional pathway for activation. In Figure 2.4, multiple levels of representation - phonological, phonetic or orthographic - create an extra route from *tall* to *tree* and this may account for the mnemonic advantage over the comparable word-pair *tall man*. The use of bidirectional arrows indicate that activation can reverberate or resonate back, and so top-down stimuli (at the word or sentence level) or bottom-up acoustic or graphemic features of the input can all trigger activation and mediate lexical access.

Figure 2.4 Network representation of an elaborative structure generated to bias recall of 'tree' to the cue 'tall'.



One of the first, and most influential, spoken language processing models to use evidence from the priming paradigm was the Cohort model, first proposed by Marslen-Wilson (Marslen-Wilson, 1984; Marslen-Wilson & Welsh, 1978). The Cohort model concerns the way in which the speech signal is mapped, in a bottom-up manner, onto representations in the mental lexicon during the real-time recognition of spoken words. The main assumptions are that when the word-initial acoustic-phonetic information is first heard, in the first few hundred milliseconds of the signal, multiple competing candidates that share this acoustic onset sequence[37] are automatically activated in parallel in the mental lexicon of the hearer – these are the word-initial cohorts of the word. For example, according to Cohort theory, the spoken word /ɛlʌfənt/ (*elephant*) activates all similar sounding word forms in memory that begin with /ɛ/ (e.g. *elephant*, *elevator*, *escalator* etc). As more of the word is heard and more acoustic-phonetic input is processed, the accumulating input will diverge

[37] Here, the term onset is used in the conventional sense of syllabic division: onset (the consonants before the nucleus) and rime (consisting of a vowel nucleus and a coda – the consonants after the nucleus) (Keuleers & Brysbaert, 2010) - see Figure 2.5 in 2.5.3.

from the form of an increasingly higher proportion of the cohort and when the patterns fail to match, the level of activation starts to decay. Thus, once information for /l/ in /ɛlʌfənt/ becomes available to the listener, words such as *escalator* are no longer viable candidates for recognition, though it is assumed deactivation takes some time and that the eliminated candidates "may remain activated for a short period thereafter" (Marslen-Wilson & Welsh, 1978, p. 56).

This process of reduction continues until only one candidate remains that still matches the speech input. At this point, the form-based selection process is complete and the word-form that best matches the speech input can be identified (Lahiri & Marslen-Wilson, 1991, p. 255). Thus, the model is essentially a pre-lexical, form-based process based on acoustic-phonetic analysis of the speech signal. If the signal is ambiguous or insufficient due to, for example, background noise, then top-down contextual constraints derived from the discourse context operate to select between possible candidates, thus allowing even highly reduced forms to be identifiable (Lahiri & Marslen-Wilson, 1991, p. 289). Central to the model is the concept of a 'recognition point', defined as "the sequential point in a word at which it becomes uniquely and securely identifiable – that is, the point where the word diverges from the other members of its word-initial cohort" (Marslen-Wilson, 1993, p. 200). In the example here, the recognition of the word *elephant* occurs roughly around the /f/. Thus, the acoustic input need only be minimally specified, just enough to discriminate it from its competitors. For monosyllabic words, the recognition point often comes at the end – it is only when the final /t/ is heard that *street* can be identified from *streak* or *stream*, for example[38].

The notion of cohort activation predicts phonological priming and takes into account evidence that even pseudowords can prime: as there is no lexical representation for a pseudoword in the mental lexicon, any activation / priming effect is held to be 'prelexical' as no semantic processing is involved. Hence, the presentation of *tall* will activate words beginning with t, including *tree*, this activation

---

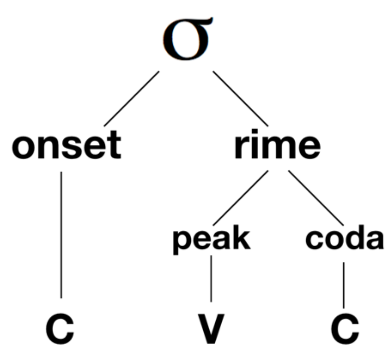
[38] Strictly speaking, there could be subtle phonetic differences linked to, for example, backwards assimilation from the /k/ to increase the backness of the vowel, or from the /m/ to increase nasality.

will start to decay as more of the prime (tall) is heard but if sufficient residual activation remains when *TREE* is presented as a target, facilitation will be faster.

2.5.3 Evidence

The literature on form or perceptual priming is extensive, so when looking for evidence that alliteration or assonance can aid lexical processing, it may help to first determine how such phonological patterns are operationalised within the priming paradigm. In effect, the question of interest is whether hearing a word beginning with or containing a particular sound will make it easier to process a subsequent word with the same sound in the same position. The form-based similarity of a prime and its target is frequently manipulated in terms of matching or overlapping particular segments of the two words, and/or manipulating the number of segments that match. This segmentation is often based on a model of syllabic division[39] seen in Figure 2.5.

Figure 2.5 Segmentation of a Consonant-Vowel-Consonant monosyllable.



Thus, a facilitatory priming effect for alliteration, bearing in mind the working definition in section 1.3.3, would involve test items with matching phonological consonantal onsets, and no other relationship - morphological, syntactic, or

[39] The cogency of this segmentation as a valid psychological construct will be addressed in Chapter 3.

semantic, since they could themselves account for any observed priming effect. Likewise, a priming effect for assonance (defined in 1.3.4) would be based only on the single aspect of a matching stressed vowel sound (the peak, or nucleus).

However, in Zwitserlood's (1996) overview of the form priming experimental paradigm, the stimuli overlap is described as "word-initial or rhyme" (1996, p. 590), with no scope for a priming effect based on assonance alone. Rhyme is usually operationalised as phonological similarity of both the syllabic peak and the coda (the rime), or in other words, similarity starting from the stressed vowel sound to the word offset. As we shall see below, perceptual priming based on assonance alone has received very little attention. As the findings for word-initial similarity are inconsistent, those based on rhyme will be considered first. Indeed, much of the most robust evidence is found in studies where the prime - target relationship is that of rhyme. Thus, participants will identify the target letter string *BEAN* more accurately and faster after being exposed to the prime *mean* than, say, after the prime *pink*. This systematic priming effect for rhyme has been found in both monosyllabic and bisyllabic prime-target pairs, irrespective of frequency, for both words and pseudowords, and independently of the task performed: lexical decision tasks (LDTs) (for example, Norris, McQueen, & Cutler, 2002), shadowing (for example, Dumay et al., 2001), and identification in noise (for example, Slowiaczek, Nusbaum, & Pisoni, 1987).

In contrast to the evidence that rhyme can prime, even under different experimental designs, the situation for a priming effect for word-initial similarity is more complex and inconsistent, with facilitatory, inhibitory and null effects being reported. Although the same dependent variable of response latency is usually the focus across experimental designs, task differences can introduce different variables, so it may be helpful to focus on just one task, the lexical decision task (LDT), to glean insight from the results. When looking at word-initial overlap that forms an alliterative relationship between the prime and the target, a set of "rather messy results" (Dumay et al., 2001, p. 121) can be found in the literature, as seen in five representative studies in Table 2.6.

Table 2.6 Examples of LDT studies which include alliterative word-initial overlap.

Study	Examples of *primes: TARGETS*	Effect [1]
1) Slowiaczek & Pisoni (1986) (Experiment 1)	*black, bland, bleed, burnt, /blæt/, /blim/, /brεm/*: *BLACK*	No effect [2]
2) Radeau, Morais & Dewier (1989) (Experiment 2)	*palais, poulet, rouler: PARURE* (French L1 speakers)	No effect [2]
3) Goldinger, Luce, Pisoni & Marcario (1992) (Experiment 3)	*bang: BONE*	Facilitation [3]
4) Monsell & Hirsh (1998) (Experiment 1)	*broom: BRUISE*	Inhibition
5) McQueen & Sereno (2005)	*zeep: ZOON* (Dutch L1 speakers)	No effect

[1] Slowiaczek & Hamburger (1992) propose facilitation is due to activation / excitation at a prelexical phoneme level, and inhibition is the result of competition between words at the lexical level.
[2] A facilitatory effect, in terms of faster reaction times, was found only when the prime and the target were identical (a repetition effect).
[3] An auditory priming technique was used and a facilitatory effect was found only when the targets were presented in white noise.

From the examples given in the table it is clear that the nature of some of the prime-target relationships goes beyond a simple alliterative pattern and encompasses a matching vowel nucleus (for example, broom – BRUISE in Monsell & Hirsh, 1998).

In Studies 1 and 2 no priming effect was found when primes shared one, two or three phonemes with the target, results which are inconsistent with predictions based on cohort theory (Marslen-Wilson, 1984, see section 2.5.2). The use of real words as primes further confounds the issue, as the relationship between prime and target is no longer purely form-based: for example, in Study 1, if *burnt* had shown a priming effect for *BLACK*, it might have been due to semantic association rather than form-based similarity.[40]

It might seem that the best way to avoid semantic priming (which might sometimes be rather indirect and individual) would be to use pseudowords. However, these too can create difficulties and the construction of the pseudowords may also be a factor in the lack of clear results. For example, in Study 1, some of the pseudoword targets are pseudo-homophones (a non-word that sounds like a real word), such as /*stik*/ /*skot*/ /*bæns*/ and /*slæk*/ (it is unclear how these were spelt in the visual presentation): Jiang (2012, p. 82) advises against the use of such items as several studies have shown that pseudo-homophones generally take longer to reject than pseudowords, and these could affect the mean RTs in the different conditions.

The experiments in Table 2.6 also incorporated different timing elements in the designs, specifically the interstimulus intervals (ISIs), the interval between the offset of the prime and the onset of the target, and differences in stimulus onset asynchrony (SOA), the interval between the onset of the prime and the onset of the target. Lengthening these two variables is often claimed to increase strategy use (Radeau et al., 1989). While none of these was itself uncontrolled within its study, they are all potential variables determining the opportunity for an observable effect, and between them these variations dilute the strength of evidence for initial consonant priming.

The failure to find a facilitatory effect with reaction time measures could be due to differences in methodologies, dependent measures and/or stimuli and, as reiterated by Goldinger et al., (1992), null results do not support any definitive conclusions. Rastle & Brysbaert's (2006) exhaustive meta-analysis of the relevant priming literature led them to conclude that phonological primes *do* have an effect, and some experimental results cannot be attributed to task-based biases or strategy use. However, the evidence for a priming effect based solely on alliteration is ambiguous.

---

[40] Aside from the *burnt-BLACK* example given in the paper, the authors only list the target items in the appendix, not the items used as primes, so it is not possible to ascertain the overall number of potential semantic associations between the test items.

Furthermore, as mentioned previously, there is scant data to make any reliable assumptions about an equivalent effect for an assonating pattern. These issues will be explored further in Chapter 5.

2.6 Conclusion

The overarching goal of this thesis is to understand the extent to which phonological patterns such as alliteration and assonance can help English L2 learners with the substantial task of learning FSs, here operationalised as word-pairs. The evidence reviewed in this chapter indicates that such patterns do have a variety of effects on how we use and process language:

- The studies in section 2.2.1 indicate that phonological patterns impact on how we, as consumers, positively evaluate brands and advertisements, and ultimately our choice of products.
- Phonological patterns can affect not only how we recall poetry, prose, and narrative text, but can also influence the language we produce (section 2.2.2).
- Alliteration and assonance can help English L2 learners with the recall and retention of the written form of (mostly familiar) collocations and compounds (section 2.3).
- Phonological similarity plays a role in the cognitive phenomenon of priming (section 2.5).

These claims generate many potential research questions, of which the following four will be addressed in this thesis:

RQ 1: Do Japanese learners of English approach phonological patterns in the same way as other learners?
RQ 2: How replicable is the work of Boers et al.?
RQ 3: Is the mnemonic effect for alliteration and assonance based on a priming mechanism?

RQ 4: Do alliteration and assonance support recall for English L2 learners when item meaning is controlled for?

Further questions that arise will be dealt with in the respective chapters, and in the extensive discussion in Chapter 7. So far, the literature has suggested a mnemonic advantage for alliteration and assonance for English L2 learners, but few have questioned the role of the L1 in perceptions of L2 phonological similarity. As the vast majority of the participants in the research reported in this thesis are Japanese, it is important to establish if their L1, and the phonological patterns associated with it, might trigger different responses to those in the participants of previous research, who have mostly been speakers of English and Dutch. This is the focus of Research Question 1, and the topic of the next chapter.

Chapter 3 Do Japanese learners of English approach phonological patterns in the same way as other learners?

3.1 Introduction

This chapter explores the question of whether the learners' L1 has any influence on the potential mnemonic effect of a phonological pattern, and in particular, if Japanese as an L1 creates any fundamental differences in phonological awareness and recall of a patterned word-pair. If there are such effects, this needs to be considered when relating the body of empirical work (Chapters 3 - 6) to the existing literature.

It may be recalled that in the majority of the classroom based quasi-experiments outlined in Table 2.4, the L1 of the participants was standard Dutch. The one experiment that did have a sample of Japanese L1 participants (Boers, Lindstromberg, et al., 2014, reviewed in Section 2.3.3), employed a very different methodology compared to the other studies, in that it looked at the acquisition of idiomatic word-pairs embedded in a graded reader. Although that investigation did find the participants made statistically significant gains on the recognition of the form of the alliterative word-pairs, there are two key reasons why it may be prudent to avoid basing a conclusive finding on this study. Firstly, only three target items alliterated – a small sample from which to make a strong inference. Secondly, as noted in 2.3.3, there may have been a test-taking effect that was not completely controlled for, as the participants had multiple exposure to the intact items prior to the delayed tests. The question remains, therefore, if the findings from the body of research primarily based on Dutch participants are generalisable to a wider population.

3.2 What might make L1 Japanese learners of English different?

By way of introducing the notion of how a Japanese L1 speaker might perceive a phonological pattern, an anecdote might prove illustrative. On running into a student of mine on campus, I could not immediately recall her name, but was aware that I knew it - a typical tip-of-the-tongue phenomenon. It is well-known that the first letter of a word is a useful retrieval cue in such situations, so I asked her for the first letter of her name as a clue; she promptly replied /tʃi:/, as in Chi(ka). That is, rather than segment her name into individual elements such as the initial letter <C> or the first sound /tʃ/ as a clue, Chika spontaneously divided her name it into two larger units and offered the first.

If Chika, in identifying the ‘first sound’ of her name as a consonant plus following vowel, was illustrating a general tendency in L1 speakers of Japanese, then it might affect the extent to which they would behave like Dutch L1 speakers given alliterative and assonating word-pairs. Therefore, it makes sense to examine why Chika might have segmented her name in that way, by looking at features of the Japanese phonological system. To this end, there follows a brief overview of the relevant aspects of Japanese phonology and writing, with particular attention to the mora (a prosodic unit) and its relationship to the syllable. With this information in place, we can then consider how disparities with English and Dutch might impact the mnemonic effect of the phonological patterns under discussion.

3.2.1 A brief description of the Japanese language

Modern written Japanese uses multiple, concurrent, and typologically distinct writing systems in the same text, resulting in “undoubtedly the world’s most complicated orthography” (Vance, 1987, p. 2). In addition to Kanji for content words (morpheme-based, Chinese ideographic characters), there are two forms of phonographic Kana: Hiragana, a syllabary for syntactic or functional words, and Katakana, a syllabary for foreign or loan words - see Example 3.1.

Example 3.1 Breakdown of a simple Japanese sentence.

Sentence:	私は毎朝コーヒーを飲む。						
Translation:	I drink coffee every morning.						
Scripts:	Kanji	Hiragana	Kanji	Katakana	Hiragana	Kanji	Hiragana
	私	は	毎朝	コーヒー	を	飲	む
Romaji [1]:	watashi	wa	mai asa	kōhī	wo	no	mu
Syntactic and semantic equivalents:	*I*	Topic marker	*every morning*	*coffee*	Object marker	*drink*	Non-past, plain form

[1] Although not part of the traditional writing system, *romaji*, a transliteration system based on the Roman alphabet, is also used.

In Japan, the Ministry of Education's guidelines stipulate which scripts are taught and when, with formal literacy instruction starting at around the age of six with Hiragana, and Katakana taught over three years from the age of seven (Koda, 2017, p. 59). These Kana both encode rhythmic units of sound called morae (singular: mora, often represented by the Greek letter μ), which Ladefoged (1982, p. 226) defines as "a unit of timing" of equal length. Hence, Japanese is often referred to as a 'mora-timed language'. It should also be noted that the kana system is transparent, with a one-to-one relationship between one kana and one mora.

3.2.2 Morae

The description of Japanese as a mora-timed language is not entirely uncontested and the relationship between morae and syllables within the standard Japanese phonological system is not without controversy. While some researchers (such as McCawley, 1968) have claimed that Japanese must be a syllabic language if the syllable is held as a phonological universal, others have disputed this claim and hypothesise syllable-free accounts of phonological structure for Japanese (for example, Labrune, 2012). In between these two positions, researchers posit prosodic models in which the mora acts as a sub-syllabic unit (for example, Kubozono, 1989) – for a comparison of models, see Lee (2016).

Perhaps this debate stems partly from the fact that, on the surface, there seems to be little difference between the two units. Take, for example, the Japanese word *namida* 涙 'teardrops'. Both an English L1 speaker and a Japanese L1 speaker would most likely agree that the word can be broken down into three segments: in the case of the English speaker, three syllables /$_{\sigma}$na $_{\sigma}$mi $_{\sigma}$da/, and for the Japanese speaker three morae, /$_{\mu}$na $_{\mu}$mi $_{\mu}$da/. Both the mora and the syllable here have one consonant onset with one vowel nucleus, giving the appearance of a direct parallel between these two units of segmentation. It should be noted that the morae in *namida* are prototypical: the CV structure consisting of a single consonant followed by a short vowel is the predominant type in modern standard Japanese[1] – hence the considerable amount of overlap with a syllable. There are also morae consisting only of a vowel sound, as in the first and last sounds of /$_{\mu}$i $_{\mu}$ri $_{\mu}$e/ 入江 'inlet', which, again, could be perceived as having three syllables.

However, when we depart from the prototypical morae and consider codas, long vowels and geminates, the contrast with syllabic units becomes more apparent. Starting with codas, if we take the example *hon* 本 'book', it appears to consist of a single segment. But in Japanese phonology, consonants[2] can only appear in the initial position of a mora, meaning that the /n/ in *hon* must be detached. This is accommodated by Japanese having a moraic nasal[3] /N/ (ん or ン depending on which Kana is called for) – this is in fact the only legal consonantal coda. As a result,

---

[1] Otake, Hatano, Cutler & Mehler (1993) estimate between 60 - 70% of all possible morae have a CV structure.

[2] Consonant clusters and consonant codas (except /N/) are not part of the legal phonemic inventory, so when loanwords are adopted, vowel epenthesis is employed: for example, the bi-syllabic English word *Christmas* becomes penta-moraic in Japanese - /$_{\mu}$ku $_{\mu}$ri $_{\mu}$su $_{\mu}$ma $_{\mu}$su/ (クリスマス).

[3] The moraic nasal /N/ has a wide range of phonetic realisations determined by the phonetic context, and can be represented by the sounds /n/, /m/ or /ŋ/,for example /$_{\mu}$ko $_{\mu}$N $_{\mu}$bu/ /kõm:bɯ/ 昆布 'kelp'.

hon can be broken down into two morae of equal length /$_{\mu}$ho $_{\mu}$N/. Thus, the differences between mora and syllable can be more readily seen in a word such as *London,* which is bi-syllabic in English, but has four morae in Japanese: /$_{\mu}$ro $_{\mu}$N $_{\mu}$do $_{\mu}$N/ (ロンドン).

A further discrepancy between syllables and morae can be seen in words with long vowels and geminates, or 'doubled' consonants, which both count as two morae. An example of a long vowel, here represented by /H/, can be seen in the word *kinō* 昨日 'yesterday', which could be segmented into two syllables but is three morae in length /$_{\mu}$ki $_{\mu}$no $_{\mu}$H/. The same applies to geminates (represented as /Q/), for example *kekka* 結果 'result' with three morae /$_{\mu}$ke $_{\mu}$Q $_{\mu}$ka/ [kʲekkɑ].

For the present purposes, the debate whether morae and syllables coexist in a single hierarchical structure is of secondary import; of greater concern is how a Japanese L1 speaker might approach the segmentation of an English word string. The following account will adopt the classification of a mora as a sub-syllabic rhythmic unit, as this seems to be most widely accepted in the literature (Vance, 2017, p. 21), plus, as we shall see in section 3.2.3, evidence suggests that the mora takes precedence over the syllable when Japanese L1 speakers segment a word-string.

That being said, it has proved notoriously difficult to define the boundaries of any speech segment (be it a mora, syllable, or phoneme) in just the acoustic signal that corresponds to some phonetic reality. Recordings, whether by microphone, electromagnetic articulography, X-ray or ultrasound, reveal a continuous flow with peaks and troughs of energy or movement (John Clark, Yallop, & Fletcher, 2007, p. 58). Thus, the idea that speech can be segmented into discrete sounds is to a considerable extent a reflection of conceptual abstraction, rather than a direct property of language. Traditional phonological theory holds that segments are phonological constructs and, as such, they can be used to describe any language in principle.[4]

---

[4] Having said that, proponents of Exemplar theory (for example, Ambridge, 2020; J. Pierrehumbert, 2001; Port, 2007) argue against stored abstractions, though it is beyond the scope of this thesis to discuss the differences.

It has been argued (for example, John Clark et al., 2007; McMahon, Heggarty, McMahon, & Maguire, 2007; Otake et al., 1993) that phonological constructs condition not only how speakers perceive the acoustic input, but also how they produce it (see 3.2.3). This conditioning is partly due to knowledge of the L1 orthography, which mediates intuitions about segmentation and facilitates an analytical awareness of structure (as claimed by Beckman, 1982 ; John Clark et al., 2007). If this is indeed so, then might the mora, as a salient linguistic unit for literate L1 speakers of Japanese, affect their perceptions, learning behaviour and judgements when encountering English as an L2?

3.2.3 Evidence for moraic segmentation

Justification for the existence of morae as psychologically salient segments of speech for Japanese L1 speakers is partly based on evidence that begins to appear in early childhood. Research on children's oral wordplay shows segmentation at points that correspond to mora boundaries rather than syllables. For instance, in the traditional word game *Shiritori* (literally 'buttock taking'), two players take turns saying a word that begins with the last sound unit of the word given by the previous player. The game is over when one player gives a word that ends with the moraic nasal /N/, as no Japanese word begins with this sound, indicating the game obeys Japanese word structure conditions – see Example 3.2:

Example 3.2 Sample of Shiritori (from Katada, 1990).

tubame ('swallow')
medaka (a type of fish)
kao ('face')
ongaku ('music')
kusuri ('medicine')
ringo ('apple')
gohan ('meal')
(game over)

Evidence that favours a segmentation strategy based on morae over syllables can be seen in words that end or begin with a unit containing a long vowel, geminate, or syllabic nasal. In Example 3.2, the word *kao* ('face') consists of a single syllable, but is treated as two units /$_{\mu}$ka $_{\mu}$o/, the latter being the same as the initial unit of the following word *ongaku* - /$_{\mu}$o $_{\mu}$N $_{\mu}$ga $_{\mu}$ku/.

Furthermore, if the operating unit was a syllable, the game could continue after *gohan*:

gohan ('meal')
hantai * ('objection')

Further studies suggest that cumulative experience with the writing system and literacy biases perception and production towards a moraic segmentation. For example, Inagaki, Hatano & Otake (2000) compared the speech segmentation strategies of 4- to 7-year old children. The researchers adopted a motor-vocal segmentation methodology from prior research on assessing children's syllabic awareness, in which children were shown pictures and asked to move a doll across a series of circles while they articulated the words represented by the pictures. Pictures were chosen to represent Japanese words containing an assortment of prototypical morae and the nasal codas, geminates, and long vowels.

The researchers found that the younger children, who were taken as representing a lower level of literacy, adopted a mixture of mora-based and syllable-based strategies. For example, shown a picture of an aeroplane (in Japanese *hikōki* /$_{\mu}$hi $_{\mu}$ko $_{\mu}$H $_{\mu}$ki/), or a crayon (/$_{\mu}$ku $_{\mu}$re $_{\mu}$yo $_{\mu}$N/), the younger children tended to move the doll three circles, which was interpreted as a mix of moraic and syllabic segmentation. In contrast, the older children tended to move the doll four circles. This was construed as indicating that the older children's phonological awareness was shaped by their instruction in Kana.

Evidence from research on adult Japanese L1 speakers also suggests an inclination to segment speech according to morae as basic units, rather than syllables. For example, in the first and third experiments reported in Otake et al. (1993), the researchers used a target monitoring task with Japanese L1 undergraduates and analysed their reaction times and the proportions of hits and misses, to determine how they segmented Japanese words. Participants were given a target such as *TA* (visually presented in Roman alphabet[5] in Experiment 1, auditorily presented in Experiment 3) and were then asked to listen to a list of real words (which would normally be written in kanji), one of which might begin with the specified target, and to press a response key as soon as they had detected an occurrence of the target in the auditory speech signal.

Stimulus words differed in syllable structure, the first syllable was either CV (such as *ta*) or CVC (*tan*). Note that for the former, the first syllable is analogous to one mora, /$_{\mu}$ta/, but in the latter, the first syllable comprises two morae, /$_{\mu}$ta $_{\mu}$N/ – see Table 3.1.

Table 3.1 Sample of stimulus item pairs from Otake et al. (1993).

Stimulus Item	Syllable Structure	Mora Structure	Gloss
tanishi	CV – CV - CV $_{\sigma}$ta $_{\sigma}$ni $_{\sigma}$shi	3 morae /$_{\mu}$ta $_{\mu}$ni $_{\mu}$shi/	a type of pond snail
tanshi	CVC - CV $_{\sigma}$tan $_{\sigma}$shi	3 morae /$_{\mu}$ta $_{\mu}$N $_{\mu}$shi/	‘terminal’

Participants had fast reaction times and low miss rates (less than 8%) when they were asked to identify a CV target (*TA*) in a CVCVCV word (such as *tanishi*), or a CVNCV word (like *tanshi*). Participants had low miss rates but longer reaction times when they had to identify a more complex two-mora CVN target (such as *TAN*) in a CVN word. However, when participants had to identify the same CVN target within a CVCVCV word, the miss rate was 64.3%, and this difference was statistically

[5] Using the Roman alphabet was justified on the grounds that it would remain neutral with regards to a syllabic or moraic representation. Targets could not be presented in kana script as this would pre-empt the participants’ decisions. It was also held that the undergraduates would have had enough prior exposure to the Roman alphabet to easily map the letters (e.g. *TA*) into sound sequences.

significant.[6] In other words, the Japanese participants had great difficulty in identifying moraic units across a mora boundary and this pattern of results did not hold for non-Japanese L1 speakers (English L1 participants and French L1 participants were tested on the same materials in Experiments 2 and 4 respectively).

This series of experiments only used materials with CV mora and the nasal coda /N/; to fully test the hypothesis of mora segmentation would require further trialling with words containing the long vowels and the geminates. However, there are two other lines of evidence that lend further credence to the notion of moraic segmentation. Firstly, the analyses of spontaneous speech errors made by Japanese speakers reveal a very different pattern to English L1 transposition errors (reversals or spoonerisms). English speech errors, such as '*three cheers for our queer old dean!*' rather than '*our dear old Queen*' (example from Lederer, 2014), follow a typical pattern of exchange such as an onset with an onset as in the example just given. There is also evidence of exchanges involving a nucleus with a nucleus ('*Wing's babliography*' rather than '*Wang's bibliography*'), or a coda with a coda, ('*cuff of copy'* rather than '*cup of coffee*') (the latter two examples taken from Fromkin, 1971). Such errors are taken as reflecting the psycholinguistic validity of the syllable as a unit of segmentation. In contrast, Japanese slips of the tongue reveal segmentation on the peak / coda boundary, or long vowels and geminates splitting into two morae (Kubozono, 1989). Indeed, Tsujimura (2007, p. 64) asserts that all attested speech errors in Japanese are consistent with moraic segmentation.

One final source of evidence to support the hypothesis of moraic segmentation is the fact that the only metrical unit of traditional Japanese poetry is moraic and not syllabic (Vance, 1987, p. 67). Perhaps the most familiar example of this is the haiku, composed strictly on the basis of 17 morae, in a 5-7-5 moraic pattern.[7]

In sum, this evidence helps support the assertion that the psychological reality of morae for Japanese speakers is "beyond dispute" (Vance, 2017, p. 21). It may also

[6] The authors note that because the miss rate was so high in that condition, there was insufficient reaction time data for a balanced and valid comparison across conditions (Otake et al., 1993, p. 265).

[7] Non-Japanese speakers tend to describe haiku in terms of the number of syllables. When written in English, haiku follow a 5-7-5 syllabic pattern.

explain why Chika did not give the initial letter or sound of her name as a clue. As they were not salient units, she intuitively offered the mora /μtʃi:/.

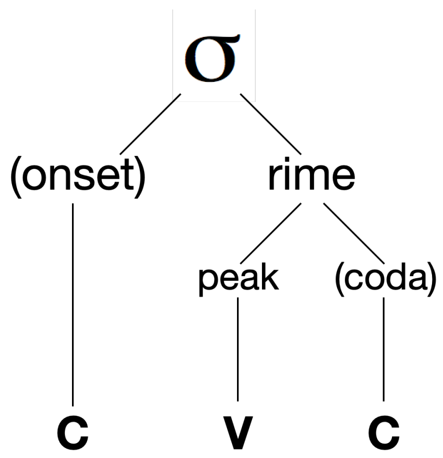
Having established that word segmentation for Japanese L1 speakers is likely to feature morae, the contrast can now be made with English and Dutch L1 speakers.

3.2.4 English and Dutch segmentation strategies

Just as Japanese is often described as a 'mora-timed' language, English and Dutch are usually deemed 'stress-timed' languages, with the syllable being the unit of rhythm. As with the mora, attempts to demarcate syllables in terms of acoustic properties (based on the notions of sonority or prominence) or chest pulses, rarely correlate with speakers' intuitions (Ladefoged, 1982).

However, a large body of linguistic and psycholinguistic evidence suggests that the English syllable has a hierarchical internal structure with two[8] main parts – an onset (an initial consonant or consonant cluster) and a rime (a vowel and any following consonants) – see Figure 3.1. Parentheses indicate that the onset and the

Figure 3.1 Tree diagram depicting the postulated structure of a syllable.



[8] In some models, the syllable may also contain an optional appendix of inflectional suffixes (for example, the /s/ in *sixths* after the coda /ksθ/).

coda are optional, while the vocalic nucleus, or peak, is an obligatory component of the syllable.

MacKay (1970) was one of the first researchers to provide behavioural evidence that onsets and rimes are cohesive units for speakers of English, after examining 'slip-of-the-tongue' errors produced in spontaneous speech. The preference for English speakers to divide words on the onset / rime boundary is further supported by behavioural evidence from experiments using nonwords. Treiman (1983) found, over a series of seven experiments, that adult English L1 speakers consistently preferred word-game rules that treated onsets and rimes as units, rather than rules that segmented words according to phonemes or (imagined) spelling. It may be useful to describe some of these experiments in a little more detail as the results highlight how English L1 speakers conceptualise the syllable.

In Experiment 1, a Training Phase involved an oral demonstration of how a CVC stimulus nonword could be transformed into two response syllables – see Example 3.3.

Example 3.3 Part of the training phase in Experiment 1 (Treiman, 1983).

/kɪg/ → /kæz/ and /ɪg/
/buːf/ → /bæz/ and /uːf/
/tep/ → /tæz/ and /ep/
/nəʊl/ → /næz/ and /əʊl/

In the following Test Phase, the same twelve participants heard a list of 20 nonwords, ten of which had a CCVC structure, and were asked to apply the same rule from the Training Phase. Of the possible solutions for these nonwords, one was a syllabic segmentation and one a phonemic segmentation – see Table 3.2.

Table 3.2 Part of the testing phase in Experiment 1 (Treiman, 1983).

Example Stimuli	Response under syllabic segmentation	Response under phonemic segmentation
/skef/	/skæz/ and /ef/	/sæz/ and /kef/
/glɔːʃ/	/glæz/ and /ɔːʃ/	/gæz/ and /lɔːʃ/

For the ten CCVC items, there was an overwhelming preference for retaining an intact onset rather than dividing the nonword after the initial consonant. This preference was shown by all the participants, at a statistically significant level.

One possible confound was that participants might have been focusing on the VC structure of the second response syllable they had learned in the Training Phase. Experiment 2, which employed a methodologically similar Training Phase then Testing Phase, ruled out this confound: in the participants' responses there was a preference for keeping the entire coda intact over a rule that referred to the final consonant. For example, after being taught /fug/ + /vɪ/ → /vɪg/ and /fu/, most participants generalised the rule so that /bɪld/ + /vɪ/ = /vɪld/ /bɪ/, rather than /vɪd/ /bɪl/).

Experiment 3 looked at the cohesiveness of the rime and asked: if participants are taught one rule for CV stimuli, and a different rule for VC stimuli, then how would CVC stimuli be segmented? Participants were taught two different rules and practised applying them to CV and VC stimuli before the Testing Phase in which they were presented with CVC nonwords - see Table 3.3.

Table 3.3 Parts of the training phase and testing phase in Experiment 3 (Treiman, 1983).

	Training Phase	
Rule	Response Syllables	Example
A (CV stimuli)	C + /əʊ/ and /d/ + V	/tʃaɪ/ → /tʃəʊ/ + /daɪ/
B (VC stimuli)	V + /d/ and /əʊ/ + C	/uːʃ/ → /uːd/ + /əʊʃ/
	Test Phase	
CVC target	Rule A response	Rule B response
/fuːg/	/fəʊ/ + /duːg/	/fuːd/ + /əʊg/

The results showed that Rule A responses significantly outnumbered Rule B responses, a pattern shown by 11 of the 12 participants. This was taken as evidence that participants conceptualise a CVC syllable as C+VC, rather than CV+C.

In other versions of the experiments, the researcher asked participants to learn rules for two new word-games: one set of rules that divided nonwords at a postulated constituent boundary, and one set of rules that divided the same syllable at some other point.[9] Participants consistently made more errors under the rule that broke the internal structure of the syllable (see previous Figure 3.1).

This preference to treat onsets and rimes as units has also been found when the participants were 8-year old children (Treiman, 1985) and also applies to real words (Treiman, 1986). Thus, if asked to blend two monosyllabic CVC words, an English L1 speaker would most likely do thus:

b(ig)	+	*(p)at*	→	*bat*
C(VC)	+	(C)VC		

Although there seems to be little corresponding research on Dutch, as Dutch L1 speakers also segment word-strings into syllables (Verhoeven, 2017; Vroomen, Van Zon, & De Gelder, 1996), it seems reasonable to infer that the onset and the rime would be perceived as distinct units.

Having seen that English L1 speakers divide syllables according to an internal hierarchy, the comparison can now be made as to how a Dutch L2 learner and a Japanese L2 learner might identify phonological patterns in English word-pairs.

3.2.5 Segmentation and perceptions of alliteration

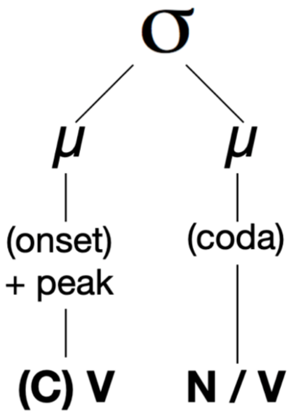
This section considers how a speaker's L1 may affect the perception of a phonological pattern, looking specifically at Dutch speakers and Japanese speakers and their awareness of alliteration (assonance will be dealt with in Chapter 4).

The evidence outlined in 3.2.4 suggests that a Dutch L1 speaker would distinguish onsets and rimes in an English word-pair, for example: *b(ig) b(and)*. However, in light of the findings seen in 3.2.3, it seems reasonable to ask whether a Japanese L1 speaker, for whom the only relevant prosodic unit is the mora – see Figure 3.2 – would do the same.

---

[9] The two sets of rules were presented in different training sessions approximately one week apart. The experiments used a counterbalanced measures design.

Figure 3.2 Tree diagram depicting the postulated structure of a prototypical mora as a sub-syllabic unit (parentheses indicate optional components).



The influence of the mora means the onset and the peak are inextricably linked, resulting in the perception of a strong bond between the word initial consonant and the following vowel. Thus, a Japanese speaker would most likely combine *big* and *pat* thus:

bi(g)	+	*(pa)t*	→	*bit*
CV(C)	+	(CV)C		

Clearly then, it is conceivable that if alliteration is defined in terms of similarity of the initial unit, a Japanese speaker would not notice a phonological pattern because the two units are perceived as dissimilar. Applying a moraic segmentation to the preceding example, *big band*, would result in *(bi)g (ba)nd.*[10] This raises the question, would a Japanese L1 speaker perceive $/_{\mu}bi/$ *and* $/_{\mu}ba/$ as alliterative?

Indeed, Kubozono & Ota (1998) report that Japanese native speakers recognise alliteration in a phrase such as *Kudo no ku,* but do not recognise alliteration in *Kato no ku*, which suggests a linkage between the consonant and the vowel in the Japanese speakers' minds. Though there is not a great deal of current research published in English on this matter, similar claims can be found elsewhere. For example, in the context of Japanese marketing, Tamori, Tatsuki & Tominaga (2008,

[10] As this consonantal coda is illegal in Japanese phonology, vowel epenthesis would be used $/_{\mu}N\ _{\mu}do/$.

p. 68) claim that alliteration and assonance are “not important or relevant” in advertising due to the moraic influence on the perception of patterns. One empirical study concluded that “the particular characteristics of Japanese phonology may prevent Japanese listeners from being able to notice the echoing effects in English rhymes” (Otaka, 2009, p. 2). Otaka drew this conclusion after conducting a discrimination task with Japanese university students (*N* = 199), in which they were asked to mark which English and Japanese phrases in a list contained a rhyme (operationalised as covering both alliteration and prototypical rhyme). None of the participants identified the phonological repetition in *Matsuoka Misako,* and only 9% of the participants identified the alliteration in *Mickey Mouse,* but participants were more successful when there was a CV match, as in *kill the king*, which was correctly identified by 50% of the subjects.

If Japanese L1 speakers cannot intuitively perceive alliteration in English, it follows that alliteration has no mnemonic advantage for them, and is therefore of little benefit as a pedagogical aid. This raises the question, could Japanese speakers be taught to segment according to the stress-units of English, rather than the mora? Cutler et al. (1986) found that English speakers listening to French, a syllable-timed language, could not exploit syllabic rhythm to aid segmentation. Furthermore, Cutler & Butterfield (1992) cite evidence which suggests even highly proficient French-English bilinguals only have one segmentation procedure available to them – either the syllabic segmentation typical of French monolinguals or the stress-based segmentation of English monolinguals.

The observations made so far strongly predict that Japanese learners of English as an L2 would find alliterative patterns masked by their dominant sensitivity to morae. The experiment reported below tests this prediction. The research question is:

RQ: Will L1 Japanese learners show an alliteration effect in familiar word-pairs, in keeping with previous research on non-Japanese speakers?

This question was operationalised by using a dictation task, in line with those used in the suite of experiments outlined in Table 2.4, in which participants were asked to identify alliteration in a decontextualised spoken word-pair. The experiment used high-frequency word-pairs that were judged to be known to the participants. Immediate and delayed free recall and cued recall tests were used to test retention.

It will be helpful here to comment on the tension between using an experimental design that predicts difference and the underlying prediction outlined above, that Japanese L1 speakers will be relatively insensitive to that difference, because their dominant perception of morae will somewhat mask the alliterative patterns in the stimuli.

There is no easy way to design an experiment to demonstrate 'no effect'. But nor would it be desirable, for two reasons. Firstly, Japanese L1 learners of English are still more like Dutch L1 learners of English than they are unlike them, and accordingly, the previous research offers a basis for predicting that there will be an alliteration effect. Secondly, it would be too extreme to predict that Japanese participants will not be sensitive to alliteration *at all.* Rather, it is a question of whether they are *as* sensitive and whether that translates into enhanced recall. Independently of the phonological considerations, there would probably be some sort of effect from the spelling, which would undermine a prediction of no advantage for alliterating words.

For these reasons, the experimental hypothesis is that there *will* be an advantage for alliterating items, and this is how the statistical analysis will be conducted. Following the report of the method and results, there will be discussion about the relationship between the observed patterns and the proposal that morae might mask the ability to notice and exploit alliteration.

3.3 To what extent can Japanese learners identify alliteration: Method

3.3.1 Participants

The participants were a convenience sample of over 100 first and second year undergraduates from four faculties (Economics, Humanities, Law and Politics, Science and Technology) at a private university in Tokyo. Their first language was Japanese, and they had approximately 6-7 years of prior English instruction. The participants all had compulsory English classes twice a week for three hours in total, and the experiment was conducted in regular class time with their class teacher (myself). In terms of English language proficiency, the participants' mean TOEIC score was 527 (*SD* = 112.5, min. = 250, max. = 810), which roughly correlates with the upper A2 level of the CEFR. Additional testing using the Vocabulary Levels Test (VLT) (Schmitt et al., 2001) gave mean scores of 29.1/30 on the K-1 band, 28.4/30 on the K-2 band, and 22.4/30 for the Academic Word List.[11] In the initial data gathering, there were 171 participants. However, due to absences some participants did not complete all the stages of the experiment, so their data was omitted from the analysis, leaving a final sample size of 124. All the participants gave written informed consent,[12] an example of which can be found in Appendix 1.

3.3.2 Materials

Although many of the issues associated with lexical recall and post-tests can be solved by using nonwords or pseudowords (Nation & Webb, 2011), this could negatively influence the behaviour of the participants as the treatment was not part of their compulsory course syllabus. Therefore, it was deemed important that

[11] A 'pass' score on the VLT is often defined as 27/30 (Nation, 1990, p. 143).

[12] All the experiments reported in this thesis conformed to Cardiff University's ethics requirements as stipulated at the time of the investigation.

authentic word-pairs were used. A set of twenty adjective + noun collocates was compiled in the way described below. Half of the phrases alliterated, and the others showed no alliteration. There were four additional filler word-pairs that were used to help counter primacy and recency effects on serial recall: these extra items did not form part of the post-tests or analysis. The word-pairs were matched as far as possible to control for confounding variables that can affect recall, such as orthographic and syllabic length, corpus frequency, concreteness of meaning, ease of translation and syntactic structure, as follows.

To compile the set, a pool of potential high-frequency monosyllabic noun lemmas was chosen from the first 2000 word families of the New General Service List (Browne, 2014). Corpus frequency counts and Mutual Information[13] (MI) scores from the Corpus of Contemporary American English (COCA) (Davies, 2008 to present) were then used to find suitable monosyllabic adjective collocates. Following the protocol in Siyanova & Schmitt (2008), 'high-frequency' was defined as a total phrase frequency of over 100 in the corpus, and strength of association as measured by Mutual Information scores was set at a minimum threshold of 3. Bearing in mind the participants' proficiency level and their scores on the Vocabulary Levels Test (see 3.3.1), it was held that the individual words would be familiar to most participants, though the Adj + N combinations themselves may have been novel to some of the students.

As item recall can also be influenced by concreteness of meaning, the potential word-pairs were rated by six L1 English speakers on a 5-point rating scale (ranging from 1 = concrete to 5 = abstract); inter-rater reliability was measured using Cronbach's alpha (α = .91). Another variable that has been shown to affect L2 word-pair processing is the availability of equivalent L1 word-pairs (Wolter & Gyllstad, 2011). To account for this potentially confounding variable, each prospective word-pair was rated by six L1 Japanese speakers on a 5-point rating scale (ranging from 1 = easy to translate into Japanese, to 5 = difficult to translate) and again, inter-rater reliability was high (Cronbach's α = .89).

[13] A statistical association measure of the strength of attraction between words, providing objective evidence for formulaicity (Lehecka & Tomas, 2015).

From this pool, the following ten alliterating and ten non-alliterating word-pairs were selected as stimuli for the experiment:

Alliterating	Non-Alliterating
slow speed	free lunch
state school	clear path
fresh food	clean shirt
great guy	blank page
hard hat	hot meal
big band	chief judge
warm wind	whole world
soft sound	thick smoke
tall tree	wrong word
high heat	small shop

The selection was guided by the aim to balance[14] the potential confounding variables in Table 3.4. It can be seen from Table 3.4 that the differences in means, other than those for orthographic length, should confer an advantage to the word-pairs in the non-alliterating condition in a test of recall, thus favouring the null hypothesis.

Table 3.4 High-frequency word-pairs: features that affect recall.

	Alliterating Word-Pairs Means *(Standard Deviations)*	Non-Alliterating Word-Pairs Means *(Standard Deviations)*
Orthographic Length	8.40 *(SD = 1.17)*	9.20 *(SD = 0.92)*
COCA Phrase Frequency (per million words)	440.30 *(SD = 473.09)*	451.10 *(SD = 434.53)*
Mutual Information Score	4.44 *(SD = 2.26)*	5.47 *(SD = 2.36)*
Concreteness Rating [1]	2.35 *(SD = 0.43)*	1.87 [3] *(SD = 0.35)*
Translatability Rating [2]	1.83 *(SD = 0.45)*	1.48 [3] *(SD = 0.29)*

[1] 1 = concrete, 5 = abstract.
[2] 1 = easy to translate, 6 = difficult to translate.

[14] This was done by choosing those items from the mid-points of the scales for the features that can affect recall.

[3] Nonparametric Mann Whitney U tests indicated no statistically significant difference between alliterating and non-alliterating word-pairs for concreteness and translatability ratings.

Controlling for the variables in Table 3.4 meant choosing some words with consonant clusters in the onset position as there were insufficient target item candidates in the pool of potential word-pairs that started with single consonants. In addition, there were no word-pairs in the potential pool whose orthography might conceal an alliterative pattern (for example, *cute kid*, *small circle*, *single cell*) but which also met the monosyllabic and / or frequency criteria. Thus, only three word-pairs in the non-alliterating condition could be used to gauge if participants' perception of alliteration was affected by the orthographic form (*wrong word, whole world, small shop*). This limitation will be returned to in Chapter 7.

3.3.3 Procedure

The experiment was administered separately to several intact classes, and in each class the same procedure was respected. Participants were informed of a dictation task and shown examples of alliterating word-pairs via the classroom projector. Written instructions appeared on the projector screen in both English and Japanese, and were also given verbally in English. In the dictation activity, the teacher said each item aloud twice and the participants were told to repeat it sub-vocally. By asking the participants to say the word-pair to themselves it was hoped the phonological pattern would be more salient. Participants then ticked either 'Yes' or 'No' on the test paper if the item alliterated or not. As this is essentially a binary assessment, it was held that participants would consider each item for an equal length of time. Participants then wrote down the word-pair on the test paper. It was thought that ticking before writing would help alleviate any orthographic effects that could erroneously suggest an alliterative pattern (e.g. *wrong word*). The word-pair then appeared on the projector screen and the participants ticked a box to indicate if they had spelled it correctly, but were told not to change their original spelling. Finally, participants ticked a rating scale according to how familiar the word-pair was to them (1 = *I know how to use this phrase*, 2 = *I can guess how to use this phrase*, 3

= *I don't know how to use this phrase*). It was thought that this additional task would promote further semantic processing and thus aid retention (Craik & Lockhart, 1972), as well as confirming if the items were familiar or not.

The test papers were then collected, and the participants immediately did an unannounced free recall test: on a new piece of paper, they wrote down all the word-pairs they could remember from the previous dictation activity. One week later, there was another unannounced free recall test that followed the same procedure. As a free recall test may not be sensitive enough considering the strength of the effect of the treatment (Nation & Webb, 2011, p. 279), a cued recall test was then administered to compensate for any floor effect in the previous test. In the cued recall test, the teacher said the initial adjective of the word-pair and the students were asked to write down the corresponding collocating noun. For all dictations and cued tests, the order of the items was jumbled, including the four filler items before and after the block of target phrases.

3.4 Results and Analysis

The empirical investigation reported in this chapter is essentially twofold, encompassing not only the core element of discrimination but also recall, with the latter dependent on the former being performed successfully. As this research is fundamentally about vocabulary learning, it is important to establish if the capacity to discriminate operationalises into anything of pedagogic value. To recap, the research question is:

Does alliteration in familiar word-pairs have a mnemonic effect for Japanese L1 learners, in keeping with previous research with non-Japanese speakers?

However, before analysing the results of the recall tests, the following two issues need to be addressed. Firstly, were the word-pairs indeed familiar? Secondly, could the participants identify the phonological pattern? These will be considered in turn, before looking at the post-test results (section 3.4.3).

3.4.1 Word-pair familiarity

In the absence of pre-tests, selecting target items likely to be familiar or not was based primarily on three factors: the participants' L2 proficiency in terms of TOEIC scores, the corpus frequency of the word-pairs, and the participants' scores on the Vocabulary Levels Test (VLT). The participants' high VLT scores for the 1K and 2K bands suggested they would know the individual high-frequency words, but the VLT test did not assess knowledge of the intact word-pairs. Therefore, further support was sought, via the subjective 3-point rating scale, which participants completed as part of the dictation phase. The rating scale meant that each target word-pair had an ordinal score which could range from a minimum of 124 (if every participant ticked '1 = *I know how to use this phrase*') to a maximum 372 (if every participant ticked '3 = *I don't know how to use this phrase*').

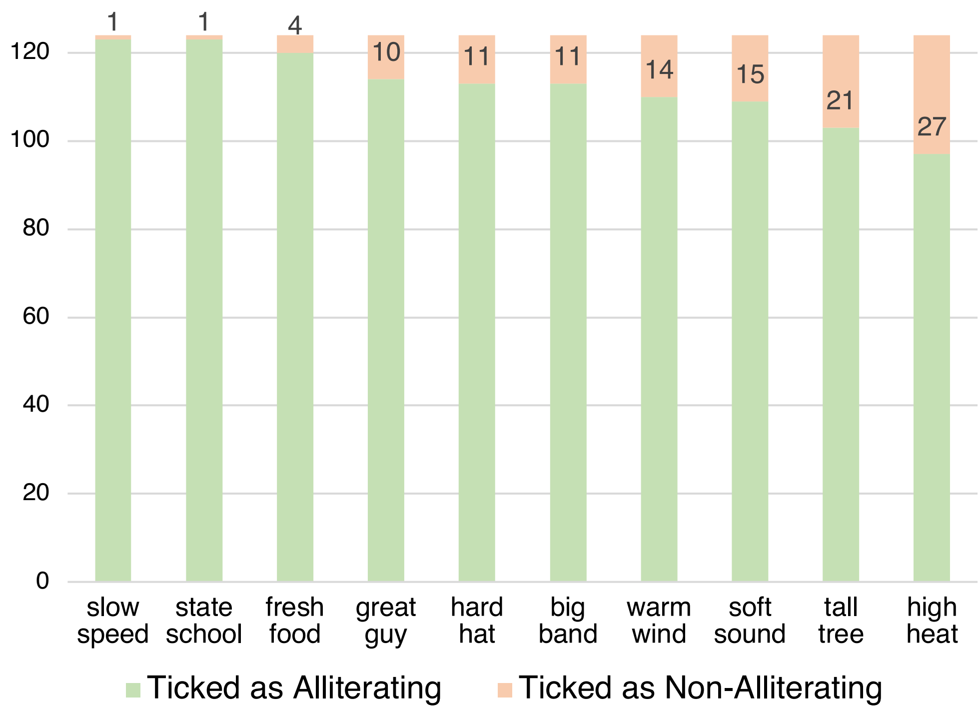
An indication of overall familiarity would be if the mean score were lower than the halfway point (248). At 173.4 (*SD* = 42.9), it was well below it, with *tall tree* reported as the most familiar word-pair, and *clear path* the least familiar (128 and 270 respectively). These findings do seem to confirm that the items were indeed known to most of the participants.

3.4.2 Identifying alliteration

As mentioned previously (in section 3.3.3), the alliteration recognition test phase occurred during the dictation; the participants were asked to tick the appropriate Yes / No box on the test paper to indicate if the phrase alliterated or not, before writing it down. However, if a participant misheard the target phrase, for example if they heard *thick smoke* as *sick smoke*, wrote *sick smoke* and ticked it as alliterating (5 instances), this was recorded as a correct response. Conversely, if they misheard *hard hat* as *old hat* (1 instance) and wrote it down as a non-alliterating phrase, this too was accepted as a correct data point. Although this marginally impacted on the achievable totals, the effect was far smaller than the quantity of missed opportunities for identifying alliteration (115 instances) and for identifying alliteration that was not present (173 instances), and so the results below retain the assumption of 10 word-pairs with and 10 word-pairs without alliteration.

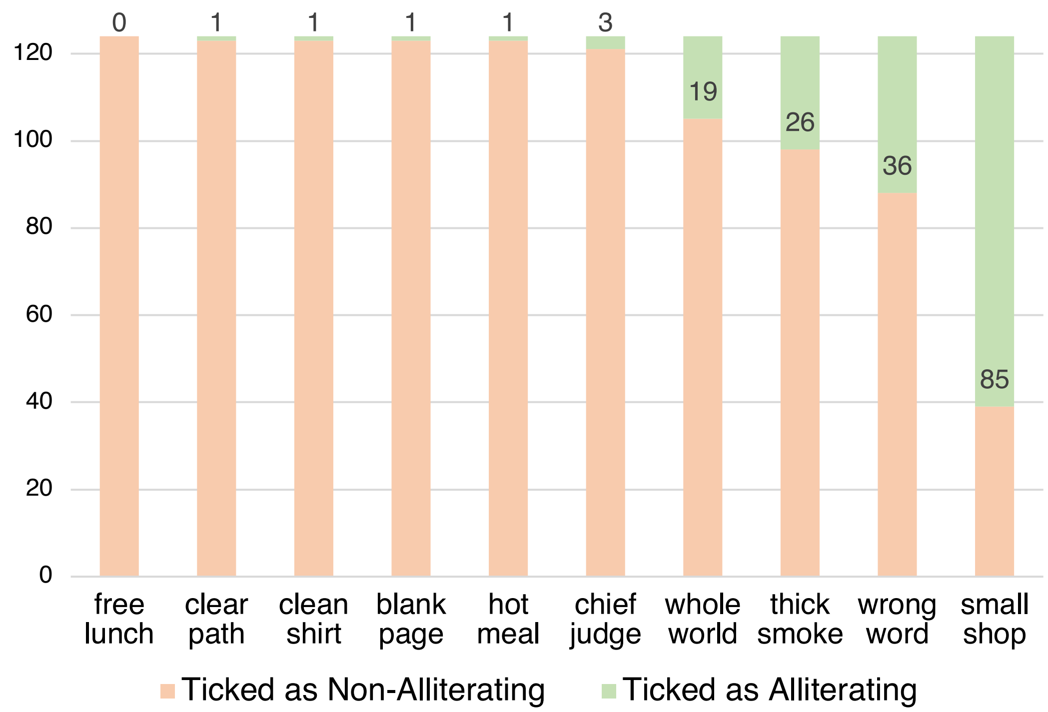
For the alliterating word-pairs, participants correctly identified the phonological pattern 90.7% of the time. Nearly all the participants (123 out of 124) recognised *slow speed* and *state school* as alliterating, whereas 27 participants marked *high heat* as non-alliterating – see Figure 3.3.

Figure 3.3 Alliterating items: recognition tally.



For the non-alliterating word-pairs, participants correctly identified the absence of a phonological pattern 86% of the time – see Figure 3.4. The greater number of misidentifications indicates that participants were more likely to assign alliteration to a non-alliterating word-pair than to miss it, but it should be noted that the result was skewed by more than half the participants mistakenly ticking *small shop* as alliterating. This seems to be a clear case of orthographic interference due to the two letters <s>, as both words exist as loan words in Japanese, yet are expressed with different kana (ス and ショ respectively). Spelling may also account for alliteration being misattributed to *whole world* and *wrong word*.

Figure 3.4 Non-alliterating items: recognition tally.



In sum, most of the Japanese L1 participants in this sample were able to identify the presence or absence of an alliterative phonological pattern: the participants' mean score for correctly identifying the presence or absence of alliteration in the word-pairs was 17.5 out of 20 (N = 124, min. = 11, max. = 20, SD = 1.9).

3.4.3 Post-test results

The recall tests were conducted to cast light on the issue of whether or not there was a mnemonic effect for alliteration. These tests necessitated a marking protocol to ascertain what was an acceptable written response. Bearing in mind the proficiency level of the participants, answers were marked as correct if they had no more than two errors per word-pair. Homophones were accepted as correct (for example, **grate* for *great*), as were all instances of /l/ and /r/ phoneme confusion (**flesh* for *fresh*); as one phoneme represents both sounds in Japanese, it was held that such spelling errors were acceptable. Acceptable errors also included grapheme substitutions (**free lanch* for *free lunch*), insertions (**high heate* for *high heat*), and deletions (such as **wole world* for *whole world*).

On average, participants recalled more alliterating word-pairs than non-alliterating word-pairs across the tests (see Table 3.5). In the immediate free recall test,

Table 3.5 Summary results for recall tests.

	Alliterating mean	Non-Alliterating mean	Analysis	Effect Size
Immediate Free Recall	3.85	3.82	Wilcoxon Signed Ranks z = -3.78, p = .353 (1-tailed)	
Delayed Free Recall	1.43	0.94	Wilcoxon Signed Ranks z = -3.74, p = < .001 (1-tailed)	r = .42 (medium effect size) R^2 = .18
Delayed Cued recall	2.84	1.98	Wilcoxon Signed Ranks z = -4.76, p = < .001 (1-tailed)	r = .47 (medium-large) R^2 = .22

participants recalled, on average, 7.7 of the 20 phrases (*SD* = 3.44). For the alliterating phrases the mean was 3.85 (*SD* = 1.90), and for the non-alliterating phrases the mean was 3.82 (*SD* = 2.04). Visual representations of the data indicated the presence of outliers and skewed distributions, so a non-parametric Wilcoxon Signed Ranks test was used. This showed no statistical difference between the alliterating and non-alliterating means (z = -3.78, p = .35, 1-tailed[15]). In the one-week delayed free recall test, more alliterating phrases (M = 1.4, *SD* = 1.2) than non-alliterating phrases (M = .9, *SD* = 1.2, z = -3.74, p < .001, 1-tailed) were recalled, and a medium effect size was observed (r = .42).[16] In the delayed cued recall test, more alliterating phrases (M = 2.84, *SD* = 1.81) than non-alliterating phrases (M = 1.98,

[15] The default prediction for this design is that alliteration is mnemonic. Furthermore, a directional test was deemed appropriate in light of the fact that the population might suppress any mnemonic tendency because of their L1, and there is no rationale for testing whether non-alliterating word-pairs will be better recalled.

[16] Effect size descriptors are based on Plonsky & Oswald's (2014) field-specific bands.

SD = 1.63) were recorded, and a medium-large effect size was observed (*p* = <.001, *r* = .47).

3.5 Discussion

The experiment found that the participants were sensitive to alliteration, despite being L1 speakers of a language based on morae. Several questions arise, however. One, addressed presently, is whether they were *as* sensitive as speakers of other L1s. The findings are broadly in line with those from other studies on alliteration: Table 3.8 shows the results from comparable research that looked at the mnemonic effect of alliteration in known word-pairs. The studies in Table 3.6 used a similar dictation-based methodology and the participants' L1 was Dutch, except for Study 3 which used Malay and Chinese L1 participants.

It seems the Japanese speakers were *as* sensitive to alliteration as their counterparts in the Boers et al. studies, if the benchmark interpretations of the effect sizes can be compared (as noted in section 2.3.3, effect sizes show the magnitude of the impact of the independent variable, alliteration, on the dependant variable, recall[17]). The large effect sizes in Study 1 (Lindstromberg & Boers, 2008a) in Table 3.8 have not been replicated elsewhere, and this will be discussed in Chapter 7.

The second question is whether the results of this experiment can reliably be generalised beyond this sample. The participants were a convenience sample, in that they were members of intact English language classes. As such, they were of differing ability levels within the scope of their year-group and stage though they all had the same class teacher and were taught the same curricula. Given the variation in individual scores, the sample size of 124 participants x 20 responses (= 2480 data points) helped mitigate the effects of extreme scores and potential extraneous factors and goes some way to overcoming the problems inherent in non-probability sampling procedures.

---

[17] In other words, they answer the question "Is it enough to care?" (McGrath & Meyer, 2006, p. 386).

The data was not consistent with the contention that the mora obscures English alliteration (Otaka, 2009), as the participants in this experiment were successful, almost 90% of the time, in correctly identifying the presence or absence of alliteration, and many of the errors can be attributed to orthographic interference.[18]

Table 3.6 Previous findings on the mnemonic effect of alliteration in known word-pairs.

Study	Tests[1]	Statistical test	Reported statistics	Effect Size (*Cohen's d*)
Study 1 (Lindstromberg & Boers, 2008a)				
Exp. 1	IR	Wilcoxon Signed Ranks	p = 0.0041, W = 174, $N_{s/r}$ = 23, z = 2.64	.088[2]
	2-wk DR	Wilcoxon Signed Ranks	p = 0.0018, 1-tailed, W = 75, $N_{s/r}$ = 12, z = 2.92	.93
Study 3 (Boers et al., 2012)				
	IR	correlated t-test	p = .04, 1-tailed	.48[2]
Study 4 Boers, Lindstromberg & Eyckmans (2014a)				
Exp. 1	ICR	paired t-test	MD = .39, p = .02, 1-tailed	*.34*
	1-wk DCR	paired t-test	no statistical difference	
Exp 2	IR	paired t-test	no statistical difference	
	ICR	paired t-test	MD = 0.3, p = .02, 1-tailed	.25
	1-wk DR	paired t-test	no statistical difference	
	1-wk DCR	paired t-test	no statistical difference	
Study 7 (Boers, Eyckmans, et al., 2014)[3]				
Exp. 1	ICR	paired t-test	no statistical difference	
	1-wk DCR	paired t-test	no statistical difference	

[1] IR = Immediate Free Recall Test, ICR = Immediate Cued Recall test, DR = Delayed Free Recall Test, DCR = Delayed Cued Recall test, DRec = Delayed Recognition Test, wk = week.
[2] Effect sizes were calculated retrospectively.

[18] It is interesting that the alliterating word-pairs with the most correct recognition responses have two consonant clusters in the onset positions. As these are not tolerated within the Japanese phonological system, Japanese speakers employ epenthetic vowels when faced with a cluster. Kubozono (2002) states that in the context of loanword adaptation, the default epenthetic vowel is /u/, as this results in an output that is more perceptually similar to the input: compare /$_\mu$mi $_\mu$ru $_\mu$ku/ ('*milk*') to */$_\mu$mi $_\mu$ra $_\mu$ko/ or */$_\mu$mi $_\mu$re $_\mu$ki/. If the participants *are* applying epenthesis and a moraic segmentation, where the nucleic vowel is attached to the consonantal onset, then the initial units may appear more similar than a syllabic segmentation would suggest: *slow speed* = /$_\mu$su $_\mu$ro/ /$_\mu$su $_\mu$pi $_\mu$du/, and *state school* = /$_\mu$su $_\mu$te $_\mu$to/ /$_\mu$su $_\mu$ku $_\mu$ru/. It is speculated that this accounts for the high recognition scores in the pattern identifying task.

[3] Half the target items in this study displayed consonance and half alliteration.

A further question arises from this set of results: were more proficient participants better able to identify alliteration than participants with lower English proficiency? It seems reasonable to suggest that as L2 proficiency increases, a learner becomes more accustomed to the patterns of the language, and furthermore, part of what determines who becomes more proficient is their capacity to perceive patterns that might be relevant to their learning.

A visual inspection of scatterplots suggested that a linear model of correlation was not the best fit for the data, that is, the assumption of a linear relationship was not met, so a non-parametric rank correlation was conducted (Spearman's ρ correlation). The correlation between Noticing Alliteration Scores and Proficiency (as measured by TOEIC) was not statistically significant, with a negligible effect size ($\rho = .032$, $p = .36$ (one-tailed), $R^2 = .001$), suggesting there was no correlation between the variables.

Before it can be concluded that alliteration conferred a mnemonic advantage to the phrases displaying it, an alternative explanation must be eliminated – namely, that phrase familiarity resulted in greater recall. Were word-pairs better recalled because they were more familiar?[19] There was no statistically significant correlation between self-reported familiarity ratings and total recall of the word-pairs ($\rho = -.38$, $p = .10$ (2-tailed), $R^2 = .12$). If all confounding and extraneous variables have been accounted for, it can be said that the results support the finding from other studies and have increased the generalisability. Thus, the experimental outcome suggests that alliteration in English word-pairs does facilitate recall and retention of word-pairs, and that this applies even to speakers of languages that do not have the same syllabic structure as English.

[19] Of course, it is not possible to tell from this, at the individual level, if not knowing a word-pair made it harder to recall it or not.

3.6 Conclusion

Even though the phonological structure of Japanese is very different to that of Dutch, this experiment provides preliminary evidence that alliteration has a mnemonic effect on the recall of familiar word-pairs. The fact that this effect was only found in the delayed tests will be discussed in Chapter 7. Now that we can be confident that the participants' L1 should not make a significant difference to the experimental findings, the next chapter will look for a mnemonic effect of assonance and attempt to replicate one of the studies in Table 2.4.

Chapter 4 Do learners of English recall assonating word pairs better than non-assonating ones?

4.1 Introduction

The previous chapter looked at the role of English L2 learners' L1 phonology in shaping their ability to discriminate an alliterative phonological pattern, and asked whether this would affect the memorability of a word-pair. In Japanese phonology the vowel nucleus is bonded to the consonantal onset to form a prosodic unit, the mora; so the question was whether this might impact on the speaker's perception of similarity – is /$_{\mu}$bi/ the same as /$_{\mu}$ba/ in *big band*? The results from the experiment described in Chapter 3 found no clear evidence of a 'mora-effect' on the participants' ability to perceive an alliterative pattern in the Adjective + Noun word-pairs. This allays any concerns that Japanese L1 speakers might be a 'special case' as learners of English as an L2. On that basis, the remainder of the studies reported here will begin with the assumption that Japanese participants are representative of, and comparable to, other learner groups, such that the findings can be generalised. Nevertheless, it is appropriate to be alert for any indications that the nature of Japanese is affecting the results and so that question will be considered as it arises. As a means of anchoring this assumption of compatibility, the study reported in this chapter is a replication. By minimising changes in the parameters in an existing design, it is easier to pinpoint the cause of any differences in outcome.

The study selected for replication is the second experiment reported in Boers, Lindstromberg & Eyckmans (2014b) (Study 5 Exp. 2 in Table 2.4). There are two key reasons why this particular study was chosen as a suitable candidate for replication. Firstly, the methodology is representative of the body of research summarised in Table 2.4. Secondly, the experiment focuses on the remaining phonological pattern under discussion, assonance.

Doing this replication will not only evaluate the internal and external validity (Mackey, 2012) of the specific study, but also give a good indication of the extent to which the body of work is generally replicable. As already noted, Chapter 3 suggests that the use of a different L1 group is not likely to introduce a new variable. However, this replication also uses learners of a different proficiency level, which does add a variable of interest. Another aim is to verify that the original results are sufficiently robust for drawing conclusions, and that possible confounds have been accounted for (Abbuhl, 2012). Thus, by assessing the validity of previous findings the replication will add to the totality of evidence accumulated over prior experiments.

In brief, this study continues the investigation whether or not phonological repetition can facilitate the processing of word-pairs. More explicitly, it asks if the presence of assonance in a familiar word-pair leads to better recall and retention, compared to a word-pair that has no assonance. The research hypothesis to be tested, based on the findings from the original study, is that salient assonance in word-pairs will result in better recall of those word-pairs, compared to recall of non-assonating word-pairs. The next section contextualises the account by introducing the original study.

4.2 Summary of the Boers et al. (2014b) study

4.2.1 Participants

The participants consisted of 44 language major undergraduates in their first year at a university college in Belgium. Their first language (L1) was Dutch and, after approximately six years of English at secondary school, their English proficiency was stated by their teachers to be at the B2 level of the Common European Framework of Reference for Languages (CEFR), that is, at what many a classroom teacher would consider to be 'upper intermediate' level. The experiment was conducted during regular class time with their class teachers of English.

4.2.2 Materials & Procedure

The word-pair stimuli consisted of either Adjective + Noun collocates or Noun + Noun compounds. Ten of the word-pairs showed assonance and ten were non-assonating (see Table 4.1). Two additional items (*tool box* and *firm hold*) were included to help counter primacy and recency effects on recall, but these were not included in the analysis. The word-pairs were matched as far as possible to control for confounding variables such as syntactic structure, length, frequency, and concreteness of meaning. To control for novelty effects, all the word types were high frequency; in the top 2000 word families of the British National Corpus with 31 tokens in the first 1000 (K1) and 9 tokens in the second 1000 (K2) list. Whole phrase frequency data was collected from the COCA corpus (Davies, 2008 to present) in February 2013 and can be seen in Table 4.1.

As item recall can also be influenced by its concreteness of meaning, potential word-pairs were screened for this factor before being included in the study. This was achieved by having eight L1 English speakers rate a pool of word-pairs on a five-point Likert scale questionnaire, ranging from 1 (abstract) to 3 (concrete). The mean per word-pair ratings of four randomly selected raters and the mean ratings of the remaining raters were correlated. This was done ten times, and the resulting mean correlation between the ratings of each rater and the other seven was strong ($r = .80$). The averaged word-pair ratings were then used to guide selection of all the target word-pairs to control for concreteness.

Table 4.1 Stimulus items and COCA phrase frequencies.

Assonant Phrases	Phrase Frequency	Non-Assonant Phrases	Phrase Frequency
town house	556	*town square*	616
deep sea	232	*deep hole*	160
soft cloth	57	*soft ground*	52
gift list	74	*check list*	65
quick trip	164	*quick stop*	107
small talk	841	*plain talk*	36
safe place	894	*nice place*	390
high price	847	*high rate*	684
fair share	1086	*fair deal*	105
main gate	233	*main road*	753
Means	498.4		296.8

The imbalance of frequency in the composition of the list of target word-pairs (as indicated by the means in the bottom row of Table 4.1) was not held as a confounding variable on the grounds that item frequency tends to correlate negatively with recall in similar test conditions. That is, a free recall advantage for low frequency items is often observed (for example Merritt, DeLosh, & McDaniel, 2006) so if the assonant word-pairs were better recalled from episodic memory that would be *despite* their higher frequency. This point will be returned to in the discussion (see section 4.5).

The experimental procedure is similar to that described in Chapter 3 for the mora experiment. The classroom teacher began the treatment phase by explaining what assonance was and giving an example of an assonating item and a non-assonating item on the board (*black cat* and *red fox*). There was then a teacher-led dictation of the randomised test items: the teacher read aloud each test item, the participants repeated it aloud and wrote the item on a piece of paper. Participants were asked to write a plus or minus sign next to each test item according to whether they thought the item displayed assonance. This component was included because in a previous version of the experiment (Study 1 in Boers et al., 2014b) there was no awareness-raising element, and this was held to be the main reason no statistically significant result was found in the test phase.

The dictation papers were collected and there was an unannounced immediate free recall test – on a new piece of paper, the participants wrote down all the items they could recall. Approximately one hour later there was a delayed free recall test, and then a one-week delayed cued recall test,[1] in which the teacher read aloud each of the ten shared first words of the word-pairs and the students wrote down the second word, or both words if they could. For example, if the teacher said the cue *town,* it was hoped this would elicit *house* and / or *square* as the written response. It was held that such a format was more congruent with the initial dictated input in the

[1] This was misreported in *The European Journal of Applied Linguistics and TEFL* as a one-day delayed test (Seth Lindstromberg 2015, personal communication).

treatment phase. The results will be reported later (see 4.4) when they can be compared to the results of the replication.

4.3 Method

4.3.1 Participants

The participants were all Japanese L1 undergraduate students in their second semester at a private university in Japan. In the initial data gathering there were 125 participants (81 female and 44 male) and their ages ranged from 18 to 19. All of them had completed approximately six years of English at high school and they were all enrolled on Modern Language degree courses, which involved about nine hours per week of classes conducted in English. Their mean TOEFL ITP score was 459 (N = 115, SD = 40.7), which approximately correlates with the B1 level of the CEFR. Thus, these participants were at a lower level of proficiency than those in the original study. The participants formed a convenience sample of five intact classes and, as in the original study, the experiment was conducted in their scheduled class time with their regular class teacher.

4.3.2 Materials

The same set of word-pairs compiled for the original study was used in the replication. The same coding scheme for concreteness was used, with eight L1 English speakers, all experienced language teachers. The concreteness scores for the twenty items in the original study and the replication were very similar, as can be seen in Table 4.2. A Pearson's *r* correlation found the correlation to be statistically significant and the effect size to be large ($r = .92$, $p = .001$ (Sig. 2-tailed), $R^2 = .85$).

Table 4.2 Concreteness ratings for experimental stimuli.

Assonant	Original	Replication	Non-Assonant	Original	Replication
town house	2.75	2.44	*town square*	2.87	2.31
deep sea	2.75	2.63	*deep hole*	2.75	2.44
soft cloth	2.75	2.5	*soft ground*	2.5	2.25
gift list	2.5	2.44	*check list*	2.87	2.44
quick trip	1.63	1.81	*quick stop*	1.63	1.75
small talk	1.37	1.63	*plain talk*	1.00	1.25
safe place	1.63	2.13	*nice place*	1.50	1.94
high price	1.75	2.31	*high rate*	1.50	1.56
fair share	1.00	1.56	*fair deal*	1.00	1.5
main gate	2.50	2.56	*main road*	2.87	2.56
Mean	2.06	2.2		2.05	2.0

4.3.3 Procedure

The experimental phase followed the same dictation activity and series of unannounced recall tests. The classroom teacher used one different example to explain assonance *(big fight* instead of *red fox*), partly to emphasise the fact that despite having the same spelling for the vowel sound, the word-pair did not assonate, and also because *fox* might have been unfamiliar and thus a distraction. There was then the dictation exercise with the awareness-raising component. As Japanese students often use circles in lieu of ticks, this was taken into account. The full set of word-pairs was read out again, in a different order, and the participants checked their judgements. The dictation papers were then collected and the immediate free recall test was administered; on a new piece of paper the participants wrote down as many of the dictated word-pairs they could remember. The papers were collected and normal class resumed. Approximately one hour later, the delayed free recall test followed, in which the students again wrote down as many of the complete word-pairs as they could.

Because of timetable constraints, the replication study differed slightly from the original study for the final stage of the procedure. Due to differing class schedules, some of the participants (N = 45) took a two-day delayed cued recall test, while the remainder took a one-week delayed cued recall test (N = 68). Twelve students were absent for the delayed tests, but their data was retained for the analysis of the Immediate Free Recall Tests. The test format was exactly the same as the original study, namely, the teacher read aloud the first word of each word-pair and the participants wrote down the second word, or both words if they could recall them.

4.4 Results and Analysis

In the original study, the data from the assonance-identification task showed that assonating word-pairs were identified by nearly all the participants. In the replication, the identification rate was similar: the mean number of correctly identified instances of assonance was 9.02, out of a maximum ten (SD = 0.95). Two participants had not followed instructions and failed to indicate which phrases showed assonance. As this did not impinge on the three tests, their participation was continued.[2]

In terms of spelling, the dictated word-pairs in the original study were all correctly spelt. However, in the replication the spelling accuracy was lower: there were 385 spelling errors out of 5000 written words (125 participants x 40 words), with an average of approximately three spelling errors per participant (N = 125, M = 3.08, SD = 1.88).

The total number of spelling errors did not include 96 instances of homophones (for example, *whole* for *hole, plane* for *plain, fare* for *fair*), nor the 76 instances of /l/ and /r/ phoneme confusion (such as *rist* for *list, prace* for *place, load* for *road*). As one phoneme represents both sounds in Japanese, it was held that such spelling errors were acceptable, bearing in mind the level of English proficiency of the participants. The quantity of spelling errors, 7.7% of the total number of dictated words, signals the first salient difference between the original study and the replication, and this outcome will be discussed below (see section 4.5).

[2] This failure to do the awareness-raising element did not result in lower test scores compared to the averages.

In the original study, participants recalled more assonating word-pairs than non-assonating word-pairs at a greater than chance likelihood across all three tests (results can be seen in Table 4.3).

Table 4.3 Summary results from Boers et al. 2014b.

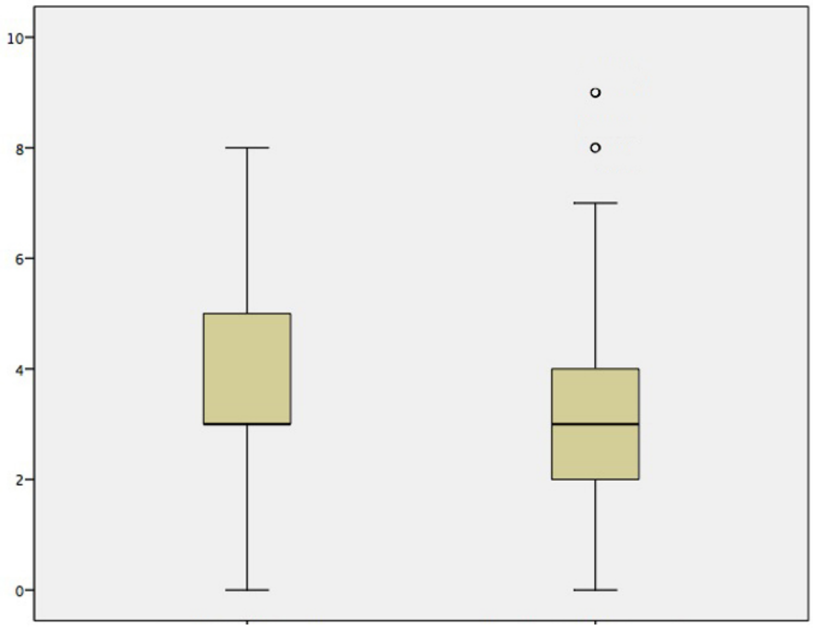
	Immediate Free Recall	1-hour Delayed Free Recall	1-week Delayed Cued Recall
N	44	44	44
M^{As}	3.43	2.73	3.68
$M^{Non\text{-}As}$	2.55	2.04	1.95
MD	.89	.89	2.17
t-test	$t\,(43) = 3.03$	$t\,(43) = 3.03$	$t\,(34) = 5.6$
alpha level (1-tailed)	.002	.002	$< .001$
effect size (Cohen's *d*)	.45	.45	1.15

^{As} = Assonating word-pairs
$^{Non\text{-}As}$ = Non-Assonating word-pairs

A paired-samples t-test showed these differences in recall to be statistically significant: in both the Immediate Free Recall Test and the 1-hour Delayed Free Recall Test, a medium effect size ($d = .45$) was observed (the authors used Cohen's guidelines). In the 1-week Delayed Cued Recall Test, a large effect size was observed ($d = 1.15$).

In the replication study, an initial visual representation of the data with boxplots indicated the presence of outliers and skewed distributions (see Figure 4.1 for an example). The skewness ratio (skewness level divided by standard error of skewness) was calculated to be in excess of 2.0 for all three post-tests, indicating a significant departure from a normal distribution (Larson-Hall, 2010, p. 74 ff). As this violates the assumptions for *t*-tests, a non-parametric test was used (Wilcoxon Signed Ranks test). As the original experiment already indicates a direction for the effect, one-tailed *p* values are reported.

Figure 4.1 Boxplots for immediate free recall test (Replication Study).



In the replication study, participants, on average, recalled more assonating phrases than non-assonating control items across all the post-tests (results can be seen in Table 4.4).

Table 4.4 Summary results from replication study.

	Immediate Free Recall	1-hour Delayed Free Recall	2-day Delayed Cued Recall	1-week Delayed Cued Recall
N	125	125	45	68
M^{As} (SD)	3.58 (1.76)	3.34 (1.87)	2.96 (1.83)	2.31 (1.59)
$M^{Non\text{-}As}$ (SD)	3.12 (1.86)	2.72 (1.72)	1.98 (1.32)	2.14 (1.56)
Wilcoxon Signed Ranks	z = 3.12	z = 3.92	z = 3.02	z = .61
alpha level (1-tailed)	.001	< .001	.002	.27
effect size (*r*)	r = .32	r = .41	r = .50	r = .08

^{As} = Assonating phrases
$^{Non\text{-}As}$ = Non-Assonating phrases

As with the original experiment, more assonating word-pairs were recalled in the immediate free recall and one-hour delayed tests, with medium effect sizes observed (r = .32 and .41 respectively). The 2-day delayed cued recall test also resulted in more assonating word-pairs being recalled and a large effect size was observed (r = .50). However, while in the original there was an increase in recall for assonating, though not non-assonating, word-pairs after one week, no such increase occurred in the replication: the difference in recall between assonating and non-assonating items was not statistically significant and the magnitude of the impact of assonance on recall was very small (r = .08).

4.5 Discussion

As with the original study, there was a recall advantage for word-pairs displaying assonance in the immediate, one-hour and 2-day tests. There was no effect in the one-week delayed test, though recall was still slightly lower for non-assonating word-pairs. However, before it can be asserted that the phonological repetition conferred a mnemonic advantage to those word-pairs displaying the pattern, alternative explanations must be eliminated. Were certain word-pairs better recalled due to their greater concreteness of meaning? This can be answered in the negative since there was neither a statistically significant correlation (at α = .05 level) between concreteness rating and recall, nor a substantial effect size (Spearman's ρ = .14, p = .56 (2-tailed), R^2 = .04). Nor was there a statistically significant correlation between corpus frequency and target item recall across the tests (Spearman's ρ = -.018, p = .94 (2-tailed), R^2 = .003). This is taken to indicate that the experimental outcome was neither an artefact of phrase frequency nor of concreteness.

A further potentially confounding variable that must be accounted for is that phrases may have been better recalled because they had Japanese translation equivalents. Evidence for this comes from experiments such as Wolter and Gyllstad (2011), in which it was found that participants in a primed lexical decision task processed and recognised L1-L2 collocations more effectively than collocations with no translation equivalents. Thus, participants in the replication could have remembered phrases because they had straightforward translation equivalents rather than because of any intrinsic phonological patterning. To determine whether or not this extraneous variable had an effect on the data, a post-hoc analysis was carried out (see Appendix 2.1). Seven L1 Japanese teachers of English ranked the twenty word-pairs on a 5-point Likert scale (1 = easy to translate, 5 = difficult to translate) and the sum of the score for each phrase was correlated with its total recall across the tests. A Pearson's *r* correlation between translation difficulty and total recall was not statistically significant, and there was only a small effect size (r = -.126, p = .60 (2-tailed), R^2 = .016). Clearly, this was not an in-depth analysis of translation equivalency and the number of Japanese L1 informants was small. However, it must be noted that this potential confounding variable was not considered in the original study, where it is arguably likely to have had more of an effect. As a result, the slight negative correlation and effect size indicate that this could be a confounding variable, which was not controlled for in either the original study or the replication, and thus we must be cautious in how far both sets of results can be interpreted and generalised.

It can be said that the results largely support the original findings. The word-pairs with assonance were recalled to a greater degree than those with no phonological pattern. The difference was statistically significant for the immediate, 1-hour delayed and the 2-day delayed tests, and the effect sizes ranged from medium to large. The disconfirming results of the one-week delayed test should not be dismissed out of hand. The increased recall after one week in the original study and the large effect size (d = 1.15) is rather striking, whereas the reduced recall and small effect size (r = .08) in the replication is perhaps closer to what would generally be expected in a memory test.

Explanations for this difference could involve external factors such as motivation, and more local ones such as familiarity. It is not possible to ascertain if the Japanese students recalled fewer word-pairs because there was insufficient interest aroused by participating in the experiment.[3] Another explanation could be that participants had difficulty recalling the items because their knowledge of the target phrases was less developed than the more proficient participants in the original study. The number of spelling errors in the initial dictation stage does seem to suggest that the words were not entirely familiar to the less proficient participants in the replication, despite them being high-frequency items. To explore this further, a post-facto questionnaire was administered to the participants during class time, asking them if they had seen / heard the phrases before the experimental phase or not (see Appendix 2.2). The results of this self-report questionnaire revealed that seven of the word-pairs were novel to the participants. Albeit a somewhat superficial instrument for data gathering, the questionnaire results do suggest that some of the target items were new combinations for some of the participants. This differs from the original study in which it was assumed that the participants knew the word-pairs, something that the lack of spelling errors supported.

4.6 Conclusion & Limitations

The aim of this approximate replication study was to increase the explanatory power and generalisability of the findings from Study 2 in Boers et al. (2014b), namely, that assonating word-pairs are better retained than non-assonating ones. The experimental outcome suggests that assonance in L2 word-pairs does facilitate recall if the phonological pattern has been made explicit. Thus, converging lines of evidence do seem to indicate a mnemonic effect for phonological patterning and this appears to be a useful tool to add to the classroom teacher's toolkit. It remains to be seen what kind of attention-directing activity is most beneficial; studies to date have used paired dictations (for example, Boers et al., 2014a), card sorting (for example, Lindstromberg & Boers, 2008b), and oral dictations (for example, Boers et al., 2012), and further research is needed to clarify what kind of engagement fosters better

[3] Of course, other aspects of motivation could counter this: the motivation to please the teacher, to do well in a task, to see what happens in a memory test, and so forth.

recall. Additional inquiry is needed on word-pairs that are not part of the subjects' existing lexicon, and on a broader range of sound patterns.

In terms of limitations, a further consideration is that the post-hoc questionnaires on translation equivalence and pre-existing phrase knowledge (Appendices 2.1 and 2.2) were not piloted. It is not claimed here that their resulting data are sufficiently reliable or valid, but inferences can be drawn, and they can be seen as indicating improvements for future experimental design, namely, controlling for L1 equivalency when choosing the target language and pre-testing the subjects' receptive knowledge of such language. As the experiment was constrained by the classroom environment, it is not possible to ascertain if the participants encountered the target items in the time between the input phase and the delayed tests. Finally, the experiment provides no information as to what cognitive styles benefit most from raising awareness of phonological patterning and this could also be addressed in further studies.

To recap, it appears that both alliteration and assonance in known word-pairs confer a small mnemonic advantage for Japanese L1 speakers. As mentioned in Chapter 2, Boers et al. have stated that this mnemonic effect can be accounted for in terms of one part of the word-pair 'priming' recall of the other part, due to orthographic and / or phonological similarity (Boers et al., 2014a, p. 295; Eyckmans et al., 2016, p. 129; Eyckmans & Lindstromberg, 2017, p. 3; Lindstromberg & Boers, 2008a, p. 206; Lindstromberg & Eyckmans, 2017, p. 3). The next chapter will report on an investigation into this potential priming effect for the phonological patterns under discussion.

Chapter 5 Do phonologically-patterned primes impact on Lexical Decision Task performance?

5.1 Introduction

The results of the experiment in Chapter 3 and the replication study in Chapter 4 contribute to converging evidence of a mnemonic effect for word-pairs that have matching phonological / orthographic elements corresponding to alliteration or assonance. This chapter explores one possible explanation for the facilitation of lexical retrieval of word-pairs displaying phonological patterns, that is, the short-term memory process of priming, an explanation which Boers and his colleagues have posited in the past (for example, Boers et al., 2014a, p. 295; Eyckmans et al., 2016, p. 129; Eyckmans & Lindstromberg, 2017, p. 3; Lindstromberg & Boers, 2008a, p. 206; Lindstromberg & Eyckmans, 2017, p. 3). As seen in Chapter 2, the term 'priming' can be defined in many ways: as a component, a process, a property, and a general characteristic of cognitive subsystems. Broadly speaking, it is used here as a term to describe how prior experience with language can unconsciously influence the processing of subsequent language (see section 2.5.1). One reason phonological patterns such as alliteration and assonance are thought to be mnemonic devices is that form-based aspects of a given word privilege the processing of a following word that shares the same form-based aspects, be they phonological, orthographic, or phonetic.

How might one set about testing the role of priming? One challenge facing the novice researcher is that although priming experiments have a long history in some form or another (see section 2.5.1), "research labs accumulate a lot of informal knowledge about how to run particular [priming] experiments, which is rarely published" (Rastle & Brysbaert, 2006, p. 185). As a result, it can be difficult to ascertain exactly how such experiments are designed and implemented from previously published studies alone. Fortunately, for one of the standard methods used in priming research, namely reaction time studies, there exists a body of more prescriptive work, with more explicit guidelines on the protocols (for example, Jiang, 2012; McDonough & Trofimovich, 2009). Though there are a wide range of experimental tasks for measuring and quantifying priming effects, perhaps the most widely used for investigating the mental lexicon is the Lexical Decision Task (LDT). In a standard LDT, two stimuli are presented successively on a computer screen, first the 'prime' and then the 'target'. The experimental task requires the participant to respond only to the target, by either pressing the 'Yes' button on a keyboard if they think the target is a real word, or the 'No' button if they decide the target is not a real word. Participants are asked to respond as fast and as accurately as possible to each target. The reaction times, mediated by the participant's motor responses, are thought to provide indirect evidence for underlying cognitive processes. Priming is said to occur if the prime facilitates the response to the target in terms of faster reaction times (measured in milliseconds) and increased accuracy.

As noted in Chapter 2, facilitation can occur because of a conceptual or semantic relationship between the prime and the target: for example, participants react faster to the visual string *NURSE* if they have just been presented with *doctor*. However, as this experiment seeks to establish if there is a facilitatory effect due solely to a perceptual, or form-based, relationship between the prime and the target, semantic variables need to be controlled for (discussed presently).

Priming is thought to be an implicit, automatic process (Tulving & Schacter, 1990) that precludes the use of strategies in language processing. To tap into this process, the initial visual stimuli can be hidden, or 'masked'. In the priming research paradigm, masking is often accomplished in two ways. Firstly, the prime stimulus is presented so briefly on screen (for example, 50 - 60 ms) that the participant is unaware of it. Secondly, a series of characters, often a string of hash keys as long as the target string (######), can be used as a forward mask preceding the prime, and the target word-string used as a backward mask (Jiang, 2012, p. 104). These masking techniques are thought to block retinal after-images and pixel overlap, overwriting any visuo-sensory representations. Such steps help to ensure that the participants cannot employ conscious strategies such as attempting to guess upcoming targets, trying to find the relationship between the primes and targets, or tuning in to experimental 'bias' such as word length for primes and targets.

Before outlining the procedures adopted for this study, a further challenge needs to be addressed. There is growing empirical evidence that the bilingual lexicon is non-selective (see, for example, Brysbaert, 2003; Nakayama, Sears, Hino, & Lupker, 2012). This premise means that the automatic activation of a phonological representation by a visual word stimulus is not limited to a representation specific to the language being read. This leads to the conclusion that if a participant has knowledge of more than one language – and if, as supposed, the first stages of word recognition are indeed language independent - it is possible to prime a target word in the L2 by a homophonic stimulus of the L1. There is an increasing amount of support for this notion in the literature; for example, for French-Dutch bilinguals, a French target (such as *OUI*) can be primed by a phonologically similar Dutch word (*wie*) (example from Experiment 1 in Brysbaert, Van Dyck, & Van de Poel, 1999). There is also evidence of cross-language cognate priming across disparate writing systems. For example, Nakayama et al. (2012) found significant priming effects with Japanese-English bilinguals, where a masked stimuli such as ガイド /μga μɪ μdɒ/ (guide) primed the English target *GUIDE*.

These findings raise four intriguing questions. Firstly, is this priming effect uni- or bi-directional? That is, would the L2 stimulus *guide* prime the L1 target ガイド (/μga μɪ μdɒ/) for Japanese-English bilinguals? Secondly, what is the role of proficiency? Are the same priming effects found with less-proficient bilinguals such as adult L2 learners? Thirdly, does the priming effect only apply to alphabets and syllabaries or also to logographic scripts such as Japanese Kanji? Finally, is there a cross-language priming effect when the primes and targets are not cognates and therefore have no semantic relationship? For example, with Japanese-English bilinguals would the L2 prime *guy* facilitate the processing of the L1 ガイド /μga μɪ μdɒ/ and vice versa?

These issues would seem to suggest that it is therefore paramount to rigorously control for pseudohomophone and cognate status in a cross-language LDT. However, a considerable number of other lexical characteristics can also affect the processing of stimuli in such tasks, and thus impact on the participant's reaction time and accuracy. According to Jiang (2012), the following properties have all been found to affect lexical processing:

> frequency, familiarity, word length (in terms of letter, syllables, phonemes or morphemes), neighbourhood density, neighbourhood frequency, concreteness, imageability, age of acquisition, spelling-sound regularity, affixation, polysemy, bigram frequency, number of associates, lexicality, nonword legality, pseudohomophone and cognate status (p. 80-82).

A further challenge is that inconsistent results are not uncommon in reaction time studies, and the exact effect of these variables is not without controversy (Jiang, 2012, p. 82). Furthermore, different theoretical positions can complexify the explanation of apparently simple effects. Word frequency, for example, is often considered one of the most robust predictors of word recognition performance. It has long been known that participants respond more quickly and accurately to high-frequency than low-frequency words across virtually all lexical processing tasks (Whaley 1978, cited in Yap & Balota, 2015). Although there are various potential models that explain the word frequency effect, most reflect a version of frequency affecting baseline activation. For instance, according to the interactive activation model (McClelland & Rumelhart, 1981), high-frequency words are responded to faster because they have a higher resting activation level (or lower threshold), so require less stimulus information to be recognised. As Marsden (2009) notes,

> As listeners (or readers) are exposed to a certain form in the input … then the neurological representations of that form are activated … via the resources available in working memory. Over the course of time, a representation may receive increasing amounts of activation, and this is thought to lower its activation threshold i.e. it can be activated with less energy (p. 11).

Taking into account the various complications of LDTs, two key decisions were taken, to attempt to minimise uncontrolled complexity. Firstly, to avoid the interaction of a different language as L1 with the language of the experiment, it was decided to conduct this experiment on English L1 speakers. Secondly, it was decided to use pseudowords, that is, strings of letters that conform to legal English orthographic and phonological constraints,[1] for example, *nirk*. By using pseudoword primes, it is therefore possible to restrict the prime-target relationship to form-based variables only and avoid semantic priming effects. Of course, even English L1 speakers might encounter priming from another language they know, but as they will perceive the experiment to engage with their L1, it should minimise the effect to background noise. Secondly, it is only the masked primes that are pseudowords, while the targets are the necessary mixture of real words and pseudowords. As a result, this experiment should come across as fully about the participants' L1.

5.2 Method

5.2.1 Participants

Data was collected from 24 English L1 speakers, 20 from the United States, three from the UK, and one from Canada (average age 24 years and 8 months, 15 female and 9 male). The participants were all university students on an exchange programme at a private university in Japan. A self-report questionnaire indicated a low-intermediate knowledge of Japanese. The participants had normal (or corrected-to-normal) vision and no history of language impairment. The questionnaire also asked if the participants were left- or right-handed. The participants provided informed consent and received a raffle ticket to win a pair of headphones in exchange for their participation. Data was collected over several weeks of university

[1] The term 'pseudoword' is used rather than 'non-word', as the latter can refer to an illegal letter string, such as *mgfa*.

term time. Although 24 participants may appear to be a small sample, Jiang (2012, p. 46) notes that many reaction time studies are done with similar sample sizes, as increasing the number of participants has little effect on the RT data.

5.2.2 Compiling the stimuli items for the Lexical Decision Task

In a Lexical Decision Task, two equally sized sets of stimuli are needed for the target items: a set of real words and a set of pseudowords. The inclusion of pseudowords is a necessary part of the design because it prevents the participants from simply pressing the 'Yes' key to every single target and developing a bias of responses.

For the set of primes, pseudowords were used to avoid any semantic priming effects. Furthermore, in a task aiming to establish the impact of a feature on how quickly words are recognised, there need to be subsets of words with and without the feature. In this experiment, the set of real word stimuli needed to consist of words used to test the effect of alliteration, words used to test the effect of assonance, and a control set that had neither feature. The following section outlines how the real words were chosen for the LDT and how the pseudowords were generated for the primes and the 'No' response targets.

Real word stimuli (targets)

As Adjective-Noun collocations and Noun-Noun compounds were the most commonly used stimuli in the work of Boers et al. (reviewed in Chapter 2), it was decided that nouns would also be used as the real word stimuli in the LDT. In an attempt to control for possible frequency, familiarity and length effects, a pool of words was assembled which consisted of 290 monosyllabic nouns of 4-5 letters in length, from the first two frequency sub-lists of the British National Corpus / Corpus of Contemporary American English (BNC-COCA) corpus (Nation, 2012). Although it is unlikely that any corpus matches exactly the linguistic experience of any particular speaker, this approach was deemed valid in that corpora are representative of the types of input the English L1 participants are likely to have encountered.

The items were then cross-referenced with the Medical Research Council (MRC) database (M. Coltheart, 1981) for concreteness and imageability ratings. This database contains up to 26 linguistic and psycholinguistic attributes for over 150,000 words collated from previously published sources. The words in the database have subjective ratings on an integer scale from 100 to 700. There were twenty-four nouns in the initial pool that had no ratings in the MRC database, so these were eliminated from the pool (for example, *bike*, *dish*, *farm*). The mean and standard deviations were calculated for concreteness and imageability of the remaining 266 nouns. This dataset was then trimmed of any nouns which had ratings of more than ± 2 SD from the mean. This resulted in the removal of 11 nouns from the pool (for example, *bunch, place, thing*), with the aim of producing a range of concreteness and imageability as narrow as possible. The remaining group of 255 potential stimuli had mean concreteness ratings of 576.9 (SD 36.7) and mean imageability ratings of 572.2 (SD 36.1). In comparison, the mean concreteness and imageability ratings in the MRC database are 438 and 450 respectively, suggesting the target stimuli for the LTD were slightly more concrete / imageable than the database average.

The pseudowords in an LDT play two different roles: as primes, and as targets for the 'No' responses. These will be dealt with in turn.

Pseudoword stimuli (primes)

The nature of a LDT, in which participants have to make rapid decisions as to whether the visual letter string is a real word or not, means that the characteristics of the pseudowords become an essential part of the experimental design. Not all pseudowords are the same. There is evidence that the types of pseudowords used in a priming experiment have a strong effect on reaction time performance. Keuleers and Brysbaert (2010, p. 627) state that “the more dissimilar the nonwords are to the words [in the task], the faster are the lexical decision times and the smaller is the impact of word features such as word frequency, age of acquisition, and spelling-sound consistency.” Clearly, features such as frequency, and age of acquisition, apply only to real words, though the form of the pseudoword can affect the speed of all decision processes in an LDT. In Gibbs and Van Orden (1998), for example, mean reaction times to reject nonwords were the shortest when the nonwords were illegal letter strings (e.g. *ldfa* – 496 ms), longer when the stimuli were legal letter strings (e.g. *dilt* – 558 ms) and longer still when the nonwords were pseudohomophones sounding like real words (e.g. *durt* – 698 ms).

To construct matching pseudowords for the 255 nouns, the Wuggy pseudoword generator was used (Keuleers & Brysbaert, 2010) to create monosyllabic stimuli of equivalent length (4-5 letters) with sound-spelling consistency. The Wuggy algorithm produced ten candidate pseudowords for each real word noun. As words that are orthographically similar to many other words are recognised faster (Yarkoni, Balota, & Yap, 2008), neighbourhood size and density plays an important role when developing stimuli lists. The Wuggy algorithm uses the Levenshtein edit distance of orthographic similarity (OLD20), where similarity includes neighbours generated by insertion, deletion, substitution, and transposition of letters to produce perceptually similar words. By referring to the OLD20 values in the Wuggy output, it was possible to select the pseudowords that remained as close as possible to real words, that is, with OLD20 values as close to 1.0 as possible. An OLD20 value of 1.8 was arbitrarily set as a cut-off point for pseudoword candidates.

The ideal pseudoword candidate would not cause any inadvertent triggering of semantic or associative representations in a participant’s mental lexicon. This proved challenging to control for, as a considerable proportion of the 2550 pseudowords could be construed as a proper name, brand name, acronym, pseudohomophone, informal or archaic or variant spelling of a common noun (examples to follow). In addition, closer scrutiny of the Wuggy output unearthed several words attested in on-line dictionaries (such as *blog*, *mage*, and *bling*).

The pseudowords were cross-referenced with the COCA database (Davies, 2008 to present). Any items that appeared as proper names with more than 10 instances in the corpus were discounted (for example, *crowe*, *imes*, *gide*). Similarly, any items that occurred more than 10 times as acronyms (*pacs*, *facs*, *mact*) were also deleted. Pseudowords were also discounted if they appeared in the corpus as brand names (such as *fage*), abbreviations or variant spellings (for example, *hols*, *dept*, *lite*, *nite*), or slang (*crip*, *shart*).

The Google search engine was then used to check if the remaining pseudowords elicited any dictionary references, especially in Merriam Webster, as participants were most likely to be from a North American background given the educational context; this led to two more deletions (*thang*, *fleed*). Finally, all pseudohomophones were deleted (for example, *bocs*, *coaks*, *ceal*, *bild*, *tode*, *wead*).

The experimental design meant that for each of the 255 noun targets, there needed to be one pseudoword prime in each condition – alliteration, assonance, and no phonological pattern. However, after removing all pseudoword candidates with OLD20 values higher than 1.8, pseudohomophones, and those attested in corpora or dictionaries, some nouns did not have equivalent pseudowords in all three conditions, so these nouns were removed as test items. Their pseudoword equivalents that met the criteria were retained to use as filler items (see below).

This process of elimination resulted in a final pool of 102 target nouns, each with one alliterating pseudoword prime, one assonating pseudoword prime, and one pseudoword prime with no phonological / orthographic overlap, an example of which can be seen in Table 5.1.

Table 5.1 Example of a stimulus item across three conditions.

Prime	Target	Condition
mouch	*MILK*	Alliteration
skift	*MILK*	Assonance
drate	*MILK*	Control

One limitation of the Wuggy algorithm is that it does not generate the pronunciation for the pseudowords. To check that the prime-target pairs did alliterate and assonate, three English L1 speakers were asked to read aloud the list of items to check the researcher's intuitions of phonological similarity. No discrepancies were found in this regard.

Pseudoword stimuli (targets)

To avoid response bias in the LDT, there had to be an equal number of 'Yes' and 'No' responses. Without such filler items, participants would only need to press the 'Yes' key to produce a correct response. Furthermore, such filler items help disguise the critical stimuli so it is less likely a participant will notice and consequently develop a processing strategy while performing the task. One hundred and two pseudoword targets, that would hopefully elicit a 'No' response on the decision task, were selected from the remainder of the Wuggy output. For the 102 pseudoword primes the mean OLD20 value was 1.34 (min. 1.0, max. 1.7, S.D. 0.22) and the mean value for the 102 pseudoword fillers was also 1.34 (min. 1.0, max. 1.75, S.D. 0.24).

In sum, there were 102 noun targets, designed to elicit a 'Yes' response in the LDT. The following variables were controlled for: frequency, familiarity, word length, neighbourhood density, concreteness, and imageability. Each noun target had three pseudoword primes, one in each condition. There were an additional 102 pseudoword targets designed to elicit a 'No' response (see Appendix 4 for a list of the test items). The primes for these filler items were chosen from the remainder of the Wuggy output. Pseudowords complied with the phonotactic constraints of English and were controlled for in terms of length and orthographic similarity.

5.2.3 Setting up the LDT programme

The Lexical Decision Task was done using the free stimulus presentation software DMDX[2] (Forster & Forster, 2003). An item file was created in a Rich Text Format, which basically instructed the software on what items to present and how to present them. This allowed the 102 nouns plus 102 pseudowords to be combined into 204 trials, which were then divided into 17 blocks. Each block contained two alliterating prime-target pairs, two assonating prime-target pairs, two prime-target pairs with no form overlap and six pseudoword filler items. Items were scrambled within each block and blocks were also scrambled for each participant. This pseudo-randomisation avoided long successions of words or pseudowords appearing by chance. It also ensured extraneous serial effects (such as practice or fatigue) were more evenly distributed across conditions and no two subjects were likely to receive the same sequence of items. In addition, it avoided the introduction of systematic errors of measurement arising, when difficult items (on which an error is likely) can affect the reaction time for following items. The three conditions were counterbalanced across three presentation lists in a Latin square design, such that in each list, one third of the stimuli (34 items) appeared in each condition. This allowed for direct comparisons across conditions and avoided list effects. Participants were randomly assigned to one of these lists and responded to an equal number of trials in all three conditions, but never responded to the same target more than once. Two versions of the item file were created to allow for different keys being allocated to 'Yes' and 'No' responses, one for right-hand dominant participants and one for left-hand dominant participants. The experiment was run on a PC desktop computer under Microsoft Windows XP (1920x1080 resolution, 60Hz refresh rate). TimeDX, a component of the DMDX software suite, was used to verify that the hardware features of the computer were satisfactory.

[2] Available at http://www.u.arizona.edu/~kforster/dmdx/download.htm

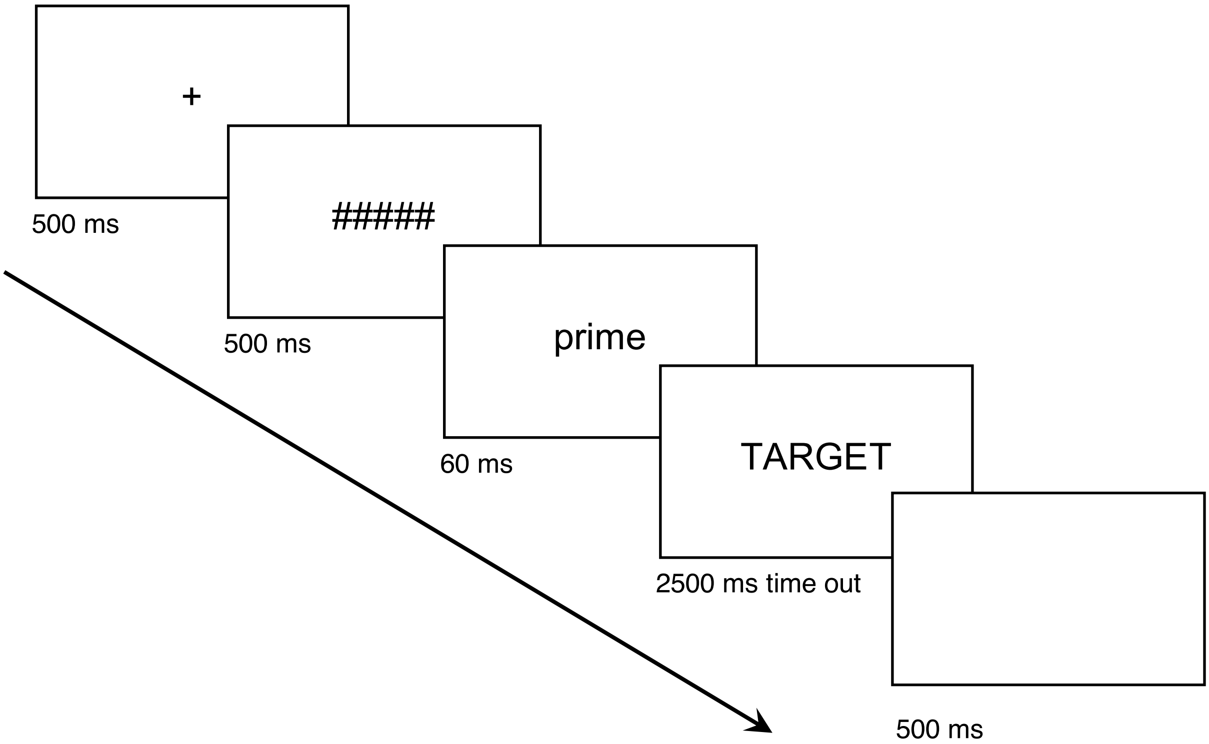
5.2.4 Procedure

Participants were given written instructions (adapted from those used in Ellis, Simpson-Vlach, & Maynard, 2008) in which they were told they were participating in a vocabulary experiment and they would see a letter string on the computer screen and to press the key labelled ‘Yes’ if they thought it was an English word, and the key labelled ‘No’ if they thought otherwise (see Appendix 5). Participants were asked to respond as quickly and accurately as possible. Participants were not informed of the presence of the masked stimuli, nor the three critical conditions.

The participants were tested individually and sat approximately 50 cm in front of the eye-level computer screen, wearing noise-cancelling headphones in a quiet and dimly lit room. The index finger of their dominant hand was assigned to the ‘Yes’ key and their other index finger to the ‘No’ key. The experiment began with ten practice trials, followed by the 204 experimental trials. There was a programmed break in the middle of the experiment for participants to rest, and the experiment resumed when participants pressed the space key. Trials were randomised for each participant and the experimental phase took approximately 17 minutes. The stimuli were presented in the centre of the computer screen in black Arial font, size 14, on a light grey background using the DMDX software. An explanatory schema of a trial can be seen in Figure 5.1.

Each trial began with the presentation of a fixation point (+) for 500ms, to direct the attention of the participant to the appropriate location on the screen. A forward mask (a meaningless string of five hash marks) appeared for 500ms centred at the same location, followed by the presentation of the prime for 60ms, an amount of time deemed safe for the item to be processed but without the participant becoming aware of it. Then the target appeared on the screen, which served as a backward mask for the prime. The target remained on the screen until the participant’s

Figure 5.1 Sample trial schema.



judgement or a 2500ms timeout, followed by a 500ms inter-trial interval (a blank screen). Response times (RTs) were measured from target onset until the participants' response on the keyboard. After the experiment was completed, participants were asked to complete a short exit questionnaire asking them to self-report on how well they followed task instructions, and whether or not they were aware of the primes.

5.3 Results

5.3.1 Data Analysis

The raw data from the response latencies and associated accuracy were saved as a data file and entered into the *Analyze* program, another component of DMDX, for which an input specification file was written. This text file instructs the programme how to treat the raw data and which item goes to what condition. The script parameters also specify how outliers and incorrect responses are to be processed, as described below.

Outliers, defined as responses less than 200ms or more than 3 standard deviations above the participant's mean, were replaced by the mean value for that participant. This appears to be standard practice in an LDT (McDonough & Trofimovich, 2009, p. 152). It eliminates the need to perform a log transformation on the subsequent data set and ensures a normal distribution, thus safeguarding from the unwanted effects of outlier response latencies.

The script parameters instructed the programme to exclude incorrect responses from the analysis. They also instructed the programme to identify and automatically remove items that generated error rates of 40% or more, together with the counterpart items in the other two conditions, so that test materials remained matched across conditions. The file also instructed the programme to reject data from participants with an error rate of 20% or over.

5.3.2 Results

Inspection of the completed data showed that no items generated error rates of 40% or more, so no items were eliminated on these grounds. No participants had error rates of 20% or over. The mean error rate per participant was 3.79% (SD = 2.40), resulting in the removal of 186 incorrect responses. A total of 55 responses qualified as outliers and were replaced by the mean value for that participant (the mean number of outliers per participant was 2.29, SD = 1.23). Furthermore, in the exit-questionnaire, no participants reported seeing the primes, and all the participants indicated that they had carried out the task exactly as instructed.

A paired-samples t-test was conducted to compare the reaction times in the real word and pseudoword conditions. There was a significant difference in the scores for real word (M = 566.14 ms, SD = 33.20) and non-word (M = 671.8 ms, SD = 76.32) conditions; $t(23) = 6.7$, p <0.001 (2-tailed). These results suggest that the experiment design and choice of pseudoword matches was sound in that words and pseudowords elicited the expected responses (as noted previously, in Gibbs & Van Orden, 1998).

To recap, the research question is whether phonologically patterned prime-target items will result in faster reaction times than prime-targets in the baseline condition with no phonological overlap. Descriptive results from the LDT can be seen in Table 5.2:

Table 5.2 Descriptive results.

	Mean Reaction Time (RT) [1]	SD
Alliteration	571.01 ms	34.45
Assonance	585.46 ms	39.07
No Phonological Pattern	575.05 ms	39.92

[1] in milliseconds (ms)

Table 5.2 shows that alliterating prime-target items produced the fastest mean reaction times, and assonating prime-target items the slowest mean reaction times. A RM ANOVA was used to analyse the results with response times as the dependent variable and alliteration, assonance, or no pattern as the categorical independent variable. Preliminary examination of the data satisfied the assumptions of homoscedasticity and normality of distribution, and Mauchly's test of sphericity showed that the assumption of sphericity was not violated (χ^2 (2) = 2.93, p = .86). There was a significant main effect of Phonological Pattern on the reaction times ($F_{2, 46} = 7.04$, $p = .002$, $\eta_p{}^2 = .23$, power = .91).

Pairwise post hoc comparisons using t-tests and Bonferroni corrected levels of significance showed that participants' RTs were significantly faster in the Alliteration condition compared to the Assonating condition (mean difference = 14.45, $t = 3.69$, $p = 0.004$) with a medium effect size (Cohen's $d = 0.76$), according to Plonsky &

Oswald's (2014) interpretation. However, there was no statistically significant difference between reaction times in the Alliteration condition and reaction times in the baseline condition (mean difference = -4.03, t = -1.06, p = 0.90, d = 0.22), nor was there a statistically significant difference between the Assonance reaction times and the No Phonological Pattern reaction times (mean difference = 10.41, t = 2.49, p = 0.06, d = 0.51).

Thus, in answer to the Research Question, phonologically patterned prime-target items did not lead to faster reaction times than prime-targets in the baseline condition with no phonological overlap.

5.4 Discussion

The experiment in this chapter sought to test the hypothesis raised by Boers et al. that word-pairs displaying alliteration and assonance have a processing advantage compared to equivalent word-pairs with no such phonological or orthographic overlap. Previous priming experiments have found reliable evidence that rhyme produces a priming effect, though the evidence is inconclusive for primes and targets that alliterate, and there is scant data on assonance alone. The results reported here do not contribute evidence of a facilitatory priming effect with a data set which used masked pseudoword primes and noun targets. The absence of a statistically significant result could be a consequence of a methodological shortcoming, or it could be evidence of the null hypothesis, that is, alliteration and assonance (as operationalised in this experiment) do not facilitate lexical processing. These potential explanations will be considered in turn.

The experiment described in this chapter used a standard Lexical Decision Task with prime-target items composed of pseudoword primes and monosyllabic high-frequency nouns in three conditions. No individual items generated large error rates and the participants' mean error rate (5.6%) is in line with other published masked LTD's (for example, a 6% mean error rate in Perea, Vergara-Martinez, & Gomez, 2015). The number of participants (N = 24) was deemed appropriate (Jiang, 2012, p.

46) as was the number of items.[3] Thus, the 24 participants responded to 34 prime-target items in each of the three conditions, generating 816 observations per condition. However, in light of the fact that any mnemonic advantage for alliteration and assonance has tended to be quite modest in the classroom-based experiments reported in Chapters 3 and 4,[4] this number of observations per condition may have been insufficient to properly investigate a small effect. Indeed, for a repeated-measures reaction time study, Brysbaert & Stevens (2018, p. 16) recommend a minimum of 1600 observations per condition. This suggests that future experiments with sufficient power to detect a small effect would require approximately double the number of items per condition, or twice the number of participants. Recruiting a larger body of participants could perhaps be done more easily via a browser-based application like the *Gorilla Experiment Builder* (Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020), though there are potential drawbacks to internet-based reaction time data gathering (see, for example, Woods, Velasco, Levitan, Wan, & Spence, 2015). To conclude, one possible explanation for the lack of a statistically significant finding is that, although alliteration and assonance do facilitate lexical processing, the experiment in this chapter was underpowered, and therefore could not capture the effect.

Another possible reason for not finding a priming effect for phonological patterns relates to the thorny issue first mentioned in Chapter 1: what counts as a 'similar sound'? In this experiment, alliteration and assonance were operationalised as sounds that match at the phoneme level. This may not be a safe assumption. At this level, allophonic differences created by the phonological environment are not taken into account. Thus, for example, the /ɪ/ in '*skift'* and '*milk'* would be considered the same, even though the latter vowel would be darker before the velarised /l/. This point will be returned to in Chapter 7. The following chapter extends the use of

[3] Jiang (2012, p. 48) states "It is desirable to have ten to twenty items in a condition whenever it is possible." The experiment in this chapter had 34 items in each condition.

[4] For example, in the replication study reported in Chapter 4, participants recalled on average one more assonating word-pair compared to non-assonating word-pairs across the tests.

pseudowords to explore the mnemonic effect of phonological patterns, but from a different methodological approach.

Chapter 6 Do alliteration and assonance support recall when item meaning is controlled?

6.1 Introduction

Previous experiments by Boers et al. (as described in Chapter 2), on the mnemonic effect of alliteration or assonance in English word-pairs, controlled as far as possible for the many variables that can affect recall from memory, such as frequency, orthographic length, concreteness, and L1 cognate status. If this is not done, it is hard for researchers to claim with confidence that a word-pair was successfully recalled based on the presence of a phonological pattern alone.

These studies commonly used word-pairs that were assumed to be familiar to the participants. But foreign language learning involves acquiring copious amounts of *new* vocabulary. Therefore, this chapter looks at the extent to which phonological patterns can aid this process. Before doing so, however, the assumption that previous studies used 'known' test items warrants further attention as it directly impinges on how to operationalise an 'unknown' test item. Section 2.3.5 touched on some of the difficulties in ascertaining if a word-pair is 'known' or not, but it is worth re-visiting some of these ideas. Chapter 7 picks up on this aspect of the suite of experiments, in particular, in those that pre-tested the participants' knowledge, Studies 10 and 11 in Table 2.4 (Eyckmans et al., 2016; Eyckmans & Lindstromberg, 2017 respectively).

In the studies outlined in Table 2.4, the validity of the belief that a word-pair is 'known' is predicated on the participants' L2 proficiency, in combination with the corpus frequency of the word-pair. These two premises will be briefly reviewed in turn. The participants' proficiency descriptors, for example, as 'Upper Intermediate' or 'Advanced', is based on their classroom teachers' assessment and/or English proficiency test scores (such as IELTS or TOEFL), or reference to the Common European Framework of Reference for Languages (CEFR), or results from the

Vocabulary Size Test (Nation & Beglar, 2007). Corpus-frequency counts, most often from the COCA database (Davies, 2008 to present), were used to establish that the word-pairs were high frequency.

Although there is evidence that, perhaps unsurprisingly, more highly proficient L2 learners tend to have larger vocabulary sizes than lower proficiency learners (for written receptive vocabulary sizes, see Beglar, 2010), assuming a learner knows a word or word-pair by virtue of its frequency alone is not so straightforward. As Horst (2013, p. 177) notes, knowledge of the first two thousand words is rarely complete, and learners often have a mixed profile in terms of word knowledge over a range of frequencies. In addition, vocabulary acquisition is a cumulative process and word-knowledge is multi-faceted, covering various aspects of meaning, form and use, indicating that the status of 'known' is rarely a yes / no proposition.

These issues are not confined to the work of Boers et al.; two previous experiments in this thesis (reported in Chapters 3 and 4) also used 'known' word-pairs. Again, this assumption was based on the corpus frequency of the stimuli, plus the participants' TOEIC scores, their results from the Vocabulary Levels Test (Schmitt et al., 2001), as well as the results of a self-report on word-pair familiarity.

As familiarity affects recall (in that it is hard for a participant to recall a word-pair that is unfamiliar to them) then this must be controlled for in the experimental design. Therefore, to ensure tighter control of potentially confounding variables, including familiarity, the experiment detailed in this Chapter uses pseudowords, defined in Section 5.1 as nonwords that conform to English orthographic and phonological constraints. Potential drawbacks to the use of pseudowords will be addressed in section 6.2.3. A phonological string with no meaning would not have a semantic representation in the mental lexicon, and so it was necessary to attach an invented meaning to simulate real vocabulary learning, for example *geed feek* = '*tall child'*.

6.1.1 Research Questions

Three research questions directed the design of the experiment:

Research Question 6.1: Are English L2 learners better able to recall pseudoword-pairs that have a sound pattern (alliteration or assonance) than comparable pairs that do not?

The research hypothesis is that the alliterating and assonating items will be recalled more readily than the items with no phonological pattern, in line with previous studies.

Research Question 6.2: Is a given stimulus more memorable if learners have had the sound pattern drawn to their attention than if they have not?

As discussed previously (see Chapter 2), in prior studies explicit awareness raising of the phonological pattern has been associated with greater recall of experimental items; in contrast, where the pattern is not highlighted, the mnemonic effect is often reduced (for example, Boers et al., 2012) or absent. Thus, the research hypothesis is that participants who have had the phonological patterns highlighted will recall more items than participants who have no patterns brought to their attention.

Research Question 6.3: Do phonological patterns affect the recall of the form and recall of the meaning differently?

Previous studies have asked participants to recall the written form of a given word-pair or, when presented with a cue, to write down the remaining half of a word-pair. This holds true for those studies which used ‘unknown’ stimuli (Studies 10 and 11 in Table 2.4). The participants’ recall of the *meaning* of a novel stimulus has not been thoroughly tested. Therefore, it is currently not clear if a mnemonic advantage applies to the meaning of a new phonologically patterned word-pair.

To sum up, this study has three variables of interest. RQ 6.1 will help to address the potential criticism of past studies that the effects found were an artefact of variation in the extent of the participants’ knowledge of the words used as stimuli. RQ 6.2 concerns the effect of awareness-raising within the experimental procedure. RQ 6.3 investigates whether participants can more accurately produce the meaning of a given pseudoword-pair in a phonologically patterned condition compared to the meaning of a pseudoword-pair with no phonological pattern.

6.2 Method

6.2.1 Participants

The participants were a convenience sample of 93 Japanese L1 undergraduates (51 male, 42 female; 19 years old) in their first year of study at a private university in Japan. The participants had, on average, six years of English classes in High School, and were from four intact classes in three different university faculties: Law, Economics and Humanities. The participants met twice a week for English classes with two different teachers, for a total of three hours per week, and none of the participants were language majors. As regards English proficiency, participants all took TOEIC Reading and Listening tests approximately one month before the start of the experiment. The mean TOEIC Listening score was 277 (minimum = 175, maximum = 385, $SD = 41.8$), and the mean TOEIC Reading score was 247 (minimum = 170, maximum = 345, $SD = 38.9$). These mean scores can be mapped onto the Common European Framework of Reference (CEFR) bands of B1 (pre-intermediate for listening skills) and A2 (elementary for reading skills) (Tannenbaum & Wylie, 2013).

6.2.2 Materials

The experiment would involve the use of flashcards with the stimuli on one side and their 'translation' on the other. In developing the materials, the aim was to create stimuli that would be as close as possible to legitimate new word-pairs of English, without using real words. The items used were pseudoword Adjective + Noun sequences based on English lemmas from the first 1000 most common adjectives and 1000 most common nouns in the COCA corpus (Davies, 2008 to present). A pool of potential pseudowords was created using the Wuggy non-word generator (Keuleers & Brysbaert, 2010), and thus the items follow the lexico-syntactic and phonological patterns of English. Monosyllabic pseudowords were chosen to ensure that the sound patterns under investigation would be sufficiently proximate to have the intended effect, and all the pseudo-words were uniform in length (four letters), as word length can affect recall in tests of memory (for example, Baddeley et al., 1975).

Duplicate pseudowords were removed from the Wuggy output, as were any items that could be construed as homophones (e.g. *wite*, *dood*, *facs*), slang (e.g. *dork*, *lite*, *naff*), abbreviations (e.g. *fash*, *fave*, *tash*) or proper names (e.g. *hugh*, *luke*, *ruth*). Any pseudowords with word-initial consonant clusters were also removed (e.g. *craf*) as Japanese speakers have a tendency to insert vowel sounds into clusters. Regarding clusters in the final position of a pseudoword, the final list of assonating target pairs contained three clusters, while the alliterating and no pattern items had five each. In terms of consonant digraphs,[1] however, the assonating list had nine, the alliterating had seven and the no pattern list had five consonant digraphs. These differences were not considered important, since neither alliteration nor assonance are concerned with the post-vocalic coda.

To confirm the researcher's intuition as to the pronunciation of the pseudowords, four English L1 speakers (2 from Canada, 1 from the USA and 1 from the UK) were asked to read the items aloud. Any items that were pronounced differently were removed from the pool: of 432 utterances only 2 items were pronounced differently: *mive sipe* was pronounced once as /mivsaip/ with no assonance, and the pseudo-adjective in *tich goak* was pronounced once as /tik/, resulting in an additional phonological pattern of consonance.

Next, three Japanese L1 speakers were asked to confirm if any of the remaining pseudowords were congruent with Japanese words, to avoid L1 cognate equivalents which would be easier to recall. Each pseudoword was rated on a 5 - point Likert scale (from 1 = this does not sound like a Japanese word, to 5 = this sounds like a Japanese word). Any items that scored 4 or 5 points from any rater were deleted from the pool. Further enquiry revealed, for example, that *bape* was held to be

[1] Two consonant letters representing a single sound (for example, *sh* for /ʃ/)

similar to ベープ, as in Vape, a brand of mosquito repellent, *blee* was rated similar to ブリー, the French cheese brie, and *weck* was found to be the name of a brand of Japanese kitchen appliances.

Following a pronounced floor effect in a pilot test, the number of test items was reduced from 45 to 30: 10 of the pseudoword Adj+N pairs showed prototypical alliteration (e.g. *virt veed*), 10 assonated (e.g. *jush mund*) and 10 had no phonological similarity (e.g. *coof mide*).

The reverse side of each flashcard needed to have a 'translation' in Japanese - that is, a plausible, and easily understood phrase that the participants would be given as the 'meaning' of the pseudo-word pair. In the absence of access to a suitable Japanese vocabulary profiler tool, the best way to ensure that all the Japanese translations were made up of high-frequency words was to create them from high-frequency English phrases. To that end, the English meanings that would be translated into Japanese came from the words which formed the initial input for the Wuggy non-word generator - that is, high-frequency English adjectives and nouns from the COCA corpus (for example, *long list, poor man*). To further check that the Japanese translations were frequent words in plausible combinations, the Japanese L1 speaker who did the translations was asked to identify any potentially problematic items. Any Japanese translations that alliterated (in terms of similar initial morae[2]) were removed from the pool, for example 厳しい規則 (*kibishii kisoku*) – *a strict rule*.

Two more variables needed to be controlled for in the translations: length and concreteness, as both of these can impact on how easily an item can be remembered. Length was calculated according to the number of morae in the translation, and a value for concreteness was calculated via a Likert-scale questionnaire administered to three Japanese L1 speakers. Each Japanese phrase was rated on a scale of 1 (most concrete) to 5 (most abstract), resulting in a minimum score of 3 (most concrete) and a maximum score of 15 (most abstract).

[2] Defined in Section 3.2.2 as a sub-syllabic unit of timing, prototypically consisting of a single consonant followed by a short vowel.

The information about Japanese phrase length and concreteness can be seen in Table 6.1.

Table 6.1 Mean length and concreteness of Japanese phrases.

Condition	Mean Number of Morae (*SD*)	Mean Concreteness Rating (min 3, max 15) (*SD*)
Alliteration	5.9 (1.20)	6.7 (2.71)
Assonance	5.9 (1.20)	6.8 (3.39)
No Pattern	5.2 (0.79)	5 (2.40)

It can be seen from Table 6.1 that the translations for the pseudoword-pairs in the 'No Pattern' condition were, on average, slightly shorter and more concrete, and thus should have an advantage in any test of recall from memory.

The final pool of thirty target pseudoword-pairs, their designated Japanese translations, and the high-frequency word-pairs used as the basis for the translations can be seen in Table 6.2.

6.2.3 Procedure

Although the use of pseudowords cannot be considered ecological in the context of English language learning, it was a necessary part of the design. In order to enhance the value of taking part in the study, the experiment took the form of a meaningful activity. The pseudoword-pairs and their Japanese translations were uploaded onto a free web-based study application, Quizlet (Sutherland, 2019), which uses flashcards for paired-associate learning.[3] The stimuli were presented via the Quizlet app and studied using the Vocabulary Card Technique (Nation, 2008, p. 106), which was due to be subsequently adopted as a learning technique for the remainder of the academic year with the participants' university course vocabulary. On that basis, the study played the role of introducing the students to the app and the Vocabulary Card Technique.

[3] The method of paired-associate learning involves asking the participant to associate the word-pair on one side of a flashcard to the L1 meaning on the reverse side.

The students were told that using pseudowords in the first instance was a way for them, and their teacher, to see the effectiveness of the method without interference from previously known words. This was a genuine benefit in relation to the use of this new learning method. Thus, combining the research with this teaching enhancement built up the ecological value of the study and helped ensure that the participants engaged with it.

Table 6.2 Thirty target items and translations.

Alliteration	Japanese	Meaning
virt veed	公正な法	fair law
coth cank	混んだ場所	full place
tull tesh	勇敢な集団	brave class
fane fipt	低い声	deep voice
goot gark	長い一覧表	long list
husp haff	格好いい仕事	cool job
jull jeed	暖かい光	warm light
soff sest	円形の競技場	round field
muge mipe	賢い女性	smart girl
nift nood	安全な薬	safe drug
Assonance	**Japanese**	**Meaning**
bame fafe	輝く星	bright star
coob woot	わずかな危険	slight risk
wesh telk	巧みな技術	neat skill
dard nart	短い試験	short test
thip kiff	ゆっくりな映画	slow film
leff nend	難しい目標	tough goal
jush mund	立派な家	fine house
nirk serm	急な土手	steep bank
sish bick	貧しい男性	poor man
geed feek	背の高い子ども	tall child
No Sound Pattern	**Japanese**	**Meaning**
neam shoy	激しい戦争	fierce war
coof mide	脇役	small role
dage kimp	細い線	thin line
vank lism	荒れた試合	rough game
seck doob	暗い部屋	dark room
junt woil	まっすぐな道路	straight road
riss woft	壮大な劇	grand show
goor hean	大きな扉	large door
niss coom	冷たい食事	cold food
pake losh	分厚い本	thick book

For the classroom-based intervention, the same procedure was followed in all four intact classes. The participants were randomly divided into three treatment groups, and seated separately. The researcher told the participants that they were going to practise the Vocabulary Card Technique using the app on their smartphones, then they would be tested. Participants in Group 1, the alliteration group, received a handout with instructions on how to use flashcards according to the Vocabulary Card Technique (see Appendix 6). The instructions included an additional step telling the participants to pay special attention to phrases in which the constituent words started with the same sound, for example, *warm wind*, as this could help them memorise the phrases. The instruction sheet had a link to the website where they could practice the Vocab Card Technique. The set of flashcards at this link consisted of the full set of thirty target stimuli, though the ten alliterating phrases had the word-initial consonants in bold and underlined, for example, ***v**irt **v**eed*.

Participants in Group 2, the assonance group, received a similar instruction sheet but were told to pay special attention to the phrases that assonated, for example*, deep sea*. Their instruction sheet included a link to the app, where they could access the full set of flashcards, with all the assonating patterns in bold and underlined, such as *b**a**me f**a**fe*. Participants in Group 3, the control group, received an instruction sheet without the step alerting them to phonological patterns, and the flashcards at their link had no sound patterns highlighted.

The participants accessed the flashcards on their smartphones and practised the Vocabulary Card Technique in class. After twenty-five minutes, their phones were put away and Test 1 was administered. The test instrument was divided into two parts. The first, which was the Focus on Form test (FonF), targeted the written form of the pseudoword-pairs and was in the format of a pen-and-paper cued-recall test. The participants were given the pseudo-adjective and had to write down the collocating pseudo-noun. A cued recall test was chosen over a free recall test, in which participants are asked to write down both words of a phrase, because a free recall test had produced a marked floor effect in the pilot study.

After the papers were collected, the second part of the test, the Focus on Meaning test (FonM), was administered: the participants were given a handout with the thirty pseudoword-pairs and had to write the appropriate Japanese translation, recalled from the reverse of the flashcards. The participants were asked to supply the Japanese translations by looking at the pseudo-word pairs, rather than the reverse, because the pilot study results revealed that asking participants to write the pseudoword pairs for the Japanese translations was too demanding and had resulted in a floor effect.

After these papers were collected, the participants were asked to work individually and study the flashcards in their own time for a test the following week, when the same tests would be given. Participants were asked to study the flashcards individually, in the hope that there would be no leaking of the instructions to pay attention (or not) to the particular sound patterns. The number of times a participant studied the set of flashcards could be tracked via the website, though this only gave an approximate picture; studying half the set then closing the internet browser would not result in a recorded session, nor would flicking through the set without turning the card, though there was no real incentive to engage superficially as the participants were not told to reach a specific goal for the number of times the set was to be studied. After the second test, the following week, access to the site was blocked so that there could be no further study for the unannounced Test 3 one week later. Tests 1, 2 and 3 followed the same format, a cued recall test (Focus on Form) and then a translation test (Focus on Meaning), but all the test items were randomised for each test time.

During the one-week interval between Test 1 and Test 2, when participants had been asked to study the set of thirty flashcards in their own time, the amount of study varied. The mean number of times a set of word-cards was studied in full was 7.6 (min = 1, max = 27, *SD* = 6.1), though as noted previously, this is a somewhat crude interpretation of the amount of deliberate study. The amount of word-card study across the three different conditions can be seen in Table 6.3.

Table 6.3 Amount of inter-test flashcard study.

	n	mean	*SD*	Min	Max
Group 1 (Alliteration)	26	8.73	7.02	1	27
Group 2 (Assonance)	25	6.84	5.32	1	20
Group 3 (No Phon. Pattern)	27	7.82	6.06	1	26

To see if there was a statistically significant difference between these means, an Omnibus One-way ANOVA was performed; this showed that the main effect of group was not statistically significant ($F_{2,75} = .59$, $p = .55$ (2-tailed)) with a negligible percentage variance (PV) effect size (.015).

6.3 Results

Regarding participant attrition, of the initial sample of 93 participants, 15 failed to complete all three parts of the experiment so their data was excluded from the analysis, resulting in a participant sample size of 78 (Group 1 = 26, Group 2 = 25, Group 3 = 27).

For the Focus on Form cued-recall tests, in which the participants were given the first part of the word-pair and had to write the second word, the pseudo-noun, one full point was awarded for an exact answer. Given the challenge of the recall task, a lenient marking scheme was adopted whereby half points were awarded if there were no more than two wrong letters per pseudoword – this included deletions, insertions, and substitutions. However, for the phrases with phonological patterns, half points were not given if the error interfered with the sound pattern, see Table 6.4 for examples.

Table 6.4 Marking rubric for pseudowords.

	Alliterating pseudowords	Assonating pseudowords	No Pattern pseudowords
Cue	*virt*	*leff*	*dage*
Target (1 point)	*veed*	*nend*	*kimp*
½ point	*vend*	*send*	*kifp*
0 point	*meed*	*nund*	*mide*

For the Focus on Meaning tests, in which the participants were asked to write the Japanese translation for a given pseudoword-pair, a Japanese L1 user marked the test papers and gave 1 point for correct answers and half a point if the participant had written the correct translation for either the adjective or the noun. No points were deducted if a participant used a kanji character for a translation that had used hiragana script on the flashcard, and vice versa.

The mean scores, standard deviations, and the minimum and maximum scores for each test can be seen in Table 6.5. Shaded cells indicate where the highest score would be predicted (for example, alliteration for Group 1), and the highest mean scores for each type of test, across the three groups, are in bold.

Figure 6.1 shows the mean scores for each group at the three different test times for the Focus on Form test, in which the participants were given the pseudo-adjective cue and were asked to write down the appropriate pseudo-noun. Similarly, Figure 6.2 shows the means scores at each test time for the Focus on Meaning test, in which participants wrote the Japanese translations for the pseudoword-pairs.

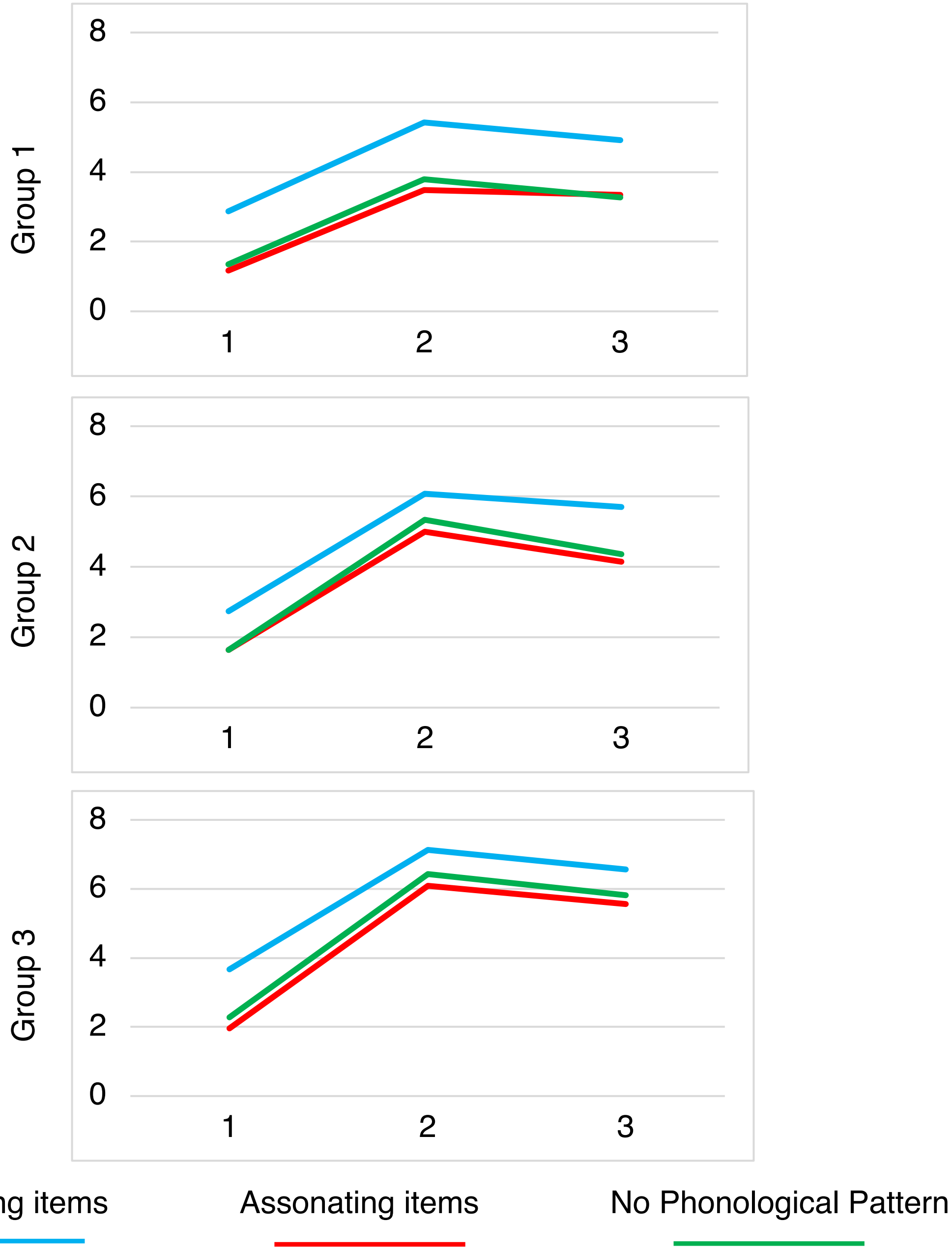
It can be seen that on average, participants in all the groups scored highest on the alliterating items, then the items with no sound pattern, and then the assonating items. With one week of deliberate study, mean recall scores improved for all types of test item (mean gains = 14.12 items, SD = 15.27), with more noticeable gains on the Focus on Form test, and then dropped only slightly in the intervening week without further study (mean losses = 4.54 items, SD = 4.24), suggesting sustained performance above the initial test (the exception being Group 1 on the Focus on Meaning test). In addition, participants in Group 3, who received no awareness-raising of the phonological patterns, scored more highly than participants who had the patterns highlighted on the electronic flashcards, particularly on the Focus on Form test. This unexpected finding will be discussed in section 6.4, after first analysing the results in relation to each of the RQs.

Table 6.5 Descriptive results.

	Focus on Form (FonF)				Focus on Meaning (FonM)			
	Test 1 (after 25 minutes of deliberate study)							
	Allit[1]	Assn[2]	NoP[3]	Total	Allit	Assn	NoP	Total
Group 1[4] *n* = 26	2.87[7]	1.17	1.35	5.39	**6.37**	**5.15**	**6.04**	17.56
	(2.09)[8]	(1.62)	(1.57)	(4.84)	(2.40)	(2.64)	(2.92)	(7.33)
	0[9]	0	0	0	1.5	2	1	4.5
	7.5[10]	7	5.5	20	10	10	10	30
Group 2[5] *n* = 25	2.74	1.64	1.64	6.02	5.26	4.62	4.98	14.86
	(2.31)	(2.20)	(2.30)	(6.34)	(3.07)	(3.37)	(3.41)	(9.48)
	0	0	0	0	0	0	0	1
	9.5	9	9	27.5	10	10	10	29
Group 3[6] *n* = 27	**3.67**	**1.96**	**2.28**	7.91	4.44	3.11	3.48	11.04
	(2.44)	(1.83)	(2.15)	(5.75)	(2.57)	(2.41)	(2.65)	(7.11)
	0	0	0	0	1.5	0	0	2
	10	6.5	8	23	10	8	8.5	26
	Test 2 (after 1 week of deliberate study)							
	Allit	Assn	NoP	Total	Allit	Assn	NoP	Total
Group 1	5.42	3.48	3.79	12.69	**6.62**	5.42	**6.67**	18.71
	(2.98)	(3.56)	(3.67)	(9.84)	(3.14)	(3.28)	(3.21)	(8.97)
	0	0	0	0	0	0	0	1
	10	10	10	30	10	10	10	30
Group 2	6.08	5.00	5.34	16.42	6.50	**6.08**	6.16	18.74
	(2.96)	(3.56)	(3.66)	(9.95)	(3.68)	(4.15)	(4.08)	(11.76)
	1.5	0	0	2.5	0	0	0	0
	10	10	10	30	10	10	10	30
Group 3	**7.13**	**6.09**	**6.43**	19.65	6.30	5.72	6.33	18.35
	(2.34)	(3.18)	(3.07)	(8.31)	(3.16)	(3.68)	(2.91)	(9.53)
	2.5	0	0	3	0	0	1	1
	10	10	10	30	10	10	10	30
	Test 3 (after a further week unannounced)							
	Allit	Assn	NoP	Total	Allit	Assn	NoP	Total
Group 1	4.92	3.35	3.27	11.54	**5.62**	4.60	5.17	15.39
	(3.27)	(3.66)	(3.16)	(9.75)	(2.72)	(2.85)	(2.87)	(7.70)
	0	0	0	0	0	0	0	2
	10	10	10	29.5	10	10	10	30
Group 2	5.7	4.14	4.36	14.2	5.50	**5.38**	**5.46**	16.34
	(2.89)	(3.06)	(3.47)	(9.02)	(3.85)	(4.14)	(4.10)	(11.91)
	1	0	0	3.5	0	0	0	0
	10	10	10	30	10	10	10	30
Group 3	**6.56**	**5.56**	**5.82**	17.93	5.52	4.83	5.20	15.56
	(2.76)	(3.15)	(3.22)	(8.71)	(3.11)	(3.59)	(2.97)	(9.38)
	1	0	0	2.5	0	0	1	1
	10	10	10	29.5	10	10	10	30

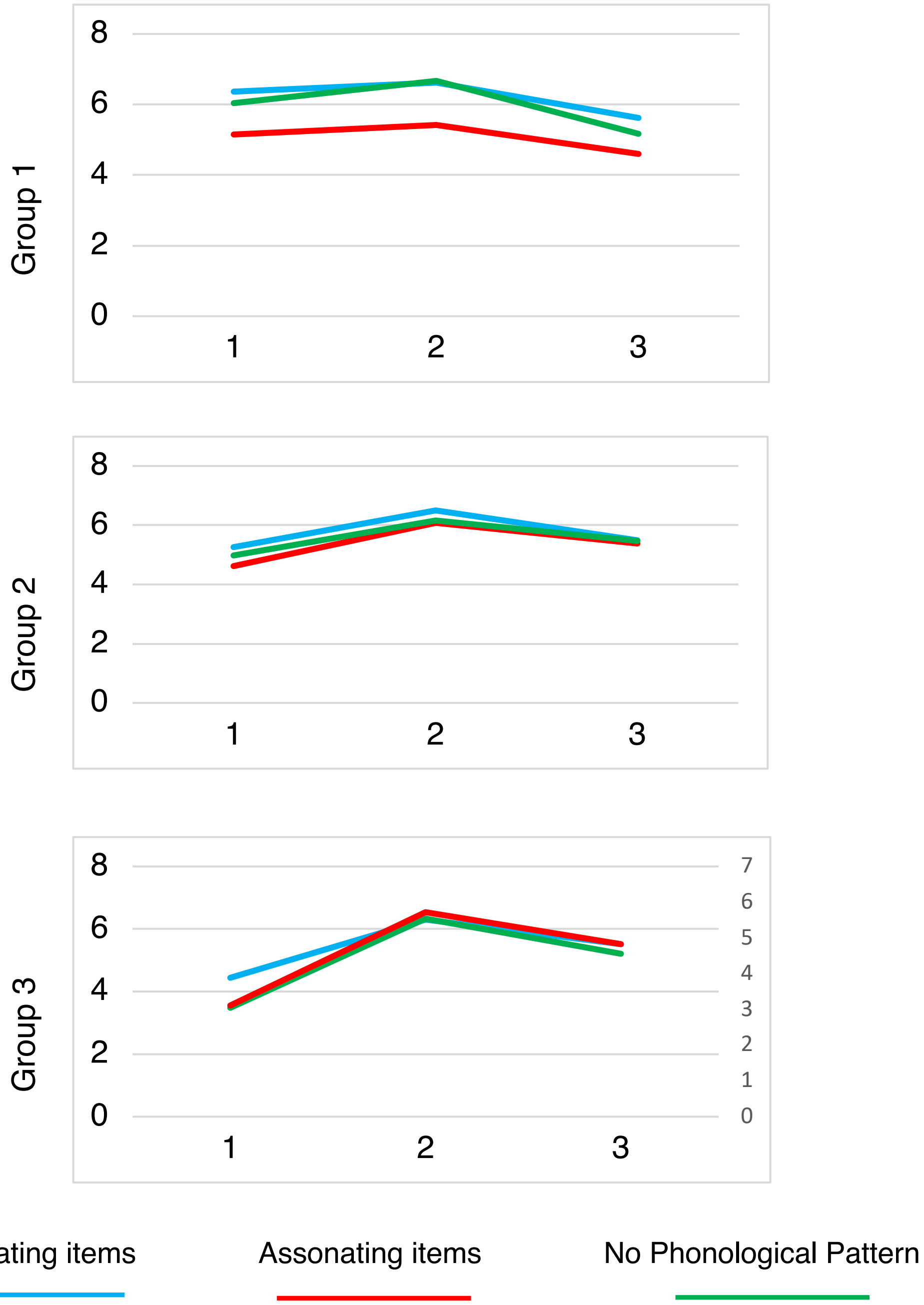
[1] Alliterating Items
[2] Assonating Items
[3] No Phonological Pattern Items
[4] Alliteration Highlighted
[5] Assonance Highlighted
[6] No Phonological Patterns Highlighted
[7] Mean score out of 10
[8] (Standard Deviation)
[9] Minimum score
[10] Maximum score

Figure 6.1 Mean scores for the Focus on Form test at three test times (1 - 3) for each group. [1]



[1] Group 1 had alliteration highlighted, Group 2 had assonance highlighted, Group 3 had no patterns highlighted.

Figure 6.2 Mean scores for the Focus on Meaning test at three test times (1 - 3) for each group. [1]



[1] Group 1 had alliteration highlighted, Group 2 had assonance highlighted, Group 3 had no patterns highlighted.

6.4 Analysis

A 3 x 2 x 3 mixed between-within Repeated Measures Analysis of Variance (RM ANOVA) was performed to compare the effect of treatment, test type, and phonological pattern on recall scores at each of the three different test times, to establish if any differences were being masked within the overall dataset.[4]

There were no statistically significant 3-way interactions between the variables of treatment, test-type, and phonological pattern, at any of the three test times[5] (see Figures 6.1 and 6.2 for a visual indication of this). In addition, the interactions between treatment and phonological pattern were not statistically significant at any of the three test times.[6] Given that this is the case, in what follows the test scores will be combined, since in general, any test items recalled at time three had also been successfully recalled at test times one and two. Figure 6.3 shows the mean scores from the combined tests (that is, the sum of scores for test times 1, 2 and 3), for each type of test item and for each experimental group. The same pattern of results can be discerned, namely, higher recall scores for alliterating items, then items with no phonological pattern, then assonating items.

To recap, Research Question 6.1 was: are English L2 learners better able to recall pseudoword-pairs that have a sound pattern (alliteration or assonance) than comparable pairs that do not? As this question has no attention-raising component, it can be answered using the data from the control group, Group 3, who received no

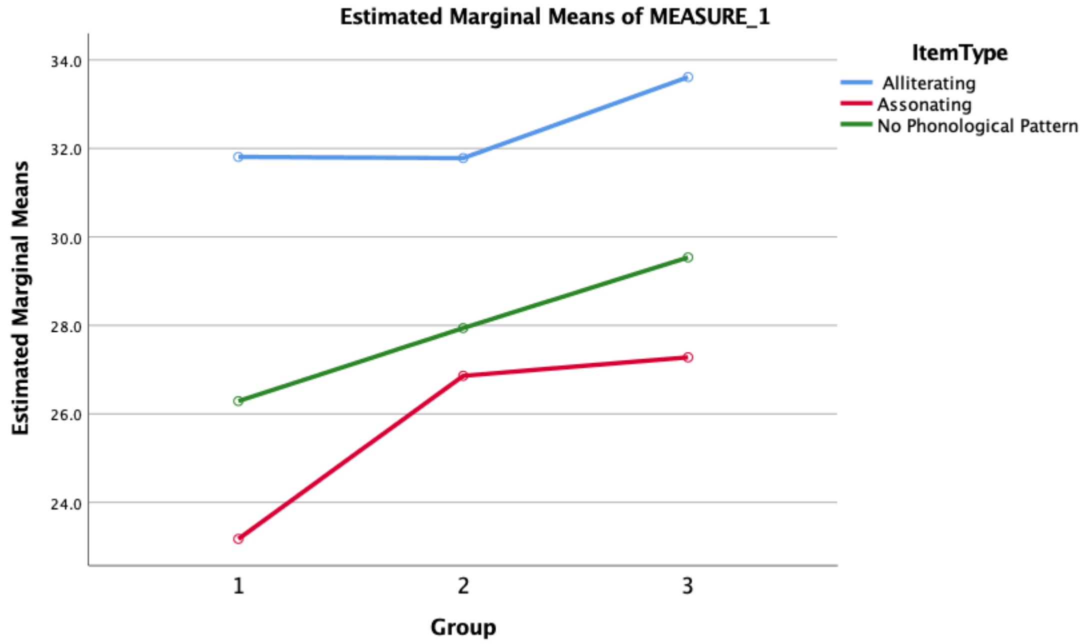
---

[4] All the ANOVA results in this experiment were found to be robust against model assumptions (homoscedasticity, normality of distribution, and sphericity) unless otherwise stated.

[5] Test time 1: ($F_{4,150}$ = .55, p = .70, η_p^2 = .01, power = .18). Test time 2: ($F_{4,150}$ = 1.90, p = .12, η_p^2 = .05, power = .60). Test time 3: ($F_{4,150}$ = 1.35, p = .25, η_p^2 = .04, power = .41).

[6] Test time 1: ($F_{4,150}$ = 1.27, p = .29, η_p^2= .03, observed power = .39). Test time 2: ($F_{4,150}$ = 2.16, p = .07, η_p^2= .05, observed power = .63). Test time 3: ($F_{4,150}$ = .95, p = .44, η_p^2= .03, observed power = .30).

Figure 6.3 Mean scores combined from test type (meaning and form) and test time, item type, and group. [1]



[1] Group 1 had alliteration highlighted, Group 2 had assonance highlighted, Group 3 had no phonological patterns highlighted.

information regarding the phonological patterns present in the stimuli. A One-Way RM ANOVA was used to compare the mean scores collected from the different conditions. For Group 3 (n = 27), the means and standard deviations can be seen in Table 6.6.

There was a significant main effect of Pseudoword-pair Type on the items that participants recalled ($F_{2, 52} = 15.40$, $p < .001$, $\eta_p^2 = .37$, power = .9). Pairwise post hoc comparisons using t-tests and Bonferroni corrected levels of significance showed that participants recalled significantly more stimuli from the Alliterating

Table 6.6 Group 3 (No Patterns Highlighted): mean and standard deviations for recall across all tests.

Pseudoword-pair Type	Mean	*SD*
Alliterating Items	33.61	13.25
Assonating items	27.28	15.93
No Phonological Pattern Items	29.54	14.27

condition compared to the Assonating condition (mean difference = 6.33, $p < .001$) and the No Phonological Pattern condition (mean difference = 4.07, $p = .002$). However, there was no statistical significance in the comparison between the Assonating condition and the No Phonological Pattern condition (mean difference = -2.26, $p = .203$). Thus, in answer to Research Question 1, the presence of alliteration, but not assonance, significantly improved the recall of the stimuli for the participants in the control group (Group 3).

Research Question 6.2: Is a given stimulus more memorable if learners have had the sound pattern drawn to their attention than if they have not?

What was the effect of having the sound patterns highlighted (or not)? Participants assigned to Group 1 (n = 26) had alliteration highlighted on the word-cards. The means and standard deviations can be seen in the Table 6.7.

Table 6.7 Group 1 (Alliteration Highlighted): mean and standard deviations for recall across all tests.

Pseudoword-Pair Type	Mean	*SD*
Alliterating Items	31.81	13.09
Assonating items	23.17	14.12
No Phonological Pattern Items	26.29	14.13

A One-Way RM ANOVA was used to compare the mean scores collected from the different conditions. Mauchly's test of sphericity again showed that the assumption of sphericity was met (χ^2 (2) = 6.18, $p = .73$) and again there was a significant main effect of Pseudoword-pair Type on the items that participants recalled and a large effect size ($F_{2, 50} = 30.95$, $p < .001$, $\eta_p^2 = .55$, power = 1.0). Bonferroni post hoc tests showed that participants recalled significantly more stimuli from the Alliterating condition compared to the Assonating condition (mean difference = 8.64, $p < .001$) and the No Phonological Pattern condition (mean difference = 5.52, $p < .001$). There was also a statistically significant difference in the comparison between the Assonating condition and the No Phonological Pattern condition (mean difference = -3.12, p = .04).

Therefore, regarding alliteration in Group 1, drawing the participants' attention to the pattern was associated with statistically greater recall of the items. This seems to be a convincing result in its own right, but in light of the control group's similar advantage, it cannot be concluded that the reason for Group 1's superior performance for alliterative items was because attention was drawn to them.

For the participants in Group 2, highlighting the assonance appeared to have no beneficial effect on recall as alliterating items were recalled more frequently, see Table 6.8.

Table 6.8 Group 2 (Assonance Highlighted): mean and standard deviations for recall across all tests.

Pseudoword-pair Type	Mean	*SD*
Alliterating Items	31.78	16.23
Assonating items	26.86	17.89
No Phonological Pattern Items	27.94	18.49

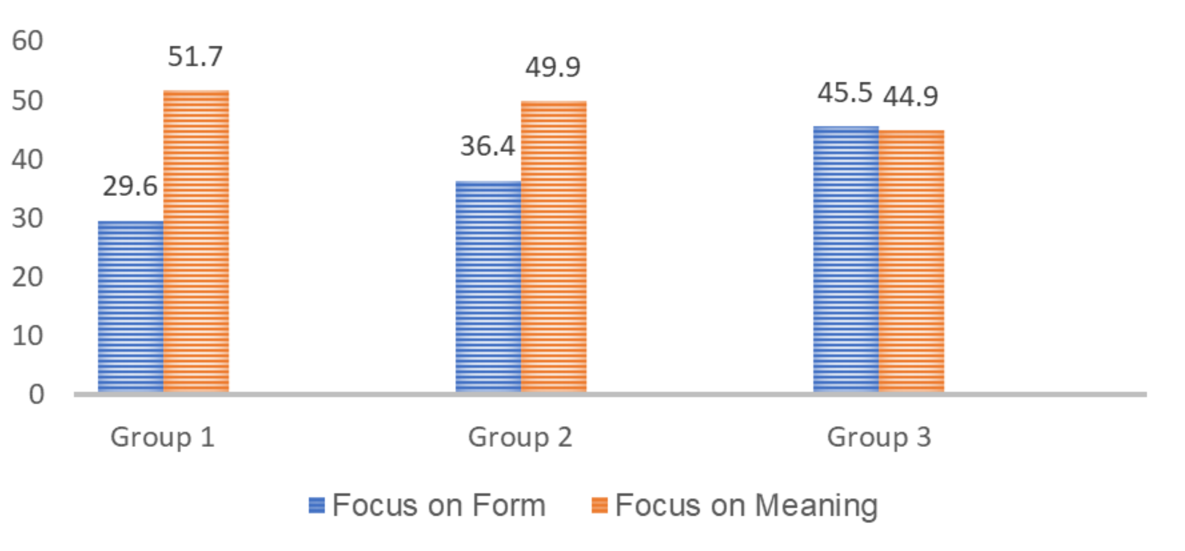
A One-Way RM ANOVA showed a significant main effect of Pseudoword Type on the items that participants recalled (F (2, 48) = 18.36, p < .001, η_p^2 = .43, power = 1.0). Bonferroni post hoc tests showed that participants recalled significantly more Alliterating items than Assonating items (mean difference = 4.92, p < .001) and the No Phonological Pattern items (mean difference = 3.84, p = .001). There was no significant difference in the comparison between the Assonating condition and the No Phonological Pattern condition (mean difference = -1.08, p = .77). This corroborates the previous finding that alliteration is memorable all on its own, irrespective of any awareness-raising task.

As noted already, the participants who had neither type of pattern drawn to their attention performed better than those who did. A closer look at this finding reveals no statistically significant difference in the recall scores of Group 3 compared to Groups 1 and 2. A mixed 'between-within' RM ANOVA with treatment as a between-groups independent variable and the recall scores for the different test items (Alliterative, Assonant, No Phonological Pattern) as a within-groups independent variable, showed no statistically significant interaction between Group and Phonological Pattern ($F_{4,150}$ = 1.58, *p* = .18, η_p^2 = .04, power = .48). In sum, drawing the participants' attention to the phonological pattern in question has a negligible mnemonic effect.

RQ 6.3: Do phonological patterns affect the recall of the form and recall of the meaning differently?

As previously noted, the test instrument was divided into two parts. The first was a written cued-recall test (the Focus on Form test) in which the participants were asked to supply the corresponding pseudo-noun for a given pseudo-adjective prompt. The second, the Focus on Meaning test, asked them to write the Japanese translations which had been allocated to the 30 pseudoword-pairs. Figure 6.4 shows the mean scores from the two parts of the test instrument, across all three test times, for the three different experimental groups.

Figure 6.4 Mean scores for form and meaning from the three combined tests.



It can be seen in Figure 6.4 that participants who had the phonological patterns brought to their attention performed better when they had to write the L1 meaning compared to when they had to recall the pseudoword. For Group 3 the difference was negligible. Starting with the combined results from the three Focus on Form tests, descriptive results can be seen in Table 6.9:

Table 6.9 Descriptive statistics for the focus on form results (3 tests combined).

Test	Group 1[2] (n = 26)		Group 2 (n = 25)		Group 3 (n = 27)		Total (N = 78)	
Items[1]	*M*	SD	*M*	SD	*M*	SD	*M*	SD
L	13.2	7.39	14.5	6.89	17.4	6.35	15.1	7.02
S	8.0	7.92	10.8	7.57	13.6	7.64	10.8	7.96
N	8.4	7.75	11.3	8.41	14.5	7.47	11.5	8.18

[1] L = Alliterating Items, S = Assonating Items, N = No Phonological Pattern Items
[2] Group 1 alliteration highlighted, Group 2 assonance highlighted, Group 3 no phonological patterns highlighted

A repeated-measures ANOVA was performed with the Focus on Form scores for the 3 types of experimental item as the within-groups independent variable, and group assignment as the between-groups IV. Mauchly's test of sphericity showed that the assumption of sphericity was met (χ^2 (2) = 1.09, p = .58). However, there was no statistically significant main effect of Group on the participants' ability to recall the form of the target items (F (4, 150) = 1.50, p = .21, η_p^2 = .04, power = 0.45).

Regarding the results from testing the meaning of the target items, descriptive results can be seen in Table 6.10.

Table 6.10 Focus on meaning results (3 tests combined).

Test	Group 1[2] (n = 26)		Group 2 (n = 25)		Group 3 (n = 27)		Total (N = 78)	
Items[1]	*M*	SD	*M*	SD	*M*	SD	*M*	SD
L	18.6	7.41	17.3	9.86	16.3	7.99	17.4	8.41
S	15.2	7.39	16.0	10.82	13.6	8.86	14.9	9.04
N	17.9	8.02	16.6	10.73	15.0	7.64	16.5	8.83

[1] L = Alliterating Items, S = Assonating Items, N = No Phonological Pattern Items
[2] Group 1 alliteration highlighted, Group 2 assonance highlighted, Group 3 no phonological patterns highlighted

A repeated-measures ANOVA was performed with the Focus on Meaning scores as the within-groups independent variable, and group assignment as the between-groups IV. Mauchly's test of sphericity again showed that the assumption of sphericity was met (χ^2 (2) = 1.39, p = .49). Yet again, there was no statistically significant main effect of Group on the participants' ability to recall the meaning of the target items (F (4, 150) = 1.82, p = .13, $\eta_p{}^2$ = .05, power = 0.54). These results indicate that assigning participants to a condition in which the phonological pattern was highlighted or was not, had no statistically significant effect on their ability to recall either the form or the meaning of the target item.

In sum, there was a difference in the recall of the form and meaning of the pseudoword-pair stimuli according to their type (alliteration > no pattern > assonance), and this difference spans all groups including the controls, irrespective of any attention-raising measures.

6.5 Discussion

In looking for explanations, we can consider both inherent differences from the previous research that underpinned the hypotheses, and possible unintended confounds in the present data.

The present results show that the alliterating items were easier to recall at all three test times for all three groups, though, to recap, only the participants in Group 1 had access to flashcards with the alliteration highlighted. This is interesting, given that previous experiments with L2 learners usually found a mnemonic effect of phonological patterns only after explicit awareness-raising. A further finding of note is that the assonating pairs were recalled less than the pairs in the No Pattern condition, a finding which seems hard to reconcile with previous experimental data on the mnemonic effect of assonance.

There are several possible reasons for this. Firstly, it could be that assonance is not a reliable pattern for this mnemonic effect and the results from previous studies are impacted by confounding variables such as familiarity.

Another possible explanation could be that there was ‘leakage’ of information from Group 1 to Groups 2 and 3; perhaps some participants saw the flashcards with the highlighted alliterative patterns, though this would not explain the results of Test 1, when participants had had no opportunity to see what other groups had access to on the Quizlet website.

A third potential explanation could be that the marking scheme penalised spelling errors on assonating items more harshly than items with no phonological pattern (examples can be seen in Table 6.4). A single spelling error in an assonating item might well annul the pattern, and thus be given a score of zero, while a single spelling error in an item with no sound pattern was given a half mark. Although this might go some way in explaining the differences between assonating and no pattern items, it does not account for the alliteration advantage, as alliterating items were also marked more severely.[7]

The test items in the No Pattern condition were, on average, the shortest in terms of the number of mora (in the Japanese translations), and were rated as more concrete (see Table 6.1). This was considered acceptable as it favoured the null hypothesis. This slight advantage could have neutralised any marginal advantage for assonance.

Another potential explanation lies in the fact that participants in Group 2 logged the least amount of flashcard study time (see Table 6.3 above), though as noted previously, the differences in mean study time were not statistically significant. One factor that could not be controlled for was that students could have copied the target items onto paper and studied from this in lieu of using the flashcard app (this was attested by the researcher in one case from a participant in Group 3). This would have impacted on the data in two ways: the reliability of the information about time spent on personal study, and the results of the unannounced Test 3. The smaller

[7] Recalculating the scores based only on completely correct items did not produce enough data as there were too many zero scores.

amount of study time by Group 2 does not, however, account for the Test 1 results, nor for the fact that both Groups 1 and 3 also recalled fewer assonating pseudoword-pairs compared to No Pattern items.

Was assonance more difficult to perceive, despite the highlighting for Group 2 participants? One assonating pair had dissimilar orthography, *nirk serm*, and this could definitely have had a negative impact on the recall scores, particularly if the phonological effects are, in fact, at least partly orthographical (see section 7.4). In addition, perhaps the greater number of consonant digraphs in the assonating list of items made them more difficult to recall and / or spell correctly in the Focus on Form cued-recall tests.

Another possible disadvantage for the assonating items is connected to how people perceive written words: Aitchison (1987) cites evidence for the 'bathtub' effect, whereby people can remember the beginning or ending of words when induced into metacognitive Tip-of-the-Tongue states. The 'bathtub' effect indicates that the beginnings and endings of a lexical item tend to be more prominent in the mental lexicon, but would this phenomenon apply with Japanese L1 speakers? Akamatsu's (2003) findings suggest that L1 logographic readers do not analyse L2 intraword components as carefully as alphabetic readers do, though Japanese L1 speakers are both logographic readers of kanji (characters) and syllabic readers of Hiragana and Katakana (syllabaries).

One final point to raise is the possibility that despite the random division of participants into groups, socio-psychological factors were not equally matched. It could be that the participants in Group 3 had greater motivation, more positive attitudes and task engagement, and / or less language anxiety than the participants in the other treatments. This notion will be returned to in Chapter 7.

Perhaps the most likely explanation for the overall results is that drawing the participants' attention to certain features induced a learning burden and was in fact a distraction from the task at hand, that of making novel form-meaning links in memory. If so, there is a potential lesson here for language teachers, whose keenness to intervene in the natural cognitive processes of learning might in fact generate problems.

It is possible that asking participants to learn a new app plus a new vocabulary learning technique was challenging enough, and so drawing the participants' attention to a phonological pattern (for Groups 1 and 2) may have imposed too much of a cognitive load and / or proved distracting from the task of memorising the stimuli. Group 3 did not have this extra burden and thus could perhaps focus more on the task of learning the novel items. As for why there was little mnemonic effect for the assonating items, one possibility is that the pattern was more difficult to discern for the Japanese L1 speakers, despite it being highlighted: Akamatsu's (2003) finding that L1 logographic readers are less sensitive to L2 intraword information compared to L1 alphabetic readers may have some bearing here. The use of Kanji on the reverse of the electronic flashcards would entail recognising the whole character and retrieving the complete L1 phonological form from memory, as the series of strokes in a character represent neither meaning nor pronunciation. In contrast, the pseudoword-pair on the other side of the flashcard would require analysis of constituent graphemes. One could speculate that such cross-linguistic differences, rooted in the nature of the L1 and L2 orthographies, mean qualitative differences in the recognition processes for each side of the flashcard; if the dominant process is visual recognition for the logographs then could this impinge on the processing of the intraword information and L2 word recognition? If this conjecture has any substance, it would predict better scores for the information on the logographic side of the word-card (the Focus on Meaning tests) than the pseudoword-pair side (the Focus on Form tests). The answer to Research Question 6.3 does seem to bear this out for Groups 1 and 2 though the differences were not statistically significant.

6.6 Conclusion

In sum, the answers to the research questions posed were inconclusive. But insofar as the finding that alliteration was easier across the board is reliable and replicable, it raises some interesting new questions. Placed in the context of the full set of investigations reported so far, it seems clear that what might have seemed a rather straightforward capacity for phonological patterns to support learning and recall is far from simple. Chapter 7 thus takes a step back and considers some broader questions around the mnemonic effect of phonological patterns.

Chapter 7 Taking stock

7.1 Introduction

The findings so far raise a number of interesting questions, and in this chapter, six are explored. Some of these questions in turn will involve the reporting of subsidiary data that were not relevant to the respective chapters but need to be addressed as matters arising.

First of all, in Chapter 3, it was noted that some researchers (such as Kubozono & Ota, 1998; Otaka, 2009; Tamori et al., 2008) have claimed that the influence of the mora in Japanese phonology will disguise any alliterative or assonating patterns in English for a Japanese L1 speaker. However, the Japanese participants in the experiments reported in this thesis successfully recognised such phonological patterns on the whole. This begs the question, why did the moraic nature of the participants' L1 not mask the phonological patterns, as suggested by these researchers? This will be addressed in section 7.2.

The studies in Chapters 3 and 4 explored whether word-pairs were easier to recall if they had alliteration or assonance. However, the stimulus words were likely to be known to the participants already. Therefore, section 7.3 considers whether pairs of real words that were unknown to participants would give the same outcome.

In the phonological pattern identification tasks, the accuracy rates were generally quite high, but there are many instances where the role of orthography can be called into question. Section 7.4 thus looks at the possibility that spelling might affect people's perception of alliteration and assonance and, more broadly, what role literacy might play in perceiving phonological patterns.

Section 7.5 returns to an issue brought up at the beginning of this thesis: in defining a phonological pattern, what counts as a 'similar' sound? This section will explore three topics: a) whether phonological similarity is a binary yes / no distinction, b) if the construct of similarity can be graded according to sub-phonemic features, and c) whether the phoneme is a reliable construct for looking at alliteration and assonance.

Further questions focus on a methodological theme that runs throughout the studies, not only in this thesis but also in those by Boers et al. reviewed in Chapter 2: in looking for evidence of a mnemonic effect, participants were tested at various times post-treatment. Section 7.6 considers what influence the timing of the tests might play in the findings so far and what this signifies in terms of permanent learning gains. More broadly, this section also considers to what extent the underlying assumptions about brand name memorability, which are prevalent in the field of advertising, and the mechanisms that underpin it, are applicable to the L2 classroom. Finally, Section 7.7 will consider what role might be played by variation in natural abilities between individual participants.

7.2 Why did the moraic nature of the participants' L1 not mask the phonological patterns, as suggested by Otaka (2009)?

In general, the majority of the Japanese L1 participants in the experiments could identify the alliteration (90.7% of the time in Chapter 3) and the assonance (almost 90% in Chapter 4) in the word-pairs. Why were the findings from Otaka (2009) not replicated? In Otaka (2009), the Japanese L1 undergraduates had very low rates of accurately identifying phonological patterns in both Japanese and English: for example, none of the 199 participants indicated that *Matsuoka Misako* alliterated, and only 18 participants identified the alliteration in *Mickey Mouse*. Accuracy rates tended to be far higher when there was some form of CV repetition: 60% of the participants successfully noted the pattern in *Makino Makiko*, and 50% noticed the pattern in the phrase *kill the king*. As stated previously, Otaka suggested that the moraic[1] properties of the L1 were masking the patterns in the target word-pairs, particularly when the vowel nucleus was dissimilar (see Chapter 3).

Why were the participants in the experiments outlined in Chapters 3 and 4 better able to identify the phonological patterns than the participants in the earlier Otaka (2009) study? One possible explanation is that the participants in Chapters 3 and 4 had a greater degree of phonological awareness of English (if phonological awareness is defined as the ability to identify and discriminate the characteristics of sounds, including sound similarities). Why might this be so? Much has been written about the "massive influx" of English into Japanese (I. Taylor & Taylor, 1995, p. 290), as a result not only of loanwords but also more frequent use of the Roman alphabet in the mass media and popular music. For example, Holmquist & Cudmore's (2013) analysis of 1,598 advertisements from 30 Japanese magazines across a range of genres found 96% used English in some form: almost three quarters of the company names, and almost half of the product names, were written in English. In a similar vein, there is an extensive literature on the increase in the use of English in the lyrics of Asian popular music, especially Japanese J-pop, Korean K-pop, and Cantopop from China. Moody (2000) analysed J-pop lyrics from the Top 50 chart of that year and found 62% of the songs used English, ranging from fillers (for example, *yeah*) to entire sentences.

Detey & Nespoulous (2008, p. 68) go so far as to suggest that this increase in the use of English has led to an "evolution of the Japanese phonological lexicon", at least in urban areas such as Tokyo.[2] It is difficult to systematically assess such a premise as there is little longitudinal quantitative evidence of this increase. However, it is also worth bearing in mind that the Otaka study preceded the widespread adoption of social media and the penetration of the biggest platforms into Japanese

[1] To recap, a mora is a sub-syllabic rhythmic unit; a prototypical mora consists of a single consonant plus one short vowel.

[2] And presumably, by extension, the Kansai conurbation around Osaka, where the participants in the replication study (Chapter 4) resided, but also where the Otaka (2009) study was conducted.

society. The most popular social media app in Japan, LINE, was introduced in 2011, and Instagram in 2014. According to the Pew Research Center (Poushter, Bishop, & Chwe, 2018), 82% of young Japanese adults (ages 18 to 36) are on social media. Taking all of this into account, any growth in the participants' exposure to English in their L1 environment may be one possible explanation for their greater phonological awareness compared to those participants in the Otaka study a decade earlier.

Another explanation as to the disparity in identifying phonological patterns may be found in the experimental design of the Otaka (2009) study. According to the published procedure, the participants were given a "short lecture about rhymes" (Otaka, 2009, p. 12), presumably in Japanese,[3] before being asked to identify patterns (operationalised to cover rhyme, alliteration and assonance) in 30 Japanese phrases – it could be argued that at this point in the procedure, the participants are in 'mora-mode', that is, their cumulative learned experience with the Japanese writing system predisposes them to a moraic C+V segmentation strategy, as was suggested in Chapter 3. When they are later asked to identify patterns in ten English phrases there might therefore have been a 'mora effect' which carries over and biases their perception of patterns. In contrast, the participants in the experiments in Chapters 3 and 4 were in an L2 classroom environment where the language of communication was solely English, the experiment was conducted predominantly in the L2, as were the examples of alliterating and assonating word-pairs. Thus, perhaps these participants were not in 'mora mode', as the mora was not psychologically salient at the time, so there was no 'mora effect'.

What could be done to explore this issue further? Although it is beyond the scope of this thesis, it would be interesting to run an alliteration or assonance experiment with a group of participants whose phonological abilities have not been shaped by their exposure to their L1 writing system. This could be done with either a group of Japanese pre-school children or, better still, with monolingual illiterate[4] Japanese

---

[3] There is no information given regarding the participants' English proficiency.

[4] The United Nations Educational, Scientific, and Cultural Organisation (UNESCO) defines an illiterate person as an adult who cannot read or write a single short sentence consistent with their daily life (UNESCO, 1997). However, it should be noted that the concept of literacy

adult speakers. If illiterate adults, with no formal literacy training in the mora-based kana syllabary, are not conditioned into linking a word initial consonant with the following vowel, then perhaps they would easily perceive an alliterating pattern in the absence of any 'mora effect'. There is ample evidence in the literature that levels of literacy can affect phonological awareness. In one early example, Morais, Cary, Alegría and Bertelson (1979), the researchers compared illiterate and literate Portuguese adults on a battery of metaphonological tasks, such as deleting the initial phoneme from a word or inserting one into it. Illiterate participants underperformed in such tasks compared to the literate participants. The authors held that awareness of speech as a sequence of units is strongly dependent on learning to read and write. There is little available data on Japanese illiterates, but there is some evidence for other logographic readers: Read, Yun-Fei, Hong-Yin, & Bao-Qing (1986), using the same types of phonological manipulation tasks as Morais et al. (1986), found that Chinese L1 participants who had learned to read alphabetic *pinyin* (the romanisation system for Standard Mandarin Chinese) in addition to the traditional logographic characters outperformed participants who only knew the characters. This sort of investigation could help to interpret what are currently rather ambiguous findings.

7.3 What role does the level of familiarity with stimuli play in patterns of recall?

As pointed out at the start of Chapter 6, when designing recall experiments of the kind used here, one important decision regards whether or not to use stimuli the participants are likely to already know, as frequency and familiarity of stimuli can strongly affect memory recall (see section 2.3.1). In the previous published research,

---

is not a simple dichotomy; there are many who would be considered 'literate' according to the UNESCO definition yet display great difficulties in reading and writing in a predominantly automatic way (see Morais & Kolinsky, 2019 for further discussion). For the purposes of this experiment, it would be useful if the participants were fully illiterate but finding a suitable population would be challenging in Japan due to its high levels of literacy.

high frequency words were typically used, so that the participants would recognise the individual items, but not necessarily the pairing. This meant that the experiment was not focussed on learning new phonological forms and their previously unknown meanings, so much as learning to associate two known forms and retrieving the pairing from episodic memory.

Accordingly, in the experiments reported in Chapters 3 and 4, when compiling lists of familiar word-pairs, monosyllabic noun stimuli were chosen from the first 2000 word families of the New General Service List (Browne, 2014), and the monosyllabic adjectives were chosen based on high corpus frequency and Mutual Information scores. Based on the participants' L2 proficiency levels and their scores on the first 2K of the Vocabulary Levels Test, it was deemed likely that the experimental word-pairs would be familiar to them. These assumptions were confirmed by the results of subjective familiarity-rating tasks incorporated into the procedures.

In the experiments reported in Chapters 3 and 4, immediately after input and also later, participants more successfully recalled word-pairs that alliterated or assonated than matching word-pairs that did not, though not all differences were statistically significant. The general pattern of results in the experiments which used familiar word-pairs can be seen in Table 7.1 (the test time aspect will be discussed in section 7.6).

Because word-knowledge is multi-faceted, it may be hard to gauge if a word is completely 'familiar' and how well it is anchored in the learner's L2 lexicon. If familiarity can be seen as a continuum, with high-frequency familiar word-pairs at one end, then at the other extreme we can situate pseudoword-pairs such as those

Table 7.1 Results with familiar word-pairs.

Chapter	Pattern	Test Time[1]	Results	Statistically Significant	Effect Size
3	alliteration	Immediate free recall	alliterating > non-alliterating	No	
		1-week delayed free recall	alliterating > non-alliterating	Yes	r = .42
		1-week delayed cued recall	alliterating > non-alliterating	Yes	r = .47
4	assonance	Immediate free recall	assonating > non-assonating	Yes	r = .32
		1-hour delayed free recall	assonating > non-assonating	Yes	r = .41
		2-day cued recall	assonating > non-assonating	Yes	r = .50
		1-week delayed cued recall	assonating > non-assonating	No	

[1] all tests were unannounced.

used as stimuli in the experiment reported in Chapter 6. In Chapter 6, it was suggested that the low recall scores could be accounted for on the grounds that the experimental task (to memorise 30 pseudoword-pairs in 25 minutes) was too cognitively demanding. This raises the question of whether or not the mnemonic effect of a phonological pattern applies to word-pairs in the middle-ground between the two extremes of highly familiar word-pairs and pseudoword-pairs.

As noted in Chapter 2, only two previously published studies had used pairs of real but unfamiliar words: Study 10 in Table 2.4 (Eyckmans et al., 2016), and Study 11 (Eyckmans & Lindstromberg, 2017). Although both studies reported a mnemonic advantage for those test items that displayed a phonological pattern, concerns were expressed in section 2.3.5 as to how robust their evidence was. These concerns included questions regarding the suitability of the test tool used in Study 10 by Eyckmans et al., which was a cloze exercise with a heavy lexical burden considering the low L2 proficiency of the participants. Regarding Study 11, it was argued in Chapter 2 that the format of the test tool meant that it could be completed using form-based knowledge only, without recourse to recalling the meaning of the idiomatic items.

In effect, both of these studies assessed participants' ability to produce the written form of the test items rather than any semantic features, leaving open the question of whether a phonological pattern can aid recall of a previously unknown meaning.

As it happens, data was collected as part of the study on alliteration reported in Chapter 3, and it will be useful to briefly explore that data here, as a means of gaining some insights into this question.

7.3.1 Materials

In choosing unfamiliar word-pairs, a pool of noun lemmas was taken from the last third[5] of the New Academic Vocabulary List (AVL) (Gardner & Davies, 2013). While it would be simplistic to assume that all the participants would know items only in the first 2K list and none outside of it, it was judged likely that their familiarity with words in this part of the AVL would be much lower. The basis for this judgement was that they had been tested on the Academic Word List section of the Vocabulary Levels Test and had scored a mean of 22.4 out of 30, beneath the threshold for 'mastery', generally defined as 27 / 30 (Nation, 1990, p. 143). The nouns from the AVL were orthographically longer and polysyllabic, but were all stressed on first syllable, as this is the most frequent stress pattern in English (Cutler & Carter, 1987).

Whole-phrase corpus frequency counts and Mutual Information scores were used to select suitable adjectival collocates (essentially the same procedure as described in section 3.3.2), in an effort to balance the items for potentially confounding variables that can affect recall. These constraints meant that it was not possible to find enough collocates that the participants would find unfamiliar, a methodological issue that will be discussed presently. The final pool of twenty target word-pairs was therefore composed of adjectives that would most likely be familiar,[6] paired with an unfamiliar noun, see Table 7.2.

---

[5] That is, the last 1,000 lemmas in the 3000 lemma AVL.

[6] The only adjectives beyond the 2K frequency level of the BNC-COCA corpus are: *radical, moral, global, negative, primary* (3K), and *sacred* and *precious* (4K). In the alliterating list, 6 adjectives are within the 2K band, and in the non-alliterating list 7 adjectives are within the 2K band.

Table 7.2 Target word-pairs.

Alliterating	Non-Alliterating
modern mores	moral betterment
sacred stricture	global affluence
famous forebear	negative prefix
precious progeny	broad panoply
monthly metrics	basic postulate
radical rupture	cultural dictate
certain synergy	primary funder
timeless truism	prime requisite
huge hindrance	simple dictum
quiet consistency	sudden upsurge

Again, the selection was guided by the aim to balance any potentially confounding variables, as seen in Table 7.3.

The inter-rater reliability for the concreteness and translatability ratings was lower than for the familiar word-pairs reported in Chapter 3 (Cronbach's alpha = .77 and .60 respectively), though this could be due to variation in the English L1 raters'

Table 7.3 Unfamiliar word-pairs: features that affect recall.

	Alliterating word-pairs Mean	Non-alliterating word-pairs Mean
Orthographic Length (letters)	13.8	13.7
COCA Phrase Frequency (per million words)	2.6	4.2
Mutual Information Score	5.9	5.5
Concreteness Rating [1]	14.9	15.8
Translatability Rating [2]	13.4	17.1

[1] A minimum score of 6 = most abstract, maximum 30 = most concrete

[2] A minimum score of 6 = most difficult to translate, 30 = easiest to translate

own familiarity with the low-frequency AWL nouns. According to Table 7.3, the ratings do slightly favour the non-alliterating word-pairs in terms of recall likelihood. However, this discrepancy was not held as a confounding variable on the grounds that if the alliterating word-pairs were better recalled from episodic memory, that would be despite them being slightly more abstract and less frequent.

Finally, a set of twenty vocabulary cards was made for each participant,[7] with the English word-pair on one side and the Japanese translation on the other. All participants had received prior training on using the word card strategy (c.f. Nation, 1990) and had had several weeks of practice with the technique using lexical items from the assigned EFL textbooks in their undergraduate classes.[8]

7.3.2 Procedure

The classroom-based procedure was the same as the one described in Section 3.3.3: firstly, there was a dictation in which the teacher said each word-pair aloud twice and the participants were told to repeat it sub-vocally. Then there was the identification task in which participants were asked to tick either 'Yes' or 'No' on the test paper according to whether the phrase alliterated or not. Participants then wrote down the word-pair on the test paper. The word-pair then appeared on the projector screen at the front of the classroom and the participants ticked a box to indicate if they had spelled it correctly. Finally, participants ticked a rating scale according to how familiar the phrase was to them (1 = *I know how to use this phrase*, 2 = *I can guess how to use this phrase*, 3 = *I don't know how to use this phrase*).

This was followed by an immediate free-recall test. Although this test, and all subsequent tests, were unannounced, some of the more cognisant participants may have expected them, as this data collection followed the high-frequency word-pair data collection the participants had participated in earlier in the semester. One difference in the procedure, however, is that following the immediate free-recall test,

---

[7] The same cohort of Japanese L1 participants outlined in Chapter 3.3.1 (N = 124).

[8] None of the target nouns from the AVL appeared in these materials.

the participants were given the pack of word cards with the twenty target collocations and Japanese translations and asked to learn them. No further instructions were given. Although this introduced the uncontrolled variable of how much time was devoted to learning the items, in principle any amount of time would be equally divided between alliterating and non-alliterating word-pairs. Approximately three weeks later, there was an unannounced free-recall test followed by a cued-recall test. This was then followed by a productive translation test; the participants were given the Japanese translations and asked to write the word-pair in English. The recall tests adopted the same marking protocol[9] as that used with the high-frequency word-pairs. However, bearing in mind the proficiency level of the participants, answers were marked as correct if they had no more than three errors per word-pair for the longer polysyllabic low-frequency word-pairs. For example, *presious propony* (target: *precious progeny*) with three substitutions was accepted, whereas *precious panocy* with one deletion and four substitutions was not.

7.3.3 Results

Results of the identification task

For the unfamiliar word-pairs, the majority of the participants correctly identified the presence or absence of a phonological pattern in the twenty word-pair stimuli (mean = 17.6 out of 20, min = 8, max = 20, *SD* = 2.36). The results of this identification task were very similar to those studies which used familiar test items. However, there were considerably more cases of participants not noticing alliteration (270 instances) than of mistakenly labeling it as such (33 instances), as seen in Figures 7.1 and 7.2 respectively.

---

[9] Homophones were accepted as correct, as were all instances of /l/ and /r/ phoneme substitution, grapheme substitutions, insertions, and deletions.

Figure 7.1 Recognition tally for alliterating word-pairs.

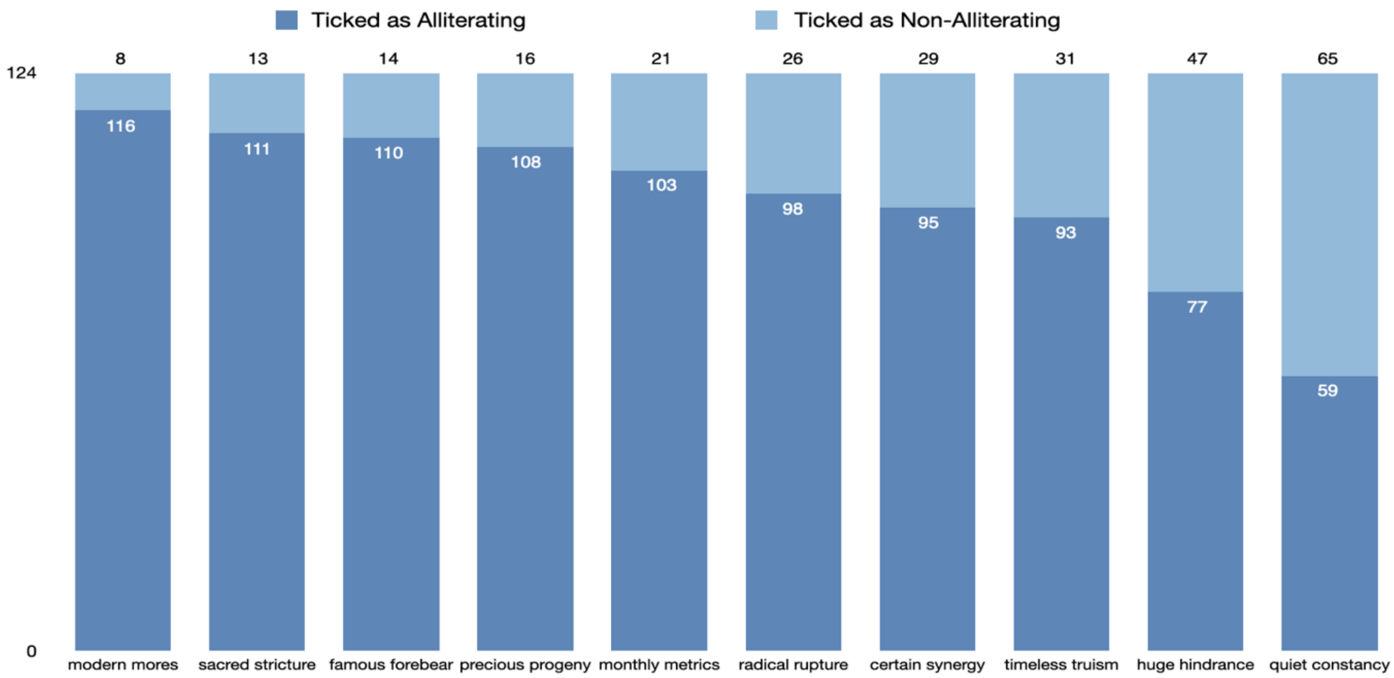
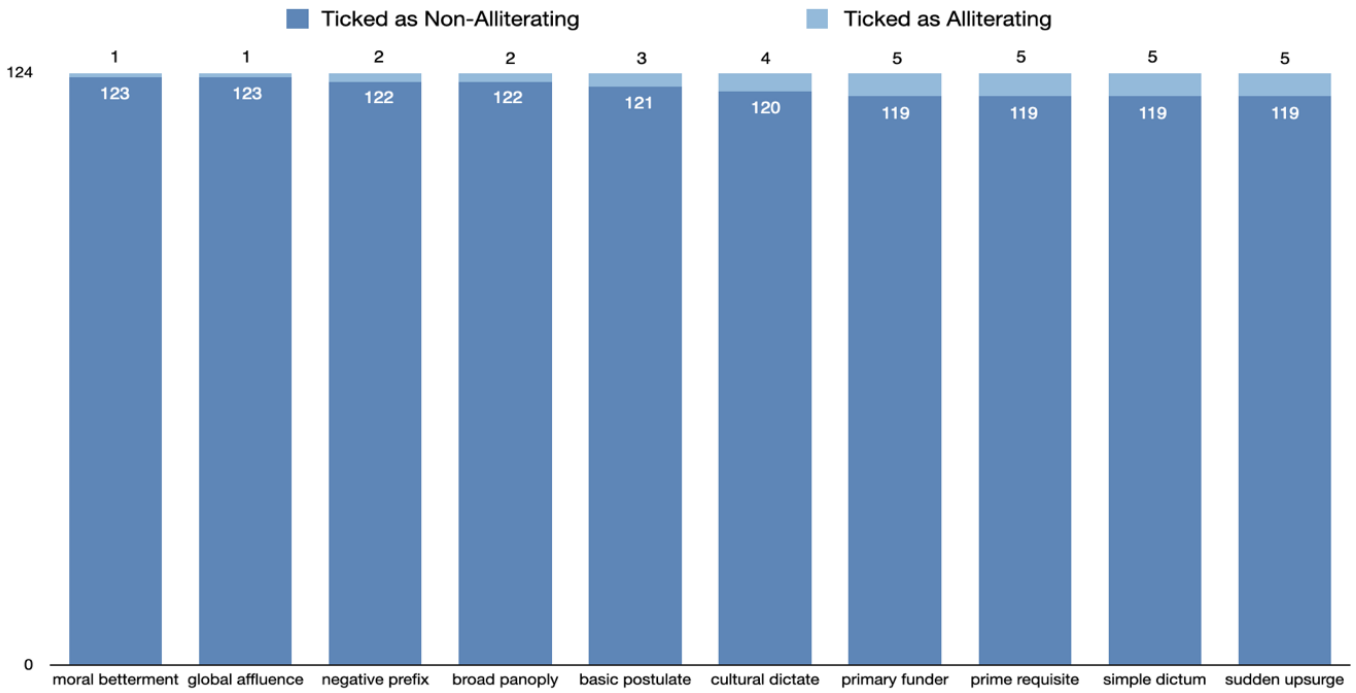


Figure 7.2 Recognition tally for non-alliterating word-pairs.



It could be argued that the orthography is an aggravating factor, especially where the phonological pattern is masked by non-prototypical spellings seen in phrases such as *quiet constancy*[10] (/k/ – /k/) and *certain synergy* (/s/ – /s/), a point that will be considered in Section 7.4.

Results of the familiarity task

Because each participant rated every word-pair on a scale from 1 to 3 (1 = *I know how to use this phrase*, 3 = *I don't know how to use this phrase*), an indication of unfamiliarity would be if the mean score were higher than the halfway point (248) and closer to the maximum 372 (124 participants x 3). For these low-frequency word-pairs, the mean score was 354.1 (SD = 15.1), with *primary funder* reported as the most familiar (320 points), and *precious progeny* the least familiar (370 points). The alliterating mean score was 357.2 (min = 336, max = 370, SD = 12.8) and the non-alliterating mean was 351.0 (min = 320, max = 370, SD = 17.1). This seems to confirm that the items were unknown to most of the participants.

Results of the recall tests

For these low-frequency unfamiliar items, there was a clear floor effect in the immediate recall test with many participants returning blank test papers: only 99 word-pairs out of a possible 2480 were recorded. The highest number of word-pairs correctly recalled by a participant was four out of 20. Table 7.4 summarises the findings. More alliterating word-pairs were recalled (alliterating mean = .42, SD = .71, non-alliterating mean = .32, SD = .69), though there was no statistically significant difference (z = -1.31, p = .10, 1-tailed). In the unannounced 3-week delayed free recall test, after having worked with the items on word-cards,[11] participants recalled more alliterating phrases (m = 3.04, SD = 2.77) than non-alliterating phrases (m = 2.52, SD = 2.64, z = -3.26, $p < .001$, 1-tailed) and a small-medium effect size was

[10] In Chapter 1, the definition of 'alliteration' included word-pairs with one consonant + one consonant cluster. It is a point of discussion whether /kw/ - /k/ is as alliterative as /k/ - /k/.

[11] Because the word-cards were paper based and not on an app, there was no way to quantify the amount of study short of a *post hoc* self-report or interview.

observed (r = .37). Similarly, on the delayed cued recall test, more alliterating items were recalled (m = 4.5, SD = 3.71) than non-alliterating phrases (m = 4.1, SD = 3.5, z = -2.74, p = .02, 1-tailed) and a smaller effect size was found (r = .3).

In the translation test, participants were given the Japanese translation of the word-pairs and asked to write the English meaning. By this stage, participants had

Table 7.4 Recall test results.

	Alliterating mean	Non-Alliterating mean	Analysis[1]	Effect Size
Immediate Free Recall	0.42	0.32	Wilcoxon Signed Ranks z = -1.31, p = .10 (1-tailed)	
Delayed Free Recall	3.04	2.52	Wilcoxon Signed Ranks z = -3.26, p < .001 (1-tailed)	r = .37 (small-medium) R^2 = .14
Delayed Cued Recall	4.5	4.1	Wilcoxon Signed Ranks z = -2.74, p = .02 (1-tailed)	r = .30 (small-medium) R^2 = .09
Translation test	5.1	4.67	Wilcoxon Signed Ranks z = -2.74, p = .003 (1-tailed)	r = .31 (small-medium) R^2 = .09

[1] As the data did not fit the assumptions for a parametric test (due to outliers and skewed distributions), a non-parametric test was done.

had the sets of words cards for approximately three weeks and had been instructed simply to try to learn the items. As can be seen in the bottom row of Table 7.4, participants scored more highly on the alliterating phrases (m = 5.10, SD = 3.81) than non-alliterating controls (m = 4.67, SD = 3.67, z = -2.74, p = .003, 1-tailed) and a small-medium effect size was found (r = .31).

Although the word-pairs in the non-alliterating condition were reported as slightly more familiar, and were more frequent, concrete, and translatable, it was the alliterating condition for which participants were more successful in recalling the forms and meanings of the word-pairs. In sum, these findings seem to be consistent with the notion that alliteration does have a mnemonic effect, even for less well-known test items, though the effect size is modest. It could be said that this lack of familiarity involves greater cognitive pressure, resulting in lower rates of recall compared to the high-frequency word-pairs.[12]

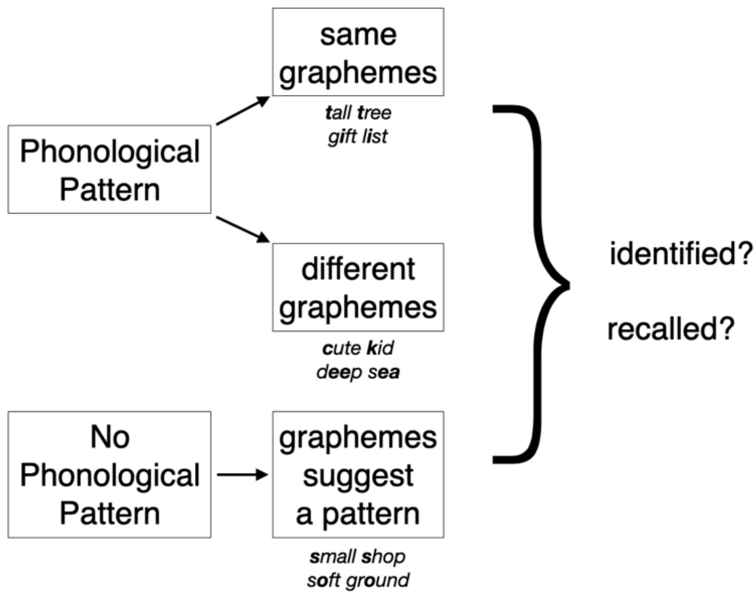
The study described above is the only one to date to furnish evidence that a phonological pattern may confer an advantage to the recall of the meaning of a real, but unfamiliar, word-pair. Of course, a logical follow up would be to conduct an experiment using real word-pairs in which both parts are novel to the participants. However, there might be methodological difficulties compiling such a set of stimuli, if the existing criteria for matching were to be retained.

7.4 What role might orthography play in 'phonological' tasks?

This section will attempt to answer two questions, which are presented on the right-hand side of Figure 7.3. As noted previously (Section 7.2), as part of the input leading up to measuring recall, participants were asked to tick or circle the word-pair on their dictation answer sheet if they thought there was a sound pattern. The participants successfully identified the phonological patterns approximately 90% of the time. However, in the majority of the word-pairs the pattern was realised through the same grapheme (for example, ***t**all **t**ree* or *g**i**ft l**i**st*). As a result, it is not possible to be sure whether they were influenced by the orthographic pattern. Therefore, this section focuses on the situations in which the pattern is expressed through different graphemes (for example, ***c**ute **k**id* or *d**ee**p s**ea***), or where the spelling may suggest a pattern that is not there (such as ***s**mall **s**hop*).

[12] There was no statistically significant correlation between self-reported familiarity ratings and total recall of the stimuli (r = -.25, p = .29 (2-tailed), R^2 = .07).

Figure 7.3 Interplay of orthography and phonology.



In research that was aiming to answer this question, an experiment could be designed with a substantial pool of target items that isolated the phonological pattern of interest from the orthography of the word-pair. In practice, however, this is difficult to accomplish. It is an intrinsic challenge in any language where the orthography represents the sounds, and especially in one that has a shallow writing system with a direct spelling-sound correspondence. It would not be possible to do so in Spanish, for example, due to the close one-to-one mapping of graphemes to phonemes (Defior & Serrano, 2017, p. 271). In English, where the grapheme-phoneme correspondence is more complex, however, there is some scope for manipulating this variable.

Although the present data were not collected with attention paid to the role of orthography, a small number of examples within the data do offer a glimpse of the potential role it had in the participants' engagement with the stimuli. In the data reported in Chapter 3, for instance, three word-pairs in the non-alliterating condition can be used to gauge if orthography affected the results of the identification task (in which the participants were asked to tick if the word-pair alliterated or not). Of the 124 participants, 69% ticked *small shop* as alliterating, 29% ticked *wrong word* as alliterating, and 15% ticked *whole world* as alliterating. In the same cohort, 52%

ticked *quiet constancy* as non-alliterating, and 23% missed the repetition in *certain synergy* (see Section 7.3). It is highly plausible that these judgements were an artefact of the spelling, even though participants had been asked to indicate if a pattern existed *before* writing down the word-pair in the dictation activities. The point at issue here is that instructing participants to indicate the absence or presence of a pattern prior to writing down the dictated target item (an instruction used in many of the studies listed in Table 2.4) might not block the unconscious activation of an orthographic representation on hearing a word. If there is unconscious activation, it would render this element of the experimental procedure (tick *before* writing) unreliable and, effectively redundant. In any case, from a practical point of view, in a classroom environment it is difficult to verify if the participants are adhering to such a directive.

As regards assonance, data from the study in Chapter 4 also suggests that spelling can impact evaluation: four of the ten assonating stimuli had a vowel nucleus with a different spelling across the target word-pair: *town house, fair share, deep sea*, and *main gate*. These four target items had, on average, 18.2 errors (min. = 3, max. = 34, SD = 12.7) on the identification task. In comparison, the six target items with identical graphemes in the nucleus had on average only seven instances of misidentification (min. 1, max. = 12, SD = 4.1). There is no data here to suggest a tendency to tick as assonating those word-pairs that had a similar spelling in the nucleus but a different pronunciation.[13]

Would it be possible to design an experiment that could tease this apart? In attempting to construct a large enough set of real word stimuli, such as *bound through* or *rough cough* for assonating word-pairs, or *cute kid* for an alliterative word-

---

[13] One anomaly does occur in this data set: over 40% of the participants ticked *nice place* as assonating. Had they misheard it as *nice price*? Neither I nor the original team of researchers (Boers, Lindstromberg, & Eyckmans, 2014b) noticed the consonance at the end of the words. This additional pattern does not seem to have improved recall: in the original study, *nice place* was recalled four times in the immediate free recall test, compared to the mean of 11.2 (Boers et al., 2014b, p. 102), and 30 times in the replication, compared to the mean of 41.9.

pair, the researcher would soon encounter problems finding enough target items that complied with the matching criteria used to balance word-pairs across the conditions, for example, frequency, length, concreteness ratings, or statistical measures of non-randomness such as Mutual Information scores. An extreme example of potential stimuli, though one that neatly illustrates the difficulties involved, are words that start with a silent first letter. The corpus frequency of such items would pose a challenge in finding a sufficient sample size. For instance, in a vocabulary profile of a 64-item list of English words starting with a silent first letter, almost one fifth are classed as off-list, that is, beyond the 25K level of frequency in the BNC / COCA corpus (for example, *psilophyte, gnomonic, pterion*). Such low frequency items are not words which learners are likely to know (or need), and even many L1 English speakers would probably struggle with some of the off-list items. A further difficulty would be in constructing meaningful and credible word-pairs (a *gnarly knight* or *mnemonic knickers*?), though of course one could use a mix of words (for example, a *gnarly guy*?). In addition, it would be problematic to balance any such items for other variables that have been shown to affect recall, such as length and concreteness. This reflects an inherent tension in experimental design – how to isolate the variable of interest from the many extraneous factors that may affect the outcome and still balance the items across conditions.

Having reviewed how the occasional phoneme-grapheme mismatch in the data affected *identification*, we turn now to how it affected *recall*. As there are so few experimental stimuli in which the phonological patterns were realised through different graphemes, or where a pattern was suggested by graphemes, the following results should be taken as suggestive only. In the data from the experiment in Chapter 3, the mean recall score[14] was 89.6, with the word-pair *small shop* falling below the mean (at 77), and two word-pairs above the mean (*wrong word* at 104, and *whole world* at 157). In the study in Chapter 4, there was no apparent tendency for the assonance to be a less effective mnemonic when the graphemes were not identical. The mean recall for word-pairs with the same spelling was lower than recall

[14] Scores were combined across the three tests: the immediate free recall, the 1-week delayed free recall, and the 1-week delayed cued recall test.

for word-pairs with different spelling (mean = 94.3, SD = 26.1 vs. mean = 146.0, SD = 76.7).[15]

What might these meagre indications suggest about the relationship between spelling and memorability? Perhaps with such word-pairs, particularly 'false-friends' that look like they have a phonological pattern due to shared orthography, it is reasonable to suppose that the spelling will play a role in the memorability and thus, from a language learner's point of view, it could be predicted that both features may confer a recall advantage. If so, then there might be value in considering such mismatches as an aid in learning.

7.5 What counts as a 'similar' sound in a phonological pattern?

This section attempts to answer the following question: to what extent can we reliably assume that phonological patterns are processed at the phoneme level? In Chapter 1, it was held that alliteration and assonance could be operationalised as 'linguistic sounds recurring', a cognition-centric definition based on what the speakers / listeners "think they are saying and hearing" (McMahon, 2002, p. 3).

The suite of experiments by Boers et al., reviewed in Chapter 2, adopted a binary approach to defining alliteration and assonance: the word-pairs either had a similar sound or they did not. However, perhaps the presence of a phonological pattern is not a simple Yes / No dichotomy, and there is in fact a gradation of similarity. If this is so, then perhaps word-pairs displaying a 'strong' pattern have a greater mnemonic advantage than word-pairs with a 'weak' pattern. To investigate this in more depth, data collected for the experiments reported in Chapter 3 and Section 7.3 was examined and the target word-pairs were analysed in terms of similarity of the alliterative onsets.

---

[15] The reason for the high mean and large standard deviation is that one particular word-pair, *deep sea*, proved to be especially memorable across the three tests.

Alliteration, as defined in this thesis and in the studies by Boers et al., has been operationalised as the repetition of the same consonant sound in an initial stressed syllable in two consecutive words. In this context, 'same' was aligned with the phoneme, such that two occurrences of word initial /t/ would be considered the same, even if the following vowels caused different amounts of lip rounding through backward assimilation. Thus, designing experiments at the phonemic level ignores certain differences in pronunciation assuming that they are not likely to have an effect on the perception of alliteration. Moving in the other direction, it could be argued that contrasting at the phoneme level ignores certain similarities between sounds, such as the fact that both a /t/ and a /k/ are unvoiced stop consonants.

The degree of similarity between words could be measured by adapting a simple approach such as the Levenshtein distance, or string-edit distance (Levenshtein, 1966). In contrast to our needs here, Levenshtein was interested in contrasts at the level of orthographic letters, such that similarity between two words is calculated according to the shortest possible distance between changing one word into another by inserting, deleting, or reversing letters. For example, the Levenshtein distance between the words *kitten* and *sitting* is three, since it requires a minimum of three edits to change one word into the other:

1. **k**itten → **s**itten (substitute < s > for < k >)
2. sitt**e**n → sitt**i**n (substitute < i > for < e >)
3. sittin → sittin**g** (insertion of < g >)

However, here, the similarity measure will be applied within the phoneme. Phonemes are abstract conceptualisations of sounds in a language, that serve to organise the phonology in one particular way. Essentially, they are units that are capable of distinguishing two words that are otherwise identical (minimal pairs). This means that regular and predictable differences in pronunciation that do not distinguish minimal pairs, are absorbed as allophones of the phoneme. For example, the phoneme /t/ can be phonetically expressed in several different ways in English, according to where it falls in the word or phrase, and who is speaking, including an aspirated alveolar stop, an unaspirated alveolar stop, a flap, and a glottal stop. At the phonemic level, for example, the two /t/s in *stops talking* are considered identical, even though, phonetically, they are not. Indeed, that is how we can detect whether someone has said 'stops talking' or 'stop stalking': in the former, the /t/ of 'talking' is aspirated (even though there is an /s/ just before it) and in the latter, the /t/ of 'stalking' is unaspirated. Similarly, a glottal stop version of /t/ as heard in, say, a London pronunciation of *butter* would be considered to display the phonological pattern of consonance with an RP pronunciation of, say, *mutter*, realised as an aspirated alveolar stop. Whether such pairings are indeed perceived as having the 'same sound', and if so, whether because of pronunciation or spelling, is an empirical question.

But it opens up a related question - whether two sounds might alliterate because they share some, though not all, of their articulatory features. For example, /p/ and /b/ are more similar than they are different, only varying in their voicing, a feature that is realised on a continuum in English, with assimilation and position significantly affecting it, for example, final voiced stop consonants are often devoiced, as in *lead* [li:d̥].

In the spirit of Levenshtein, it was decided to find a method for calculating similarity at the phonemic level. In order to attribute a numerical value to the target word-pairs as a proxy for a 'strong' or 'weak' alliterative pattern, a method for quantifying judgements of relative closeness and distance between two words was adapted from Grimes & Agard (1959). Their framework was originally used for evaluating the distance between the sound inventories of two languages, and as part of that, it involves quantifying phonemes in terms of the following six independent variables as applicable:

A: Point of articulation: 1 labial, 2 denti-alveolar, 3 palatal, 4 velar

B: Manner of articulation: 1 nasal, 2 stop, 3 fricative, 4 approximant

C: Voicing: 1 voiced, 0 voiceless

D: Vowel height: 3 High, 2 Mid, 1 Low

E: Back: 1 Front, 2 Central, 3 Back

F: Word-initial cluster: 1 cluster, 0 no cluster

Thus, for instance, a voiced velar fricative /ɣ/ scores 4 + 3 + 1 = 8, whereas a voiceless labial nasal /m/ scores 1 + 1 + 1 = 3. This scoring system, when values for phonemes are summed, can give an indication of how similar two sounds are: /m/ and /b/ vary only by one point (manner of articulation), for instance. By adopting this approach, it may be possible to calculate the degree of similarity between potentially alliterating or assonating elements in a word-pair. If this is applied to an alliterating word-pair, then only the onsets of the words need to be compared, and the information regarding the peak vowel (variables D and E above) can be discarded, as in Example 1:

Example 1: Quantifying the initial consonants in *tall tree*

	A	B	C	F	
[tʰɔːl]	2	2	0	0	
[tʰriː]	2	2	0	1	
numerical difference	0 +	0 +	0 +	1 +	**= 1**

In Example 1, *tall* and *tree* can be said to differ by a factor of 1. A useful effect of this method is that it pins down the level of similarity between a simple consonant and a cluster, on account of category F. Thus, for example, in *tall tree*, although the initial /t/s constitute alliteration formally, the method recognises the difference between them on account of the following /r/ in *tree,* something that most speakers would naturally sense to be a 'compromise' to 'full' alliteration. In comparison, the words in a non-alliterating word-pair, such as *clean shirt*, should differ by a larger factor, as can be seen in Example 2:

Example 2: *clean shirt*

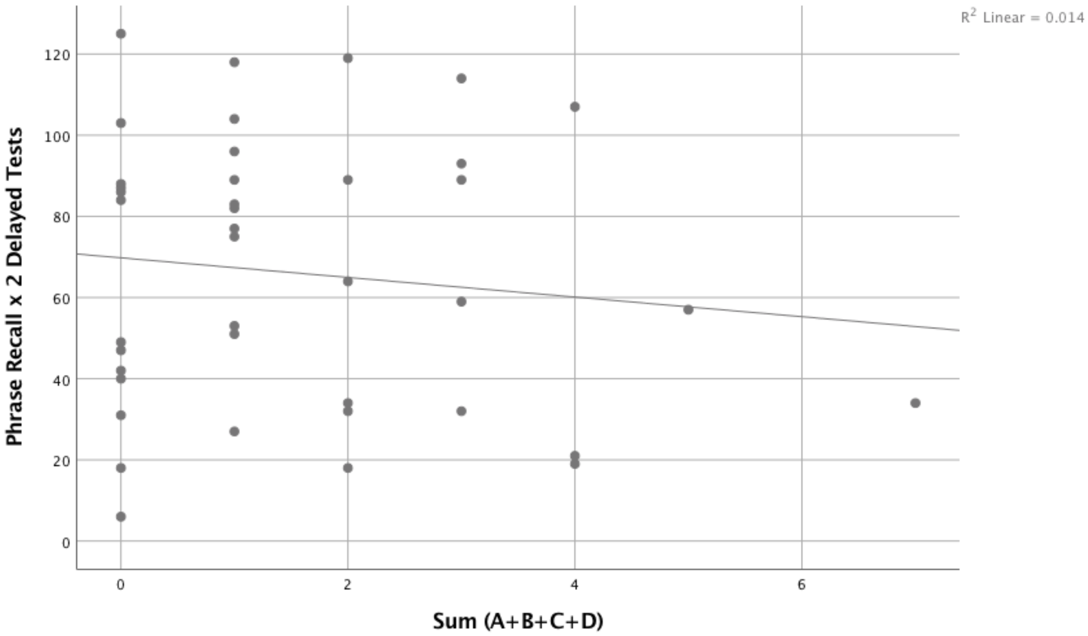
	A	B	C	F	
[kliːn]	4	2	0	1	
[ʃɜːt]	3	3	0	0	
numerical difference	1 +	1 +	0 +	1 +	**= 3**

If participants are sensitive to articulatory similarities between sounds that are not identical, then any advantage that alliteration affords for recall might be weakly replicated in words that do not alliterate at the phoneme level but have a low score on difference on this scale. These numerical values might then afford some insight into whether there is a linear relationship between similarity and the rates of recall: the hypothesis being that word-pairs with a lower difference in feature-score (indicating similarities in the articulatory process arising from the phonological pattern) will have higher recall scores. Conversely, word-pairs with a larger difference between scores (indicating different onsets) will have lower recall scores.

A correlation was done to see if there was a linear relationship between the feature-score difference and recall.[16] Although the scatterplot (see Figure 7.4) indicated a general downward trend in the data, consistent with the hypothesis, a visual inspection of the regression line suggested the assumption of linearity had not been met and a linear model was not the best fit for this data set. A Pearson's *r* correlation between similarity and recall scores was not statistically significant and there was a negligible effect size (p = .22 (1-tailed), R^2 = .02).

[16] Combined recall scores from three tests (the immediate free-recall, the 1-week delayed free-recall and the 1-week delayed cued-recall tests) for the twenty high-frequency word-pairs used in the study in Chapter 3, plus the twenty unfamiliar word-pairs from Section 7.3

Figure 7.4 Scatterplot of feature-score difference and recall.



The fact that these results are not statistically significant may be due to a lack of power, as there are only 40 items. An *a posteriori* power analysis using *R* (pwr package v. 1.2-2) showed very low power at 12% (n = 40, *r* = -.125, sig.level = 0.05, power = 0.124, alternative = two.sided), far short of the 80% judged to be adequate (Larson-Hall, 2010, p. 105).

Therefore, although this preliminary analysis does not support the notion of graded phonological similarity, it could be that it simply lacked the power to find an effect. Future experiments could perhaps tease this apart by using a larger pool of target items.

7.6 Long-term recall and phonological patterns.

The question of a phonological pattern aiding long-term retention seems almost axiomatic, especially in the fields of advertising and marketing where vast amounts of resource are invested with the aim of persuading a customer to buy a particular product; even the most cursory internet search will bring up countless hits extolling the virtues of alliteration or rhyme when choosing a brand name or creating an advertisement. If *Coca Cola* is so memorable, why not a word-pair? If people are sensitive to alliteration in a brand name, does it follow they will be sensitive to alliteration in a language learning environment? Leaving aside the question, for now, of whether a brand name is equivalent to a word-pair, the evidence in this thesis suggests that sound patterns can make a difference to how language is noticed and remembered. This section looks at one of the parameters for this phenomenon namely, what impact these patterns have on memory over time.

Many of the experiments seen in Table 2.4, and many of those reported in this thesis, featured a treatment phase followed by tests at different intervals of time. These included free recall tests given (more or less) immediately after the treatment, and delayed tests ranging from one hour up to three weeks after the treatment, sometimes with further intermediary study. Test timing was not built in as a major variable, rather a way to track learning in a manner consistent with what language learners need to do. As a result, little attention has been paid to the detail of recall over time in the previous chapters. However, now it might prove informative to examine how recall and retention is affected by the passage of time.

In those studies in this thesis which looked for evidence of a mnemonic effect for alliteration, an effect appeared only in the delayed tests. This applied to both familiar word-pairs (Chapter 3) and unfamiliar word-pairs (section 7.3) as can be seen in Table 7.5

Table 7.5 Results of delayed recall of alliterative word-pairs.

Chapter	Test Time	Results	Statistically Significant	Effect Size
3	Immediate free recall	alliterating > non-alliterating	x	
	1-week delayed free recall		✓	r = .42
	1-week delayed cued recall		✓	r = .47
7	Immediate free recall	alliterating > non-alliterating	x	
	3-week delayed free-recall		✓	r = .37
	3-week delayed cued recall		✓	r = .3
	3-week delayed translation		✓	r = .31

In the study which looked at assonance, delayed responses increased the contrast between patterned and non-patterned word-pairs, but the effect has dissipated after one week, as shown in Table 7.6:

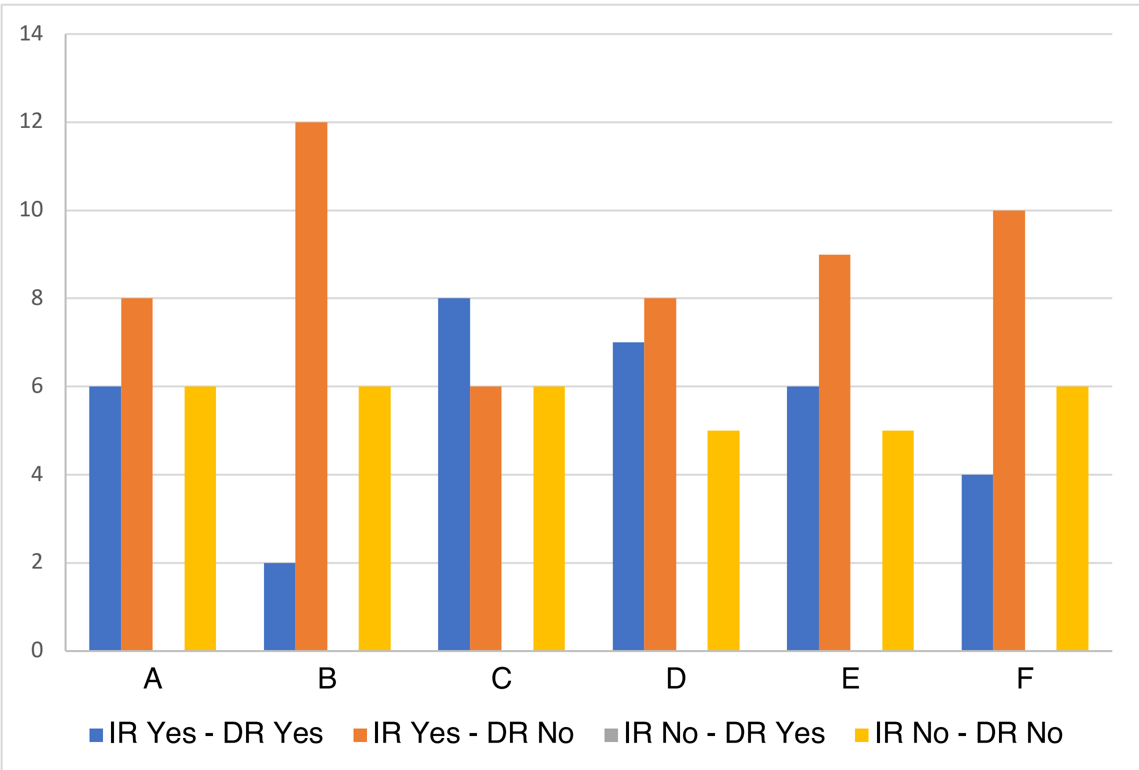
Table 7.6 Results of delayed recall of assonating word-pairs.

Chapter	Test Time	Results	Statistically Significant	Effect Size
4	Immediate free recall	assonating > non-assonating	✓	r = .32
	1-hour delayed free recall		✓	r = .41
	2-day cued recall		✓	r = .50
	1-week delayed cued recall		x	

In the study reported in Chapter 6, delayed testing after the study phase improved scores and the extra week of delay without study (when access to the flashcard app was blocked) sustained performance above the initial test. This held true even for those participants whose attention had not been drawn to the phonological patterns, a result which is hard to reconcile with the majority of previous studies in which attention-raising played an important role – this exception will be discussed in the following section.

So far, we have only seen that participants' overall recall held up well over time. However, there is a further question to consider: whether they are recalling the same items across tests. Figure 7.5 shows the results from the immediate free-recall test (IR) and the one-week delayed free-recall test (DR) for six participants (A to F) from the alliteration experiment in Chapter 3. The scores from these six participants were chosen because their scores on the immediate free-recall test were more than two Standard Deviations above the mean (M = 7.7 out of a maximum 20), ensuring sufficient data for further analysis (a random selection of participants would have included scores of zero which would not offer much insight).

Figure 7.5 Comparison between immediate (IR) and delayed recall (DR) of familiar word-pairs for six participants (A – F).



IR Yes - DR Yes = the same word-pair was recalled in the immediate free-recall test and the one-week delayed recall test.

IR Yes – DR No = the word pair was recalled in the immediate free-recall test but not in the one-week delayed free recall test.

IR No – DR Yes = the word pair was not recalled in the immediate free-recall test but was recalled in the one-week delayed free recall test.

IR No – DR No = the same word-pair was not recalled in either test.

Taking the results from participant A as an illustration, Figure 7.5 shows that of the word-pairs recalled in the immediate test, six were also recalled one week later (blue column). Eight other word-pairs were successfully recalled in the immediate test but could not be recalled one week later (the orange column). And the six remaining word-pairs could not be recalled in either test (the yellow column). There were no recorded instances of any of the 124 participants failing to remember a word-pair in the immediate test but then successfully recalling it the following week (the grey *IR No – DR Yes* column).

In other words, Figure 7.5 shows that if a word-pair was successfully recalled in the delayed test, then it had been recalled in the immediate test. This lends weight to the idea of a post-encoding process in which the memory is retrieved, consolidated and strengthened - for example, the participant thinks about what they did in class that day, or tells a friend about taking part in a fascinating word-pair experiment, or reactivation occurs during a sleep state. Each time an episodic memory is retrieved, it is subsequently re-encoded, resulting in stronger memory traces. What are the neural underpinnings that can account for this? The hippocampus will have played a role in laying down the new episodic memory, and over time the information could have been transferred to different parts of the neocortex as semantic memory (as discussed in Chapter 2). In this way, if a participant has forgotten the episode, they may nevertheless still have access via semantic memory to information extracted from the context-dependent episode (Eichenbaum, 2010, p. 482). The inability to recall a word-pair could be due to inadequate attention: familiar word-pairs may not be of any significance to the participants beyond the test situation, and thus the episodic memory decays. A further cause could be a lack of familiarity with the material, which is of particular relevance to the pseudoword study and the low-frequency academic word-pairs which both showed low recall test scores.

In Chapter 2 it was posited that word-pairs with a phonological pattern are privileged because of the form-based similarity (see Chapter 2), this allows one word to activate or prime the other and so they are more amenable to retrieval and re-encoding.

7.7 How might individual variation play a role in the findings?

If the extraneous variables that can affect the memorability of a word-pair, such as length, concreteness, and L2 translatability, are controlled for successfully, then the target items will only be considered to differ in the presence or absence of a phonological pattern. The experimental hypothesis is that such patterns affect the recall of the form and meaning of the target word-pairs during the testing phase. But perhaps there are additional variables, relating to the participants themselves. Classic variables that research considers include motivation, task engagement and other affective factors (Dörnyei & Schmidt, 2001). In what follows, however, two other potential variables will be considered: firstly, variation in musical aptitude, and secondly, variation in the capacity of the phonological loop.

7.7.1 Variation in musical aptitude.

One finding from the literature that may have some bearing on the results of the experiments described in this thesis is the participant's musical ability, in particular the extent to which they can perceive repetitive patterns imparted by phonological repetition.

A study that touches on this variable is Argo, Popa and Smith (2010), previously noted in section 2.2.1, because the authors found that invented brand names containing phonological patterns were evaluated more positively than brand names that had no phonological repetition. What was not mentioned in Chapter 2 is that in their second study, the authors found that this evaluation was moderated by the participant's sensitivity to repetition (STR). As a proxy for assessing sensitivity to phonological repetition, the authors measured sensitivity to repetition of musical sounds. They justified this by reasoning that a musical note is comparable to a phoneme. The experimental hypothesis was that individuals with high STR would rate a brand name containing phonological repetition more highly than an individual with low STR.

To test this claim, 125 English L1 user undergraduates sat the rhythm section of the computer-based Primary Measures of Music Audiation test (PMMA) (Gordon, 1979), part of a commercially available battery of tests that purports to measure musical aptitude. The rhythm section of the PMMA consists of 40 pattern recognition tasks; in each task the participant listens to a short sequence of sounds of various duration, some of which are combined to generate repetitive patterns. Then they listen to a second sequence of sounds and are asked to indicate if the two sequences are identical. Based on their responses, the software generates an overall score.

In the experimental phase of the 2010 study, the participants were randomly assigned to a condition in which they either heard brand names containing a sound pattern, or they heard brand names with no pattern, for example, *Zanozan* vs. *Zanovum*. The participants tested product samples,[17] which unbeknownst to them were from the same product, and rated them on a series of Likert scales. The authors found that the participants who scored highly on the PMMA test,[18] which the authors interpreted as having high sensitivity to repetition (STR), evaluated products more favourably when the brand name contained a sound pattern, than when the name had no sound pattern. In contrast, in the case of the participants with low STR, there was no statistically significant interaction and brand evaluations were undifferentiated.

The Argo et al. (2010) authors took an interesting and novel approach to exploring the moderating role of individual differences in STR as an explanatory factor in brand evaluation, and the authors acknowledge that there are few other studies that have adopted this methodology. If a novel brand name were indeed comparable to a new

[17] Participants were asked to taste samples of ice-cream, or physically touch samples of cat litter.

[18] 'High' and 'low' STR were defined as at least one standard deviation above or below the mean score.

L2 lexical item, then differences in STR could help account for some of the variation in how participants cognitively react to a new word-pair in the experiments in this thesis. In assessing the validity of the claims made in Argo et al. (2010), two questions arise. Firstly, to what extent is sensitivity to musical rhythm analogous to sensitivity to a rhetorical device such as alliteration? Secondly, is the PMMA test a valid instrument to gauge if a participant 'has an ear' for these devices?

Regarding the question of whether musical rhythm is comparable to a sound pattern found in language, Argo et al. (2010) note that there exists no current instrument to measure a person's sensitivity to a phonological pattern. They justify the use of the PMMA test on the grounds that a phoneme and a musical note are analogous. There is support in the literature for this view. From the perspective of the cognitive psychology of music, Sloboda (1985) argues that music is a cognitive skill with close parallels to language, having its own syntax, semantics and phonology, with the musical note as the basic phoneme (Sloboda, 1985, p. 24). Some educators have also noted the connection between music and literacy: for example, Hansen, Bernstof & Stuber (2014, p. 51) suggest that some aspects of L2 reading, such as phonological awareness, have parallel processes in music – "the syllables we sing are blended phonemes at the single *sound* level instead of the single *meaning* level." (original emphasis). Further support for the comparison can be found in neuroimaging data which shows an overlap of brain activity between language and music processing (for a review of ERP studies see Patel, 2012). In sum, it seems that the use of a rhythm test does stand as a reasonable surrogate for assessing STR.

The second question concerns the validity of the Preliminary Measures of Music Audiation (PMMA) test used in Argo et al. (2010). The PMMA, together with the Intermediate and Advanced Measures of Music Audiation tests (IMMA and AMMA respectively) (Gordon, 1979, 1986, 1989), were designed to measure musical 'audiation', which Gordon (1989) equates with musical aptitude. It should be noted, however, that there seems little agreement about what terms such as *musical aptitude* and *STR* actually mean (Murphy, 1999). It seems odd that Argo et al. (2010) chose the PMMA to test their undergraduate participants, as the PMMA was designed for school children aged five to eight, while the AMMA was designed for college-age students (Stamou, Schmidt, & Humphreys, 2010, p. 76). Indeed, even the intermediate test (IMMA) might have been too basic for college students. Degé, Patscheke, & Schwarzer (2017) report skewed results on the IMMA, which they interpreted to indicate that their 89 participants (mean age = 10 years 9 months) found the rhythm section too easy. Argo et al. (2010) do not report any descriptive statistics for the results of the PMMA so perhaps this was not an issue for their participants. However, Degé et al. (2017) also ran a multiple regression analysis and found that the IMMA scores accounted for only 4% of the variance, other contributing factors included socioeconomic status (based on the parents' education) and the number of music lessons the participants had received. In sum, though the grounds for using a musical aptitude test seem sound, the test itself potentially suffers from validity issues.

In order to assess the extent to which musical aptitude may play a role in the findings reported in this thesis, data was collected from the 78 undergraduates who took part in the study reviewed in Chapter 6. The participants were given a questionnaire (see Appendix 6) in which there were nine statements relating to their musical ability, for example, *I can read music（楽譜）*. Some of the key words were glossed in Japanese. Each question was answered on a Likert scale with descriptive anchors ranging from 1 = *no / not at all* to 7 = *yes / very well*. In this way, each participant had a value that was held to approximate a level of musical ability.[19]

It may be recalled that in Chapter 6, the participants were randomly assigned to one of three conditions: those in the alliterating condition were asked to study electronic word cards with the alliteration highlighted, those in the assonating condition had the assonance highlighted, and the participants in the third condition had no phonological patterns highlighted. The participants in this last group, who received no awareness-raising of the phonological patterns, in general scored more highly on the recall tests than the participants who had the patterns highlighted on the electronic flashcards. If high and low musical ability was equally distributed

[19] Due to costs, it was not possible to bulk purchase one of the Measures of Music Audiation tests.

across the groups, then it would not have any capacity to explain the findings. However, if it happened to be the case that, for instance, the assonance group had disproportionately low musical ability, that might explain why they did not display an assonance advantage. The results from the questionnaire could thus be used to check the distribution of musical ability across the conditions (see Table 7.7).

Table 7.7 Music aptitude scores across conditions.

	n	Mean	SD	Min[1]	Max[2]
Group 1 (Alliteration)	26	18.54	8.96	10	51
Group 2 (Assonance)	25	22.60	13.34	9	53
Group 3 (No Patterns Highlighted)	27	27.30	12.17	13	50
Total	78	22.87	12.03	9	53

[1] minimum possible = 9

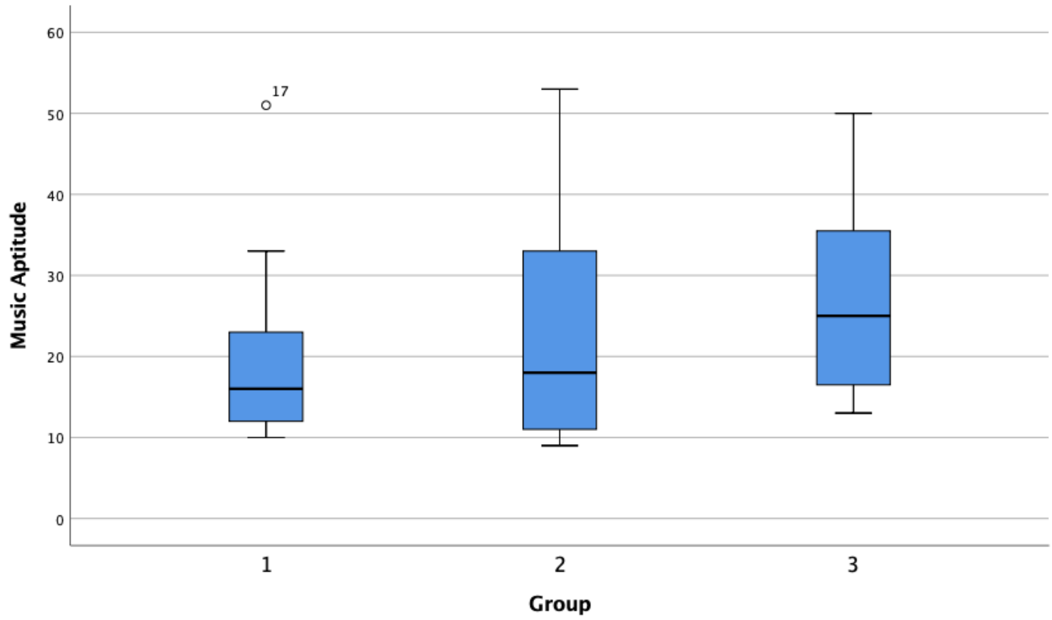
[2] maximum possible = 63

To test whether the differences between the mean scores were large enough to not be ascribable to chance fluctuations, a One-Way analysis of variance was conducted. A visual inspection of the data showed the presence of outliers and unequal variance, confirmed by Levene's test of homogeneity of variance (p = .03) – see Figure 7.6.

For the variable of Music Aptitude, the omnibus ANOVA shows a statistically significant difference between groups ($F_{2,\ 75}$ = 3.77, p = .027). Post-hoc comparisons using Games-Howell[20] tests found a statistically significant difference and a large effect size between Groups 1 and 3 (mean difference = 8.76, 95% CI = 1.68, 15.84, p = .01, Cohen's d = 0.81), but no statistically significant differences between Groups 1 and 2, nor between Groups 2 and 3.

[20] These tests were thought the most suitable due to the unequal sample sizes and unequal variance (Larson-Hall, 2010, p. 282).

Figure 7.6 Boxplot of music aptitude.



These results suggest that although the participants were randomly assigned to groups, Group 3 had higher Music Aptitude scores than the others, which we would infer is due to chance. Of course, participants would potentially vary in how they interpreted and responded to the statements on the questionnaire, but there is no reason to believe that those in Group 3 would have a consistent difference, given the random allocation to groups. It is of course not entirely safe to infer from how participants respond to questions about musical experience that the scores they get are an exact map for their musical aptitude measured in other ways. All the same, the simplest explanation for the results is that the participants in Group 3 happened to have an overall greater level of musical aptitude and that this explains their higher performance on the tests, even though they were the control group. Other potential explanations for their superior performance remain possible (for example, that the experimental conditions hindered rather than helped recall) and in such cases, the musical ability might not be the determining factor at all, or only partly so. In the absence of further data (which could have been gleaned from individual *post-hoc* interviews) there is no clear way to resolve this issue.

The next question is whether or not there is a linear relationship between music aptitude and word-pair recall, the hypothesis being that those students with high music aptitude scores may ‘have an ear’ for phonological patterns that students with low music aptitude scores do not. However, a non-parametric correlation (Spearman's *rho*) between musical aptitude scores and total recall scores found no statistically significant correlation and negligible effect size (N = 78, $p = .60$ (two-tailed), $R^2 = .004$, 95% CI = -.28, .16).

7.7.2 Variation in the phonological loop

The second factor that could account for individual variation concerns the phonological loop. In brief, the phonological loop, part of the multicomponent model of working memory (WM), deals with auditory material. First proposed by Baddeley & Hitch (1974), it is often conceived as being divided into a phonological store (the 'inner ear' which temporarily holds speech input from the outside world or recalled from long-term memory), and an articulatory rehearsal mechanism (the subvocal 'inner voice' which allows us to repeat and maintain decaying representations). The most common example given of the phonological loop in use is when a person is told a telephone number and they repeat the string of digits to keep it in working memory until they can write it down or make the call.

It is thought that the phonological loop plays a fundamental role in both L1 and L2 language learning in that it is key to the formation of long-term representations of novel material, providing a direct link between temporary storage and lasting retention. Longitudinal studies and correlational studies suggest there is a strong association between measures[21] of WM and vocabulary acquisition (Baddeley, Gathercole, & Papagno, 1998). For example, children with strong phonological memory skills have been found to have greater L1 vocabulary knowledge compared to their peers with weaker memory skills (see for example, Gathercole & Adams,

---

[21] Two common methods for assessing memory function are the digit span task (a measure of the maximum sequence of spoken letters, words, or digits that can be reliably recalled) and non-word repetition; in both tasks accurate performance requires the storage of the sequence in temporary phonological memory.

1994). Furthermore, greater phonological memory function is also associated with increased L2 acquisition (for example, Ellis, 1997; Service, 1992).

Three variables are known to impact on the phonological loop. The effective capacity of the loop is diminished when items to be recalled are long words rather than short words (the word length effect) due to the limited time available for maintenance rehearsal. This empirical effect has some explanatory power for the participants' low recall scores for the polysyllabic word-pairs seen in section 7.3. Secondly, the rehearsal mechanism can be easily interrupted: telling somebody a phone number and then requiring them to engage in irrelevant articulation (saying 'blah blah blah, for instance) interferes with the process (the articulatory suppression effect). The third variable which impairs the phonological loop demands more in-depth consideration as it has a direct bearing on the evidence presented in this thesis: words that are phonologically similar are more difficult to remember (the phonological similarity effect or PSE).

How can the evidence of a mnemonic effect for word-pairs that alliterate or assonate be reconciled with the PSE, one of the most well-known and robust findings in memory research (e.g. Conrad & Hull, 1964, Salamé & Baddeley, 1986)? Briefly, the PSE commonly occurs in memory span experiments where participants are presented with lists of known words, or letters, on a computer screen, then are immediately asked to recall them in the correct order, either verbally or in writing. Such experimental paradigms are thought to measure working memory capacity, and test performance is often found to be worse when participants are asked to recall items that are acoustically homogeneous (e.g. *bat, flat, mat* or *b, v, p, c*), than when the items are acoustically heterogeneous (e.g. *dirt, pen, hug* or *f, a, m, t*) (e.g. Conrad & Hull, 1964). If alliteration and assonance can be viewed as instances of acoustic homogeneity, then why does the PSE not apply to the experiments reported here as well as those of Boers and his colleagues, resulting in diminished recall?

One reason why the effect may not impinge on the recall ability seen in the experiments is that the PSE decrement tends to be limited to strict 'serial recall', where only the item-in-position satisfies the correctness criterion. In memory span tests which adopt an order-free criterion of correctness, 'item recall', phonological similarity can often facilitate recall (see Gupta, Lipinski, & Aktunc, 2005 for a review). Why would this be so? The studies which purport to test working memory span and the PSE often use rhyming items as stimuli. Fallon, Groves & Tehan (1999) suggest that the rhyme ending acts as an effective retrieval-cue because the rhyming items are all members of a taxonomic category, a rhyme category, which facilitates short-term item recall where the order is not important, in effect, superseding the PSE. Gupta et al. (2005, p. 1012) define such category cues as "commonalities between the items in a list that can be extracted and used as a cue", giving the example of "all the numbers are multiples of 11" for the list 11, 55, 77, 22, 88. But does it follow that the notion of a taxonomic category can be expanded to encompass alliteration (for example, "all the words start with a /b/ sound") or assonance (such as, "all the words have an /æ/ sound in the middle"), which then themselves also act as retrieval cues and facilitate item recall? This question is not one that can be easily answered without further research. If so, the absence of the Phonological Similarity Effect in the alliteration / assonance experiments may be attributed to the fact that the testing phases often used free item recall in which the participants were permitted to answer in any order.

So far, we have seen that acoustic homogeneity appears to *impede* accurate recall when recall must be in the correct order (the Phonological Similarity Effect), and that it appears to *assist* recall when the order does not matter (where it acts as a retrieval cue). However, there is also evidence that phonological similarity impedes recall even when ordering is not required. Such evidence is contrary to the patterns found in this thesis and by Boers et al. There is experimental data showing a detrimental effect in free recall tests of phonologically similar items (e.g. V. Coltheart, 1993; Fallon et al., 1999; Gupta et al., 2005), where the items do not rhyme, but have "high phonemic overlap" (Fallon et al., 1999, p. 303), for example, *can mad cap mat*; in other words, lists of words displaying assonance and non-sequential alliteration. Indeed, Fallon et al. (1999) argue that alliterating / assonating types of phonological similarity do *not* provide an effective category cue and therefore do not facilitate item recall. They tested this hypothesis by examining serial recall plus item recall of rhyming lists, assonating / alliterating lists and phonologically dissimilar lists. In the first experiment, they found that item recall was greater for rhyming lists, consistent with the idea that rhyme produces a facilitatory category-cuing effect. Of particular relevance, however, is the finding that item recall was greater for the dissimilar lists than the alliterating / assonating lists. The authors suggest that the phonological similarity observed in alliterating lists and assonating lists detrimentally affects recall due to phonological confusion, that is, the classic PSE. Moreover, Gupta et al. (2005) cite further evidence along the same lines.

However these contradictory findings are to be explained, it is clear that there are many differences between the PSE experiments described above and the core focus of this thesis, namely, how vocabulary learning in the L2 classroom can be enhanced. Nonetheless, it could be said the underlying themes are essentially the same: word processing and memory. It should be remembered that in the task of recalling words from memory, phonological similarity is but one of many key variables that can affect the outcome (for example, word length, measured in the number of letters, and word frequency), and these other variables must be controlled for if the results are to be interpreted unambiguously. The same sets of word lists were used in Fallon et al. (1999) Experiments 1 and 2, Coltheart (1993) and one of Baddeley's (1966) experiments, and word length is clearly matched across all the conditions in these sets. However, the stimulus items used by Gupta et al., (2005) Experiments 1- 5, range from 3 to 7 letters in length across the conditions. For example, in Experiments 1 and 2, the ten lists in the "Dissimilar" condition consist only of three-letter words. This set of lists is supposedly 'matched' with the ten "Alliterative" lists which contain words of three, four and five letters in length. This disparity in word length could perhaps explain why recall was poorer for the alliterative lists.

Word frequency is another important factor in cognitive processing: it is well-established that high-frequency words are produced more quickly and more efficiently than low-frequency words (e.g. Balota & Chumbley, 1984). Looking at the lists used in the Fallon / Coltheart / Baddeley sets, and bearing in mind that memory span tests utilise 'known' words, there are some that seem intuitively less frequent (e.g. *cox, dux, gab, sup*), and a quick check on the *lextutor* vocabulary profiler bears this out (Cobb, n.d.). From the Gupta et al. (2005) experiments come *samp, fain, yam,* and *rand*. If frequency was matched across these lists of 'known' words, then there is no cause for alarm, but was it? The frequency data for the words used in the Fallon / Baddeley sets, and in the five experiments reported in Gupta et al. (2005), are based on the Kučera and Francis frequency norms (Kučera & Francis, 1967). Frequency counts are subject to measurement error arising from factors such as the size of the corpus and the sample of texts used in generating the corpus. The Kučera and Francis frequency norms are based on a corpus of just over one million words, which is small by today's standards, and, more importantly, these norms consistently underperform in validation studies (for a review, see Brysbaert & New, 2009). If the Kučera and Francis (1967) measure provides a "poor estimate of cumulative frequency" (Zevin & Seidenberg, 2002, p. 12), then it could be claimed that the effects of the phonological similarity variable are, in fact, a frequency effect in disguise.

In sum, the stimuli matched across conditions in terms of frequency and / or length in the above experiments (Coltheart, 1993, Fallon et al. 1999, Gupta et al., 2005) are unlikely to be confound free because the variable of interest, recall, is so closely correlated with word frequency and length.

This section looked at whether differences in the phonological loop could explain the findings, but it seems some of the studies that might have shed light on that are potentially unreliable. To resolve the question, future studies would need to control for confounding variables such as orthographic length and frequency.

7.8 General conclusions

This chapter has attempted to take stock of what the accumulated findings of the previous chapters might mean, by examining additional considerations. These considerations suggest that phonological patterns do seem to have a mnemonic effect, which is not contingent on the pattern being made salient. The effect is small and dissipates over time.

Chapter 8 Conclusion

8.1 Introduction

This thesis began with a consideration of the claims made by Nation (2001, 2014) and Szudarski (2017) that L2 learners should pay attention to the phonological patterns characteristic of English formulaic language, which itself plays a major role in lexical competence. The empirical work reported in this thesis has focussed on the specific element of phonological patterns in ‘word-pairs’, an experimental tool operationalised as two adjacent word strings or pseudoword strings.

Before summarising the assembled evidence from this thesis, it might prove constructive to first query how the findings might reflect back more widely to formulaic language as a whole.

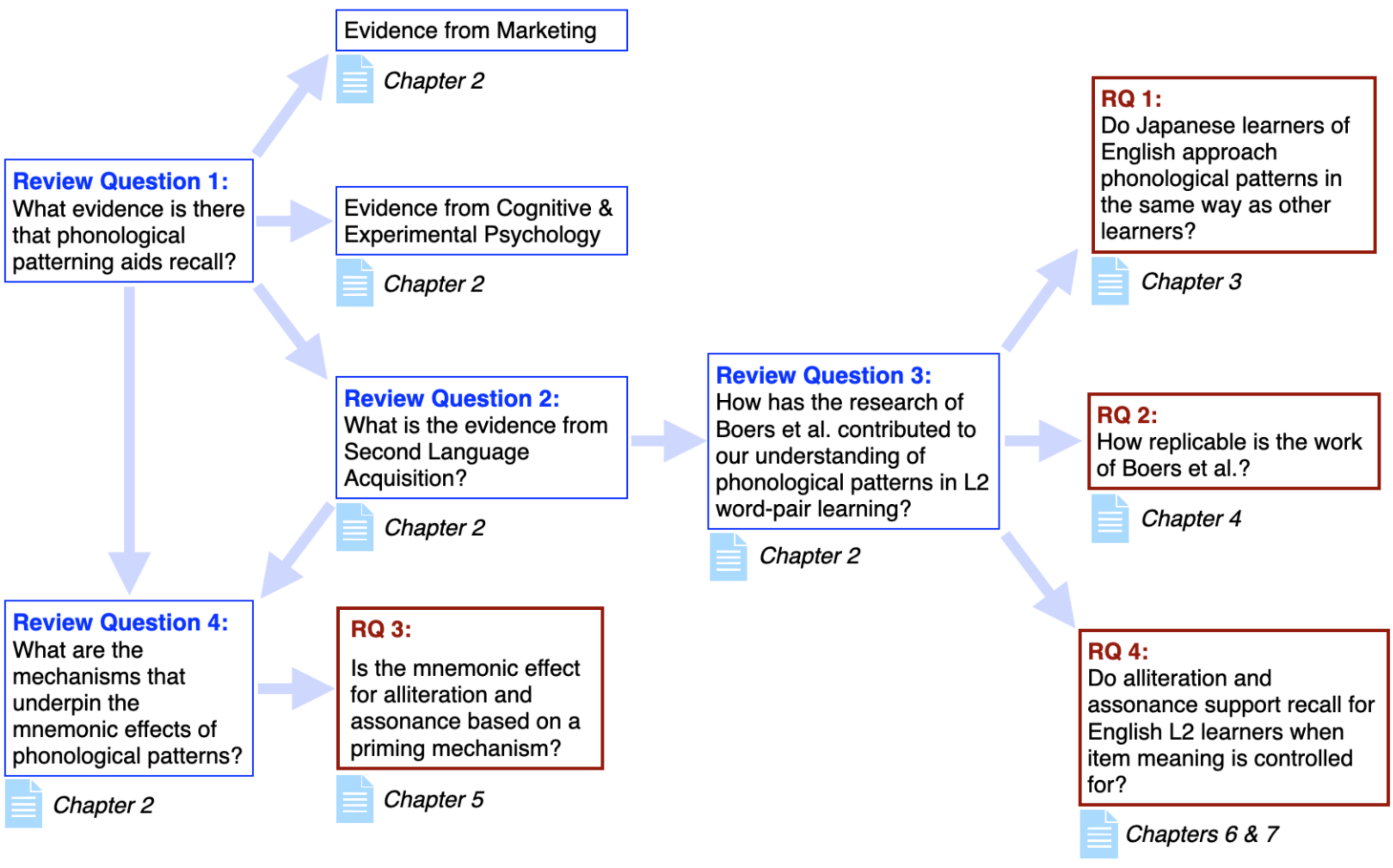
It seems reasonable to propose that any mnemonic advantage that holds true for a word-pair is applicable to a typical FS which might have intervening material. There is also a possible similarity between intervening words, like *the* as in *run the risk*, and intervening syllables in polysyllabic word-pairs. It is held that any interceding material will not violate the constraint of what constitutes “close enough … for the ear to be affected” (Greene et al., 2012, p. 40). However, the exact parameters of proximity and the limits of working memory span are future avenues to be explored. On the basis of the empirical work in this thesis, it is not possible to make strong claims about FSs more generally, but it should be noted that many researchers do count word-pairs as a subset of FSs. Furthermore, phonological patterns are a feature of many FSs (see Chapter 1), so the work herein constitutes a modest addition to understanding one aspect of learning L2 formulaic language.

In this final chapter, the findings will be summarised from the perspective of the contribution they make in understanding the role phonological patterns may play in L2 learning.

8.2 Towards a deeper understanding of the role of phonological patterns in vocabulary learning

The research reported here investigated ways of supporting Japanese learners of English into increasing their retention of multi-word strings, by testing their capacity to attend to phonological features within them. A map of the review questions (in blue), and the research questions they generated (in red), can be seen in Figure 8.1. The themes, while presented in linear order in the thesis, relate in certain less linear ways, such that, for example, Research Question 3 (RQ3), addressed in Chapter 5, derives from a different part of the Chapter 2 literature review than RQs 1, 2 and 4 (Chapters 3, 4, 6 and 7 respectively). This is not uncommon in research, which often benefits from developing a web of enquiry rather than just a single line.

Figure 8.1 Thesis map with review questions and research questions (RQ).



The review of the literature in Chapter 2 began with two general observations: that phonological patterns form one aspect of ludic language, and that they are often implicated in theories of L1 acquisition. This suggests an innate human predisposition not only towards the perception of sound patterns, but also towards the value of such patterns in verbal play, poetry and prose. Phonological patterns also appear as poetic devices in theories pertaining to the oral transmission of epics and ballads, in which they act to restrict choice and serve as memory cues. More specific evidence from the literature on advertising and marketing was assessed, suggesting that rhetorical devices play a role in affective factors involved in language processing; that is, people tend to evaluate brand names and advertisements more favourably when phonological patterns are present.

Findings from cognitive and experimental psychology were reviewed, indicating that phonological patterns function as part of the resonance process, activating matching elements and influencing both memory processes and speech production. Indeed, given the pervasiveness across a wide variety of academic fields of the notion that phonological patterns influence linguistic processing, it is somewhat surprising that they only sparked interest in Second Language Acquisition research so recently. The experimentation by the team of researchers led by Boers has substantiated claims that at least some rhetorical devices aid memory, by showing that English L2 learners recall the forms of collocations and compounds more readily when there are form-based similarities. The notion of an intrinsic responsiveness to phonological similarities is bolstered in those studies in which there is no element of drawing the learners' attention to the patterns, yet a mnemonic effect is still in evidence. The empirical work in this thesis has extended these findings to a wider population, whose L1 phonology some said would mask any such patterns. A natural sensitivity to phonological patterns was observed in the pseudoword-pair experiment (Chapter 6). The evidence presented here also demonstrates a mnemonic advantage not only for the written forms of novel language but also for the meanings.

The cognitive process of priming was considered a credible explanatory mechanism underpinning the mnemonic effect. The priming paradigm has produced robust evidence of a priming effect for rhyming stimuli, though the evidence for alliteration is inconclusive and assonance alone has not been the focus of much study. In the end, the data in this thesis did not provide solid support for the conjecture that the mnemonic effect is grounded on a priming process. Two reasons were suggested: the study may have been underpowered, and the experimental design, which operationalised alliteration and assonance at the phoneme level, may not have been optimal. This line of enquiry invites further examination.

8.3 Towards some new insights for L2 teaching and learning

The empirical work in this thesis has practical implications for L2 pedagogy. It may come as no surprise that studying word-pairs improves recall. The literature on memory indicates that consolidation processes recur over time with repetition and rehearsal, so learning is more than just a question of input. Furthermore, material can only be consolidated in memory if it has been stored sufficiently to be built on, and there is always a risk with vocabulary learning that the initial input leaves an insufficient memory trace. On that basis, any kind of feature of the material that helps hook it into memory, even for the short term, will give it more chance of that consolidation. Though it might be tempting for teachers to draw their students' attention to overlapping phonological or orthographic features, the evidence in Chapter 5, suggests that learners need spare cognitive capacity. Perhaps it is better if the learners already know the meaning of at least one constituent word, as this can reduce the learning burden and makes the use of phonological / orthographic features easier to deploy. Without previous semantic knowledge or partial knowledge of the stimuli, it is too much of a cognitive burden and attention to phonological features may hinder rather than help (see Chapter 6).

In sum, notwithstanding the capacity for phonological patterns (aided perhaps by orthographic ones - see Chapter 7) to support learning, there are limits. If a multiword string contains too much unfamiliar material, the effects of phonological patterning may not be sufficient for effective learning.

8.4 Reflections on the processes and outcomes of this research

> One of the most important factors which influences what happens in the classroom is the totality of ideas, knowledge and attitudes which represent the teacher's mind-set. (Lewis, 1993, p. 32)

It was noted at the outset of this thesis that *The Lexical Approach* (Lewis, 1993) informed much of my early approach to vocabulary teaching. Irrespective of the specific influence of the research in this thesis on my daily practice as an EFL teacher, there have been significant benefits for my general learning and development as an academic researcher, which will stay with me for the rest of my career. Research is sometimes said to be a journey, and much is learnt on the way. In my case I have benefitted from learning progressively about the challenges of empirical design; the vagaries of language; the unpredictability of participants; the opaqueness of much previous research; and the complexities of statistical analysis. Not every experimental procedure carried out was perfectly designed and executed, and as pointed out in the relevant chapters, that may sometimes explain the pattern of results. However, the thesis as a whole presents a set of findings that generally match previous research and help to extend our understanding of some under-researched questions. In relevant places, particularly in Chapter 7, an attempt has been made to tease apart some of the issues, to explore alternative explanations and suggest some fruitful directions for future research. In these regards, it is hoped that the assembled work does offer a significant contribution to new knowledge.

References

Abbuhl, R. (2012). Why, when, and how to replicate research. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (First ed., pp. 296-312). Oxford: Blackwell Publishing Ltd.

Ackermann, K., & Chen, Y.-H. (2013). Developing the Academic Collocation List (ACL). A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*, 235-247.

Adelman, J., & Estes, Z. (2013). Emotion and memory: a recognition advantage for positive and negative words independent of arousal. *Cognition, 129*(530-535).

Aitchison, J. (1987). *Words in the mind: An introduction to the mental lexicon*. Oxford: Basil Blackwell Ltd.

Akamatsu, N. (2003). The effects of first language orthographic features on second language reading in text. *Language Learning, 53*, 207-231.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 101-122). Oxford: Clarendon Press.

Ambridge, B. (2020). Against stored abstractions: a radical exemplar model of language acquisition. *First Language, 40*, 509-559.

Anderson, J. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior, 22*, 261-295.

Argo, J. J., Popa, M., & Smith, M. C. (2010). The sound of brands. *Journal of Marketing, 74*(4), 97-109.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). *The CELEX lexical database (CD-Rom)*.

Baddeley, A. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*(4), 362-365.

Baddeley, A. (2001). The concept of episodic memory. *Philosophical Transactions of the Royal Society of London. Series B, 356*(1413), 1345.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173.

Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation, 8*, 47-89.

Baddeley, A., Thomas, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior, 14*, 575-589.

Baldick, C. (1996). *The concise Oxford dictionary of literary terms*: Oxford University Press.

Balota, D., & Chumbley, J. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 340-357.

Barcroft, J. (2002). Semantic and Structural Elaboration in L2 Lexical Acquisition. *Language Learning, 52*, 323-363.

Beckman, M. (1982). Segment duration and the 'mora' in Japnese. *Phonetica, 39*, 113-135.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*, 101-118.

Benczes, R. (2013). The role of alliteration and rhyme in novel metaphorical and metonymical compounds. *Metaphor and Symbol, 28*, 167-184.

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research, 18*(1), 54-74. doi:10.1177/1362168813505395

Boers, F., Eyckmans, J., & Lindstromberg, S. (2014). The effect of a discrimination task on L2 learners' recall of collocations and compounds. *International Journal of Applied Linguistics, 24*(3), 357-369. doi:10.1111/ijal.12033

Boers, F., & Lindstromberg, S. (2005). Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. *System, 33*(2), 225-238. doi:10.1016/j.system.2004.12.007

Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*: Springer.

Boers, F., Lindstromberg, S., & Eyckmans, J. (2012). Are alliterative word combinations comparatively easy to remember for adult learners? *RELC Journal, 43*(1), 127-135. doi:10.1177/0033688212439997

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014a). Is alliteration mnemonic without awareness-raising? *Language Awareness, 23*(4), 291-303. doi:10.1080/09658416.2013.774008

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014b). When does assonance make L2 lexical phrases memorable? *The European Journal of Applied Linguistics and TEFL., 3*(1), 93-107.

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014c). Some explanations for the slow acquisition of L2 collocations. *Vigo International Journal of Applied Linguistics, 11*.

Boers, F., Lindstromberg, S., & Webb, S. (2014). Further evidence of the comparative memorability of alliterative expressions in second language learning. *RELC Journal, 45*(1), 85-99. doi:10.1177/0033688214522714

Boers, F., & Stengers, H. (2008). Adding sound to the picture: Motivating the lexical composition of metaphorical idioms in English, Dutch and Spanish. In M. S. Zanotto, L. Cameron, & M. C. Cavalcanti (Eds.), *Confronting metaphor in use: An applied linguistic approach* (pp. 64-78). Amsterdam: John Benjamins Publishing Company.

Bower, G., & Bolton, L. (1969). Why are rhymes easy to learn? *Journal of experimental psychology, 82*, 453-461.

Brendl, C., Miguel, E., & Higgins, E. (1996). Principles of judging valence: What makes events positive or negative? *Advances in Experimental Social Psychology, 28*, 95-160.

Browne, C. (2014). A new general service list: The better mousetrap we've been looking for. *Vocabulary Learning and Instruction, 3*(1), 1-10.

Brysbaert, M. (2003). Bilingual visual word recognition: Evidence from masked phonological priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: State-of-the-art* (pp. 178-189). Hove, U.K.: Psychology Press.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of cognition, 1*(9), 1-20.

Brysbaert, M., Van Dyck, G., & Van de Poel, M. (1999). Visual word recognition in bilinguals: Evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 137-148.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*, 904-911.

Burton-Roberts, N., Carr, P., & Docherty, G. (2000). *Phonological knowledge: conceptual and empirical issues*. Oxford: Oxford University Press.

Cattell, J. M. (1887). Experiments on the association of ideas. *Mind, 12*, 68-74.

Charteris-Black, J. (1999). The survival of English proverbs: A corpus-based account. *De Proverbio, 5*, 262-288.

Church, B., & Fisher, C. (1998). Long-term auditory word priming in preschoolers: Implicit memory support for language acquisition. *Journal of Memory and Language, 39*, 523-542.

Clark, J., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review, 3*, 149-210.

Clark, J., Yallop, C., & Fletcher, J. (2007). *An introduction to phonetics and phonology* (3rd ed.): Blackwell.

Clopper, C. (2002). Frequency of stress patterns in English: A computational analysis. *IULC Working Papers Online, 2*(2), 1-9.

Cobb, T. (n.d.). VocabProfile (Version 8). Available from <http://www.lextutor.ca>.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33*(A), 497-505.

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term-memory recall of repeated and novel word lists. *Memory & Cognition, 21*(4), 539-545.

Conrad, R., & Hull, A. (1964). Information, acoustic confusion and memory span. *British journal of psychology, 55*(4), 429-432.

Conway, M. A., & Pleydell-Pearce, C. (2000). The construction of autobiographical memories in the self memory system. *Psychology Review, 107*, 261-288.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11*(6), 671-684.

Crystal, D. (1997). *The Cambridge encyclopedia of language* (2nd ed.). Cambridge: Cambridge University Press.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31*(2), 218-236.

Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2*(3-4), 133-142.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*(4), 385-400.

Davies, M. (2008 to present). The corpus of contemporary American English. *Available at* <http://www.americancorpus.org>.

Davis, D., Bagchi, R., & Block, L. (2016). Alliteration alters: Phonetic overlap in promotional messages influences evaluations and choice. *Journal of Retailing, 92*(1), 1-12.

de Groot, H. (1999). *Van dale idioomwoordenboek: Verklaring en herkomst van uitdrukkingen en gezegden*. Antwerp: Van Dale Lexicografie.

de Saussure, F. (1916). *Course in general linguistics (Roy Harris, Trans.)*. London: Duckworth.

Defior, S., & Serrano, F. (2017). Learning to read Spanish. In L. Verhoeven & C. Perfetti (Eds.), *Learning to read across languages and writing systems* (pp. 270-298): Cambridge University Press.

Degé, F., Patscheke, H., & Schwarzer, G. (2017). Associations between two measures of music aptitude: Are the IMMA and the AMMA significantly correlated in a sample of 9 to 13-year-old children? *Musicae Scientiae, 21*(4), 465-478.

Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.

Demb, J., Desmond, J., Wagner, A., Vaidya, C., Glover, G., & Gabrieli, J. (1995). Semantic encoding and retrieval in the Left Inferior Prefrontal Cortex: a Functional MRI study of task difficulty and process specificity. *The Journal of Neuroscience, 15*, 5870-5878.

Detey, S., & Nespoulous, J.-L. (2008). Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua, 118*, 66-81.

Dörnyei, Z., & Schmidt, R. (2001). *Motivation and second language acquisition* (Vol. 23): National Foreign Language Resource Center.

Dowker, A. (1989). Rhyme and alliteration in poems elicited from young children. *Journal of Child Language, 16*(1), 181-202.

Dumay, N., Benraïss, A., Barriol, B., Colin, C., Radeau, M., & Besson, M. (2001). Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. *Journal of Cognitive Neuroscience, 13*, 121-143.

Eichenbaum, H. (2010). Memory systems. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 478-490.

Ellis, N. C. (1997). Vocabulary acquisition: Word structure, collocation, grammar, and meaning. In M. McCarthy & N. Schmidt (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 122-139). Cambridge: Cambridge University Press.

Ellis, N. C. (2003). Constructions, chunking and connectionism: the emergence of second language structure. In M. Doughty & M. Long (Eds.), *Handbook of second language acquisition*. Oxford: Pergamon Press.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus and TESOL. *TESOL Quarterly, 42*(3), 375-396.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29-62.

Eyckmans, J., Boers, F., & Lindstromberg, S. (2016). The impact of imposing processing strategies on L2 learners' deliberate study of lexical phrases. *System, 56*, 127-139. doi:10.1016/j.system.2015.12.001

Eyckmans, J., & Lindstromberg, S. (2017). The power of sound in L2 idiom learning. *Language Teaching Research, 21*(3), 341-361.

Fallon, A., Groves, K., & Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT, but not MAN. *International Journal of Psychology, 34*(5-6), 301-307.

Ferrand, L., & Grainger, J. (1994). Effects of orthography are independent of phonology in masked form priming. *The Quarterly Journal of Experimental Psychology, 47*, 365-382.

Fitzpatrick, T. (2006). Habits and rabbits: word associations and the L2 lexicon. *EUROSLA yearbook, 6*, 121-145.

Fitzpatrick, T., & Thwaites, P. (2020). Word association research and the L2 lexicon. *Language Teaching, 53*, 237-274.

Fliessbach, K., Weis, S., Klaver, P., Elger, C., & Weber, B. (2006). The effect of word concreteness on recognition memory. *NeuroImage, 32*, 1413-1421.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers, 35*(1), 116-124.

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27-52.

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305-327.

Gathercole, S., & Adams, A. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language, 33*, 672-688.

Gerrig, R. J., & O'Brien, E. J. (2005). The scope of memory-based processing. *Discourse Processes, 39*(2-3), 225-242.

Gibbs, P., & Van Orden, G. (1998). Pathway selection's utility for control of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1162-1187.

Giegerich, H. J. (1992). *English phonology: An introduction*: Cambridge University Press.

Goldinger, S., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: the roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1211-1238.

Gordon, E. E. (1979). Developmental music aptitude as measured by the Primary Measures of Music Audiation. *Psychology of Music, 7*(1), 42-49.

Gordon, E. E. (1986). *Intermediate measures of music audiation: A music aptitude test for children, grades one through four*. Chicago: GIA Publications.

Gordon, E. E. (1989). *Manual for the advanced measures of music audiation*. Chicago: GIA Publications.

Greenberg, D., & Verfaellie, M. (2010). Interdependence of episodic and semantic memory: evidence from neuropsychology. *Journal of the International Neuropsychological society 16*, 748-753.

Greene, R., Cushman, S., Cavanagh, C., Ramazani, J., Rouzer, P., Feinsod, H., . . . Slessarev, A. (2012). *The Princeton encyclopedia of poetry and poetics*: Princeton University Press.

Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition*. London: Wiley.

Gries, S. (2006). Corpus-based methods and cognitive semantics: The many senses of *to run*. In S. Gries & A. Stefanowitsch (Eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis* (pp. 57-100). Berlin: De Gruyer Mouton.

Gries, S. (2011). Phonological similarity in multi-word units. *Cognitive Linguistics, 22*(3), 491-510.

Grimes, J., & Agard, F. (1959). Linguistic divergence in Romance. *Language, 35*(4), 598-604.

Grossberg, S., & Stone, G. (1987). Neural dynamics of word recognition and recall: attentional priming, learning, and resonance. In S. Grossberg (Ed.), *The adaptive brain II: vision, speech, language, and motor control*. The Netherlands: Elsevier Science Publishers.

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Re-examining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition, 33*(6), 1001-1016.

Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. Cambridge: MIT Press.

Hansen, D., Bernstof, E., & Stuber, G. (2014). *The music and literacy connection* (2nd. ed.): Rowman & Littlefield.

Henson, R. (2003). Neuroimaging studies of priming. *Progress in neurobiology, 70*, 53-81.

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *The Quarterly Journal of Experimental Psychology Section A, 58*, 193-233.

Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 47-67). Hove: LTP.

Holmquist, J., & Cudmore, B. (2013). English in Japanese and Philippine advertising: An exploratory comparison. *Journal of Global Business Management, 9*, 82-91.

Horner, A. J., & Henson, R. (2008). Priming, response learning and repetition suppression. *Neuropsychologia, 46*, 1979-1991.

Horst, M. (2013). Mainstreaming second language vocabulary acquisition. *Canadian Journal of Applied Linguistics, 16*, 171-188.

Hudson, S. B., & Tanenhaus, M. K. (1985). Phonological code activation during listening. *Journal of psycholinguistic research, 16*, 557-567.

Hughes, C., Filimonov, M., Wray, A., & Spasić, I. (2021). Leaving no stone unturned: Flexible retrieval of idiomatic expressions from a large text corpus. *Machine Learning and Knowledge Extraction, 3*, 263-283.

Inagaki, K., Hatano, G., & Otake, T. (2000). The effect of Kana literacy acquisition on the speech segmentation unit used by Japanese young children. *Journal of experimental child psychology, 75*, 70-91.

Irujo, S. (1986). A piece of cake: Learning and teaching idioms. *ELT Journal, 40*, 236-242.

Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 2*, 2-36.

Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E.-J. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp. 133-165). New York: Psychology Press.

Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York: Routledge.

Jusczyk, P., Goodman, M., & Baumann, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language, 40*, 62-82.

Kaneko, M. (2013). Estimating the reading vocabulary-size goal required for the Tokyo University entrance examination. *The Language Teacher, 37*, 40-45.

Katada, F. (1990). On the representation of moras: evidence from a language game. *Linguistic inquiry, 21*, 641-646.

Katkov, M., Romani, S., & Tsodyks, M. (2014). Word length effect in free recall of randomly assembled word lists. *Frontiers in computational neuroscience, 8*, 1-4.

Kelley, K., & Preacher, K. (2012). On effect size. *Psychological Methods, 17*, 137-152.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627-633.

Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience, 28*, 1224-12230.

Kikuchi, K. (2013). Demotivators in the Japanese EFL context. In M. Apple, D. Da Silva, & T. Fellner (Eds.), *Language learning motivation in Japan* (pp. 206-224). UK: Mulitilingual Matters.

Koda, K. (2017). Learning to read Japanese. In L. Verhoeven & C. Perfetti (Eds.), *Learning to Read Across Languages and Writing Systems* (pp. 57-81). Cambridge: Cambridge University Press.

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech, 32*(3), 249-278.

Kubozono, H. (2002). Prosodic structure of loanwords in Japanese: Syllable structure, accent and morphology. *Journal of the Phonetic Society of Japan, 6*, 79-97.

Kubozono, H., & Ota, S. (1998). On-in kozo to akusento (Phonological structure and accent). *Tokyo: Kenkyusha*.

Kučera, H., & Francis, W. (Eds.). (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Labrune, L. (2012). Questioning the universality of the syllable: evidence from Japanese. *Phonology, 29*, 113-152.

Ladefoged, P. (1982). *A course in phonetics* (2nd ed.): Oxford Univeristy Press.

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38*(3), 245-294.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Lea, R., Rapp, D., Elfenbein, A., Mitchel, A., & Romine, R. (2008). Sweet silent thought: Alliteration and resonance in poetry comprehension. *Psychological Science, 19*(7), 709-716.

Lederer, R. (2014). *Get Thee to a Punnery: an Anthology of Intentional Assualts Upon the English Language*.

Lee, Y. (2016). A Lexicon and Behavioral Study on the Status of the Syllable in Japanese. *Language Research, 52*, 115-137.

Lehecka, & Tomas. (2015). Collocation and colligation. In *Handbook of pragmatics online*: Benjamins.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady, 10*(8), 707-710.

Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Hove: Language Teaching Publications.

Lindstromberg, S. (2012). Consonance in English proverbs and collocations. *Humanising Language Teaching, 14*, 16-25.

Lindstromberg, S. (2017). Surplus interword phonological similarity in English multiword units. *Corpus Linguistics and Linguistic Theory, 16*, 95-118.

Lindstromberg, S., & Boers, F. (2008a). The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics, 29*(2), 200-222. doi:10.1093/applin/amn007

Lindstromberg, S., & Boers, F. (2008b). Phonemic repetition and the learning of lexical chunks: The power of assonance. *System, 36*, 423-436.

Lindstromberg, S., & Eyckmans, J. (2014). How big is the positive effect of assonance on the recall of L2 collocations? *International Journal of Applied Linguistics, 165*(1), 19-45.

Lindstromberg, S., & Eyckmans, J. (2017). The particular need for replication in the quantitative study of SLA: A case study of the mnemonic effect of assonance in collocations. *Journal of the European Second Language Association, 1*(1), 126-136.

Liu, D. (2010). Going beyond patterns: Involving cognitive analysis in the learning of collocations. *TESOL Quarterly, 44*(1), 4-30.

Logan, G. (1990). Repetition priming and automaticity: common underlying mechanisms? *Cognitive Psychology, 22*, 1-35.

Lowel, S., & Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science, 255*, 209-212.

Lowrey, T., Shrum, L., & Dubitsky, T. (2003). The relation between brand name and linguistic characteristics and brand name memory. *Journal of Advertising, 32*(3), 7-17.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review, 7*(4), 618-630.

Luce, P. A., Goldinger, S., Auer, E., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & psychophysics, 62*(615-625).

MacKay, D. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*(3), 323-350.

Mackey, A. (2012). Why (or why not), when, and how to replicate research. In G. Porte (Ed.), *Replication research in applied linguistics*. Cambridge: Cambridge University Press.

Marsden, E. (2009). What is 'priming' and what might priming techniques be able to tell us about L2 learning and proficiency. In A. Benati (Ed.), *Issues in second language proficiency* (pp. 9-23).

Marslen-Wilson, W. (1984). Function and process in spoken word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes.* Hillsdale: Erlbaum.

Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting.*: Psychology Press.

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

McCarthy, M. (1998). *Spoken languages and applied linguistics*. Cambridge: Cambridge University Press.

McCawley, J. D. (1968). *The phonological component of a grammar of Japanese.* The Hague: Mouton & Co.

McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: An account of basic findings. *Psychological Review, 88*(5), 375.

McDonough, K., & Trofimovich, P. (2009). *Using priming methods in second language research*. New York: Routledge.

McGrath, R., & KMeyer, G. (2006). When effect sizes disagree: The case of r and d. *Psychological Methods, 2*, 386-401.

McLean, S., Hogg, S., & Kramer, B. (2014). Estimations of Japanese university learners' English vocabulary sizes using the vocabulary size test. *Vocabulary Learning and Instruction, 3*, 47-55.

McMahon, A. (2002). *An introduction to English phonology*. Edinburgh: Edinburgh University Press.

McMahon, A., Heggarty, P., McMahon, R., & Maguire, W. (2007). The sound patterns of Englishes: Representing phonetic similarity. *English Language and Linguistics, 11*(1), 113-142.

McQuarrie, E., & Mick, D. (1992). On resonance: A critical pluralistic inquiry into advertising rhetoric. *Journal of consumer research, 19*(2), 180-197.

McQuarrie, E., & Mick, D. (1996). Figures of rhetoric in advertising language. *Journal of consumer research, 22*(4), 424-438.

McQueen, J., & Sereno, J. (2005). Cleaving automatic processes from strategic biases in phonological priming. *Memory & Cognition, 33*, 1185-1209.

Merritt, P., DeLosh, E., & McDaniel, M. (2006). Effects of word frequency on individual-item and serial order retention: Tests of the order-encoding view. *Memory & Cognition, 34*(8), 1615-1627. doi:10.3758/BF03195924

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of experimental psychology, 90*(2), 227.

Minkova, D. (2002). Ablaut reduplication in English: The criss-crossing of prosody and verbal art. *English Language and Linguistics, 6*, 133-169.

Monsell, S., & Hirsh, K. (1998). Competitor priming in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1495-1520.

Moody, A. J. (2000). Beyond ‘‘shooby-dooby-doo-wah’’: An examination of English lyrics in Japanese pop music. *Gengo Bunka [Language and Culture], 8*, 1-8.

Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*: Oxford University Press.

Moon, R. (1999). Needles and haystacks, idioms and corpora: Gaining insights into idioms, using corpus analysis. In T. Herbst & K. Popp (Eds.), *The perfect learner's dictionary (?)* (pp. 265-282): De Gruyter.

Moon, R. (2015). Idioms: A view from the web. *International Journal of Lexicography, 28*, 318-337.

Morais, J., Cary, L., Alegría, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323-331.

Morais, J., & Kolinsky, R. (2019). Phonological abilities in fully illiterate adults. In D. Perin (Ed.), *The Wiley handbook of adult literacy* (pp. 41-61): wiley Blackwell.

Murphy, C. (1999). How far do tests of musical ability shed light on the nature of musical intelligence? *British Journal of Music Education, 16*(1), 39-50.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*(2-3), 131-157.

Nakayama, M., Sears, C., Hino, Y., & Lupker, S. J. (2012). Cross-script phonological priming for Japanese-English bilinguals: Evidence for integrated phonological representations. *Language and Cognitive Processes, 27*, 1563-1583.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Heinle and Heinle Publishers.

Nation, I. S. P. (2001). *Learning vocabulary in another language*: Cambridge : Cambridge University Press.

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston: Heinle.

Nation, I. S. P. (2012). Information on the BNC/COCA lists.

Nation, I. S. P. (2014). *What do you need to know to learn a foreign language?* Available at http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/foreign-language_1125.pdf.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*, 9-13.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*: Heinle, Cengage Learning.

Nesselhauf, N. (2005). *Collocations in a learner corpus* (Vol. 14). Amsterdam: John Benjamins Publishing.

Norris, D., McQueen, J., & Cutler, A. (2002). Bias effects in facilitatory phonological priming. *Memory & Cognition, 30*, 399-411.

Otaka, H. (2009). Effects of the bimoraic foot on Japanese perception of rhymes from Japanese and English sequences. *Ex: エクス: 言語文化論集*(6), 1-22.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32*(2), 258-278.

Parkinson, D. (Ed.) (2006). Oxford University Press.

Patel, A. D. (2012). Language, music, and the brain: A resource-sharing framework. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 204-223). Oxford: Oxford University Press.

Patterson, K., Nestor, P., & Rogers, T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature reviews neuroscience, 8*, 976-987.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191-226). New York: Longman.

Perea, M., Vergara-Martinez, M., & Gomez, P. (2015). Resolving the locus of case alternation effects in visual word recognition: Evidence from masked priming. *Cognition, 142*, 39-43.

Perfetti, C., & Liu, Y. (2005). Orthography to phonology and meaning: comparisons across and within writing systems. *Reading and Writing, 18*, 193-210.

Philip, G. (2005). *Discovering clines of variation: The location and analysis of non-canonical forms in general reference corpora*. Paper presented at the Phraseology 2005. The many faces of Phraseology. An interdisciplinary conference, Louvain, Belgium.

Pierrehumbert, J. (2001). Exemplar dynamics: word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (Vol. 45, pp. 137-158): John Benjamins Publishing.

Pierrehumbert, J., Beckman, M., & Ladd, R. (2000). Conceptual foundations of phonology as a laboratory science. In N. Burton-Roberts, P. Carr, & G. Docherty (Eds.), *Phonological knowledge: Conceptual and empirical issues* (pp. 273-304). Oxford: OUP.

Plonsky, L., & Oswald, F. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912.

Port, R. F. (2007). How are words stored in memory? Beyond phones and phonemes. *New ideas in psychology, 25*, 143-170.

Poushter, J., Bishop, C., & Chwe, H. (2018). Social media use continues to rise in developing countries but plateaus across developed ones. *Pew research center,* (22), 2-19.

Radeau, M., Morais, J., & Dewier, A. (1989). Phonological priming in spoken word recognition: Task effects. *Memory & Cognition, 17*, 525-535.

Rapp, D., & Samuel, A. (2002). A reason to rhyme: Phonological and semantic influences on lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*(3), 564-571.

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology, 53*, 97-145.

Read, C., Yun-Fei, Z., Hong-Yin, N., & Bao-Qing, D. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24*, 31-44.

Rey, A., & Chantreau, S. (1979). *Dictionnaire des expressions et locutions*: Le Robert.

Roediger, H. (1980). Memory metaphors in cognitive psychology. *Memory & Cognition, 8*, 231-246.

Rubin, D. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballads and counting-out rhymes*. New York: Oxford University Press.

Salamé, P., & Baddeley, A. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society*, *24*(4), 263-265.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*, 26-43.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88. doi:10.1191/026553201668475857

Schwanenflugel, P. J., Akin, C., & Luh, W.-M. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition, 20*, 96-104.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology, 45*, 21-50.

Shallice, T., & Cooper, R. P. (2011). *The organisation of mind*. New York: Oxford University Press.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review, 64*(3), 429-458.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2019). Formulaic language: Setting the scene. In A. Siyanova-Chanturia & A. Pellicaer-Sánchez (Eds.), *Understanding formulaic language: A second language acquistion perspective* (pp. 1-15). New York: Routledge.

Sloboda, J. (1985). *The musical mind: The cognitive psychology of music*: Oxford University Press.

Slowiaczek, L., Nusbaum, H., & Pisoni, D. B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 64-75.

Slowiaczek, L., & Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory & Cognition, 14*, 230-237.

Sneve, M., Grydeland, H., Nyberg, L., Bowles, b., Amlien, I., Langnes, E., . . . Fjell, A. (2015). Mechanisms underlying encoding of short-lived versus durable episodic memories. *Journal of Neuroscience, 35*, 5202-5212.

Sorhus, H. B. (1977). To hear ourselves: Implications for teaching English as a second language. *ELT Journal, 31*(3), 211-221.

Stamou, L., Schmidt, C. P., & Humphreys, J. T. (2010). Standardization of the Gordon primary measures of music audiation in Greece. *Journal of Research in Music Education, 58*(1), 75-89.

Strong, B., & Boers, F. (2019). Weighing up exercises on phrasal verbs: Retrieval versus trial-and-error practices. *The Modern Language Journal, 103*, 562-579.

Sutherland, A. (2019). www.Quizlet.com.

Sweegers, C., & Talamini, L. (2014). Generalization from episodic memories across time: a route for semantic knowledge acquisition. *Cortex, 59*, 49-61.

Szudarski, P. (2017). Learning and teaching L2 collocations: Insights from research. *TESL Canada Journal, 34*(3), 205-216.

Tamori, I., Tatsuki, D. H., & Tominaga, H. (2008). Rhetorical devices in Japanese advertisements: Towards a taxonomy. *Journal of cultural science, 43*, 59-115.

Tannenbaum, R., & Wylie, E. C. (2013). Mapping TOEIC® and TOEIC Bridge™ test scores to the Common European Framework of Reference. *The Research Foundation for the TOEIC Tests: A compendium of studies, 2*(6-1), 1-11.

Taylor, I., & Taylor, M. M. (1995). *Writing and literacy in Chinese, Korean and Japanese* (Vol. 3): John Benjamins Publishing Company.

Taylor, J. (2012). *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press.

Thomas, D. (1940). *Portrait of the artist as a young dog*. London: J. M. Dent & Sons.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition, 15*(1-3), 49-74.

Treiman, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of experimental child psychology, 39*(1), 161-181.

Treiman, R. (1986). The division between onets and rimes in English syllables. *Journal of Memory and Language, 25*(4), 476-491.

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: implications for pronunciation instruction. *The Modern Language Journal, 90*, 519-535.

Tsujimura, N. (2007). *An introduction to Japanese linguistics* (Second ed.). Singapore: Blackwell Publishing.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). New York: Academic Press.

Tulving, E. (2001). Episodic memory and common sense: How far apart? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 356*, 1505-1515.

Tulving, E., & Schacter, D. (1990). Priming and human memory systems. *Science, 247*(4940), 301-306.

UNESCO. (1997). Standard classification of education (ISCED).

Vance, T. (1987). *An introduction to Japanese phonology*. New York: State University of New York Press.

Vance, T. (2017). *The Japanese syllable debate: A skeptical look at some anti-syllable arguments.* Paper presented at the GLOW in Asia XI.

Verhoeven, L. (2017). Learning to read Dutch. In L. Verhoeven & C. Perfetti (Eds.), *Learning to Read across Languages and Writing Systems* (pp. 323-346). Cambridge: Cambridge University Press.

Verhoeven, L., & Perfetti, C. (2017). Operating principles in learning to read. In L. Verhoeven & C. Perfetti (Eds.), *Learning to read across languges and writing systems*: Cambridge University Press.

Vroomen, J., Van Zon, M., & De Gelder, B. (1996). Cues to speech segmentation: evidence from juncture misperceptions and word spotting. *Memory & Cognition, 24*, 744-755.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning, 63*(1), 91-120.

Wilks, C., & Meara, P. (2002). Untangling word webs: Graph theory and the notion of density in second language word association networks. *Second Language Research, 18*, 303-324.

Wilks, C., & Meara, P. (2007). Implementing graph theory approaches to the exploration of density and structure in L1 and L2 word association networks. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 167-181). Cambridge: Cambridge University Press.

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics, 32*(4), 430-449. doi:10.1093/applin/amr011

Woods, A., Velasco, C., Levitan, C., Wan, X., & Spence, C. (2015). Conducting perception research over the internet: A tutorial review. *PeerJ, 3*.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21*(4), 463-489.

Wray, A. (2002). *Formulaic language and the lexicon*. New York: Cambridge University Press.

Wray, A. (2015). Why are we so sure we know what a word is? In J. R. Taylor (Ed.), *The Oxford handbook of the word* (pp. 725-750). Oxford: Oxford University Press.

Yap, M., & Balota, D. (2015). Visual word recognition. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 26-43). Oxford: Oxford University Press.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*(5), 971-979.

Zevin, J., & Seidenberg, M. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47*(1), 1-29.

Zwitserlood, P. (1996). Form priming. *Language and Cognitive Processes, 11*(6), 589-596.

Appendix 1

Example Consent Form

I (Michael Green) would like to conduct research on English vocabulary learning. If you agree to participate in this research, your performance will NOT affect your grades for any of your university courses in anyway.

Your name will NOT appear in published results of this research. Likewise, any identifying information will NOT appear in published results. In sum, there will be no way to identify you personally as a participant.

You can withdraw from this study afterwards without giving a reason. You are free to ask questions at any time. The data collected from this research will be held confidentially.

If you agree to take part in this research, please sign the consent below:

Student Name: _________________________ Student No. ____________

Date: __________ Signature: _____________________________________

Appendix 2

Translation Equivalency

Please rate each item in terms of how easy it is to translate into Japanese.

1	**1.5**	**2**	**2.5**	**3**
Easy to translate				*Difficult to translate*

1.	gift list	1	1.5	2	2.5	3
2.	deep hole	1	1.5	2	2.5	3
3.	small talk	1	1.5	2	2.5	3
4.	soft cloth	1	1.5	2	2.5	3
5.	town house	1	1.5	2	2.5	3
6.	high price	1	1.5	2	2.5	3
7.	quick stop	1	1.5	2	2.5	3
8.	main road	1	1.5	2	2.5	3
9.	nice place	1	1.5	2	2.5	3
10.	fair share	1	1.5	2	2.5	3
11.	deep sea	1	1.5	2	2.5	3
12.	main gate	1	1.5	2	2.5	3
13.	high rate	1	1.5	2	2.5	3
14.	soft ground	1	1.5	2	2.5	3
15.	plain talk	1	1.5	2	2.5	3
16.	town square	1	1.5	2	2.5	3
17.	safe place	1	1.5	2	2.5	3
18.	check list	1	1.5	2	2.5	3
19.	fair deal	1	1.5	2	2.5	3
20.	quick trip	1	1.5	2	2.5	3

Appendix 3

Vocabulary Questionnaire

Last week you took part in a vocabulary experiment. How familiar were the phrases?

Tick (✓) one box for each phrase.

Phrase	I'd never seen / heard these 2 words together.	I had seen / heard these 2 words together.	I used these 2 words together in speaking and / or writing.
gift list			
deep hole			
small talk			
soft cloth			
town house			
high price			
quick stop			
main road			
nice place			
fair share			
deep sea			
main gate			
high rate			
soft ground			
plain talk			
town square			
safe place			
check list			
fair deal			
quick trip			

Appendix 4

Items for Lexical Decision Task

'No' Response Targets	'Yes' Response Targets	Assonating Pseudoword Primes	Alliterating Pseudoword Primes	No Pattern Pseudoword Primes
lude	band	cang	baws	frow
tode	skirt	earch	sluck	mants
pook	guide	sice	gream	hane
gide	farm	yark	finx	sove
frime	lawn	sauns	lask	chipe
pree	cream	blee	clons	parl
bofs	boss	sonk	bers	felds
crowl	mail	nade	mard	tunt
lacs	sauce	hort	speep	hool
chame	crown	floud	crope	blods
mooth	bank	slass	booh	shers
preet	slide	fike	skift	frace
fluck	truck	dunt	tast	speep
gake	shed	elts	shap	cass
noss	hand	lask	hame	freet
cews	rope	loat	rine	halm
chird	block	sogs	beft	lale
bape	frame	lape	frong	sonk
suide	cape	shain	cass	sowl
bave	food	stune	fied	bolk
creas	cloud	mouch	caze	nurge
fard	plane	snate	pless	loat
shen	bomb	frong	brac	fror
moule	nail	sace	nost	crope
cass	brain	caze	beals	polt
ross	train	paits	tals	goid
naze	mouth	sout	mool	coof

spide	lamb	tast	lubs	drig
gool	cash	sprag	cark	poth
lare	bark	lart	bace	geed
bame	path	sabs	pame	bily
dast	pole	soat	pape	sout
hase	dust	tunt	dack	scane
sorn	sink	fims	sowl	cound
lides	milk	skift	mouch	drate
tams	tape	frace	tunt	bals
nell	bell	helts	blods	hode
fage	tank	mants	traw	clons
pice	page	mape	parl	elts
shobs	salt	poth	sabs	geal
degs	cart	mard	cang	earch
plang	sheep	geed	shick	sice
tade	back	cass	bool	mibe
misk	shore	fror	sheel	dape
clead	plate	hane	plog	baws
cout	king	wrin	kire	nade
papt	bath	shap	bolk	dunt
cluff	wire	mibe	worb	clush
roat	note	stoge	noss	crief
brunk	light	fied	ludge	carn
stoil	wing	lints	wive	spood
cort	brick	pilt	bure	sauns
frake	cent	beft	cest	lart
clope	gate	lale	gaws	booh
wite	ship	gilk	shain	pless
dage	cage	bace	carn	soun
bere	whip	hins	wone	sluck
flars	meal	teast	mibe	wrin
lage	steel	freet	snate	mouls
sals	heart	carn	heaks	freen
tice	knee	speep	noke	boke
lant	coat	smole	clush	hins

baze	goal	slore	goid	hort
plub	pool	woob	polt	stam
marn	deck	bers	dest	stoge
grame	tide	hine	teast	shain
pade	home	crope	hort	pilt
whis	soil	goid	sove	noss
heach	flood	trush	frace	hine
hode	toast	noke	trush	closs
cabe	flame	dape	felds	nost
saire	chain	slame	chers	kire
kile	hedge	dest	hool	fike
sape	nest	shers	nade	woob
sheed	roof	booh	rown	blee
shis	soup	mool	stalt	wime
cood	flash	bracs	flate	stune
cags	belt	felds	bolf	stire
weet	face	tany	farn	woll
beld	cloth	plog	closs	lubs
pland	storm	gaws	slass	woot
sare	sheet	beals	shers	waps
steek	sore	boke	soun	bool
skilt	snake	pape	slore	tals
rool	brush	spum	bant	sogs
pome	cliff	shid	clice	helts
slood	pipe	mide	pilt	shick
bawn	rice	chipe	rone	teast
trame	tail	cace	tark	skift
ceal	child	gice	charp	heaks
wint	mouse	soun	mants	bant
herks	dawn	baws	dape	mide
toars	ball	traw	bals	trape
soop	lump	clush	lape	finx
sint	boat	hode	bads	slore
bope	gift	drig	gice	slass
brins	plug	ludge	poth	gice

brage	weed	freen	woob	dack
hink	lane	drate	lart	shap
blick	lamp	bads	lale	juffs
balt	hill	swip	hode	tany
tark	sand	bals	sace	beals

Appendix 5

Instructions for Lexical Decision Task

Welcome to the vocabulary experiment. In this experiment you will see a letter string (such as *wine* or *goter*) presented on the computer screen. Your task is to decide if the letter string is an English word or not. Press the YES button if it is an English word. Otherwise, press the NO button.

Once you are in the test room and ready to proceed, read the instructions on screen. After this, there will be 10 practice items. After the practice items, press the SPACE bar to begin the test.

We are interested in how quickly and accurately you can perform the task. Your response time and accuracy will be measured by the computer.

<u>**Please respond as quickly and accurately as possible.**</u>

Please do not work so fast that you make mistakes. If you find yourself making errors, slow down.

If you have any questions before, during or after the test, please feel free to ask.

Appendix 6

Vocabulary Card Technique

Step 1	Make the cards – write the useful English word or phrase or sentence on one side, and the Japanese translation on the other.
Step 2	Shuffle the cards. Look at the English, say it aloud and try to recall the translation. If you can, check, then put it on the bottom of the pack. If you *can't*, check, then put the card in the middle of the pack.
Step 3	Go through the pack for 5-10 minutes (don't forget to say the English). Put it away for half an hour, then repeat step 2.
Step 4	Keep adding useful words / phrases / sentences to the pack until you have about 50 cards. Then start a new pack. Go through the packs at increasingly spaced intervals of time.
Step 5	When you can recall *all* the translations easily, turn the cards over. Look at the Japanese and try to recall and say the English. Check. If you're right, put the card on the bottom of the pack, if you can't say the English, check, then put it in the middle of the pack.

Adapted from Nation (2008, p. 106)

Appendix 7

Music Questionnaire

Circle the number that describes you best.

1 = no / not at all **7 = yes / very well**

1. I studied music in High School	**1**	**2**	**3**	**4**	**5**	**6**	**7**
2. I was in a music club in High School	**1**	**2**	**3**	**4**	**5**	**6**	**7**
3. I was in a choir (聖歌隊) or singing club in High School	**1**	**2**	**3**	**4**	**5**	**6**	**7**
4. I played a musical instrument (楽器) in High School	**1**	**2**	**3**	**4**	**5**	**6**	**7**
5. I play a musical instrument now	**1**	**2**	**3**	**4**	**5**	**6**	**7**
6. I can read music (楽譜)	**1**	**2**	**3**	**4**	**5**	**6**	**7**
7. I can write music (楽譜)	**1**	**2**	**3**	**4**	**5**	**6**	**7**
8. I can sing	**1**	**2**	**3**	**4**	**5**	**6**	**7**
9. I think I am a musical person	**1**	**2**	**3**	**4**	**5**	**6**	**7**