The effects of different learning conditions on the development of collocational knowledge in a second language

by

Mark Toomer

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy
in Applied Linguistics

Victoria University of Wellington

2019

Abstract

Previous research has shown that incidental exposure to second-language collocations in reading texts can produce gains in declarative collocational knowledge. However, there is little evidence that incidental exposure leads to the acquisition of procedural collocational knowledge. One key study that investigated these two areas was Sonbul and Schmitt (2013). They found that advanced non-native speakers gained substantial declarative knowledge of low-frequency technical collocations after three exposures in two incidental reading conditions in one treatment session; typographic enhancement of the collocations produced more correct answers than no enhancement. However, the researchers found no evidence of procedural collocational knowledge in a primed lexical decision task.

Experiment 1 in this thesis was a conceptual replication and extension of Sonbul and Schmitt (2013). In a counter-balanced learning and experimental condition, 62 advanced adult ESL speakers were exposed to nine occurrences of 15 low-frequency technical collocations in 500-word texts in three sessions on two consecutive days. (input flooding). Three incidental learning treatments were implemented: reading-only (typographically unenhanced), bolding-only and bolding-plus-glossing. Collocational knowledge was assessed in two tests of declarative knowledge: a cued-recall (gapfill) test and a form-recognition (multiple-choice) test. Procedural collocational knowledge was operationalised as a priming effect in a primed lexical decision task. The results of the immediate cued-recall and form-recognition post-tests corroborate Sonbul and Schmitt's findings: multiple encounters with the collocations produced substantial declarative collocational knowledge, and more declarative knowledge was produced through exposure to typographically-enhanced collocations (with and without glosses) than to typographically-unenhanced collocations. Procedural knowledge was found, but, unexpectedly, only in the reading-only treatment.

Experiment 2 focused on non-technical lexical (verb + noun) collocations and grammatical (preposition + noun) collocations. Two incidental learning conditions were used: bolding and no-bolding. Seventy-eight intermediate-to-upper-intermediate-level adult native speakers of Chinese were exposed to six occurrences of each of 48 non-technical English collocations as they read twelve 170-word treatment texts on two consecutive days. The immediate post-test session comprised a gapfill (cued-recall) test,

for measuring declarative knowledge, and a self-paced reading task, for measuring procedural knowledge. The results show an increase in declarative knowledge in both learning conditions. Bolding produced more declarative knowledge of preposition + noun collocations than no bolding; however, bolding was no more effective than no bolding for verb + noun collocations. In the self-paced reading task, the *absence* of bolding of verb + noun collocations led to a tendency towards the development of procedural knowledge, but this was not the case for the typographically enhanced verb + noun collocations. For preposition + noun collocations (both unbolded and bolded) no evidence of procedural knowledge was found.

The findings of both experiments indicate that input flooding of collocations in incidental learning conditions can produce declarative collocational knowledge, and that typographic enhancement may lead to more declarative knowledge than lack of typographic enhancement. Repeated exposure to typographically-unenhanced collocations in reading contexts may produce procedural knowledge of collocations more quickly than exposure to typographically-enhanced collocations. These findings indicate that declarative and procedural knowledge of MWUs are dissociated in the sense that they follow different developmental paths. In a teaching context, I predict that the knowledge of collocations may be acquired incidentally through the use of texts such as graded readers which contain multiple instances of collocations.

Acknowledgements

I would like to thank Irina Elgort for her strong commitment to her role as my primary supervisor over the last few years. While encouraging me to be an autonomous researcher, she has also been vitally interested in the details of my research, guiding me when necessary and promptly giving me indispensable critical feedback.

I would also like to thank Averil Coxhead for advising me on the PhD process in a number of key areas such as ethics approval, brainstorming research ideas and planning the final stage of writing—not to mention general encouragement.

Thank you to other (present and past) academics at LALS at Victoria University for allowing me to discuss aspects of my research with them. Thank you also to Lisa Woods for her invaluable and patient help with the statistical analyses.

Thanks to Brian Strong for the many hours of discussion in which we grappled with the complexities of researching multiword units, and to the other LALS PhD students, especially the members of the PhD vocabulary group and my long-suffering officemates—you know who you are!

Thanks to my participants, who gave up several hours of their time on two or three days to attend my experimental sessions.

Thank you to Suhad Sonbul and Norbert Schmitt for making the materials from their experiment available to me to replicate.

I am grateful to Victoria University for providing me with a Victoria Doctoral Scholarship and Doctoral Submission Scholarship.

Finally, I would also like to thank my family, especially Chris Guerin, Rowan Toomer and Annie Toomer, for their inspiration, support and forbearance over the years.

TABLE OF CONTENTS

LIST OF TABLES

Tables

LIST OF FIGURES

Figures

1. Introduction

Several years ago I was a teacher of English for Academic Purposes (EAP) at a New Zealand university. My students, non-native speakers of English, were taking the EAP courses in order to gain direct entry into the university's undergraduate and postgraduate programmes. Many of those who were close to entering postgraduate programmes, especially PhD courses, had an advanced level of English: the grammar in their writing tended to be mostly accurate and their choices of lexical items were generally appropriate. Yet it was quite common for them to write combinations of two or more words that did not seem natural to native English speakers; in other words, the individual words which comprised the combination did not collocate. Classic examples of acceptable collocations are *powerful car* and *strong tea*, whereas we do not say *strong car.* The error with *strong car* is not grammatical but collocational.

I found errors with collocations that comprise a lexical word and a preposition to be particularly common. Learners seemed to have difficulty in choosing acceptable prepositions, a situation that I found puzzling. Why did they often write, for example, "…an increase *of* the number of…" instead of "an increase *in* the number of…"? Was it because the collocation *increase in* is uncommon in naturally-occurring language? Was it because prepositions are small words which are easily overlooked? Was it because learners confuse English prepositions with prepositions in their first languages? Was it because more than one preposition often fit acceptably into a certain slot next to a particular lexical word but with different functions (e.g., "… an increase *of* 550 ..." vs. "…an increase *in* the number of…")? What was the best way to help the students learn the correct prepositional collocates? Was it enough to correct their collocational errors when marking their writing? Was it enough to give them exercises on frequent collocation errors? These were the sorts of questions I asked myself and which motivated my interest in the learning of collocations, especially grammatical collocations, which appeared to be generally neglected in English-language textbooks (but see Prodromou, 1999). Several months later, I chose the learning of collocations as the focus of this PhD thesis, and grammatical collocations as an area of focus of Experiment 2 within the thesis.

In the remainder of this chapter, I briefly introduce the key areas of research which have been relevant to my investigation and identify gaps in the research which I

have attempted to address. I outline the overall purpose of my thesis research and its more specific aims and discuss some of the issues I needed to address in order to achieve the aims. I then define a number of the key terms in the thesis and, finally, I outline the remaining chapters.

1.1 Previous research and research gaps

A large proportion of language in spoken and written discourse is formulaic; that is, words tend to co-occur in set configurations that can be described as multiword chunks or units (e.g., *strong tea*, *increase in*, *kick the bucket*, *on the other hand*) (Hoey, 2005; Sinclair, 1991). Developing knowledge of multiword units (MWUs) is an essential aspect of learning a second language, but such development is slower than that of knowledge of individual words (Nation, 2013; Schmitt, 2010; Wray, 2012).

Research into the learning of second-language MWUs has, so far, examined mostly the acquisition of declarative knowledge of MWUs, which is conscious and analysable (R. Ellis, 1993). Only a small number of studies have investigated the development of procedural knowledge—a fluent, automatised type of knowledge (Suzuki & DeKeyser, 2017) of MWUs—or have compared the development of both declarative and procedural knowledge of MWUs. One pioneering study which did both, Sonbul and Schmitt (2013), found clear evidence for the development of explicit (declarative) knowledge of technical, lexical collocations after participant exposure to three repetitions of each of the collocations in written input but found no evidence of the development of implicit (procedural) knowledge.

Research into formulaic language suggests that knowledge and processing fluency of MWUs are increased through exposure to increased frequency (repetition) of MWUs in input—or input flooding—both in the first language and second language (e.g., Conklin & Schmitt, 2008; Durrant & Schmitt, 2010; N. Ellis, 2002; Sonbul & Schmitt, 2013; Webb, Newton, & Chang, 2013). However, more research is needed to trace the role of this frequency effect in the learning of MWUs, particularly in the development of fluency of processing of MWUs.

Further research is also required to clarify the role that different learning techniques play in the development of MWU knowledge. Typographic enhancement (e.g., bolding, italics) of MWUs in texts has been shown generally to have an advantage over no TE in declarative knowledge tests (e.g., Boers, Demecheleer, He, Deconinck,

Stengers, & Eyckmans, 2016; Sonbul & Schmitt, 2013; Szudarski & Carter, 2014), but to my knowledge no studies have yet shown a processing-speed advantage for either typographic enhancement or no typographic enhancement (see Sonbul & Schmitt, 2013). Almost no research has been conducted on the effectiveness of another technique, glossing, on the learning and processing of MWUs.

Another under-researched area in the learning of second-language MWUs is the development of knowledge of grammatical collocations (Durrant, 2009). This situation exists despite errors in grammatical collocations being common among even advanced language learners (e.g., Jimenez Catalan, 1996), but more research investigating how the knowledge of such collocations develops is needed.

1.2 Overall purpose and specific aims of research

The overall purpose of my thesis has been to address the research gaps described in section 1.1 by investigating the effects of different incidental learning techniques used in conjunction with input flooding on the development of declarative and procedural knowledge of written MWUs, specifically lexical and grammatical collocations. I did this firstly by conceptually replicating and extending Sonbul and Schmitt's (2013) study in Experiment 1, which investigated technical lexical collocations; next, I built on the findings of Experiment 1 by extending the scope of my investigation to include grammatical as well as lexical collocations in Experiment 2.

This research has had three substantive aims. The first has been to ascertain whether input flooding of written collocations in incidental learning conditions facilitates the development of either or both declarative knowledge and procedural knowledge of collocations. The second aim has been to find out which of the following types of input enhancement— typographic enhancement, typographic enhancement plus glossing and no typographic enhancement—are more effective at promoting the development of declarative and procedural collocational knowledge. The third substantive aim of the thesis has been to discover to what extent declarative and procedural knowledge of lexical (verb + noun) and grammatical (preposition + noun) collocations can be established through typographic enhancement and lack of typographic enhancement.

My research has also had a number of theoretical aims. One has been to trace the development of declarative and procedural knowledge of collocations and assess

whether, as Sonbul and Schmitt (2013) claim, the two knowledge types are dissociated—in other words, they follow different developmental paths. Another aim has been to assess the role of input flooding of collocations in texts in the development of collocational knowledge in the contexts of usage-based models and instance-based word-learning models (e.g., N. Ellis, 2007; Reichle & Perfetti, 2003). A third theoretical aim has been to identify the roles that different types of typographic enhancement and two levels of attention (with awareness and without awareness) play in the development of declarative and procedural collocational knowledge (e.g., Choi, 2017; Tomlin & Villa, 1994). The final aim has been to situate any recorded development of declarative and procedural collocational knowledge from glossing within depth-of-processing and related frameworks (e.g., Barcroft, 2002; Craik & Lockhart, 1972; Hulstijn, Hollander, & Greidanus, 1996).

With respect to methodological aims, alongside the more traditional pen-and-paper tests used to assess declarative knowledge I have employed techniques that have been used only in the last decade or so to measure the gaining of procedural knowledge of MWUs in SLA. For Experiment 1, I deployed a primed lexical decision task (LDT), which Sonbul and Schmitt had used in their 2013 study, and I used a self-paced reading task for Experiment 2.

1.3 Outline of thesis

I will now outline the remaining chapters in this thesis. In Chapter 2, I review the research and the theory relevant to my studies. In Chapter 3, I firstly outline Sonbul and Schmitt's (2013) study; then I describe the method I used for Experiment 1, in which I conceptually replicate and extend Sonbul and Schmitt's study; next I present the results of Experiment 1 and discuss the significance of the findings. In Chapter 4, I firstly describe the method used in Experiment 2, in which I focused on non-technical lexical and grammatical collocations, and then I present the results of Experiment 2 and discuss the findings. In Chapter 5, I consider the findings of the two experiments and issues arising from the findings, including theoretical and pedagogical implications, and suggestions for future research.

Chapter 2: Literature review

In this chapter, I review key research findings and theoretical perspectives related to my experiments. Firstly, I define MWUs and collocations. I then discuss the acquisition and processing of MWUs, particularly the reasons for the slowness of MWU acquisition in a second language. I outline the types of knowledge of MWUs and how they interact. I then discuss the roles of noticing and attention in second-language learning, including the facilitative role that input enhancement (of which I use various techniques in my experiments) can play. Finally, I examine the measures that have been used to assess knowledge of MWUs, some of which I deploy in my experiments.

2.1 Collocations and other multiword units

Mastery of multiword units, or set configurations of words, is a key aspect of mastery of a language. In fact, Schmitt (2010) argues that MWUs are "perhaps *the* essential element" of language (p. 146), while Howarth (1998b) claims that use of conventional collocations is essential for effective communication in academic writing. One reason that knowledge of multiword units is necessary for natural language use is that the co-occurrence of words is pervasive in spoken and written discourse (Hoey, 2005; Sinclair, 1991; Wray, 2002). Indeed, Renouf and Sinclair (1991, p. 143) claimed that "the normal use of language is to select more than one word at a time, and to blend such selections with each other." A number of researchers have estimated the extent of occurrence of multiword units in discourse. Biber, Johansson, Leech, Conrad and Finegan (1999) found that recurring lexical bundles of three words (including two-word contractions such as *don't want*) comprised 30% of their conversation corpus and 21% of their academic prose corpus. Approximately 55% of the spoken and written texts that Erman and Warren (2000) analyzed were multiword composites. Foster (2001) determined that 32% of unplanned classroom speech by 16 native speakers of English contained lexicalised sequences.

The term *multiword units* (MWUs) is just one label among dozens used in the linguistics and applied linguistics literature to describe recurring sequences of more than one word. Other terms include *formulaic sequences*, *formulaic language*, *lexical bundles*, *idioms*, *collocations*, *multi-word expressions*, *multi-word strings, formulas*, *chunks*, *clusters*, *composites*, *prefabricated routines, constructions*, *set phrases, lexicalised*

sequences, lexical phrases, *conventionalised forms*, and *unanalysed chunks of speech* (Lehecka, 2015; Wray, 2002). In this thesis, I will use the term *MWUs* as an all-encompassing term to refer to word sequences which recur in usage (Grant & Bauer, 2004; Nation, 2013; Nation & Webb, 2011). *Formulaic sequences* and *formulaic language* are frequently used as general-purpose terms for word strings (e.g., by Schmitt & Carter, 2004). A constraint of these terms is that they are frequently associated with claims of mental storage, with Wray (2002) arguing that formulaic sequences seem to be stored as unanalysed wholes. However, the evidence for holistic storage and processing of sequences is far from conclusive. Siyanova-Chanturia (2015) argues that the processing advantage of sequences over matched control phrases is evidence of phrasal frequency effects, but not necessarily of holistic storage. While I believe, as Hoey (2005, p. 7) argues, that collocation is "fundamentally a psychological concept," I make no claims about the storage of MWUs. In this thesis, one particular type of MWU, the collocation, is the focus of my research.

A major issue in investigating MWUs, including collocations, is identifying them (Schmitt, 2010). One of the two predominant approaches is the phraseological tradition (Benson et al., 1997; Cowie, 1998; Gyllstad & Wolter, 2016; Henriksen, 2012; Howarth, 1998a), which identifies MWUs primarily according to two main semantic criteria (Gyllstad, 2014; Nation, 2013; Nesselhauf, 2003; Schmitt, 2010). One is compositionality, or transparency, in other words, how clearly the meanings of the parts are related to the meaning of the whole. The other criterion is substitutability, or restriction, in other words, whether the restriction on combining elements is arbitrary or not. Different types of MWUs can be placed on a continuum (Howarth, 1998b). At one extreme, free combinations (e.g., *want a car*) are high in transparency and low in restriction, while idioms (e.g., *kick the bucket*) are at the other extreme: low in transparency and high in restriction (Bahns, 1993; Benson, Benson, & Ilson, 1986; Howarth, 1998a; Laufer, 2011; Laufer & Waldman, 2011; Nesselhauf, 2003). Collocations, sometimes called restricted collocations (e.g., *take a picture*), are in the middle of the continuum. They are habitually-occurring sequences of two or three words (Brown, 2014; Durrant, 2009; Henriksen, 2012; Laufer, 2011) which Paquot & Granger (2012, p. 136) define as "lexically constrained combinations that allow for limited substitution within a particular grammatical construction (e.g., verb-object, adverb-adjective, or adjective-noun)." There are several drawbacks to the

phraseological approach of MWU identification: transparency and substitutability are not easily operationalised; identifying MWUs is very labour-intensive and thus limited in scope; and the process is rather subjective (Schmitt, 2010).

The other major tradition of MWU and collocation identification is the statistically-grounded, frequency-based tradition, which was pioneered by Firth (e.g., 1962) and Sinclair (e.g., 1987). Collocation in the broader sense in this approach is "the relationship a lexical item has with items that appear with greater than random probability in its (textual) context" (Hoey, 1991). MWUs are identified mostly through frequency of occurrence through corpus searches. However, using frequency alone is problematic for two reasons: the most frequent units consist of function words and may not have an interesting relationship (e.g., *and the*); and the approach misses very low-frequency collocations (Biber, 2009; Schmitt, 2010). Thus, researchers using a frequency-based approach tend to also use strength-of-association measures, of which there are two types. One is asymptotic hypothesis tests, such as *z*-scores, *t*-scores and log-likelihood tests, which test whether the target words co-occur more frequently than would be expected by chance (Schmitt, 2010, p. 124). The other type of strength-of-association measure is mutual information (MI), which measures the likelihood that one word in a word pair will be found near the other member of the word pair (Hunston, 2002; Schmitt, 2010). The results of MI are often quite different from hypothesis testing, with high-frequency phrases generally scoring highly on *t*-scores and lowly on MI. Collocations in the frequency-based tradition are defined as statistically significant co-occurrences of two or more words; each collocation consists of a node (or base) and a collocate, which are found within a certain span of (distance from) each other, usually three or four words, in a large corpus (Barfield & Gyllstad, 2009; Biber & Barbieri, 2007; Byrd & Coxhead, 2010; Gyllstad, 2014; Paquot & Granger, 2012). Although some frequency-based studies define collocations purely according to frequency (e.g., Wolter & Gyllstad, 2013), many add extra elements such as strength-of-association measures.

A large amount of recent research into MWUs has used a mixed-method approach, combining aspects of the frequency-based and phraseological traditions (Henriksen, 2012). Such research includes Shin and Nation's (2008) most frequent spoken collocations, Wang and Shaw's (2008) investigation into collocations used by Chinese and Swedish learners, and Simpson-Vlach and Ellis's (2010) and Ackerman and Chen's (2013) academic phrases lists. Such a mixed-method approach allows the

strengths of both approaches to be retained while minimising their weaknesses. In other words, the statistical measures of the frequency-based tradition are employed to identify potentially useful MWUs by objective means, while the intuitive approach of the phraseological tradition helps identify MWUs which are psychologically salient and pedagogically valuable (Henriksen, 2012; Simpson-Vlach & Ellis, 2010). Ackerman and Chen (2013), in developing their Academic Collocation List, employed a number of frequency-based measures, such as frequency, mutual information, *t*- and distribution scores; their study also had a phraseological-based stage, in which experts reviewed the collocations for their pedagogical appropriateness. Both experiments 1 and 2 in this thesis use a mixed-method approach in the selection of target collocations (see Chapters 3 and 4 for more details).

2.1.1 Lexical and grammatical collocations

Collocations in the phraseological tradition are frequently divided into two types: *lexical* collocations, each comprising two content words and possibly a grammatical word (e.g., *strong tea*, *play the guitar*); and *grammatical* collocations, each consisting of at least one content word and at least one preposition or grammatical structure (e.g., *under attack; effect on, agreement that* + clause) (Bahns, 1993; Benson et al., 1997; Brown, 2014; Durrant, 2009; Gyllstad, 2014; Henriksen, 2012; Howarth, 1998a). The majority of phraseological research appears to have been conducted into lexical collocations (Henriksen, 2012), with some researchers (e.g., Ackerman & Chen, 2013; Howarth, 1998b) claiming that they are more difficult for second-language learners to master than grammatical collocations because they are less fixed and thus less predictable.

The most frequently-researched type of lexical collocations are verb + noun (V+N) combinations (Ackermann & Chen, 2013; Bahns & Eldaw, 1993; Henriksen, 2012; Laufer & Waldman, 2011; Nesselhauf, 2003; Webb et al., 2013). These may be the most communal source of lexical-collocation errors among learners. Gitsaki (1999) found that the most difficult type of collocation for her high-school students to translate were V+N combinations, which she labelled as arbitrary and unpredictable. In the one-million-word Chinese Learner English Corpus, more than half of the lexical collocation errors made by senior middle-school and college students are V+N errors, "mainly caused by literal translation and overgeneralization" (Huo, 2014, p. 91). The cause of the difficulty in learning V+N combinations may be the abstract and polysemous nature

of verbs (Altenberg & Granger, 2001; Boers, Lindstromberg, & Eyckmans, 2014; Crossley, Subtirelu, & Salsbury, 2013); these are often high-frequency verbs, which, even outside of collocations, are problematic for learners: "Despite fundamental semantic similarities across languages, high-frequency verbs have developed many language-specific differences which make them treacherous for foreign language learners" (Altenberg & Granger, 2001, p. 174). The second-most frequently researched lexical collocations are adjective + noun collocations (Henriksen, 2012; Laufer & Waldman, 2011; Paquot & Granger, 2012; e.g., Durrant & Schmitt, 2010; Kasahara, 2011; Li & Schmitt, 2010; Siyanova & Schmitt, 2008; Wang & Shaw, 2008; Webb & Kagimoto, 2009; Webb & Kagimoto, 2011. Errors with adjective + noun combinations are the second-most common type of lexical collocation error in the Chinese Learner English Corpus, accounting for 16% of the total number of errors (Huo, 2014). Errors with noun + noun collocations accounted for 12% of the errors in the same corpus (ibid.). The learning of these three types of collocations will be investigated in this thesis: (technical) noun + noun and adjective + noun collocations in Experiment 1 (although they are not separated into these two categories) and V+N (along with preposition + noun combinations) collocations in Experiment 2.

While lexical collocations are a source of difficulty for L2 learners, so too are grammatical collocations (although possibly not to the same extent) (Gitsaki, 1999; Hemchua & Schmitt, 2006). Howarth (1998b) claims that lexical collocation errors are a "much more serious challenge" for advanced learners than grammatical collocation errors (p. 163). Ackermann and Chen (2013) argue that grammatical collocations are more easily internalised in the mental lexicon because their patterns are more fixed and thus more predictable whereas lexical collocations are more variable. Nevertheless, grammatical collocations frequently cause problems for learners, especially in their writing (Hemchua & Schmitt, 2006). Schmitt (2010, p. 55) claims that, despite their extreme frequency, grammatical or function words (e.g., *of*, *and*, *the*) are often more difficult for language learners to learn than content words. N. Ellis (1994) reported that, in psychological experiments, patients with Broca's aphasia found lexical words easier to produce in speech than function words and that deep dyslexic patients had greater difficulty reading function words than content words.

Despite the difficulties that they cause learners, grammatical collocations have been largely ignored by researchers. Sinclair (1991, p. 170) described lexical

collocations as "[c]ollocation in its purest sense"—the implication apparently being that grammatical collocations are not genuine collocations—while some linguists (e.g., Rundell, 2010) do not even acknowledge them as collocations (Durrant, 2009). One possible reason for this lack of research interest in grammatical collocations is that they are very frequent (Gledhill, 2000). Durrant (2009) suggests that another reason is that they lack the "striking salience" of lexical collocations. It is perhaps unsurprising, then, that very little research has explored the learning of grammatical collocations or the comparative difficulties in the retention of lexical and grammatical collocations. In one study to have done so, Gitsaki (1999) tested 275 Greek junior high school students on English collocations that they had encountered in their textbook. She found that, in a translation task, the students translated lexical collocations less accurately than simple grammatical collocations. In a gap-fill task, verb + noun (V+N) collocations were more difficult to accurately produce than preposition + noun (prep+N) combinations; however, noun + preposition (N+prep) combinations were more difficult than V+N collocations. The collocations that students most accurately answered were two types of lexical collocations: verb + adverb and noun + noun combinations. Explaining the finding that prep+N collocations were easier for learners than N+prep collocations, Gitsaki (1999) claimed that prep+N collocations are quite fixed and rule-governed whereas N+prep collocations are "also fixed but less regular, more unpredictable (i.e., no rules can be generated for them)" (p. 141).

One of the most common types of grammatical collocation is the *prepositional collocation*, which consists of a content word and a dependent, or "bound", preposition (e.g., *account for, at school*) (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Prodromou, 1999; Sicherl, 2004). It is claimed that prepositions generally are "largely phraseological" in nature in that they are bound to another word (Cosme & Gilquin, 2008, p. 258; see also Sinclair, 1991). Errors with prepositions, and by implication prepositional collocations, are one of the most common sources of language-learner error, even among advanced learners (Dahlmeier, Ng, & Wu, 2013; Jafarpour & Koosh, 2006; Jiminez Catalan, 1996; Lindstromberg, 1998; Tyler, Mueller, & Ho, 2011). Jiminez Catalan (1996) analysed the English essays of 290 high school students in Spain of varying proficiencies and found that preposition errors were the most common, the sixth most common and seventh most common of the top ten types of error. One of the two most frequent types of errors in a one-million-word corpus of Chinese learners

English, along with article errors, was preposition errors (Dahlmeier, Ng, & Wu, 2013). Errors with prepositions sometimes lead to unintended changes to the writer's meaning, although they do not generally cause major communication problems (Hemchua & Schmitt, 2006; Jimenez Catalan, 1996; Mueller, 2011). Nevertheless, correct preposition use is important for accuracy, on which students are assessed, and the overall effectiveness of a piece of writing (Howarth, 1998b; Jiminez Catalan, 1996).

To partially fill this gap in the measurement of the development of knowledge of grammatical collocations—specifically, prepositional collocations—Experiment 2 investigates how different learning conditions (bolding vs. no-bolding) affect recall and recognition of prep+N collocations, as well as V+N collocations, by learners of English.

2.2 Acquisition and processing of L2 MWUs

2.2.1 Factors in the (slow) acquisition and processing of L2 MWUs

The acquisition of MWUs in a non-native language has often been thought to be slow when compared with the acquisition of individual words (Henriksen, 2012; Laufer & Waldman, 2011; Li & Schmitt, 2010; Peters, 2014). Knowing the individual words that comprise a particular MWU does not necessarily imply knowing the MWU as a whole. Many, if not most, MWUs are at least partly non-compositional and thus learners often misunderstand them without knowing it (Boers et al., 2014; Martinez & Murphy, 2011). Martinez and Murphy's (2011) participants overestimated their comprehension of test texts due to their misunderstanding of the target idioms.

Collocations are more semantically transparent than idioms, so when learners encounter novel collocations in texts, they are more likely to understand them. However, collocations can be "deceptively transparent", with collocations like *catch a cold* seeming to be transparent to native speakers but not to non-native speakers (Boers et al., 2014, p. 44). Even if collocations are understood, production of them may lead to errors, even for advanced learners (Laufer, 2011; Nesselhauf, 2003).

> [W]hen encountered in the input, [collocations] may not be noticed by learners and teachers as problematic. However, producing correct collocations is often difficult because 'equivalent' collocations in L1 may often include at least one word that is different from the L2 (Laufer & Waldman, 2011, p. 665).

It has been claimed that second-language students' knowledge of individual lexical items "far outstrips" their collocational knowledge (Bahns & Eldaw, 1993, p. 108). This lack of collocational knowledge was illustrated in Bahns and Eldaw's (1993) landmark study, which found that the number of V+N collocation errors made by advanced German learners of English in a translation task was twice their number of errors for individual lexical words. Alali and Schmitt (2012) taught ten unknown English idioms and ten words taken from those idioms to 12- and 13-year-old Arabic-speaking students. While the idioms and single words were recognised equally well in their delayed post-tests, the meanings of the single words were remembered better than the meanings of the idioms[1].

It appears that non-native speakers not only know fewer MWUs than individual words but, compared with native speakers, they also have "an impoverished stock of formulaic expressions" (Wray, 2012, p. 236). Durrant and Schmitt (2009) report that, compared with natives, non-natives in their study overused high-frequency collocations in written texts but underused less frequent collocations with high mutual-information scores—combinations which native speakers would likely find highly salient. Laufer and Waldman (2011) found, in a 300,000-word learner corpus of high-school students' essays, that learners at three different proficiency levels produced only about half the number of V+N collocations that native speakers did; they also found that a third of the collocations were deviant. Other corpus studies have reported similar results, finding a pattern of heavy use of familiar collocations (Granger, 1998; Hasselgren, 1994; Li & Schmitt, 2009), light use of less frequent collocations (Granger, 1998) and incorrect use (Altenberg & Granger, 2001; Nesselhauf, 2003)—even at the advanced level, where accuracy of grammar and single words is high (Laufer & Waldman, 2011). As a consequence, learners' speech and writing may be viewed by native speakers as unnatural and non-nativelike (Durrant & Schmitt, 2009; Howarth, 1998a; Siyanova & Schmitt, 2008).

Several factors likely account for the apparent slowness of second-language learners' acquisition of MWUs. One factor is the low frequency of occurrence of MWUs

[1] While the perspective just described compares the acquisition of MWUs with the acquisition of single words, it should be remembered that knowledge of a word's collocations is only one of several aspects of knowledge of a word (Nation, 2013). Pellicer-Sánchez (2017) claims that gaining collocational knowledge "may not be intrinsically more difficult than other components of lexical mastery" (p. 395). In her experiment, she found that there was no significant difference between the amount of recall of the collocates of pseudowords and the recall of the *meanings* of the pseudowords.

in the input compared with their component individual words (N. Ellis, 2002; Siyanova-Chanturia & Martinez, 2015). Repeated exposure to linguistic items is needed for learning (e.g., Sonbul & Schmitt, 2013; Webb, Newton, & Chang, 2013). The effect of frequency on memory was first researched in psychology. The experimental psychologist Arthur Reber (1976) highlighted the role of repetition and frequency of linguistic structures when he described implicit learning as "a primitive process of apprehending structure by attending to frequency cues" (p. 93). Eysenck (1982) reported that repetition of vocabulary items in psychological experiments usually produces a fairly substantial improvement in recall and recognition of the items. Reviewing neuropsychological and neuroimaging research on implicit learning, Paul Reber (2013) reports that repetition of stimuli leads to implicit learning through the extraction of statistical relationships among stimulus features, including language structures.

Within SLA, several perspectives claim to account for frequency effects. Usage-based theories of language acquisition view language as thousands of conventionalised constructions (morphemes, words, phrases and syntactic phrases); acquisition occurs through processing input as language users' language systems compare each structure with previous encounters of the same, or a similar, structure and abstract the structures' regularities (Ellis, Römer, & O'Donnell, 2016a; Wulff, 2018). These theories posit that language users are sensitive to frequency of usage, implicitly tallying the statistics as they process linguistic structures; therefore, users learn high-frequency items more easily and process them more fluently than low-frequency items (N. Ellis, 2002, 2007, 2015; Ellis et al., 2016a; Siyanova-Chanturia, 2015).

> [O]ur perceptual system is sensitive to the probabilities of occurrence of words and word sequences in English. It has expectations about four-word sequences and better perceives stimuli which meet these expectations. At a lower level, it has expectations about two-word collocations and better perceives stimuli which meet these expectations (Ellis, Römer, & O'Donnell, 2016b, p. 50).

According to usage-based theories, frequency is the key input property in the acquisition of formulaic constructions (particularly frequency of type as opposed to frequency of token) (Wulff, 2018). However, input properties other than frequency also

play roles in language-learning; these include recency of the construction in the input and the construction's salience and redundancy (in the understanding of the construction in the input) (ibid.).

Another perspective accounting for frequency effects is Reichle and Perfetti's (2003) instance-based word-learning framework, according to which each encounter with a word produces a new trace of that word in episodic memory. The traces, which include information such as the word's spelling, meaning and contextual information, are strengthened through each new encounter with the word. Multiple encounters activate memory traces consistent with previous encounters and facilitate an eventual abstraction of meaning of lexical items from context (Bolger, Balass, Landen, & Perfetti, 2008). In this way, abstracted meanings "arise from the summation of unique contexts and their effects on new encounters with the word" (p. 126).

With regards to collocations and other MWUs, *frequency effects* have been found in the visual processing of MWUs during reading (e.g., Arnon & Snider, 2010; Siyanova-Chanturia, Conklin, Heuven, 2011; Sonbul & Schmitt, 2013). Exposure of repeated occurrences of MWUs in written texts have been found in several studies to promote learning of the MWUs (e.g., Pellicer-Sanchez, 2017; Sonbul & Schmitt, 2013; Webb et al., 2013—see subsection 2.2.2). A number of studies have also tested the reading processing speeds for MWUs among native speakers and higher-proficiency second-language speakers; they have found that both categories of participants are sensitive to phrasal frequency, reacting faster to higher-frequency phrases than to lower-frequency phrases (e.g., Arnon & Snider, 2010; Siyanova-Chanturia, Conklin, & van Heuven, 2011). In one of the first studies to measure reading processing speeds of collocations, Siyanova and Schmitt (2008) employed a naming task. They found that collocations seen in the treatment phase (e.g., *common people*, *right mind*) were read more quickly than non-collocations by both natives and non-natives. Using a self-paced reading paradigm, Kim and Kim (2012) reported a processing advantage for high-frequency two-word phrasal verbs over low-frequency two-word phrasal verbs, for both native and non-native speakers. In Vilkaite's (2016) eye-tracking experiment, native speakers read V+N collocations (e.g., *achieve status*, *seek help*), both adjacent and non-adjacent, faster than control phrases. Yi (2018), using a phrasal acceptability judgement task, found that both L1 and L2 speakers of English were sensitive to the frequencies of adjective + noun collocations.

While the studies cited in the previous paragraphs examined frequency effects for lexical collocations, few have done this with grammatical collocations. In one that did, Lowie and Verspoor (2004) gave four proficiency levels of learners a 25-sentence test with a high-frequency or low-frequency preposition missing from each sentence (e.g., *He is ____ [at] work; The ashtray fell ____ [off] the table*). Most of the prepositions were components of prepositional collocations. The researchers found a clear effect of preposition frequency for the low- and moderate-proficiency groups but not for the high-proficiency group, which they attributed to a ceiling effect. In a similar study, Mueller (2011) selected 18 prepositional collocations (e.g., *on time*, *dreams of*) according to frequency (high and low) and instructed his participants to supply the correct preposition in each space in each sentence. The learners were sensitive to word co-occurrence, as shown in a large phrasal frequency effect. While the just-mentioned studies administered tests of declarative knowledge, to my knowledge, no experimental studies have examined the processing of grammatical collocations.

Frequency-related processing advantages for MWUs over non-MWUs have also been found in experiments focusing on phrases other than collocations, both compositional/literal (e.g., binomials, lexical bundles) and non-compositional/figurative (e.g., idioms). Siyanova-Chanturia, Conklin and van Heuven's (2011) eye-tracking study found that both natives and advanced non-natives processed high-frequency binomial phrases (e.g., *bride and groom*) more quickly than their reversed forms (e.g., *groom and bride*). In Arnon and Snider's (2010) two experiments involving phrasal decision tasks, native speakers of English recognised high-frequency compositional four-word phrases (e.g., *a lot of places*, *don't have to worry*) more quickly than lower-frequency phrases. Hernandez, Costa and Arnon (2016), using the same materials as Arnon and Snider (2010, Experiment 1), found a frequency effect of four-word phrases for both native speakers and non-native speakers. Also using a timed phrasal decision task, Supasiraprapa (2019) reports frequency effects for both L1 and L2 speakers of four-word compositional phrases (e.g., *out of the house*, *don't have to worry*) in an elicited oral production task. Similarly, in Tremblay, Derwing, Libben and Westbury's (2011) three SPR experiments, the sentences containing lexical bundles (e.g., *in the middle of the*) were processed by natives more quickly than the control sentences (e.g., *in the front of the*). Two studies using ditropically ambiguous idioms—idioms such as *a breath of fresh air* which have both literal and figurative meanings—also found a processing

advantage for the idioms. In Conklin and Schmitt's (2008) SPR study, both natives and proficient non-natives read idioms faster than their controls regardless of whether the idioms were in literal or figurative contexts. Likewise, Siyanova-Chanturia, Conklin and Schmitt (2011) report, in their eye-tracking study, an advantage for idioms for natives and higher-proficiency non-natives, although lower-proficiency non-natives read the idioms and the novel phrases at similar speeds.

Closely related to low frequency of occurrence as a cause of the slow acquisition of MWUs is *incongruence*, or lack of similarities between the MWUs in the learners' first and second languages. A congruent collocation has "a corresponding equivalent in the L1 in terms of the core meanings of the constituent words, in a word-for-word translation" (Gyllstad & Wolter, 2016, p. 298), while an incongruent collocation does not have that equivalent. Incongruence has been found to be a major source of collocational errors in tests of declarative knowledge, likely due to the heavier learning burden of incongruent collocations than congruent collocations. Farghal and Obeidat (1995), for example, found that, on a cued-recall test, Arabic-speaking university students received high scores for predictable (congruent) English adjective + noun and noun + noun collocations and low scores for unpredictable (incongruent) collocations. Analyses of learner corpora have found that substantial proportions of collocational errors are observed for incongruent expressions. Nesselhauf (2003) reported that just over half of the mistakes with V+N collocations made by advanced native speakers of German in their English essays were likely influenced by their first language. In their examination of a corpus of Chinese learner English, Yun and Youmei (2006) estimated that 37% of L2 V+N collocational errors made by advanced learners were likely due to incongruence. More than 60% of errors in V+N collocations in Laufer and Waldman's (2011) learner corpus of L1 speakers of Hebrew and Arabic were judged to be interlingual, including those by lower-level and advanced learners. In Lowie and Verspoor's (2004) study on prepositions, similarity (congruence) of English prepositions to their Dutch counterparts had a strong effect on test scores for the lower- and moderate-proficiency learners but little effect for the high-proficiency learners. They found that similarity was only a factor in test scores when the prepositions were less frequent.

Studies that have examined the recognition and processing of MWUs that are either congruent or incongruent in the participants' first and second languages also show sensitivity to the frequencies of occurrence across the L1 and L2, with congruent

collocations being read more quickly than incongruent collocations (e.g., Carrol, Conklin, & Gyllstad, 2016; Wolter & Gyllstad, 2011, 2013). In Yamashita and Jiang's (2010) experiment, EFL learners responded faster in a phrase-acceptability task to congruent collocations than to incongruent collocations, although there was no difference in reaction times (RTs) between the two types of collocations for ESL learners. Wolter and Gyllstad (2013) report that advanced L2 learners reacted more quickly in an acceptability judgement task to congruent collocations than to incongruent collocations. In Carrol, Conklin and Gyllstad's (2016) eye-tracking study, a processing advantage was found in native speakers of Swedish for English idioms congruent with Swedish idioms compared with their literal control phrases, but not for English idioms not found in Swedish. The authors argue that although the English-only idioms were easily understood, they were "not as well entrenched in the mental lexicon" (p. 433).

Another factor contributing to the slow acquisition of MWUs is lack of perceptual salience (N. Ellis. 2002). MWUs may be less salient to learners than individual words because they are not marked as phrases within written texts, because some are discontinuous (e.g., *as … as*), and because, when spoken, they are usually spoken quickly (Boers & Lindstromberg, 2009). Boers et al. (2014) argue that many collocations lack novelty, and thus salience, because the constituent content words tend to be frequent and the constituent function words even more frequent. The high frequency of the individual words means a lower chance of the MWU being attended to as a unit. It is also possible that many classroom teachers do little to highlight MWUs to their students (Hill, 1999). As a result of this lack of salience, post-childhood learners often fail to register MWUs as distinct phrases (Bishop, 2004), instead focusing on single words (Boers et al., 2014; Henriksen, 2012; Wray, 2002). How to promote noticing of, or greater attention to, MWUs is addressed in section 2.4.

Yet another cause is figurativeness, or non-compositionality, of MWUs. Recent research into the processing of idioms by second-language speakers has found a processing advantage for the literal meanings of idioms over their figurative meanings. Cieślicka (2006), in a cross-modal lexical priming experiment, found faster responses to the words of the literal meanings of the idioms than to the words of the figurative meanings of the idioms. Siyanova-Chanturia, Conklin and Schmitt's (2011) eye-tracking study found that literal uses of idioms were processed more quickly than figurative uses. Thus, it can be concluded that processing of figurative MWUs such as idioms may

be more difficult for L2 speakers than processing of more literal MWUs such as lexical bundles and, to a lesser extent, collocations, which tend to be semi-literal (Howarth, 1998a).

There are several factors which create barriers to learning prepositional collocations in particular. First, even advanced learners may have gaps in their knowledge of, and be confused about, the central spatial meanings of some prepositions (Lindstromberg, 2001). Second, prepositions are generally polysemous, having many different but related senses, many of them figurative (Boers & Demecheleer, 1998; Tyler, Mueller, & Ho, 2011). Third, prepositions often do not have congruent translation equivalents (Cosme & Gilquin, 2008). Fourth, there is variability in the usage of prepositions as shown in the disagreement among native speakers as to the best choices of prepositions in use (Chodorow, Gamon, & Tetreault, 2010).

The experiments in this thesis investigate two of the major factors connected with the acquisition of MWUs: frequency of occurrence in the input and perceptual salience. Input flooding of collocations in text and typographic enhancement are two treatment techniques that are employed in attempt to facilitate the development of declarative and procedural collocational knowledge.

2.2.2 Intentional and incidental learning of MWUs

Second-language acquisition of vocabulary, including MWUs, is frequently conceptualised as occurring in two main learning conditions, as defined by the presence or absence of the learner's intention to learn the target items: *intentional*, or deliberate, learning (also called *focus on forms*) and *incidental* learning (*focus on form*), respectively (N. Ellis, 1994; Hulstijn, 2003; Laufer, 2005; Long, 1991). In the context of psychology and second-language learning experiments, the dichotomy is realised as participants being told or not told before the learning phase whether they will take a retention post-test (Eysenck, 1982; Hulstijn, 2003). When there is an intentional focus on the learning of single words, the retention gains can be substantial, and tend to be larger than gains from an incidental focus (Elgort, 2011; Hulstijn, 2001; Laufer, 2005; Nation, 2013; Paribakht & Wesche, 1997). While it might also be assumed that an intentional emphasis on the learning of MWUs produces larger gains than unintentional learning, few studies have investigated this area and no clear pattern has yet emerged. Steinel, Hulstijn and Steinel (2007) found that participants made sizeable gains in their

knowledge of idioms through paired-associate learning, although their study did not include an incidental learning condition. Also without an incidental learning condition, Chan and Liou's (2005) study found significantly higher long-term retention gains of V+N collocations taught using traditional explanations with exercises than collocations taught using a bilingual concordancer and cloze tasks. Sonbul and Schmitt (2013) report that an intentional learning condition was not significantly better for long-term retention than an incidental reading condition which included three repetitions of each target medical collocation. However, the intentional learning condition may have been no more effective because the collocations were decontextualised, with no opportunity for meaning elaboration (Anderson & Reder, 1979; Laufer & Hulstijn, 2001) (see next section), whereas the collocations in the incidental condition occurred three times within meaningful (highly constrained) written contexts.

Acquisition of second-language vocabulary also occurs unintentionally, if generally less efficiently and more slowly, through reading and listening (Laufer, 2005; Paribakht & Wesche, 1999; Read, 2004). *Incidental* learning takes place within the context of a meaning-focused activity when the learner's attention to the words is the by-product of a temporary focus; that is, the learner's intention is not to learn vocabulary but there is peripheral attention to form (R. Ellis, 1999; Laufer, 2005; Hulstijn, 2001, 2003). Most second-language vocabulary, apart from the first few thousand words, is learned incidentally (Rieder, 2003). The learning gains of single words through incidental exposure, for example, in extensive reading, tend to be small, although repetition of the target items has been shown to increase the likelihood of learner uptake (Elgort & Warren, 2014; Nation & Wang, 1999; Waring & Takaki, 2003; Webb, 2007; Zahar, Cobb, & Spada, 2003). In Waring and Takaki's (2003) study, eight or more repetitions of the target words in a graded reader were required for learners to gain about 50% correct on a form-recognition test and about 40% correct on a multiple-choice meaning-recognition test three months later. In Pigada and Schmitt's (2006) case study of one learner doing extensive reading, the learner gained knowledge of the spelling of target verbs and nouns after only a few exposures, while knowledge of meaning required 20 or more exposures to achieve an uptake of 60% on an oral-recall think-aloud task.

A number of studies have also demonstrated incidental, incremental learning of MWUs in context, especially through learner exposure to repeated occurrences of the

target phrases (e.g., Pellicer-Sanchez, 2017; Sonbul & Schmitt, 2013; Webb, Newton, & Chang, 2013). Sonbul and Schmitt (2013) found that learners gained declarative knowledge of (mostly) previously unknown technical collocations after three contextual exposures to the collocations in the reading text. (For details of this experiment, see Chapter 3). In Webb et al.'s (2013) study, EFL students at three Taiwanese universities simultaneously read and listened to a modified graded reader, with the number of exposures to the 18 target collocations varying, according to learning condition, at between 1, 5, 10 and 15 exposures. The researchers found, in tests of declarative knowledge, that the more often a collocation was encountered in a graded reader, the more likely it was to be learned and that large learning gains resulted from learners encountering a collocation 15 times. However, such a frequency effect was not found in Pellicer-Sánchez's (2017) study. In that study, intermediate learners of English read a modified text of approximately 2,300 words in length containing multiple repetitions of six adjective + pseudoword phrases, four occurrences in one learning condition and eight occurrences in another. In pen-and-paper post-tests conducted one week after the reading session, on average participants recalled 11% and recognised half of the pseudowords' collocates, but there was no difference between the four-repetition and eight-repetition conditions. It may be, as Pellicer-Sánchez speculates, that a larger number of target items are required in order to find a significant frequency effect. Incidental learning approaches to MWUs, involving repetitions of the MWUs, are employed in the experiments in this thesis.

Although the intentional/incidental learning dichotomy is a commonly-used framework in second-language vocabulary-learning research, it has weaknesses. One is that, while the distinction between incidental and intentional learning is clear-cut at an operational level in experimental work, it is difficult to maintain theoretically (Eysenck, 1982; Hulstijn, 2001). According to Eysenck (1982, p. 197), it is erroneous to claim that incidental learning occurs "with the total absence of an intention to acquire information" since it is impossible to know what participants do in an incidental learning condition: they may, for example, intentionally learn items in anticipation of a test even though the experimenter has not forewarned them of the test (Bruton, Lopez, & Mesa, 2011). Another limitation of the intentional/incidental framework is that whether the learning condition is intentional or incidental appears to have little effect on the learning, with experimental evidence showing that unintentional learning can be

substantial, depending on the tasks and the strategies used (Craik & Lockhart, 1972; Logan, 1988). Hulstijn (2001) claims that what leads to greater retention is not whether the learning occurs incidentally or intentionally but the degree of elaborateness of processing of lexical information (e.g., aspects of word form and meaning)[2]. Multiple psychological studies, including studies on the learning of words, have demonstrated that the way in which material is processed is the key factor in memorising the material, not the intention or the lack of intention to learn:

> [I]f subjects process material in the same way, they will recall the same amount whether they intend to learn or not. Frequently, subjects intending to learn are in fact able to recall more material, but only because they engage in processing more conducive to learning the material (Anderson, 1995, p. 209).

In reviewing psychological studies of recall and recognition, Eysenck (1982, p. 206) argued that the intention to learn has little effect on recognition; however, the effects of intention to learn on recall memory are "more variable."

The experiments in this thesis employ incidental learning conditions in the operational sense: the participants are not informed about the tests of collocational knowledge before the treatment sessions.

2.2.3 Processing frameworks for learning vocabulary

Another group of models which have been commonly used in vocabulary acquisition research are *depth-of-processing* frameworks. The first version of this model was articulated by the cognitive psychologists, Craik and Lockhart (1972), who suggested that greater depth of processing—with more attention paid to, and a larger amount of semantic or cognitive analysis done on, the stimulus—produces more persistent memory traces that are "more elaborate, longer lasting, and stronger" (p. 675). Anderson and Reder (1979) argued that the number and types of elaboration are critical for depth of processing (or their preferred term, *breadth of processing*), with a larger number of elaborations of vocabulary items leading to better long-term retention.

[2] Also challenging the intentional/incidental distinction, Paribakht and Wesche (1999, p. 215) claim that, from the perspective of the learner, learning words through reading is "in some fundamental sense not 'incidental' … Achieving any level of input processing by drawing on knowledge sources for information on the meaning of a word requires both attention to a given new word and effort on the part of the learner to find its meaning."

They also claimed that semantic elaboration is superior to orthographic elaboration. A development of the depth-of-processing concept in SLA, Hulstijn and Laufer's (2001) *involvement load hypothesis*, states that more involvement on the part of the learner in the processing of unknown words leads to greater retention of those words. The researchers argue that depth, or elaboration, of processing is required for the retention of new words and that processing of vocabulary can be just as elaborate in an incidental condition as in an intentional condition. They found that the degree of retention of vocabulary in their experiment was connected to the amount of "task-induced involvement" (comprising three factors: *need*, or motivation, *search* for the meaning or form of the word, and *evaluation* of the information obtained); however, time on task was different for the three tasks. Controlling for time on task, Folse (2006) found that ESL students learned more vocabulary by completing three fill-in-the-blank sentences for each target word than by writing an original sentence for each word. He argues that the number of retrievals of a word is more important for learning in a vocabulary exercise than depth of word processing. Schmitt's (2008) concept of *engagement* focuses on the motivation, attitudes and strategic behaviours of learners. He lists a number of factors that have been shown to promote the learning of target lexical items, such as looking in a dictionary, increased noticing, greater intention to learn the item, a need to learn it, and the amount of interaction with the item.

Another processing model used in vocabulary acquisition is *transfer appropriate processing* (TAP), which was first outlined by Morris, Bransford and Franks (1977) in response to perceived inadequacies in Craik and Lockhart's (1972) depth-of-processing model. From the findings of their experiments, Morris et al. (1977) claimed that non-semantic (shallower) processing is not necessarily inferior to semantic (deeper) processing. Instead, they found that acquisition of rhyming target words was greater than acquisition of words that were meaningful in a sentence when measured on a rhyme test. They argued that "different modes or levels of processing may simply allow people to acquire different sorts of information, each of which may have the potential for being equally strong and durable" (p. 520). Developing TAP in the SLA context, Barcroft (2002), in his *type of processing-resource allocation* (TOPRA) model, defines elaboration as "increased evaluation of an item" (p.324). He claims that *semantic elaboration* (e.g., doing a meaning-judgement task about a word) can increase learning of the semantic properties of novel words but at the same time decrease learning of the

structural (formal) properties of the words when processing demands are high; the converse is true for *structural elaboration* (e.g., copying a word), he argues. In his experiment, the learners experienced difficulty processing meaning-related input and form-related input at the same time. His findings support the idea that semantic elaboration can inhibit coding of the formal properties of new words; he suggests that these results necessitate a re-evaluation of the use of deeper processing during the initial stages of word-learning.

The present research aims to fill gaps in the knowledge about the learning conditions of collocations. It will compare the retention of both lexical collocations and grammatical collocations in different incidental contextualised learning conditions, and in conditions of semantic elaboration and no elaboration. Semantic elaboration is operationalised in Experiment 1 in this thesis as glossing.

2.3 Explicit/declarative and implicit/procedural memory, learning and knowledge

Most studies which have measured the acquisition or learning of vocabulary knowledge, including knowledge of MWUs, have assessed gains in explicit, declarative knowledge alone, with only a few studies investigating procedural or implicit knowledge (e.g., Elgort & Warren, 2014; Sonbul & Schmitt, 2013). The investigation of the development of procedural or implicit knowledge of MWUs, however, is needed in order to predict whether learners can access MWUs during meaning-focused language use in real time.

Implicit and explicit knowledge are the outcomes of implicit and explicit learning respectively (see below for more detail). These two types of learning in turn employ what neuropsychologists assume are two distinct systems of the long-term memory: the procedural (implicit; nondeclarative) and declarative (explicit) memory systems (P. Reber, 2008, 2013; Roediger, 1990). The procedural memory (PM) system subserves the acquisition of first and second languages in natural conversations while the declarative memory (DM) system subserves the formal learning of second languages (Paradis, 1994). The implicit and procedural memory systems on the one hand and the explicit and declarative memory systems on the other hand are roughly equivalent, and, in SLA literature, the two pairs tend to be used interchangeably (Ercetin & Alptekin, 2013; P. Reber, 2008). However, there are differences between them. For example, the

implicit memory system is only one type of procedural, non-declarative memory system (P. Reber, 2008; Squire & Zola, 1996; Ullman, 2004).

Evidence from experiments with amnesic and aphasic patients indicates that the procedural and declarative memory systems may be neurologically distinct and the learning processes may be dissociated (N. Ellis, 1994, 2011; Hulstijn, 2002; Paradis, 1994; Squire & Zola, 1996). People with amnesia, for example, have great difficulty making new semantic links (through explicit learning), although they can implicitly learn new word forms (N. Ellis, 1994). Despite the assumed separation of the two memory systems, the results of neuroimaging experiments suggest that the systems act both cooperatively and competitively (Ullman, 2004). Ullman (2004, p. 247) argues that competitive interaction leads to a "see-saw" effect, with the language-learning functions in both systems not fully operating simultaneously:

> Access to a stored representation which has similar mappings to one which could be composed compositionally by the procedural system (e.g. an irregular vs. a regular past-tense form of the same verb) would block completion of the latter computation. Damage to the declarative system is expected to lead to enhanced learning and processing by the procedural system, and vice versa. Moreover, learning in one system may depress functionality of the other.

Ullman (2004) and Paradis (2009) argue that declarative memory can play a stronger role than procedural memory in early second-language learning by adults, with the balance shifting to procedural memory later in the learning process, when skill proficiency develops with practice, and when there is less dependence on declarative memory.

Explicit language learning is learning of a language during which there is "awareness at the point of learning" (Schmidt, 1994, p. 20). N. Ellis (2002) describes explicit language learning as conscious learning in which the learner creates and test hypotheses about language. Explicit *knowledge* is the result of explicit learning. It is conscious, declarative, analysable and potentially verbalisable—the type of knowledge that can be retrieved with the help of meta-cognitive and task-related strategies through controlled processing such as knowledge of abstract rules and exemplars (R. Ellis, 1993, 2006; Paradis, 1994). *Implicit* learning is learning "without awareness of what is being learned" (DeKeyser, 2003, p. 314; see also P. Reber, 2013; Schmidt, 1994,

p. 20). Implicit *knowledge* is often, if not always, the outcome of implicit learning, depending on one's theoretical perspective (see Section 2.3.1). It is frequently described as knowledge that is unconscious, intuitive, procedural, ballistic, unanalysed and not verbalisable (Anderson, 1983; R. Ellis, 1993, 2004, 2009), the type of knowledge acquired in meaning-focused interactions and through exposure to comprehensible input (Krashen, 1982; Paradis, 1994). Paradis (2009, p. 38) defines implicit knowledge as "a competence, namely the ability to do something without knowledge of the actual underlying mechanism that allows a particular performance." However, a small number of experiments have found that explicit learning of vocabulary items can lead to procedural, possibly implicit, knowledge of the items (Elgort, 2011; McLaughlin, Osterhaut, & Kim, 2004; Mestres-Missé, Rodriguez-Fornells, & Münte, 2007—see Section 2.3.1.1 for details).

The terms *explicit* and *declarative* knowledge on the one hand and *implicit* and *procedural* knowledge on the other hand are often used interchangeably in second-language research literature (Doczi & Kormos, 2016; Hulstijn, 2002; Paradis, 1994). However, as with the equivalent memory systems, a distinction can be made and is particularly relevant to my experiments. Although procedural knowledge is generally equated with implicit knowledge, it can also denote explicit knowledge (of skills or habits such as how to play a game) which is sped up, or automatised (Anderson, 1983, 1995; Suzuki & DeKeyser 2017). Suzuki and DeKeyser (2017) argue that while automatised explicit knowledge and implicit knowledge of language both involve fast access to linguistic knowledge they are distinguished from each other by awareness and lack of awareness of the linguistic forms respectively. It is possible that the procedural knowledge involved in, for example, timed experimental tests, is in fact implicit knowledge only; however, it is also possible that such procedural knowledge is sped-up explicit knowledge or, alternatively, that explicit knowledge and implicit knowledge are being retrieved simultaneously (DeKeyser, 2003; Doczi & Kormos, 2016; Suzuki & DeKeyser, 2017). As Ercetin and Alptekin (2013, p. 729) explain, "[I]t is not impossible for knowledge stored in [procedural memory] to be at least partly explicit if it happens to be the outcome of the proceduralization of declarative knowledge, as might be the case at advanced levels of L2 proficiency." Since there is currently a lack of clarity in the research about whether timed tests such as LDTs and self-paced reading tasks—both of which are employed in this thesis—test the presence of implicit knowledge only and/or

sped-up explicit knowledge—I will primarily use the term *procedural knowledge* and its opposite term *declarative knowledge* to refer to the types of knowledge measured in my, and others', studies.

2.3.1 The interface debate

How to develop fluent, procedural, implicit linguistic knowledge has been a central question in second-language research for several decades, with most debate and research focusing on the acquisition of grammatical structures (R. Ellis, 1993, 2002; Krashen, 1981, 1982; Rieder, 2003). In the interface debate, three main theoretical positions on how to gain implicit, procedural knowledge of a language exist: the no-interface position, the strong interface position and the weak interface position. According to the no-interface position, a dual-system explanation of learning first proposed by Krashen (1981), "learning does not 'turn into' acquisition" (p. 83). In other words, declarative, explicit knowledge cannot be directly converted into procedural, implicit knowledge through direct instruction and practice; it remains declarative, explicit knowledge only (R. Ellis, 1993, 2009). Further, Krashen (1981) saw only a limited role for explicit learning in second-language acquisition: to act as a "monitor," or editor. Krashen's no-interface position is these days largely rejected by researchers (N. Ellis, 2011).

The strong interface position holds that declarative, explicit knowledge is able to be converted into procedural, implicit knowledge through direct instruction and practice (DeKeyser, 2003; R. Ellis, 1993; 2009). An early proponent of this position in psychology was Anderson (1995), whose adaptive control of thought (ACT) theory states that declarative knowledge is proceduralised through practice and that the proceduralised explicit knowledge is then converted into implicit knowledge. Within the SLA field, Sharwood Smith (1981) and DeKeyser (2003) have been major proponents of the strong interface position, arguing that explicit knowledge converts to implicit knowledge through practice and automatization as learners lose awareness of rules over time. "At that point they not only have procedural knowledge that is functionally equivalent to implicitly acquired knowledge, but even implicit knowledge in the narrow sense of knowledge without awareness" (DeKeyser, 2003, p. 329).

The weak interface position states that explicit knowledge has an indirect role in the acquisition of implicit knowledge. Several of its main proponents are N. Ellis (2005),

Paradis (1994; 2009), Hulstijn (2002, 2015) and R. Ellis (2009). They argue that the implicit and explicit memory systems are dissociable and that explicit knowledge cannot be directly turned into implicit knowledge; however, they maintain that explicit knowledge can somehow indirectly facilitate the development of implicit knowledge. Even though this process of facilitation is as-yet unclear, the weak interface position appears to accord best with the neurophysiological research reported above, which shows that the two types of memory are located in different areas of the brain. It seems to be more accurate to assert that, rather than explicit knowledge turning into implicit knowledge, "explicit knowledge forms a *pre-requisite* for implicit knowledge to come into existence" [my italics] (Segalowitz & Hulstijn, 2005, p. 378).

In spite of the varying theoretical stances taken by SLA researchers, however, the majority agree on an important pragmatic point: learners will only become fluent in a second language by continuing to practise their language use in different communicative settings (Hulstijn, 2015). Making a point about speaking which could equally be applied to other linguistic skills such as reading, Hulstijn (2015, p. 36) argues that "it is *continued* practice in speaking that is ultimately causally responsible for proceduralised, automatic and largely unconscious cognition in L2 speech production, which may or may not co-exist with declarative knowledge." In other words, the development of fluency in a second language rests on proceduralization of the linguistic items, so proceduralization through practice is an important goal to aim for. Less important for the purposes of this thesis is what combination of explicit and implicit learning and knowledge of the target items is involved in the process. Schmitt (2010, p. 246) makes a similar point:

> This speeding up versus automaticity distinction may not be important if the purpose is determining if vocabulary is being processed more quickly and what teaching or input led to this speed increase. But it may well be important to researchers who are interested in understanding the mental lexicon and explaining the mechanism underlying any increase in processing speed.

One focus of my thesis is to determine the role of certain types of input in facilitating speed of access to MWUs. As such, I am more concerned with the practical effect of the treatment than conjecturing about the nature of lexical representations in the mental lexicon.

2.3.1.1 The interface debate and acquisition of vocabulary (including MWUs)

Knowledge of individual words is both explicit and implicit. Extensive work has been conducted in the last several decades into the gaining of explicit knowledge of words, yet research into implicit, procedural knowledge of vocabulary in SLA has been lacking. That may be because, in the past, vocabulary knowledge was believed to be explicit and declarative only (e.g., Ullman, 2001; Hulstijn, 2007). Now, however, it is thought that while the knowledge of word meaning draws on explicit memory structures and involves conscious learning processes, the ability to fluently process a word form in written or spoken input, as well as its collocations and constraints on use, is mostly implicit (N. Ellis, 1994; R. Ellis, 2004; Paradis, 1994; Ullman, 2001); in other words, the explicit and implicit aspects of vocabulary knowledge are dissociated (Sonbul & Schmitt, 2013). Memory studies show that implicit memory principles determine word identification, which comes solely from frequency of exposure to the words (N. Ellis, 1994). However, explicit memory of words is different because it is argued that such memory is affected by depth of processing and "the degree to which subjects analyse their meaning" (N. Ellis, 2004, p. 226; see also Paradis, 2009). Evidence of a dissociation between implicit and explicit memory of words has existed for several decades in experimental psychological research. Summarising the results of their, and other researchers', experiments on associations between pairs of unrelated words, Graf and Shacter (1989) reported that some elaborative processing is required for both implicit and explicit memory of new word associations; however, modality manipulations and associative elaboration and interference manipulations affect implicit and explicit memory differently, implying that different processes mediate implicit and explicit memory of the associations.

Few experimental studies have investigated the development of procedural, implicit lexical knowledge in a second language. However, several experiments have shown that explicit, declarative knowledge from language instruction leads to the development of procedural (implicit) knowledge before the development of explicit, declarative knowledge (e.g., Doczi & Kormos, 2016). In one experiment that did so, McLaughlin, Osterhaut and Kim (2004) used event-related potentials (ERPs) to determine the electrical activity in the brains of adult, beginning-level learners of French who were performing a LDT (lexical decision task) on French word pairs. The researchers found that the learners quickly gained information about the word forms

and then the word meanings of the second-language words and that the ERPs showed early evidence of learning, even before evidence of learning from the LDTs: "ERPs might more accurately reflect implicit learning and continuous change in knowledge than do explicit, categorical judgements" (p. 704). Mestres-Missé, Rodriguez-Fornells and Münte (2007) also found evidence of implicit lexical knowledge early in the learning process. They used ERPs to follow the process of real-time acquisition of novel words in a foreign language after one, two and three exposures in meaningful contexts in which participants were asked to derive the meanings of the new words. They found that, after three exposures, the learners' electrical brain activity—as measured by the ERPs—for the novel words was the same as that for the real words. Further evidence of the development of procedural lexical knowledge through explicit second language learning comes from Elgort (2011), who showed that both declarative and procedural knowledge of novel words can develop over time through deliberate study. Using form- and semantic priming with speeded LDTs, Elgort found that pseudo-words learned by advanced learners on a spaced-repetition schedule over a week were processed with greater precision and automaticity in a post-test than non-words and low-frequency words not learned by the participants.

Knowledge of MWUs may also be both explicit and implicit (Doczi & Kormos, 2016), and perhaps mostly implicit (R. Ellis, 2004). According to cognitive psychologists, knowledge of the form of MWUs (or which words go with which) is mostly implicit (procedural) (R. Ellis, 2004). It has been suggested that collocations are learned inductively through repeated exposure (Hoey, 2005; Siyanova-Chanturia & Martinez 2015). Durrant and Schmitt (2010, p. 165) argue that collocations are, compared with idioms and longer strings, "often relatively lacking in salience in the input, so [are] likely to be subject to more implicit processes of acquisition than other—perhaps consciously taught—sequences." One of the first studies to investigate how procedural knowledge develops through exposure to novel collocations in written input was conducted by Sonbul and Schmitt (2013), who found that three exposures to each collocation did not produce measurable procedural knowledge in a primed LDT. However, Sonbul and Schmitt (2013) also report a substantial amount of declarative knowledge of low-frequency medical collocations gained by learners in their study following three exposures in a reading text, leading them to claim that there was a "clear dissociation between explicit and implicit knowledge" (p. 15). This lack of evidence of

procedural (implicit) knowledge was likely partly due to the insensitivity of the LDT as an instrument for measuring the early stage of lexical knowledge. In Experiment 1, I will present learners with a larger number of exposures of (mostly) the same medical collocations in the hope that that will produce stronger memory traces which will register as an effect in the LDT.

2.4 Noticing and attention in L2 learning

Connected to the interface debate (see Section 2.3) is the issue of the roles in second-language learning of the key concepts of noticing, consciousness, awareness and attention. In fact, the first proponent of a major theoretical construct in this area, Schmidt (1990), outlined his noticing hypothesis in reaction to Krashen's (1981) no-interface position that only subconscious processes lead to acquisition—or implicit learning—of language (Ahn, 2014; Schmidt, 2001; Truscott & Sharwood Smith, 2011). A hybrid concept from cognitive psychology and bilingualism encompassing attention and awareness, Schmidt's noticing hypothesis stresses the importance of noticing, or awareness, of language form for language learning, with noticing being at a low level of awareness (and understanding at a high level) (Godfroid, Boers, & Housen, 2010; Schmidt, 1990, 1995, 2001; Truscott & Sharwood Smith, 2011). In the earlier, strong version of the noticing hypothesis, Schmidt (1990) claimed that noticing is the "necessary and sufficient condition for converting input to intake" (p. 129). In vocabulary learning, this would mean that for novel words and MWUs to be learned or acquired they must first be (consciously) noticed. In the later, weaker version of the hypothesis, he states that "more noticing leads to more learning" (Schmidt, 1994, p. 18).

Other SLA models of noticing and attention claim that encoding of language in long-term memory, or language learning, can occur without awareness, or consciousness (e.g., Robinson, 1995; Tomlin & Villa, 1994) but emphasise the importance of attention as a prerequisite for learning of a second language (Gass, Svetics, & Lemelin, 2003; Robinson, 1995, 2003; Schmidt, 1995; Tomlin & Villa, 1994). In fact, it appears that merely paying attention to an item is enough to encode it in the memory. The experimental psychologist, Gordon Logan, argues that once attention has been paid to an item, its encoding in, and retrieval from, the memory is an "obligatory, unavoidable consequence" (Logan, 1988, p. 493). Attention can be defined as "the process that encodes language input, keeps it active in working and short-term memory

and retrieves it from long-term memory" (Robinson, 2003, p. 63). Schmidt (1995, p. 1) claims that noticing and attention are "nearly isomorphic." Other researchers (e.g., Tomlin & Villa, 1994), however, argue that a low level of attention can be at the unconscious level and thus not meet the threshold for noticing/awareness. Tomlin and Villa (1994) separate the concept of attention into three elements: alertness, orientation and detection. Detection—similar to Gass, Svetics and Lemelin's (2003) *apperception*—is the cognitive registration of stimuli, which, Tomlin and Villa argue, is the level at which acquisition must operate and that information can be detected without an individual's awareness. As evidence of this dissociation of detection and awareness, they point to semantic priming studies in which participants report being unaware of primes. SLA researchers have investigated the development of grammatical rules in learners, but few have examined grammatical development among "unaware" learners. Leow (2013, p. 5) reports that those studies that have have clearly indicated that awareness appears to have a "facilitative effect on intake and learning", while the effects of lack of awareness are unclear.

The first attempt in SLA research to operationalise attention and noticing (awareness) was Godfroid, Boers and Housen's (2013) eye-tracking experiment. The researchers operationalised attention as a continuous variable measured by eye-fixation times during the reading process. Their vocabulary post-test was a recognition gap-fill test of the 12 target items taken under time pressure. They found that longer, and presumably deeper, engagement ("focused attention") with novel words during reading was more likely to lead to learning of those words. Godfroid et al. report that the amount of attention to the novel words and the subsequent amount of learning were directly, positively related. The researchers' results contrast somewhat with those of psychologists Williams and Morris (2004), who found in their second experiment that shorter initial gaze duration times of novel words but longer second-pass reading times of native-speaker participants led to higher scores on a synonyms post-test. Williams and Morris cannot account for this result, but Godfroid et al. offer a number of explanations for the different results between the two studies, including different text lengths and different tests. The latter researchers stress, however, that both studies show that longer later fixation times facilitated word learning.

In another eye-tracking study, Choi (2017) found that the amount of attention paid to the target collocations—specifically, total reading time and fixation count—was

increased through typographic enhancement, leading to superior recall of the collocations in a cued-recall test one week later. (More information on this study is given in section 2.4.1.) In their post-tests, Godfroid et al. (2013), Williams and Morris (2004) and Choi (2017) measured only declarative collocational knowledge, so it is unclear whether more attention to the collocations led to faster collocational processing speeds. It does seem, however, that generally the quality and amount of attention determines the quality of the encoding (Logan, 1988) and that the more attention that is paid to a vocabulary item the greater the likelihood of the item being committed to long-term memory (Doczi & Kormos, 2016).

2.4.1 Input enhancement

An effective approach to increasing retention of vocabulary entails altering the *input*—or "potentially processible language data" (Sharwood Smith, 1991, p. 167)—presented to the learner. *Input enhancement* is the manipulation of input in a way which directs the learner's attention to—and in some cases, promotes noticing of—aspects of language through a brief externally-induced focus on target words, usually in the context of meaning-focused activities (Barcroft, 2003; Boers et al., 2016; Sharwood Smith, 1991; Tomlin & Villa, 1994). Sharwood Smith (1981), applying the input enhancement approach to grammatical structures, initially called it *consciousness raising*; however, in an effort to avoid a focus on the mental processes of the learner—particularly the controversial question of the nature of consciousness—he later (1991) preferred the term *input enhancement.* Long (1991) called input enhancement of grammatical structures *focus on form*. Laufer (2005) argues that there should be a focus on form (incidental) element in vocabulary instruction (although she also highlights the superiority for retention of the *focus on form**s*** approach, or intentional learning).

Input enhancement techniques, which include input flooding, typographic enhancement and glossing, can make target vocabulary more perceptually salient to learners (Barcroft, 2003; Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; R. Ellis, 1999; Han, Park, & Combs, 2008; Hulstijn, Hollander, & Greidanus, 1996), who are able to focus on only a limited amount of information at one time since the human attentional system has a limited capacity (Barcroft, 2002; Rott, 2007; Tomlin & Villa, 1994). Barcroft (2003) identifies two key dimensions of input enhancement: how much

it changes the input and how much more than just understanding the input learners are required to do:

> More invasive varieties of input enhancement tend to involve altering the original input to a substantial degree (e.g., input flood, marginal glosses) and tend to require learners to perform tasks that go beyond input comprehension only (e.g., explicit instruction, a multiple-choice activity), whereas less invasive varieties do not alter the original input to such a degree and do not require learners to perform tasks that go beyond processing the input only (e.g., typographical manipulation) (Barcroft, 2003, p. 49).

Input flooding—or the deliberate insertion of repeated occurrences of target items in input—in both contextualised and decontextualised settings, is effective for learning both single words and MWUs (e.g., Durrant & Schmitt, 2010; Han et al., 2008; Sonbul & Schmitt, 2013; Webb, Newton, & Chang, 2013), and is often used in conjunction with typographic enhancement (e.g., Rott, 2007; Sonbul & Schmitt, 2013; Szudarski & Carter, 2017). Studies have found that a larger number of exposures to individual words in context leads to better retention than a smaller number of exposures (e.g., Horst, Cobb, & Meara, 1998; Hulstijn, Hollander, & Greidanus, 1996; Hulstijn & Laufer, 2001; Paribakht & Wesche, 1997; Rott, 2007; Szudarski & Carter, 2014; Webb, 2007; Zahar, Cobb, & Spada, 2011). For example, Horst, Cobb and Meara (1998, p. 215) found that eight or more occurrences of low-frequency target words in a graded reader tended to produce "sizeable learning gains" in a multiple-choice recognition test. Webb (2007) presented learners with words in short contexts in four learning conditions—1, 3, 7 and 10 encounters—and administered 10 tests of different aspects of declarative knowledge of the words. The larger the number of repetitions, the more knowledge was gained. He also found "sizeable learning gains" from 10 encounters but argues that gaining full knowledge of words might require exposure to more than 10 occurrences. Brown, Waring and Donkaewbua (2008) found that the larger the number of encounters with a word in a graded reader the more likely the word was to be learned, as measured in a multiple-choice form-recognition test and a meaning-translation test. They concluded that reading a word 7-9 times in a graded reader was insufficient for long-term retention (measured 3 months later) and that 30-50 encounters might be required. As for procedural knowledge of single words, it appears that as few as three repetitions of

individual words in constrained contexts may be enough to establish robust lexical representations. As noted above, Mestres-Missé, Rodríguez-Fornells and Münte (2007) found, using ERPs, that participants activated semantic knowledge of new words after only three exposures to each of the words in three meaningful sentences.

Repetition alone of words is not the strongest predictor of learning for all words, however. Crossley, Subtirelu and Salsbury (2013), examining lexical production of beginning-level second-language learners, found that, while word frequency was the strongest predictor of noun production in speech, contextual diversity was the strongest predictor of verb production. Nouns were likely learned through repetition alone because of their lexical properties such as being more concrete, more imageable and less ambiguous than verbs. "The abstractness and ambiguity found in verbs likely make them more difficult to acquire through repetition alone, and, thus, acquisition is likely aided by exposure to the words in a variety of contexts." (p. 745).

Repetition in the input is effective not only for individual words but also for MWUs. Durrant and Schmitt (2010), for example, report that participants who had been exposed to each target adjective + noun collocation twice exhibited greater recall of the items in a timed naming task than those who had been exposed to the items only once. Webb, Newton and Chang (2013) found that the more often a collocation was encountered in a graded reader the more likely it was to be learned (as measured in declarative-knowledge post-tests) and that large learning gains occurred if a collocation was encountered 15 times.

Sonbul and Schmitt (2013) found that three highly-contextualised occurrences each of 10 technical collocations in a reading text was enough to produce substantial retention in advanced non-native speakers, as measured in immediate and delayed post-tests of form recall and form recognition. Szudarski and Carter (2016) report that six or 12 repetitions of 10 V+N and 10 adjective + noun collocations in reading texts (with no typographic enhancement) read over three weeks did *not* lead to retention of the collocations on post-tests two weeks after the treatment finished. They conjecture that the lack of retention was due to the presence of low-frequency nouns (e.g., *retort*, *shortcut*) in the collocations.

A factor related to repetition in input that needs to be considered is the contextual diversity of the texts. Two types of diversity are verbatim repetition and varied repetition, both of which appear to be effective. Verbatim repetition is repetition

of each lexical item in the same textual context, while in varied repetition the contexts are different for each repetition of the item. Bolger, Balass, Landen and Perfetti (2008) exposed undergraduate native speakers to low-frequency English words with abstract meanings (e.g., *venial*, *turgid*) in written sentence contexts. In one condition, participants read each target word in four different sentence contexts; in another condition, they read each target word in the same sentence presented four times. The researchers found that exposure to varied contexts led to greater decontextualised knowledge of words, as measured in a meaning-generation task, than repetition of the same context. They argue that in varied contexts each encounter with a new word produces a "context-encoded episodic memory trace": "Abstraction over instances occurs with time, as the context portion of the memory fades … and the meaning features shared over contexts become more prominent" (p. 148). However, Durrant and Schmitt (2010) found an advantage for verbatim repetition in their experiment. They asked graduate-level non-native speakers to read low-frequency target adjective-plus-noun combinations (e.g., *cheap ball*, *effective guide*) in sentence contexts. In the fluency-oriented verbatim repetition condition, participants saw each sentence twice; in the varied repetition condition, participants saw two different sentences for each target collocation in their experiment. In a naming task, the verbatim condition was found to be superior to the varied condition (with a medium effect size). Durrant and Schmitt speculate that initial memory traces of the collocations were stronger in the verbatim condition because there was a lower cognitive burden on the participants than in the varied condition. One possible reason for the difference in the two studies' results could be related to the frequency of the target words: Bolger et al.'s (2008) target words were low-frequency items whereas Durrant and Schmitt's (2010) combinations were low in frequency but were transparent, and the component adjectives and nouns in the combinations were individually high in frequency. Both verbatim and varied repetition will be used in the experiments in this thesis.

Another type of input enhancement is the provision of contextual support of the target vocabulary items, which can facilitate understanding of their meanings. Education researchers Beck, McKeown and McCaslin (1983) categorised contexts that give clues about a word's meaning into *pedagogical* contexts, which are designed to teach specific words, and *natural* contexts, which are not intended to convey the meanings of words. They further divide natural contexts into a continuum: at one end

are *directive* (supportive) contexts, which lead readers to the meaning of a word, and at the other end *misdirective* contexts, which lead the reader away from the meaning. Researchers who have empirically investigated the link between contextual support and learning include Daneman and Green (1986) and Bolger et al. (2008). Daneman and Green (1986) found that context strength, or constraint, was a major predictor of the learning of unknown words by adult NSs. Bolger et al. (2008) included definitions of rare words in their learning conditions alongside a variety of contexts and a single repeated context. They found that the combination of definitions of the target words and sentences containing the word produced better decontextualised learning of the words, as assessed in a meaning-generation task (but not in an orthographic choice task or a forced-choice sentence completion task. This result, they argue, occurred because "the process of deriving abstract knowledge from contextual instances is enhanced with explicit exposure to dictionary-style definitions that provide core meaning features" (p.127).

Typographic enhancement, such as bolding, italics, underlining and quotation marks, is commonly used in introductory university textbooks to familiarise student with key words (see Bramki & Williams, 1984); it is also used for the same purpose in English-language text books (e.g., Schmitt & Schmitt, 2011). It has been widely used to investigate retention of grammatical form, with mixed results. For example, Jourdenais, Ota, Suffer, Boyson and Doughty (1995) found that typographic enhancement drew attention to Spanish preterit and imperfect verb forms and had a positive effect on participants' written output of the target features. However, typographic enhancement of Spanish present perfect and present subjunctive forms in Leo, Egi, Nuevo and Tsai's (2003) study had no effect on the participants' intake of the forms. Evidence from eye-tracking studies indicates that typographic enhancement of grammatical forms in texts draws learners' attention to non-salient forms, but the effectiveness of retention of these forms is mixed (Cintrón-Valentín & Ellis, 2015; Winke, 2013). In contrast, typographic enhancement in Winke's (2013) study did not produce measurable acquisition of passive verb forms in an error-correction post-test. Cintrón-Valentín and Ellis (2015) found that typographic enhancement of verb inflections during reading produced better verb comprehension and production than no treatment.

The small number of studies which have investigated the effects of typographic enhancement on the learning of the forms of discrete lexical items have produced

discouraging results. Barcroft (2003) found that typographic enhancement of single words in a vocabulary list was effective (as measured in L1>L2 and L2>L1 recall tests) only when applied to a limited number of words (3 out of 24 words) but with a larger number (9 out of 24 words) it was ineffective. That result may have been partly due to the fact that the target words were completely decontextualised so they were already likely salient to the learners. Kim (2006) found that EFL learners who read 26 typographically-enhanced unknown target words (e.g., *windfall*) in a reading text had no advantage in a form-recognition post-test over learners who saw the words unenhanced. (I assume that each target word appeared only once in the treatment text: the author does not specify the number of occurrences.) She also found that typographic enhancement alone did not help meaning recognition of the words. Kim speculates that the no-effect result may be because the comparison group reported having seen half the target words in the text. If this is true, then the low-frequency words that were noticed were already quite perceptually salient to the learners, and typographic enhancement was largely unnecessary. In a study in which typographic enhancement was part of a "focused attention" learning condition, Gass et al. (2003) found that the underlining of novel words plus several other conditions (such as being instructed to pay attention to the target words) led to a learning gain of the target words of a large effect size (d = .97) as measured in a translation test. On the other hand, the non-focused attention condition, in which the target words were not underlined and participants were told they would answer comprehension questions after reading the text, produced a medium-effect-size learning gain (d = .49).

In contrast to the uncertain effectiveness of typographic enhancement in reading texts on the retention of individual words, L2 studies that have investigated its effects on the retention of MWUs—mostly collocations—in texts (e.g., Boers et al., 2016; Peters, 2012; Sonbul & Schmitt, 2013; Szudarski & Carter, 2014) have largely found an advantage for enhanced input over non-enhanced input in tests assessing declarative knowledge (Boers et al., 2006). In Peters' (2012) experiment, in which the participants were forewarned of the vocabulary test, bolding and underlining of single German words and German MWUs (e.g., *Spenden sammeln*—to raise money) in a reading text led to higher scores on an immediate translation post-test than no enhancement. However, there was no difference between the learning conditions in the delayed post-test. Sonbul and Schmitt (2013), referred to earlier in this section, found that their

participants' gains in declarative collocational knowledge after three exposures to 10 technical lexical collocations were greater under the typographically-enhanced learning (highlighted and bolded) condition than under the unenhanced condition. As reported above, Szudarski and Carter (2016) found that input flooding alone did not lead to retention of the 20 target lexical collocations in delayed tests; however, input flooding plus underlining did produce gains in form recall and form recognition (but not meaning recall). Boers, Demecheleer, He, Deconinck, Stengers and Eyckmans (2016) report that exposure to underlined collocations (nearly all of which were lexical) that occurred once in a text fostered retention more effectively than exposure to typographically-unenhanced collocations in the same text, as measured in a form-recognition test. In Choi's (2017) eye-movement study, mentioned in Section 2.4, the enhanced treatment group read a text containing 14 bolded lexical and grammatical collocations while a comparison group read an unenhanced version of the text. The treatment group spent longer processing the collocations than the comparison group, and the treatment group performed significantly better on a one-week delayed cued-recall test of the collocations. However, the treatment group recalled less unenhanced information from the text than the comparison group. Choi (2017) suggests that there is a trade-off between learning enhanced collocations and recall of unenhanced text since "drawing attention to target language items, such as collocations, is likely to undermine the creation of a coherent mental representation of the text." (p. 418). In one study that focused on idioms, Salazar Campillo's (2015) study, 32 low-level learners of English were exposed in two short texts to eight English idioms, both transparent and opaque (e.g., *to drink like a fish*, *to kick the bucket*). Unexpectedly, the experimental group, who read the idioms underlined, did not perform as well on a recognition test as the control group, who read the idioms unenhanced. However, because of the small sample of participants, no statistical analysis was performed, so the results may be unreliable.

The difference in the learning outcomes of typographically enhanced and unenhanced single words and MWUs, as surveyed in the previous two paragraphs, is possibly because novel single words may be salient to the learner since they are immediately perceived as unknown, whereas MWUs, which are often fairly transparent, may not be recognised as being unknown if the constituent words of the units are known, and the MWUs may not be recognised as units without enhancement. Both studies in this thesis incorporate learning conditions in which repeated occurrences of

typographically-enhanced collocations are compared with repeated occurrences of unenhanced collocations in an effort to establish whether enhancement will produce an advantage in declarative and/or procedural knowledge of collocations.

Glossing, or the provision of synonyms or short definitions of vocabulary items in or near the text, compensates for lack of contextual cues to the items' meanings in the text (Ko, 2012). Glossing has been found to lead to declarative knowledge of unknown individual L2 words (e.g., Abraham, 2008; Carpenter, Sachs, Martin, Schmidt, & Looft, 2012; Hulstijn et al., 1996; Rott, 2007) by focusing explicit attention on them (Schmitt, 2008). Hulstijn et al. (1996) reported that a larger amount of incidental vocabulary learning occurred through text-reading with marginal glosses than through text-reading with no external aids as recorded in a combined recognition and recall post-test and a meaning-generation immediate post-test. Watanabe (1997) found that marginal single glosses and multiple–choice glosses produced more incidental learning of novel words than texts with no glosses, as measured in immediate and delayed translation tests. Ko (2012) found a significant difference between the scores of the glossed conditions and the no-gloss condition in both the immediate and delayed multiple-choice tests of recognition for the 16 glossed target words (which occurred once each in the text). Abraham's (2008) meta-analysis of 11 studies shows that the use of computer-mediated marginal glosses during treatments produced a large effect size of incidental vocabulary retention in immediate and delayed post-tests. It appears that single-instance glosses in texts lead to low retention of the glossed words but that there is substantially more retention when there are several repetitions of the glossed words (Hulstijn, 2003). An example of this is Rott's (2007) study, in which words enhanced in texts with four repeated L1 meaning glosses in incidental learning conditions were learned more effectively by learners of German than words which were glossed once only and then bolded three times (as measured in both immediate and delayed post-tests of active recall and passive recall). While most of the studies investigating retention of vocabulary through glossing have used incidental learning modes, Carpenter, Sachs, Martin, Schmidt, & Looft (2012) employed intentional learning. The participants in the glossing condition remembered translations of a mean 2 out of 8 glossed words in a test four days after the treatment, but participants who had inferred the meanings of the new words and had then been given the correct translations remembered more.

Besides drawing attention to words, and possibly facilitating their acquisition, glossing has several other benefits but also has several drawbacks. Among the benefits of glossing are: it aids text comprehension, allowing more difficult texts to be read; it supplies accurate meanings, preventing incorrect guessing of the glossed words; and it minimises interruptions to reading (Nation, 2013). In one of the first glossing studies, Jacobs, Dufon and Fong (1994), participants in the two glossing conditions remembered significantly more of the text than participants in the no-glossing condition in the immediate post-test. However, retention of the glossed items was not measured. There are also a number of potential weaknesses of glossing. O'Donnell's (2012) intermediate learners sometimes lost their place while reading, sometimes read the wrong gloss for a particular word and sometimes read a gloss incorrectly.

The effect of glossing on the learning and retention of MWUs is still unclear: while glosses have been shown, as reported above, to enhance the learning of individual words, very few studies have examined the learning of MWUs through glossing. In one study, Webb and Kagimoto (2009) exposed participants to 24 collocations (e.g., *lose touch*) and their first-language meanings followed by three single "glossed" sentences for each collocation. The receptive-treatment group and the productive-treatment group both substantially improved their receptive-knowledge scores of the target collocations from the pre-test to the post-test. However, it may be misleading to call the treatment technique in this experiment *glossing* since the term usually refers to external definitions of lexical items in longer texts; Webb and Kagimoto did not expose participants to longer texts. Bishop (2004) found that typographic salience produced a larger number of clicks on glosses for MWUs than for non-salient MWUs. His participants had read a 750-word text containing 20 low-frequency words and 20 ~~5~~MWUs synonymous with those words. Bishop argued that typographic salience seemed to increase text comprehension; however, he did not measure retention of the MWUs. Clearly, more research is needed into the learning of MWUs through glossing, a gap which I aim to partly address in Experiment 1 by using repeated glossing in one learning condition.

2.5 Offline and online measures of MWU knowledge

Vocabulary tests can be divided into two broad categories: those that measure the extent of vocabulary knowledge through offline measures—tasks performed without

time pressure in which participants have time to reflect on the form and meaning of the items and make conscious judgements about them (Marinis, 2013)—and those that use online measures—tasks conducted under time pressure which measure processing in real time (ibid.).[3] Most vocabulary research has measured offline, declarative knowledge, usually that of form and meaning (Schmitt, 2010). Discrete offline vocabulary tests of declarative knowledge tend to use either recall or recognition (ibid.; see also Roediger, 1990), both of which are used in my thesis. Recognition is a receptive measure in which test-takers recognise the form or meaning of the target word; for example, they select the first-language equivalent of a second-language word from multiple-choice options (Read, 2000; Schmitt, 2010). In recall tests, the test-takers produce the target word from memory in response to a stimulus; examples of this include giving the second-language word in response to the first-language word and completing a gap-fill format (ibid.). Laufer and Goldstein (2004) categorise declarative vocabulary knowledge measures into a hierarchy of four degrees of strength, or difficulty: *active recall* (the most difficult), *passive recall*, *active recognition*, and *passive recognition* (the easiest) (see Table 1). For greater ease of identification, Schmitt (2010) relabels these categories as *form recall*, *form recognition*, *meaning recall* and *meaning recognition*. Schmitt claims that form recall of a vocabulary item is the first step towards productive mastery. A commonly-used version of the form-recall test measuring initial productive knowledge of collocation is the gap-fill (cloze) technique (e.g., *I prefer ___ tea to strong tea.*), which I use in both experiments in this thesis. Another commonly-used test is the form-recognition test, a multiple-choice version of which I use in Experiment 1.

Table 1: Declarative-knowledge measures of vocabulary

Laufer & Goldstein (2004)	**Description**	**Schmitt (2010)**	**Description**
Active recall	Supply the L2 word	Form recall	Supply the L2 item
Passive recall	Supply the L1 word	Meaning recall	Supply definition/L1 translation
Active recognition	Select the L2 word	Form recognition	Select the L2 item
Passive recognition	Select the L1 word	Meaning recognition	Select definition/L1 translation

[3] It should be noted that in this thesis the terms *offline* and *online* refer to measures performed without and under time pressure respectively. They do not refer to the everyday meanings of not being and being connected to a computer network respectively.

Perhaps the first researcher to develop measures to assess declarative knowledge of L2 MWUs was Bonk (2001). He found that the gap-fill subtest for V+N collocations, in which participants supplied a suitable verb, was reliable and valid. However, the subtest for verb + preposition/particle collocations (phrasal verbs) (e.g., *move on*), in which participants supplied a particle, had low reliability. He conjectured that the latter result was due to the limited number of possible particle answers, leading to possible guessing by some test-takers: "[T]hese students seem to know that *out*, *up*, *on* and *off* are very common particles, and chose them far more than would be expected if they were choosing from the full spectrum of prepositions" (p. 134). Another commonly-used type of test is the form-recognition test, which, Bonk (2001) suggests, is only used in the real world when checking reference materials. Eyckmans (2009) developed a 50-item, form-recognition test of advanced learners' knowledge of V+N collocations. She reports that it was sensitive enough to capture the development of the collocational knowledge of advanced learners over 60 hours of English instruction in eight months. Both the gap-fill form–recall test and the (multiple-choice) form-recognition test are used in my experiments.

The other strand to vocabulary testing, the measurement of speed of processing, is important because fluency, or automaticity, is an essential component in the mastery of vocabulary in production and comprehension (Schmitt, 2010). One aspect that is necessary for fluent reading is vocabulary recognition speed (Grabe, 2009). As Schmitt (2010, p. 106) points out, "If lexical recognition is not fast enough … then reading slows down to a frustrating word-by-word (or even letter-by-letter) decoding, in which meaning construction is impaired and the overall flow of the text cannot be understood." The studies investigating online processing of MWUs have used different techniques, such as eye-tracking (Siyanova-Chanturia, Conklin & van Heuven, 2011; Underwood, Schmitt & Galpin, 2004), self-paced reading (SPR) (Conklin & Schmitt, 2008; Millar, 2011; Schmitt & Underwood, 2004; Tremblay et al., 2011) and lexical decisions (Durrant & Doherty, 2010; Ellis, Frey, & Jalkanen, 2009; Sonbul & Schmitt, 2013; Wolter & Gyllstad, 2011).

In eye-tracking studies, participants do not have to do any tasks (such as press buttons) besides reading, so the technique is very similar to normal reading (Jegerski, 2013; Schmitt, 2010). The eye-tracking studies conducted on MWUs have found a

processing advantage for MWUs over novel phrases. In one of the first such studies, Underwood, Schmitt and Galpin (2004) found that both native speakers and non-native speakers of English more quickly predicted the terminal words of 4-8-word MWUs than those of non-sequences because, based on the first words of the sequences, the participants fixated on fewer words when the words were in MWUs. However, a number of methodological limitations may have affected the reliability of the results. First, the target MWUs were a mixture of lexical phrases, transparent metaphors, proverbs and idioms for which transparency or predictability appear not to have been controlled for. Second, the terminal word of each target MWU was also used as the terminal word in a non-target MWU, possibly producing a priming effect. Third, the lengths of the terminal words in the target and non-target MWUs were not controlled for. Three further eye-tracking studies, all of which are reported above, also found advantages for MWUs over their controls: Siyanova-Chanturia, Conklin and Schmitt (2011), and Carrol, Conklin, & Gyllstad (2016), in both of which natives and higher-proficiency non-natives read idioms (although in the latter study, the advantage was only present for congruent collocations); and Siyanova-Chanturia, Conklin and van Heuven (2011), in which natives and advanced non-natives read high-frequency binomial phrases.

A number of studies have used SPR to measure online processing of MWUs. As with eye-tracking, SPR measures reading comprehension processes in real time using tasks similar to normal reading; however, in SPR the participant presses a key after each word is read in order to display the next word in the sentence (Jegerski, 2013; Juffs & Harrington, 1995; Just, Carpenter, & Woolley, 1982). Longer reading times of words are assumed to indicate processing difficulty while shorter times are thought to indicate facilitation (Jegerski, 2013). One of the first studies to use SPR with MWUs, Schmitt and Underwood (2004), unexpectedly found no difference between the reading times (RTs) of both native speakers and non-native speakers for the terminal (target) words of formulaic sequences of four to seven words (e.g., *on the other hand*, *the straw that broke the camel's back*) and the same words when they appeared in nonformulaic contexts. However, in Conklin and Schmitt's (2008) SPR experiment, both natives and non-natives read idioms faster than nonformulaic phrases. Similarly, as reported earlier, native speakers in Tremblay et al.'s (2011) SPR experiments read the sentences containing lexical bundles more quickly than the control sentences. Millar (2011)

reports that NSs' RTs were significantly longer for non-native-like collocations (learner errors) (e.g., *best partner*) than for correct native-speaker collocations (e.g., *ideal partner*). He argues that the incorrect learner collocations placed "an increased and sustained cognitive burden" on the participants (p. 144). In another experiment by Millar (2016) using SPR with natives, strongly collocating adjective + noun combinations offered a processing advantage over weakly collocating combinations. As reported above, an advantage was also found for high-frequency MWUs over low-frequency MWUs for native and non-native speakers by Kim and Kim (2012). Despite the fact that a number of SLA studies have investigated the processing of MWUs, to my knowledge, no published SLA articles have used SPR to measure *learning* of MWUs; this will be done in Experiment 2 in this thesis.

The LDT is perhaps the most commonly-used online measurement of vocabulary processing speed. A few studies have employed LDTs incorporating priming techniques (see Section 2.4.1) to measure processing of MWUs. I will discuss these studies in Chapter 3.

One important issue regarding the use of online speeded tests, such as self-paced reading tasks (SPRTs) and LDTs, as well as offline untimed tests (both of which types are used in my thesis), is the question of exactly what type of knowledge is being measured. Marinis (2010) refers to online comprehension tasks such as SPR as "implicit tasks" (p. 155):

> On-line comprehension tasks … are relatively immune to metalinguistic abilities because they measure the participants' unconscious and automatic response to language stimuli … Participants do not have time to think about the meaning of the sentence and do not use their explicit knowledge about language (p. 140).

R. Ellis (2005) would seem to agree. The features of the LDT and SPRT fulfil all, or nearly all, of Ellis's requirements for measuring implicit knowledge (e.g., time pressure, response according to feel, metalinguistic knowledge not required). In their review of 52 SPRT studies in second-language research, Marsden, Thompson and Plonsky (2018) report that while the studies make references to SPR measuring implicit knowledge or processing, they do not discuss in depth the nature of the knowledge and processing involved. The authors state that there is "a consensus that reactions in SPRs are deemed to operate below the level of consciousness, though empirical validation of this would

be useful" (p. 10). Nevertheless, as DeKeyser (2003) asserts, no test perfectly distinguishes between implicit and explicit knowledge and, while speeded tests make retrieval of explicit knowledge more difficult than the retrieval of implicit knowledge, time pressure does not necessarily produce "a pure measure" of implicit knowledge (p. 320). That is why researchers use priming in conjunction with online measures (e.g., LDTs, SPRTs), in order to ensure that differences between conditions are measures on the same item (target) and to minimise the use of conscious strategies tapping into explicit knowledge. Similarly, on the topic of measures of explicit knowledge, R. Ellis (2004) argues that even tests intended to measure declarative, explicit knowledge may not measure purely declarative knowledge since the use of implicit knowledge is always an option for a learner; for example, in gap-fill measures, such as a *C-test*, participants may use their implicit knowledge of the distributional and statistical properties of the language. With these caveats in mind, I will however refer to the two types of tests as tests of declarative and procedural knowledge for the purposes of this study.

2.5.1 Priming

Priming is a well-established experimental measure of procedural knowledge in psychology and SLA (N. Ellis, 2015). Psychological studies have, for decades, used priming techniques in LDTs, finding semantic priming effects for associated words (e.g., bread—butter) (Meyer & Schvaneveldt, 1971; Shelton & Martin, 1992). Priming techniques are used in order to reduce the involvement of conscious strategies (N. Ellis, 1994); such techniques can access natural language processing and use, processes which "for the most part [occur] with little awareness on the part of individual language users" (McDonough & Trofimovich, 2008, pp. 1-2).

The priming paradigm involves the presentation of two successive stimuli—a *prime* and a *target*—in a perceptual task or a cognitive task in such a way that previous exposure to the stimuli (relative to a baseline) improves the processing of the target (Anderson, 1995; McDonough & Trofimovich, 2008; McNamara, 2005). A priming effect occurs, for example, in a primed LDT, when participants record faster lexical decision times for the target after the prime than they do to the target preceded by a matched control word. In that case, the prime is said to prime the target, an event which is assumed to be evidence of a relationship between the words (Durrant & Doherty, 2010).

The concept of *collocational priming* was introduced by Hoey (2005) in his theory of lexical priming. He argues that collocation is a psycholinguistic phenomenon and that when language users meet the first word commonly found in a collocation they are mentally primed to anticipate the second word. That is because users' knowledge of words includes contextual knowledge of the words:

> We can only account for collocation if we assume that every word is mentally primed for collocational use. As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the context and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context (Hoey, 2005, p. 8).

In the last decade, collocational priming effects have been found in a number of second-language-acquisition experiments with multi-word units using primed LDTs (e.g., Durrant & Doherty, 2010; Ellis, Frey, & Jalkanen, 2009; Wolter & Gyllstad, 2011). The studies, employing primed LDTs, have shown that "language users are sophisticated in their knowledge of the sequential probabilities of the units of language, and that it is their usage experience that has cultivated this knowledge" (N. Ellis, 2015, p. 10). Ellis, Frey and Jalkanen (2009) employed high, medium and low-frequency lexical bundles (e.g., *the length of the*, *a wide variety of*) with native and non-native speakers. In Experiment 3, participants were asked to read aloud the terminal word of each string. Language processing in the native speakers was affected by MI, while processing in the non-native speakers was affected by frequency. Durrant and Doherty (2010), in two experiments with native speakers, found collocational priming in high-frequency collocations compared with non-collocations, while no priming effect was found for low- and moderate-frequency collocations. In Wolter and Gyllstad's experiment (2011), non-native speakers processed collocations that were congruent in English and Swedish faster than non-congruent collocations; in other words, priming was more prevalent in the processing of the congruent items. While the three studies cited above report the existence of collocational priming, they appear to share a major flaw, however. Sonbul and Schmitt (2013) argue that the studies did not follow McNamara's (2005) requirements for automatic priming and thus the priming may have been observed due to participants' use of task strategies. I planned to avoid that limitation to finding evidence of procedural collocational knowledge as I designed a primed LDT in

Experiment 1; thus, I adopted Sonbul and Schmitt's short stimulus onset asynchrony of 150 ms.

2.6 Summary of research gaps and thesis research questions

This thesis will investigate three areas in which there are gaps in the literature with respect to the development of collocational knowledge. One area concerns the extent to which repetition of collocations in written texts affects the development of procedural—as well as declarative—knowledge. Sonbul and Schmitt (2013) and other researchers (e.g., Webb, Newton, & Chang, 2013) have found evidence of declarative knowledge of collocations following repeated exposure to collocations. While Sonbul and Schmitt were one of the first, if not the first, researchers to investigate the development of *procedural* collocational knowledge, they found no evidence of its existence. Another area that this thesis will investigate is the role of several input enhancement techniques—typographic enhancement, typographic enhancement + glossing, and neither typographic enhancement nor glossing—in the development of collocational knowledge. While typographic enhancement generally produces more declarative knowledge of collocations than no typographic enhancement (e.g., Boers et al., 2016; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016), it is unclear whether typographic enhancement or no typographic enhancement produces faster processing times for collocations. It is not known how glossing affects processing of collocations; this, too, will be a subject of investigation in this thesis. A third area to be researched will be the development of knowledge (both declarative and procedural) of two types of collocations: lexical—specifically V+N—and grammatical—in this case, prep+N. Little, if any, research has been conducted into the learning and processing of grammatical collocations, although it does seem that frequency plays a role in the gaining of declarative knowledge of both grammatical (prepositional) and lexical collocations (Lowie & Verspoor, 2004; Mueller, 2011).

The overarching research question of this thesis is: Which types of knowledge of collocations are developed through different input enhancement techniques in written input?

The question will be investigated in two experimental studies that will address the following more specific questions:

Experiment 1 research question: Which types of knowledge of technical lexical collocations are developed through different input enhancement techniques in written input?

Experiment 2 research question: Which types of knowledge of non-technical lexical and grammatical collocations are developed through different input enhancement techniques in written input?

The next chapter investigates the Experiment 1 research question of how different input techniques help in the development of declarative and procedural knowledge of technical collocations.

Chapter 3: Experiment 1

Experiment 1 was both a conceptual replication and an extension of Sonbul and Schmitt's (2013) pioneering study. The purpose of a conceptual replication is to confirm the generalisability of a study's findings. This is done by changing elements of the methodology of the original study, such as its participants and/or its setting (Polio & Gass, 1997, Porte, 2013) or the method of analysis (Porte, 2013). A conceptual replication contrasts with two other types of replication: an exact replication, in which the elements of the original study are reproduced (but not necessarily repeated), and an approximate replication, in which only "one or two of the non-major variables" are repeated (Porte, 2013, p. 11). In my experiment, I attempted to confirm Sonbul and Schmitt's findings regarding declarative (explicit) knowledge by changing a number of features of their study—too many features to allow it to be defined as either an exact or an approximate replication. As an extension, or follow-up, study, I expected that changes to the methodology of Sonbul & Schmitt (2013) would show the presence of procedural collocational knowledge where the original researchers found none (Polio & Gass, 1997; Porte, 2013). Cumming (2014) argues that "[a] study that keeps some features of the original and varies others can give a converging perspective, ideally both increasing confidence in the original finding and starting to explore variables that influence it" (Cumming, 2014, p.10).

In this chapter, I describe the method I used in the experiment. Firstly, I outline Sonbul and Schmitt's study. Then I describe the aim and research questions of my study, followed by the participants, the materials, the learning conditions, the measures, the procedure and, finally, the data analysis.

3.1 Sonbul and Schmitt (2013)

Sonbul and Schmitt (2013) used a primed LDT and two offline tests to evaluate the development of procedural (implicit) and declarative (explicit) knowledge of collocations in the initial learning stages. The researchers conducted two experiments of a similar procedure, one with adult native speakers and the other with advanced adult non-native speakers from a variety of first-language backgrounds. In the latter experiment, 43 advanced-level students not majoring in medicine were exposed to 15 two-word, low-frequency medical collocations (the component words of which were in

the three thousand most common words in the British National Corpus—Leech, Rayson, & Wilson, 2001—or on the General Service List—West, 1953) in three learning conditions in one learning session. In two incidental learning conditions, participants read a passage containing 10 collocations repeated three times; in one condition (enhanced), five items were bolded, while in the other condition (enriched), five items were typographically unenhanced. In the third, an intentional decontextualised learning condition, participants were instructed to remember five collocations which were presented onscreen for 10 seconds each.

A series of immediate post-tests and, two weeks later, delayed post-tests was administered. The first test was a perceptual collocational primed LDT intended to measure implicit knowledge. In the task, the first word of each word pair was the prime and the second word (a noun) the target, and participants were instructed to make lexical decisions (word or non-word) about the target. The word pairs in this task were the collocations to which the participants had been exposed in the treatment phase (e.g., *cloud baby*) and their control pairs (e.g., *steam baby*) with the same second word as the collocations. For each target, each participant saw either the collocation or the control pair. In this task, the prime was displayed for 150 ms and was immediately replaced by the target, which remained on-screen until response. Two stimuli lists were created (one for each of the two testing groups), each containing 7 or 8 intact pairs, 7 or 8 control pairs and 53 filler pairs. Participants were each shown a fixation point for 2 seconds, then the first word of the collocation (the prime) for 150 ms and then the second word (the target). They were instructed to indicate whether the target was a real word or not by pressing YES or NO on a button box. (Non-words as well as real words were used in the LDT, both as primes and as targets.) The target remained on-screen until the participant had made his or her lexical decision. The two remaining post-tests measured declarative collocational knowledge. The second post-test was a form-recall (cued-recall) task: a summary sentence including a gap-fill and a definition in the margin; the first word of each collocation was missing but no answer options were provided. The third post-test was a form-recognition task: the summary sentence and gap-fill task from the form-recall test was used again, but this time with four answer options and an *I do not know* answer option provided for each answer. In order to avoid establishing prior knowledge of the collocations, no pre-test was given; instead, a control group was used as a baseline.

Sonbul and Schmitt found that both L1 and L2 participants gained significant declarative collocational knowledge from all three learning conditions. For native participants, highlighting collocations in context was not more effective at promoting declarative knowledge than no highlighting. The declarative-knowledge test results for the decontextualised condition were not significantly different from those for the enriched condition (i.e., no highlighting). However, the absence of significant collocational priming in the primed LDT suggests that the L1 participants did not develop procedural knowledge of the collocations as a result of one learning session containing three repetitions (regardless of the learning condition).

Similar results were observed in the second experiment with non-native participants for both declarative- and procedural-knowledge tests. The non-native-speaker participants who had undergone a treatment (in all three conditions) were significantly more accurate in declarative-knowledge tests than the control group; and highlighting of collocations was more effective at promoting declarative knowledge than no highlighting. The declarative test results for the decontextualised condition were not significantly better than those for the enriched condition, an outcome that was unexpected. Also unexpectedly, in none of the conditions did the participants process the collocations fluently, as measured by collocational priming; no significant difference was observed in the response latencies of lexical decisions for the intact collocations and control pairs, suggesting that procedural collocational knowledge was not established.

Based on their findings, the authors conjectured that there was a dissociation between explicit and implicit collocational knowledge in both experiments and that implicit knowledge is more difficult to facilitate than explicit knowledge (1) within a short one-off learning session, and (2) with three contextual encounters. I address the two points listed by Sonbul and Schmitt above in my study by lengthening the treatment period to three sessions over two days and by allowing for more recycling of the target items (nine instead of three contextual occurrences). It is also possible that another factor accounting for a lack of evidence of implicit knowledge—as well as the insensitivity of the primed LDT as an instrument—is the small number of participants. Each experimental group contained only 11 or 12 participants, likely leading to a lack of statistical power for the study. In Experiment 1, I recruited 62 participants in order to increase the statistical power of the study.

3.2 Aim and research questions of Experiment 1

As a conceptual replication, Experiment 1 aimed to confirm the generalisability of Sonbul and Schmitt's finding that repeated occurrences of low-frequency medical collocations produced declarative knowledge of these collocations and that typographic enhancement of the collocations produced more declarative knowledge than no typographic enhancement. The present study also aimed to extend the findings of Sonbul and Schmitt by increasing the number of encounters with the collocations in context and extending the learning treatment from one to two days. In particular, I investigated whether a larger number of contextual exposures would result in procedural knowledge of the collocations; such an effect was expected since previous studies had shown that high-frequency MWUs were processed faster than low-frequency MWUs (e.g., Arnon & Snider, 2010; Conklin & Schmitt, 2008; Durrant & Doherty, 2010; Siyanova-Chanturia, & van Heuven, 2011). Because the focus of my experiment was on learning MWUs from context, I did not include a deliberate decontextualised learning condition, instead replacing it with another type of input enhancement: glossing. As a conceptual replication and extension study, a number of the features of the original study were changed (for details, see Appendix A). The study experiment addressed the following research questions:

1. Does repeated exposure to technical lexical collocations in supportive contexts in written input lead to the development of declarative knowledge and/or procedural knowledge?

2. Does typographic enhancement lead to greater development of declarative and/or procedural knowledge of technical lexical collocations than no typographic enhancement?

3. Does typographic enhancement combined with glossing lead to greater development of declarative and/or procedural knowledge of technical lexical collocations than no typographic enhancement and no glossing?

3.3 Participants

Sixty-two advanced non-native speaker participants took part in the learning procedure and immediate post-tests (compared with 43 non-native-speaker participants in Sonbul and Schmitt's Experiment 2); the delayed post-tests were taken by 46 participants. As with Sonbul and Schmitt, advanced-level proficiency was defined as postgraduate level at the university, for which an overall International English Language Testing System (IELTS) score of 6.5+ was a common means of outside entry. However, undergraduate or non-student non-native speakers with an overall score of IELTS 6.5+, or a reading score of 6.5, were also accepted as suitable participants. Volunteers were initially screened for English language proficiency based on their overall IELTS test scores (6.5 or higher), i.e., upper-intermediate to advanced proficiency. Higher-proficiency learners were sought as they were likely to process vocabulary more automatically and accurately than lower-proficiency learners, which is important for the speeded LDTs (Elgort, 2011). A further measure of lexical proficiency (for the accepted participants) was the Vocabulary Size Test (VST) (Nation & Beglar, 2007). The participants' mean score on the VST was 92.34 out of 140 (= approximately 9000 word families) (SD = 23.56). Participants ranged in age from 18 to 61, with a mean age of 30.13 (SD 7.04). They reported that they had spent an average of nearly four years living in English-speaking countries; the mean age at which they were first exposed to the English language was just over nine years old (see Table 2).

Table 2: Characteristics of participants in Experiment 1

Characteristic	Mean	SD	Range
Age	30.13	7.04	18-61
Time in English-speaking countries	3.93	4.38	0.08-20
Age first exposed to English	9.20	3.54	3-18
VST score (max= 140)	92.34	23.56	40-140

As with Sonbul and Schmitt's experiment, the participants spoke various first languages (see Table 3).

Table 3: First languages of participants in Experiment 1

Participants' L1	No.	Participants' L1	No.
Arabic	2	Javanese	1
Bangla	1	Khmer	1
Cantonese	2	Korean	1
Danish	2	Malay	4
Dutch	1	Malay + Mandarin	1
Farsi	2	Mandarin	11
French	1	Mauritian	1
German	3	Russian	2
Greek	1	Sinhalese	1
Hindi	1	Spanish	4
Hindi + Tamil	1	Tagalog + Kapangpagnan	1
Iban	1	Taiwanese + Mandarin	1
Indonesian	1	Thai	1
Indonesian + Javanese	1	Vietnamese	8
Italian	4		

Participants were recruited through Victoria University of Wellington networks, through advertising posters, mass emails in two schools at the university, university Facebook pages and personal contacts. Participants who expressed interest in participating were emailed a screening questionnaire. Because the target items were medical collocations, potential participants were only accepted into the study if they

had received no medical training. A small number of advanced non-native speakers of English who volunteered were also excluded from the study because they had been living in English-speaking countries since early childhood and were assumed to be able to communicate at the level of native English speakers. Participants were rewarded with a $20 supermarket voucher and the offer of 1-1½ hours of proof-reading by me. Treatment-group participants who took part in the voluntary delayed post-test had their names placed in a draw to win a prize of $200 cash.

3.4 Materials

3.4.1 Collocations

The target items were 15 lexical collocations: technical, medical phrases. They were: *regional control*, *smooth diet*, *gap periods*, *chief complaint*, *pure absence*, *fixed end*, *gene therapy*, *shell shock*, *stone heart*, *golden hour*, *silent areas*, *cloud baby*, *iron lung*, *split hand* and *partial response*. Thirteen of these collocations had been used in Sonbul and Schmitt's (2013) experiment with second-language speakers[4]. However, I decided that two of their collocations, *specific diseases* and *principal cells*, were unsuitable for use in my experiment and replaced with the alternatives *gene therapy* and *partial response*. The reasons are outlined below.

The main problem with using *specific diseases* as a collocation in the present experiment was the potential confusion between its technical medical meaning and its general meaning. Technical definitions of the term can be found in online dictionaries. Here are examples:

-a disorder caused by a special pathogenic organism (*Mosby's Medical Dictionary*, 2009);

-a disease caused by a particular and characteristic organism (http://medicine.academic.ru/118438/specific_disease);

-a disease which produces a determinate definite effect upon the blood and tissues or upon some special tissue; a disease which is itself uniformly produced by a definite and peculiar poison or organism (www.thefreedictionary.com)

While dictionary definitions of the technical meaning of *specific disease* were available, more detailed descriptions of this medical term and uses of this term in

[4] I received the medical collocations, treatment passages and explicit-knowledge tests directly from Suhad Sonbul and then adapted them for my study.

context were not easily accessible online. This would have made it more difficult for me and the participants, who were not medical professionals, to understand its meaning and use. In fact, *specific disease(s)* seems to be used much more commonly in texts as a general, *non*-technical phrase than as a technical phrase. For example, all or nearly all of the 92 occurrences of *specific disease* or *specific diseases* in the Corpus of Contemporary American English (COCA) (Davies, 2010), and the 16 occurrences in the British National (BNC) (Leech, Rayson, & Wilson, 2001), appear to have a general meaning (see examples below). This means that participants exposed to the collocation's technical sense in the experiment would have potentially confused that sense with its non-technical sense. These are some examples of occurrences of *specific disease(s)* in the BNC:

> In the middle of the 19th century those who proclaimed that specific drugs cured specific diseases were called quacks; a century later, the traditionalists were the quacks.
>
> Many support groups have sprung up in recent years, at national and local level, to support and advise families of those suffering from specific diseases or forms of disablement.
>
> More rarely, disease products from specific diseases such as measles, chicken pox, whooping cough, syphilis, gonorrhoea, cancer etc. are used to produce the corresponding nosode.

Compounding the potential confusion is the use of *specific diseases* in the treatment texts of Sonbul and Schmitt's (2013) experiment with non-native speakers of English. As can be seen in the text below, it is rather unclear whether the authors are using *specific diseases* in a general sense, in a technical sense, or in a sense specifically related to heart disease.

> One of the most common serious illnesses in modern society is heart disease. The heart might be affected by a number of specific diseases. Many factors of modern life—such as stress, an unhealthy diet, and the use of cigarettes—contribute to the development of specific diseases to the heart muscle. The most common of these specific diseases is heart attack. Prompt emergency procedures can save

> victims from sudden death. These procedures can sometimes substitute for surgery, but surgery is unavoidable at times. (Sonbul & Schmitt, 2013)

The use of *specific diseases* within the paragraph above is also inaccurate in two instances. Firstly, the authors write about "the development of specific diseases to the heart muscle." Due to their choice of words in this phrase, it is unclear whether they mean "the development of *diseases specific* to the heart muscle" or "the development of specific diseases *in* the heart muscle", or something else. The second instance of inaccurate use of the collocation is when the authors state that "the most common of these specific diseases is heart attack." This statement is factually incorrect: a heart attack is *not* a disease but an event resulting from coronary heart disease (National Heart, Lung, and Blood Institute, n.d.; Tidy, 2017).

Principal cells was rejected since its very specific meaning (one type of cells in the thyroid gland) would have made it very difficult to place in three texts on three different topics. The following is the paragraph from Sonbul and Schmitt's (2013) experimental text containing the second problematic collocation, *principal cells*:

> Another major killer is cancer. Cancer is characterized by an unrestricted growth of abnormal cells. There are three main types of cancer: carcinoma, sarcoma, and leukemia. Cancer can appear anywhere in the body, but some common sites are the lungs, breasts, skin, colon, and blood. Some of the most common symptoms of cancer are bleeding, a thickening in any area, difficulty swallowing, or unexplained weight loss. Symptoms vary greatly depending upon the location. For example, neck swelling is often diagnosed as thyroid cancer affecting the principal cells of the gland. These principal cells are arranged in circles. The thyroid gland is responsible for metabolism through the hormone produced by these principal cells. (Sonbul & Schmitt, 2013)

As can be seen in the preceding paragraph, placing the phrase *principal cells* into a narrow context within the text has meant that:

1. the focus of the sentences containing this collocation move from the general topic of cancer to a particular type of cancer (thyroid cancer) and then focus narrowly on one type of cell in the thyroid gland (*principal cells*);

2. the authors have had to describe what principal cells are for the reader's understanding. Providing this description has meant:

a. an increase in the density of medical information in the last three sentences of the paragraph; and

b. an increase in the concentration of low frequency words (e.g. *thyroid, gland, metabolism, hormone*) in the last three sentences.

The effect of the above factors is to:

1. make the information in the sentences containing the collocation difficult for a non-specialist, particularly a non-native speaker, to understand; and to

2. make it very difficult to use the collocation in different contexts in two further texts (as was done with the other collocations) because of its narrow scope of reference. By way of contrast, an example of a collocation that can fit into more than one context is that of *split hand*. In my experiment, I inserted *split hand* into (1) a text on injuries as a condition that possibly looks like (but is not) an injury, (2) a text on diseases and injuries in a section on conditions that can have negative psychological effects, and (3) a text on modern treatments in a section on illnesses and conditions that currently do not have a cure.

3.4.1.1 Replacement collocations

To find the two replacement collocations, I largely followed Sonbul and Schmitt's (2013) selection procedure. In the first stage I consulted online medical resources (although Sonbul and Schmitt consulted paper medical resources). Collocations that were found met the following criteria of Sonbul and Schmitt (p. 130): they were transparent; their frequency in the BNC was very low; and the collocations' first words had several synonyms which could be used as distractors in a multiple-choice test and used in control pairs. In the second stage, control pairs were created for the LDT (see below). The replacement collocations were *partial response* and *gene therapy*.

3.4.2 Pseudowords

Seventy-six pseudowords, two for each of the two real words in the 15 medical collocations, plus another 16, were chosen for the experiment. I did not use Sonbul and Schmitt's (2013) pseudowords since they were not available but created my own list of

pseudowords using the word and nonword generating programme Wuggy (Keuleers & Brysbaert, 2010). Pseudowords are nonwords, or strings of letters which are not real words but which follow the phonological and orthographic conventions of real English words (ibid.). The pseudowords retrieved from Wuggy were not included in the experiment if they were low-frequency real English words (e.g., *greep, capo, vouch*) or even mid-frequency words (e.g., *teen*) (Nation, 2013). I checked word frequency in the BNC20 window of the Compleat Lexical Tutor (Cobb, n.d.).

Potential pseudowords retrieved from Wuggy were excluded from the present experiment if they had one or more of the following features:

1. They were orthographically or phonologically similar to the words in the medical collocations. I did this to avoid causing transference for the non-native-speaker participants, as alliteration has a mnemonic effect (Boers & Lindstromberg, 2009). Therefore, pseudowords with similar spellings to real words in the collocations, particularly those starting with the same consonant or two-letter cluster as the corresponding word in the real collocation (e.g., *stone* and *stune; split* and *spliy*), were excluded where possible. In several cases, however, the inclusions of pseudowords with orthographical similarities could not be avoided. In the case of one word from a collocation, *complaint*, all of the maximum 10,000 pseudowords generated by Wuggy started with *co-*; all pseudowords for *response* started with either *re-* or *de-*; all pseudowords for *absence* started with *a-*; and all pseudowords for *therapy* ended in *–y).*

2. They were likely homophones of high-frequency real words (e.g., *skeird / scared; grene / green; rize / rise*) were excluded. Pseudowords were also excluded if they had spellings similar to those of high- or mid-frequency real words (e.g., *miky* for *milky*, *hapy* for *happy*, *dredg* for *dredge*). For target words which were nouns, pseudowords that appeared to be other parts of speech (e.g., words ending in *–ed* resemble adjectives) were avoided.

3. They were frequent foreign common nouns, as checked on the internet with the Google search engine. Because the participants were to be from mixed first-language backgrounds, some pseudowords that were words in non-English languages might have been known to one or more of the participants. As well as checking the number of Google hits for each word, I checked the first four or five pages of hits to check that the word being investigated was not a frequent word in another language. For example, one

pseudoword generated for the collocation prime *golden* which was excluded from the experiment was *worten*, the German word for *words*.

4. They were frequent names or acronyms (e.g., of people, products and places), in English but particularly in other languages (e.g., *jill, puth, pisa, WAP*) (according to a Google search). I was particularly careful to exclude names and acronyms found in the native languages of possible participants.

5. They surpassed arbitrary frequency cut-off points in Google searches. Pseudowords with the smallest number of Google hits were chosen where practicable. I set the threshold for words of four letters or more at 500,000 words, although the majority of pseudowords used in this experiment had fewer than 100,000 occurrences in Google. Due to the higher frequency of three-letter words on the internet, the frequency threshold for three-letter words was set at a higher level: at about one million.

6. They occurred more than three times in each of COCA and the BNC. Only six of the 60 pseudowords selected for the experiment appeared in COCA and/or the BNC. *Drich*, *voif* and *vup* occurred in COCA once each but appeared to be typing errors. *Jix*, the nickname of a British government minister, appeared once in the BNC. *VUP* appeared twice in the BNC as an acronym. *ONJ*, an acronym for "osteonecrosis of the jaw", appeared in COCA three times.

After I had controlled for the criteria of the pseudowords, very few suitable pseudowords were available. For example, approximately 290 pseudowords retrieved by Wuggy for the word *hand* were checked before a suitable one was found. Thus, other features of the Wuggy programme frequently used to refine the choices of pseudowords, such as OLD20 (orthographic Levenstein distance) and NED1 (Neighbors at edit distance 1), were not needed for this study.

3.4.3 Control pairs

I created *control pairs* for the two new medical collocations (see Appendix B). This meant finding first words of the control pairs with meanings similar to the first words of the collocations. *Halfway response* became the control pair for *partial response* and *code therapy* the control pair for *gene therapy.* The synonyms of *gene* and *partial* were found using the online resources Macmillan Dictionary and Thesaurus (http://www.macmillandictionary.com) and Oxford Dictionaries

(http://www.oxforddictionaries.com/definition/english-thesaurus). I retained 12 of the other 13 control pairs (e.g., *plain diet* for *smooth diet*; *coat shock* for *shell shock*) used by Sonbul and Schmitt. However, the first word of the control pair for the collocation *shell shock*—*coat*—along with the other two distractors, *hard* and *jacket*, were not adequate synonyms for the specific meaning sense of *shell* in *shell shock.* These I replaced with *round* (the control word), *blank* and *shot*.

3.4.4 Filler pairs

There were three sets of 19 filler pairs (see Table 4). I selected 114 filler words (not Sonbul and Schmitt's as they were unavailable), 57 to function as the first words in filler pairs and 57 to function as the second words. Seventy-six of the 114 filler words were pseudowords generated by Wuggy. In one set, each of the 19 pairs of words comprised two pseudowords. Each pseudoword in 15 pseudoword pairs matched a word in the corresponding 15 collocations for number of letters and number of syllables, while the other four pseudoword filler pairs were chosen to approximate the average number of letters and syllables for either the first or second words of the collocations. Another set of 19 filler pairs comprised word + pseudoword while a third set comprised pseudoword + word.

Most of the 38 real words were generated by Wuggy and chosen according to three main criteria. Each word was semantically unrelated to its corresponding collocation word but matched it in: 1. number of letters, 2. number of syllables, and 3. frequency in the BNC (within the 1000 most frequent word families (1k), 2K, 3k, or, in the case of lung, a word from one of Sonbul and Schmitt's, 2013, medical collocations, 4K). Other factors were taken into account, as outlined above, such as the number of occurrences on the internet as measured by Google, and usage in other languages. Concreteness and abstractness was also loosely and subjectively matched for. For example, address was chosen as a filler word for control, and girl as a filler word for hand.

Table 4: Primed LDT trials

Number x collocation type	Item type
15 x intact collocations 15 x control collocations	1st + 2nd = word
19 x filler word collocations type 1	1st = word + 2nd = pseudoword
19 x filler word collocations type 2	1st = pseudoword + 2nd = word
19 x filler pseudoword collocations	1st + 2nd = pseudoword
TOTAL = 87 collocations	

As in Sonbul and Schmitt's experiments (2013, p. 133), in the present experiment 30 distractor primes from the declarative-knowledge form-recognition (multiple-choice) test (including the 15 control-pair primes) were included as primes in the primed LDT (although each participant saw only 15 distractor primes each since the 30 distractors were spread across two LDT item lists, one for each testing group). For example, *smoke wyha* (word + pseudoword filler) was used in the LDT as a filler for its medical collocation counterpart, *cloud baby* (see Table 3). The aim of this approach was to increase the likelihood that participants would see these distractors as possible correct answers (since, for example, *smoke* is a synonym of *cloud*) in the form-recognition test, which they would take later.

Table 5: Example of collocation, control pair and filler pair

Experimental condition	Prime	Target
Collocation	iron	lung
Control pair	metal	lung
Filler pair	smoke	wyha

3.4.5 Treatment texts

I compiled nine reading texts, each 500 words in length, to act as treatment passages. Each text contained five of the fifteen medical collocations. The nine texts were divided into three sets of three texts each: texts 1, 2 and 3 between them contained all 15 collocations, as did texts 4, 5 and 6, and texts 7, 8 and 9. Three of the reading texts were my adaptations of 1000-word texts used by Sonbul and Schmitt (2013), which were in turn Sonbul and Schmitt's adaptations of a chapter in Tiersky and Tiersky (1992, pp. 38-52). I created another six texts. Below is an example of a reading text (Figure 1). A complete set of treatment tests is in Appendix C.

Figure 1: Example of treatment materials in Experiment 1

2.1 **Name: _____**

Cancer and heart disease

Cancer and heart disease are together responsible for the deaths of more than half the New Zealanders who die each year. However, these diseases are often successfully treated.

Cancer

Cancer is the leading cause of death in New Zealand; it caused 8891 deaths, or nearly a third of deaths in the country, in 2011. Most deaths from cancer occur after the cancer has spread from its original location to other organs. Because of this, **regional control [1]** is important. **Regional control**, or stopping the spread of the cancer from its starting place to other locations, can be done through the use of surgery, drugs or radiation. The use of **regional control** can have a positive effect on the treatment's success. The use of special drugs to attack cancer cells is called chemotherapy. The drugs are often given to patients in cycles, which consist of drug treatments for several days each followed by **gap periods**. No drugs are given to patients during these **gap periods**, which allow normal cells to recover from any negative effects of the treatment. The **gap periods** are longer than the cancer treatment periods. When the treatment kills some of the cancer cells, this is called partial response. More specifically, a partial response means that the size of the cancer has decreased by at least 50% but the cancer has not disappeared. As a result of a partial response, treatment may be stopped unless the cancer starts growing again.

[1] prevention of the spread of cancer in the initial stages

Sometimes a cancer patient's body reacts with discomfort to cancer treatments, and the patient must follow a **[2] smooth diet**. On a **smooth diet**, a person eats foods containing little fibre (the part of fruit, vegetables and grains that the body does not break down) – foods such as milk shakes and soft cheeses. Because a **smooth diet** is a major change in the diet, a patient should check with their doctor before making such a change. A doctor may recommend different variations of this diet, depending on the capability of the patient's body to process food.

[2] a diet without much fibre in it

Heart disease

Heart disease is the second most common cause of death in New Zealand, accounting for 24% of all deaths in 2011. A large amount of heart disease is caused by lifestyle choices, particularly smoking, following an unhealthy diet, being physically inactive, keeping an unhealthy weight and managing stress badly. Making changes to these areas of lifestyle is important in order to prevent or treat heart disease. If such changes are not effective, medicine may be needed. If the heart disease is severe, surgery may be required. The chances of surviving heart surgery are very high, with the survival rate of one type, heart valve surgery, being over 98%. However, during heart surgery, a rare but serious complication that has occurred is the **stone heart** condition. **Stone heart** occurs when heart muscles become stiff, which usually leads to the death of the patient. However, these days **stone heart** is almost always avoided through the use of modern medical techniques.

The first three reading texts, to be read by the participants in the first session of the experiment, contained embedded explanations of each of the 15 medical collocations (e.g., *Stone heart occurs when heart muscles become stiff, which usually leads to the death of the patient*.). The purpose of the definitions was to explicitly provide the meanings of the collocations while at the same time to fit as naturally as possible within the texts. The aim of this high-context approach was to provide opportunities for participants to establish explicit connections between the forms and the meanings of the collocations as quickly as possible. The first set of three texts contained explicit definitions of all 15 medical collocations (unlike Sonbul & Schmitt's texts, which varied in explicitness), a factor which is assumed to help learning more efficiently than if fewer contextual clues were given (Carnine, Kameenui, & Coyle, 1984; Nation, 2013). The use of explicit definitions is ecologically valid, since textbooks and other academic texts often provide in-text definitions of technical terms (e.g., Clark & Randal, 2004; Field, 2013).

In order to increase the comprehensibility of the nine texts, I controlled for several factors (see Appendix D), most importantly word frequency. Each text had at least 97% lexical coverage using the most frequent 3000 words of the British National Corpus (as calculated with the Vocabprofile tool of the Compleat Lexical Tutor—Cobb, n.d.; Nation, 2006, 2012). While 98% coverage might have been optimal for promoting guessing from context (Nation, 2013, p. 352), achieving such coverage proved to be impossible without removing more occurrences of medical terms (e.g. *polio*, *tuberculosis*), terms which needed to be retained in the texts for the sake of the texts' coherence. However, the comprehensibility (lexical coverage) of my texts may have been higher than it was in the texts of Sonbul and Schmitt. In their texts, 98% lexical coverage was reached only at the K7 level.

Other major factors that were controlled for in the creation of the texts were type-to-token ratio, which was limited to between 38% and 44%, and the number of sentences in each text: between 26 and 30.

Each treatment text contained a number of marginal glosses as part of the bolding-plus-glossing learning condition. I placed the glosses in the margins in keeping with the finding that learners tend to prefer glosses to be located there (compared with, for example, at the bottoms of pages) (Jacobs, Dufon, & Fong, 1994; Ko, 2005). Most of the glosses, all of which were restricted to lengths of 6-10 words, with a mean of 8.53 words per definition, were adaptations of Sonbul and Schmitt's (2013) glosses. Each

gloss did not contain the medical collocation but was linked to the appropriate collocation in the text with a number (e.g., [1]), which appeared after the collocation in the text and the marginal glossed entry. The collocation was not repeated in the gloss. Displaying a gloss without the target lexical item is an unusual approach, yet it was necessary in order to avoid repetition of the collocations: to do so would have meant participants were exposed to the collocations on three more occasions, and in a decontextualised situation, confounding the results of the reading-in-context treatments. (See Appendix E for an example text which contains glossed entries).

The glossed definitions were different from the explicit contextual explanations embedded within the texts (and encountered by the participants on the first meeting with the medical collocations). The aim of these differences was to provide the participants with opportunities for semantic elaboration. To increase the potential for such elaboration, where possible, information not included in the embedded definitions was included in the gloss. For example, while *chief complaint* was explicitly defined in the first set of texts as "the main reason that sufferers of [an] illness seek medical help," the gloss description of *chief complaint* was "the most obvious sign of a patient's illness." Furthermore, where possible, to increase opportunities for elaboration, the glossed definitions did not contain either of the words from the collocations. For example, the glossed definition of *shell shock* was "a negative psychological condition experienced by soldiers in war," a definition which contained neither of the component words of the collocation. However, one word from each of six collocations was kept in the definition since this was more straightforward than using synonyms. For example, the glossed definition of the term *cloud baby* was "a baby who looks healthy but who rapidly spreads illness"; an alternative to *cloud baby* such as "a *human child of up to 12 months old* who looks healthy but who rapidly spreads illness" would have been a clumsy and unnaturalistic option. The use of *cloud infant* might have been a feasible option. However, while one definition of *infant* in the online Oxford Dictionaries (en.oxforddictionaries.com) —"a very young child or baby"—may have been appropriate, another, British English, definition of *infant* in the same dictionary is "a schoolchild between the ages of about four and eight", which could have caused confusion.

Each medical collocation was encountered within the texts nine times by each participant over the three learning sessions. Nine repetitions are within the range that

some researchers claim is needed to develop long-term vocabulary knowledge. A review of the literature (Nation, 2000) found that a word needs to be encountered between five and 16 times to be learned (although for a substantial amount of learning about a particular word to occur, 50 or more repetitions may be required—Nation, 2013). Elgort, Brysbaert, Stevens and Van Assche (2017) argue that at least six encounters with words in short texts and 8-12 encounters in long continuous texts are required to learn word meanings. Pellicer-Sánchez and Schmitt (2010) found that a noticeable increase in the contextual word learning of 34 target words in an English novel started with 5-8 exposures and accelerated with 10-12 exposures. In her eye-movement study, Pellicer-Sánchez (2016) found that participants reading a 2300-word story gained considerable declarative knowledge of the six target non-words after eight exposures. She also found that after eight exposures the nonwords were read at the same speed and with the same number of fixations as the previously-known real words. While there is little evidence regarding the number of encounters needed for the learning of collocations, Durrant and Schmitt (2010) suggest that gaining initial receptive knowledge of collocations would require at least 8-10 exposures. A related factor that might have enhanced retention was the proximity of the repetitions of the medical collocations to each other. This proximity could have strengthened the likelihood of successful guessing from context in the treatment texts (Carnine et al., 1984; Nation, 2013, p. 363).

3.5 Learning conditions

While reading the texts, the participants encountered the collocations in three counterbalanced treatment conditions: reading-only (no typographic enhancement and no glosses); bolding (reading with typographically enhanced collocations); and bolding-plus-glossing (reading with typographic enhancement and glossed definitions). Each participant in the treatment groups encountered the same number of collocations (n = 5) in each of the three treatment conditions. For each participant, the same five collocations were consistently presented in the same condition throughout the experiment. Across all texts, the collocations occurred in all three conditions an equal number of times.

3.6 Measures

Collocational knowledge was assessed using the three tests adapted from Sonbul and Schmitt: a primed LDT and two pen-and-paper tests of declarative knowledge.

3.6.1 Primed LDT

Procedural knowledge was tested in a primed LDT using E-Prime 2.0 (Psychology Software Tools, Inc.). A LDT is a common technique used to measure procedural knowledge. In this task, participants see strings of letters and must decide whether each string is a real word or not by pressing one of two buttons labelled Yes and No. The response time is assumed to indicate the access time of the word and the time it takes to register a lexical decision, and faster response times are thought to indicate familiarity with the words (Haberlandt, 1996).

In my experiment, a priming paradigm was used with the LDT. The second word of each medical collocation was used as the target, about which speeded lexical decisions were made. The target (e.g., *baby*) was preceded by the prime, which was either the first word of the collocation (e.g., *cloud*) or another word synonymous with the prime (e.g., *steam*), making a *control pair*. The development of procedural knowledge of the collocations was operationalised as collocational priming, i.e., faster responses to the targets in the collocations than the targets in the control pairs. In addition to the collocations and control pairs, the stimuli in the LDT included pseudowords (e.g., *woup*, *flyst*), which matched the words in the collocations for the number of letters and syllables. Also included in the trials were 16 filler pairs for the word—pseudoword (e.g., *broad—iruax*) or pseudoword—word (e.g., *crorm—size*) filler trials.

3.6.1.1 Presentation order of items

The lexical items in the immediate LDT post-test were presented to all participants in the same order through a pseudo-randomised list created in Excel using the randomising function. To achieve this pseudo-randomisation, I first entered the 76 items (the medical collocations, the control pairs and the filler pairs) into a Microsoft Excel file. I also placed two practice items before the 76 items to help participants orientate themselves to the task. I then randomised the order of the 76 items in Excel. I

checked that there were no more than three targets from the collocations or target pseudowords in a row, and that there was spacing among the control pairs; this necessitated several manual changes. I also swapped the positions of two pairs. For example, the positions of the filler pairs *key cogstirch* and *voif addicts* were exchanged because *key cogstirch* was adjacent to *ploub copshidge*; in this case, I thought that having the same two initial letters of the target (second) words of the pairs (*cop*-) could have produced a priming effect. The file containing this list became the file for participant testing group A. I then manually changed the collocations and control phrases for group B by making the first words of collocations control pairs and by changing control pairs into collocations. For example, the collocation *regional control* in one testing-group list became the control phrase *domestic control* in the other list. For the second LDT, to be used in the delayed post-test, I randomised the first list in Excel and made four changes to the order, breaking up a clump of four pairs whose second word was a pseudoword, a group of five pairs whose second word was a filler word, a clump of five collocations whose second word was a real word, and removing a collocation from next to its neighbour, another collocation. This approach meant that both testing groups of participants would see the medical collocations or the equivalent control pairs in the same order and in the same positions on the lists. I used pseudo-randomization for the ordering of trials because it eliminated the possibility of clustering of collocations and reduced the possibility of ordering effects (e.g., a practice effect or a fatigue effect) in one list but not the other.

3.6.2 Declarative-knowledge measures

3.6.2.1 Cued-recall test

For the tests of declarative collocational knowledge, I created a text of just over 500 words from the first three treatment texts (which were given to the participants in the first learning session). The rationale for this was to give the participants a shortened version of the treatment text to facilitate understanding. This followed Sonbul and Schmitt (2013), whose reading texts were also used as templates for tests. The test texts contained the 15 collocations and were split into several paragraphs. Only the second word of each collocation was provided in the text. In the cued-recall test, participants were instructed to supply the first word of the collocation by writing in a space. Two aids were given: firstly, the context of the text, which indicated the collocation's

meaning (e.g., "A baby who quickly spreads infectious diseases into its immediate environment is known as a _____ baby."); secondly, a gloss of the medical collocation, a short definition prompt which was different from the gloss for the reading texts was also given in the margin (e.g., for *cloud baby*, "baby carrying a disease"). To avoid the recency effect, the text was a summary of a text that they had read in the first learning session on the previous day. As well as the glossed definitions, participants also had strong contextual clues for the collocations from the treatment texts. This test measured participants' ability to retrieve the form of part of the collocations from a context in which the meaning and some form (the second words of the collocations) were provided.

I made several changes to the marginal-gloss definitions in Sonbul and Schmitt's cued-recall test (see Table 4). The definitions of *partial response* and *gene therapy*, like the collocations themselves, were new. I changed the definition of *gap periods* from "period when a cell cannot divide" to "periods when cancer treatments stop" as the latter seemed to better reflect the explanation in the treatment texts. I changed the definition of *stone heart* from "left side of the heart not beating" in Sonbul and Schmitt (2013) to simply "the heart is not beating" since the sides of the heart are not referred to specifically in the texts. I also changed the gloss definition of *chief complaint* from "main complaint" to "main symptom of an illness" as "main complaint" as a definition does not obviously signify a medical complaint. The cued-recall test is presented in Appendix E.

Table 4: Differences in marginal-gloss definitions in cued-recall tests between Sonbul and Schmitt (2013) and the present experiment

Medical collocation	**Sonbul & Schmitt (2013)**	**This experiment**
specific diseases	diseases with definite effect upon special body part	N/A [collocation replaced]
principal cells	fundamental cells of an organ	N/A [collocation replaced]
partial response	N/A [new collocation]	a decrease in the size of a cancer after treatment
gene therapy	N/A [new collocation]	putting biological material into a patient's cells
stone heart	left side of the heart not beating	the heart is not beating
chief complaint	main complaint	main symptom of an illness
gap periods	period when a cell cannot divide	periods when cancer treatments stop

After the experiment had started, I discovered that two of the items in the cued-recall test gave unintended clues to the participants. In particular, the sentence for *iron lung* read "An (7. [ ] lung)(can be used to help polio victims breathe." The insertion of "An" before the missing word of the collocation gave participants a clear grammatical clue to the answer. I found another, less serious, in the article "a" preceding "_____ response". The missing word, *partial*, starts with a consonant sound, as do two of the distractors; however, one of the distractors, *undone*, begins with a consonant sound. After the first eight participants had done their first two sessions, I corrected the tests: "An _____ lung" was replaced in both tests by "The ____ lung", while the article in "a ____ response" was dropped.

The marking of the test was straightforward for the vast majority of answers. Answers which were synonyms of the correct answers (e.g., *rest periods* for *gap periods*) were marked as incorrect. Answers which were misspellings (e.g., *cheif complaint* for *chief complaint*) or incorrect word forms (*silence areas* for *silent areas*; *gold hour* for *golden hour*) were marked as correct. One misspelling which was also another word—*ion lung* for *iron lung*—was marked as correct. The final questionable area was the inclusion of a compound word instead of a single word—*split-hand hand* for *split hand*—which was marked as correct.

3.6.2.2. Form-recognition test

The declarative-knowledge form-recognition test measured participants' ability to recognise the forms of the medical collocations and retrieve their meanings. It comprised the same text as that used in the cued-recall test. Instead of glossed definitions, there was a multiple-choice format: for each missing collocate the correct collocate, the control word (used as a prime in the control pair of the LDT) and two further distractors were placed in the margin (see Sonbul and Schmitt, 2013, p. 132, for a visual illustration).

I modified two sentences of Sonbul and Schmitt's instructions, which read:

> Please look at the context and choose the word that completes the phrase. If you do not know the answer and can only guess, please choose option 'e': "I DO NOT KNOW"."

This instruction appeared to discourage guessing based on partial knowledge. Because multiple-choice questions test partial knowledge (Nation & Webb, 2011), I wanted to encourage guessing based on such knowledge. Therefore, I changed the sentences to:

> Please look at the context and choose the word that you think completes the phrase. If you have no idea of the correct answer, choose "I DO NOT KNOW".

One task from Sonbul and Schmitt's experiment that I did not include in mine was the exercise asking participants to rank their confidence level that they were answering correctly. I omitted that task as it would have extended the length of the text-reading process and thus the time needed to complete it. Furthermore, it appeared to be redundant. Its purpose seemed to be to give the researchers an indication of the amount of guessing that the participants did in a test. However, it is unclear how reliable this would be as a measure; it also seems reasonable to suggest that the researchers did not see the confidence measure as producing significant results since none were presented in their article. Although I eliminated this measure from my experiment, I stressed in the instructions the importance of participants not making random guesses. As well as including Sonbul and Schmitt's "I DO NOT KNOW" option in the set of possible answers for every collocation, the instructions in the present study emphasised that participants should choose the "I DO NOT KNOW" option if they had no idea of the correct answer. This instruction seems to be clearer than only including an "I DO NOT KNOW" option since choosing answers can often involve levels of guesswork: test-takers may believe that they may know the answers but may not be confident enough to claim that they "know" the answers.

3.6.3 Rationale for no pre-test and no compulsory delayed post-test

This experiment included neither a pre-test nor a compulsory post-test session. The pre-test is commonly used in applied linguistics experimental studies. It functions as a baseline measure through which an experimenter can gauge the amount of learning of the target items achieved by the participants during the study. One risk associated with conducting a pre-test, however, is that it can lead to unintentional learning of the items. As with Sonbul and Schmitt (2013), I did not use a pre-test in the present study, since even one exposure to a collocation could have affected the results by creating an initial memory trace and leading to unintentional learning (Nation & Webb, 2011).

Furthermore, a pre-test containing the collocations could have alerted participants to the specific purpose of the study: not just to examine reading skills in general but to measure the learning of the medical collocations in particular. Instead, to provide a baseline, a control group was used in both experiments as a way of estimating the treatment-group participants' knowledge of the collocations before the treatment phase. I assumed that a finding of no (or negligible) knowledge in the control group meant that the low-frequency medical collocations were unfamiliar to the treatment-group participants since both groups were from the same population. However, a two-tailed *t*-test found that the control group's VST scores were slightly lower than the treatment group's scores. The mean VST score of the 15 control-group participants was 82.20 and the score of the 47 treatment-group participants was 95.57. The difference between the two means was marginally statistically significant ($p = .055$). This result occurred despite the participants being allocated to the control group and the treatment subgroups by pseudo-random means (see section 3.7). An added measure employed in the present experiment was self-reporting of prior knowledge of the collocations by both the treatment-group and control-group participants at the conclusion of their experimental sessions, information which was included as a variable in the data analysis.

The delayed post-test in applied linguistics measures the strength of retention of the target items a day to several weeks after the experimental treatment, allowing the researcher to argue with some confidence that his or her treatment has led to long-term retention gains (Nation & Webb, 2011, p. 279). It is standard for participants to forget some target items between an immediate post-test and a delayed post-test. This in general was the trend observed in the two explicit tests of Sonbul and Schmitt's (2013) non-native-speaker experiment. The participants in their three treatment groups forgot some target items between the immediate explicit tests and the delayed explicit tests apart from the "enriched" (no typographic enhancement) treatment group's score in the form recall tests, where the score increased slightly (from 35.3% to 37.1%). Although RTs were generally faster in the delayed priming test than for the immediate priming test, they were not statistically different.

One disadvantage of the delayed post-test is the potentially low motivation of participants to take part in another experimental session a week or two later. When the experiment is conducted outside class time, such as in a laboratory, this can make

access to the participants more difficult. In the present study, I felt that to attract participants it was necessary to make the second series of post-tests an optional element of the experiment. I thought that reducing the total mandatory experiment time from three hours to a maximum of two and a half hours and the number of sessions to be attended from three to two would be more likely to attract more participants. However, participants were also given the opportunity of taking part in the second series of post-tests.

3.7 Procedure

Each participant was allocated to either one of three treatment subgroups (created to implement the counterbalanced design) or to the control group, in the order of their volunteering for the experiment (according to the date and time that they initially contacted me either by email or by SMS). A number of participants withdrew from the study before attending any sessions. When this occurred, the participant who next volunteered after the withdrawal took both the place and participant number of the person who had just withdrawn. There were 15 participants in the control group (n =14 on the delayed post-test as one participant withdrew from the study for personal reasons); these participants did not complete the reading procedure and only participated in the immediate and delayed tests. The treatment group initially comprised 47 participants (Subgroup 1 = 16; Subgroup 2 = 16; Subgroup 3 = 15), of whom 32 returned for the delayed post-test.

The reading sessions were conducted with small groups of participants from the treatment group. Participants first completed the reading task, in which they were exposed to the medical collocations. They were instructed to focus on the meaning of the texts they were reading (incidental learning); in order to facilitate this approach to reading, they were asked to complete two reading-comprehension questions after each text. Participants were also instructed to read the marginal glosses at the points where references to glosses were included in the text. Participants were given a written sheet containing the instructions, which I read aloud, with participants following the text silently, before the first reading session. I did not mention the bolded collocations[5].

[5] In a one-off incident, one participant reading the texts in a single-person session asked if she could read the three remaining texts aloud. She had read one aloud before I decided that she should not continue the practice with the remaining two texts (as the other participants were not doing that). Also, one participant in a group of four said one of the collocations, *cloud baby*, aloud at the end of a learning

The experiment used a counterbalanced learning and experimental condition and a between- and within-participants design. While reading the texts, the participants read the medical collocations in three counterbalanced treatment conditions: reading-only (no typographic enhancement), bolding (typographically enhanced collocations); and bolding-plus-glossing (typographic enhancement and glossed definitions). Each participant in the treatment group encountered the same number of collocations (n = 5) in each of the three treatment conditions. For each participant, the same five collocations were consistently presented in the same condition throughout the experiment. Across all texts, the collocations occurred in all three conditions an equal number of times.

Following Sonbul and Schmitt (2013), the collocations were divided into three stimuli sets, each consisting of five collocations (see Figure 1). Each of the three treatment conditions was allocated to an experimental block; each of the experimental blocks contained the three stimuli sets in three different conditions (see Figure 2).

Figure 2: Primed LDT stimuli sets

Stimuli set A	**Stimuli set B**	**Stimuli set C**
1 gap periods	1 regional control	1 partial response
2 stone heart	2 smooth diet	2 golden hour
3 shell shock	3 silent areas	3 fixed end
4 chief complaint	4 split hand	4 gene therapy
5 pure absence	5 iron lung	5 cloud baby

session, which another participant later reported as having lodged it in her memory. Although she later indicated that she had not known *cloud baby* before, for the purpose of the study I counted her as having prior knowledge of it.

Figure 3: Distribution of the treatment conditions in the primed LDT experimental blocks

Experimental block 1	**Experimental block 2**	**Experimental block 3**
Stimulus set A: Unbolded	Stimulus set A: Bolded	Stimulus set A: Bolded + glossed
Stimulus set B: Bolded	Stimulus set B: Bolded + glossed	Stimulus set B: Unbolded
Stimulus set C: Bolded + glossed	Stimulus set C: Unbolded	Stimulus set C: Bolded

One of the major differences between the present study and Sonbul and Schmitt's (2013) was in the number of experimental sessions. While Sonbul and Schmitt ran one learning session for each treatment participant, in the current study three learning sessions were run over two days. The first two sessions were held on the first day, and a third session on the following day. This use of multiple sessions was intended to produce a repetition regime, a technique that has been shown to be effective in vocabulary learning studies (Elgort, 2011; Karpicke & Roediger, 2007). It was possible that the memory of the collocation meanings might be consolidated overnight, as research has provided strong evidence that lexical competition effects in the learning of written words emerge following sleep (Wang, Savage, Gaskell, Paulin, Robidoux, & Castles, 2017; Lindsay & Gaskell, 2010). In each session, my participants read different (but comparable) texts containing three occurrences of the same collocations, equating to a total of nine contextual encounters with the same collocations in all texts over the whole learning period. The use of different texts was intended to make the task more engaging while facilitating recycling of the collocations.

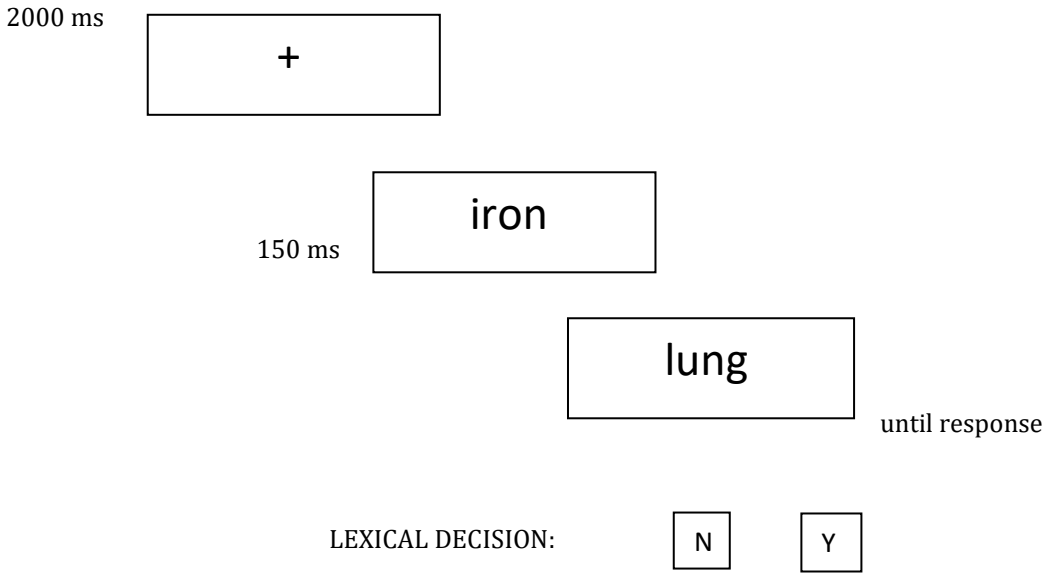
Participants' knowledge of the collocations was measured after the third learning session on day two (immediate post-tests) and again after two weeks (delayed post-tests). Both test sessions were conducted in a laboratory with no more than two participants at a time; a divider was placed between the two computers in the lab. Of the 47 treatment participants, 23 attended the optional second set of (delayed) post-tests two weeks later. These tests were the same tests as the immediate post-tests except

that the order of the items in the priming test had been changed and the texts of the tests had been re-arranged in order to avoid order effects[6].

In the primed LDT, the participants were first presented with a fixation (+) displayed in the centre of the computer screen for two seconds (see Figure 3). That was followed by the prime, either a word or a pseudoword (see the definition of pseudoword below). The prime remained onscreen for 150 milliseconds, after which it was replaced by the target (either a word or a pseudoword). (Both the prime and target were in lower-case letters.) The stimulus display terminated when the participants responded. Participants were asked to respond as quickly and accurately as possible to the target, indicating whether it was a real English word or not by pressing the "Y" (for "yes") or "N" (for "no") buttons on the response box respectively. For right-handed participants, the right button was labelled "Y" and the left button "N"; the button labels were reversed for left-handed participants. The letter strings were presented in white text on a black background (Schneider, Eschman, & Zuccolotto, 2002, p. 12). Each word and pseudoword was only seen once by each participant in these tests. In order to avoid prompting any further learning of the collocations taking place, the correct answers were not shown to participants.

[6] There were several irregularities in the administration of the declarative-knowledge tests. In the immediate post-test session, I gave two participants the *delayed* post-test version of the form-recognition test. In the delayed post-test session, these two participants were given the immediate post-test version of the form-recognition test. I assumed that this would have made no difference to the results as the tests were very similar. Clarity of instructions appeared to be an issue with the tests for some of the earlier participants. Two of the first eight participants mentioned in the debrief after the delayed post-tests that, in spite of the written instructions, they did not read the texts of the delayed declarative-knowledge tests: they went straight to the collocations and filled in the missing words. Following this revelation, I stressed to participants that before they completed the missing words in the tests that they should read the texts of the tests.

Figure 4: LDT display



This study included a control group of 15 participants, also advanced non-native students, who completed the same tests as the experimental group at the same times as the experimental group in order to control for the possibility of prior knowledge of the collocations and the test-retest effects of the immediate and delayed post-tests. The control-group participants attended two sessions two weeks apart. In the first session they sat the three immediate post-tests. The 14,000-word-family version A of the VST (Beglar & Nation, 2007) was used as an intervening task between the priming test and the two explicit tests. During the second session, two weeks later, 14 of the 15 control group participants sat the delayed post-tests, one participant having dropped out due to a personal issue.

Between the first two learning experimental sessions, participants completed a background questionnaire, giving information about their language-learning experience, such as the length of time learning English and living in English-speaking countries (see Appendix F). The questionnaire also acted as an intervening task between the two sessions.

After the participants had completed all the post-tests that they were able and willing to attend, I asked them, either face-to-face, or via email, about their prior knowledge of the collocations. Collecting this information addressed a limitation of Sonbul and Schmitt (2013, p. 143), who assumed that their Experiment 2 participants

had no prior knowledge of the collocations (based on the control group's scores of 2.2% on the form-recall test and 29.6% on the form-recognition test). Familiarity with MWUs has been shown to lead to faster processing times (e.g., Arnon and Snider, 2010; Carrol, Conklin, & Gyllstad, 2016; Durrant and Doherty, 2010; Siyanova-Chanturia, Conklin and van Heuven, 2011; Wolter & Gyllstad, 2011, 2013). Tabossi, Fanari and Wolf (2009) found in a semantic judgement task that native-speaker participants judged decomposable idioms, nondecomposable idioms and clichés more quickly than their matched controls.

For each medical collocation, a participant had one of the following levels of knowledge:

1. knew the phrase in English and its medical meaning;
2. knew the phrase in English but in a nonmedical context;
3. had heard or read the phrase in English (in a medical or non-medical context) but did not know its meaning (e.g. *My iron lung* is the name of a rock music song and album that several participants knew);
4. knew the meaning of the phrase in their L1, which has a similar translation in English (e.g. *golden hour*);
5. knew the meaning of the phrase in their L1 but translation was different to the English phrase;
6. had heard or read the phrase in their L1 (in a medical or non-medical context) but did not know its meaning;
7. was not sure if they had encountered the phrase before;
8. had not encountered the phrase before.

For each participant, each collocation was assigned a score of 1 (previously known) or 0 (not previously known). For the purposes of this analysis, all levels of collocational knowledge except number 8 (had not encountered the phrase before) were counted as prior knowledge. In vocabulary recognition tests, pseudo-words are frequently included to mitigate against over-reporting of self-knowledge (e.g., Meara & Buxton, 1987). However, in this experiment, the use of pseudo-words would have been inappropriate because participants were asked about the collocations at the end of the experiment—after the treatment phase—by which time they were quite well-known to them all.

Overall, the participants' mean prior knowledge was estimated as 11% (SD = 12%). As can be seen in Table 5, the medical collocation reported as known by the

largest number of participants, 21, was *gene therapy.* Next was *golden hour*, known by 16. The collocations previously encountered by the fewest participants were *pure absence* (1) and *cloud baby* (0).

Table 7: Participants' self-reported prior knowledge of medical collocations

Medical collocation	No. participants with prior knowledge (N = 62)
gene therapy	21
golden hour	16
shell shock	13
gap periods	11
iron lung	11
partial response	9
smooth diet	8
chief complaint	6
fixed end	4
split hand	4
silent areas	2
pure absence	1
stone heart	1
cloud baby	0

After the testing had been completed, 38 of the 62 participants reported having previously encountered one or more of the 15 medical collocations (later reduced to 14), either in English or in their first language. Two participants knew seven of the collocations, one knew six, and another four participants knew five; others knew fewer. Each of the 62 participants had prior knowledge of an average 1.52 collocations. On the basis of the prior knowledge self-report, I decided to exclude the most-known collocation, *gene therapy*, from analysis since the phrase was familiar to more than a third of the participants (21 out of 62). This left 14 collocations in the analysis. In the priming test analysis, prior knowledge of the medical collocations was included as a predictor variable.

3.7.1 Order of post-tests

The order of the three post-tests, for both the series of immediate post-tests and the series of delayed post-tests, was: 1. primed LDT; 2. cued-recall task (gapfill + glossed definitions); 3. form-recognition task (multiple-choice answers). This test order was chosen (following Sonbul and Schmitt) to reduce any potential learning effect from one test to the next. It was clear that the form-recognition task needed to be positioned last as this test supplied the collocate answer (along with three distractors) to each item and also allowed participants plenty of time to choose their answers. The more difficult

question was determining the order of the priming task and the cued-recall task in relation to each other.

The primed LDT was placed first to most effectively reduce any potential testing effect (Rowland, 2014) in the ordering of the tests. The short times of exposure to the collocations in the priming test plus the use of many distractors (including 53 filler pairs and 15 control pairs) should have meant that participants had little if any time to consciously elaborate on the forms and meanings of the collocations, thus minimising the amount of collocational learning. The test order reported above was also the sequence used by Sonbul and Schmitt (2013) in order to "avoid earlier measures affecting later ones" (p. 134). Another technique, following Sonbul and Schmitt, to help the priming task minimise learning of the collocations was the inclusion among the fillers and controls in the priming test of 30 distractors from the explicit form recognition test as primes.

It should be noted, however, that Sonbul and Schmitt (2013) have expressed some doubt about the appropriateness of the use of the primed LDT as the initial test. They report that more proficient learners in the control group in their non-native-speaker experiment recorded a significant increase in their delayed declarative-knowledge test scores, a gain they attribute to a potential learning effect from the LDT to the declarative-knowledge tests (p. 153). While this scenario is possible, the evidence that the primed LDT produced a learning effect is scant. Another, perhaps more feasible, explanation by Sonbul and Schmitt is that control-group participants may have had prior partial knowledge of a number of the collocations. A third possibility not suggested by the researchers is that the participants' exposure to the medical collocations in all three immediate post-tests—not just the LDT but also the explicit tests—may have motivated a number of them to retrieve information about the collocations from outside sources in the period between the immediate post-tests and the delayed post-tests (Nation & Webb, 2011, p. 276). There was insufficient evidence to suggest that this experiment should not replicate Sonbul and Schmitt's test order.

It was important for this experiment that the primed LDT came first for another reason. The experiment was primarily interested in establishing the acquisition of procedural knowledge of collocations, evidence of which Sonbul and Schmitt failed to find after one learning session. By placing the priming test first in the sequence, there could be no doubt that any automatic knowledge gained, at least in the first LDT post-

test, would not have been the result of a learning effect from one or both preceding declarative-knowledge tests.

To reduce the testing (retention) effect of the priming test on the declarative-knowledge tests, an intervening task was given after the LDT. Both treatment-group participants and control-group participants were given the second half of the 14,000-word version of the VST (Beglar & Nation, 2007) at that point in the set of immediate post-tests (the first half of the test had been given to participants as an intervening task between the treatment phase and the primed LDT). In the delayed post-test sessions, treatment-group participants were given the 5000- and 10,000-word versions of the Productive Vocabulary Levels Test (Laufer & Nation, 1999), while control-group participants were given the background questionnaire.

3.7.2 Order of post-test items

For all three post-tests, the orders of the items in the delayed post-tests were different from those in the immediate post-tests. The pairs of words and/or non-words in the primed test were pseudo-randomised again for the delayed test version. The text for the declarative-knowledge post-tests was changed, mostly by re-arranging the constituent paragraphs. This was done to reduce the possibility of an order effect.

3.7.3 Vocabulary tests

The 14,000-word-family Version A of the VST (Nation & Beglar, 2007), a multiple-choice meaning-recognition test (Schmitt, 2010), was given to all participants in pen-and-paper form as an independent measure of their language proficiency, since vocabulary knowledge has been found to be correlated with contextual learning of new words (e.g., Daneman & Green, 1986). The VST also functioned as an intervening task between the primed LDT test and declarative-knowledge tests. The 14,000-word version of the VST gives an approximate measure of receptive knowledge of frequent word families. I discovered that two of the first eight participants had completed an online version of the VST, one participant one year earlier and the other several weeks earlier. I considered whether to replace the VST Version A with another version. However, the tests previously completed appear to have been different versions, mostly the online version at myvocabularysize.com, which has very little (if any) item overlap with Version A. Furthermore, at that stage it would have required asking eight participants to

retake the test, a request I was unwilling to make since participants had already given me, or were about to give me, more than two hours of their time over two or three sessions. Discussion with the two participants revealed that neither participant had checked the meanings of any of the words from the test in a dictionary. The participant who had done the test only weeks earlier claimed to have learnt no words from taking the test, to which the answers are traditionally not given, only scores. To monitor the situation I asked the remaining participants after they had finished the experiment whether they had done the test before. The great majority had never done a version of the test, and none appeared to have done Version A. Thus, I decided to keep the VST as a measure of vocabulary knowledge and general proficiency.

I gave an abridged, pen-and-paper version of the Productive Vocabulary Levels Test (PVLT) (Laufer & Nation, 1999) to the treatment-group participants who sat the delayed post-tests. The 5,000 and 10,000-word levels of the test were used as an intervening task. My first version of the PVLT, given to one participant, contained four sections of the test (all sections minus the University Word List section). Completion of this version took 30 minutes, which was too long. Subsequently, I cut the test back to two (the 5,000- and 10,000-word) sections. The error was caused by my lack of piloting of the test instrument and an underestimation of the time that the longer version of the test would take. I thought that the level of difficulty of the 10,000-word part of the levels test might have been too high, negatively affecting the participant's results for the other tests. However, the 10,000-word section of the PVLT appeared to be comparable to that of the second half of the VST (8000-14,000), and the results of the test did not appear to affect the results of the post-tests.

3.8 Data analysis

The data analysis for all tests was conducted using mixed-effects modelling in R (version 3.4.4) (R Core Team, 2018). Participants and items were entered in the models as crossed random effects. A minimally adequate statistical model was fitted to the data, using a stepwise variable selection and the likelihood ratio test for model comparisons (Baayen, Davidson, & Bates, 2008). Participants' lexical proficiency (VST scores) and self-reported prior knowledge of the collocations were treated as potential covariates. The immediate and delayed results were considered within the same statistical model for each of the three knowledge measures; a two-level predictor, session

(immediate/delayed), was included in the data modelling procedure. The resulting statistical model contained only variables for which regression weights were different from zero ($p < 0.1$), or they improved the model fit, or were involved in significant interactions. The models contained random slopes supported by the data (i.e., parsimonious mixed models) because this approach improves the balance between Type I error and power (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017).

For each test, two analyses were conducted. The first (preliminary) analysis was conducted to verify that the control group did not have the knowledge of the collocations. The main analysis aimed to establish which treatments contributed to the development of declarative and procedural knowledge of the collocations[7]. The treatment type predictor had three levels: reading-only (enriched), bolding, and bolding-plus-glossing. The initial alpha level was set to .05. To control for the Type I error rate, function *glht* was used to obtain multiplicity-adjusted *p*-values for the primary interest predictor (i.e., treatment) and the corresponding confidence intervals (CI).

In the two declarative-knowledge tests (gap-fill and form-recognition), the accuracy of responses to the collocations encountered in the bolding and bolding-plus-glossing treatments was compared with that to the collocations encountered in the reading-only condition. In the analysis of the declarative-knowledge tests, responses were scored as either correct (1) or incorrect (0)[8] and analysed using mixed logit models. For each declarative-knowledge test, the odds ratios (OR) and standardised effect sizes[9] (Chinn, 2000) are reported. The odds are defined as the probability of an event occurring divided by the probability of it not occurring (Field, 2013, p. 767). When the odds are more than 1.0, a "success" (in this case, a correct answer) is more likely than a "failure" (an incorrect answer) (Agresti, 2007, p. 28).

---

[7] I did not combine the data of the control and experimental groups in one analysis because participants in the control group did not participate in the learning treatments and, therefore, the primary-interest variable—treatment— could not be applied to the control group.

[8] Marking of the gap-fill test was straightforward for the vast majority of answers. Answers which were synonyms of the correct answers (e.g., *rest periods* for *gap periods*) were marked as incorrect. Answers which were misspellings (e.g., *cheif complaint* for *chief complaint*) or incorrect word forms (*silence areas* for *silent areas*; *gold hour* for *golden hour*) were marked as correct. One misspelling which was also another word—*ion lung* for *iron lung*—was marked as correct. The final questionable area was the inclusion of a compound word instead of a single word—*split-hand hand* for *split hand*—which was marked as correct.

[9] Effect sizes in logit models were calculated by dividing the ln(OR) by 1.81, because the standard logistic distribution has variance $\boldsymbol{\pi}^2/3 = 1.81$ (Chinn, 2000; Grissom & Kim, 2012, p. 273).

In the procedural knowledge analysis, inverse-transformed response times (-1000/RT) were used because the non-transformed RTs' distribution is positively skewed and does not fit the assumption of normal distribution. Procedural knowledge was operationalised as collocation priming; therefore, we compared RTs for each of the three treatments with those for the control pairs. Effect sizes were calculated following Brysbaert and Stevens (2018) (based on Westfall, Judd, & Kenny, 2014). Minimum a priori outlier removal was performed (i.e., only extreme outliers were removed) but the final regression models were subjected to model criticism; potentially harmful outliers (i.e., data points with standardized residuals exceeding 2.5 standard deviations) were removed and the model was refitted (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Brysbaert & Stevens, 2018).

3.9 Results

In this section, I present first the analyses of the declarative-knowledge post-tests. For the primed LDT, I present the preliminary analysis—of the control group data only—followed by the main analysis—of the treatment group data.

3.9.1 Post-tests of declarative knowledge: Main analyses

The analysis of responses on the declarative-knowledge tests showed that encountering a collocation in the bolding condition was significantly more likely to result in correct responses on the cued-recall and form-recognition tests compared to the reading-only condition (tables 9 and 10). In the cued-recall task, the bolding treatment resulted in the largest percentage of correct answers (74%), followed by the bolding-plus-glossing treatment (68%) and the reading-only treatment (63%) (based on model predictions). In the form-recognition test, the bolding treatment was again the most effective, with a mean score of 96%, while the bolding-plus-glossing and the reading-only treatments had scores of 92% and 89% respectively (based on model predictions). (For descriptive statistics, see Table 8.)

Table 8: Mean percentile scores and standard deviations for response accuracy in cued-recall and form-recognition post-tests

		Cued recall		Form recognition	
Group/Condition	Session	*M*	*SD*	*M*	*SD*
Control group	Immediate	0.48	6.90	20.95	40.79
	Delayed	5.61	23.07	23.47	42.49
Treatment group	Immediate	68.54	46.47	92.40	26.52
	Delayed	65.21	47.69	92.17	26.90
Reading-only	Immediate	63.05	48.39	89.16	31.16
	Delayed	61.65	48.81	87.97	32.65
Bolding	Immediate	73.73	44.10	95.76	20.19
	Delayed	67.72	46.04	94.54	23.25
Bolding+Glossing	Immediate	68.04	46.74	91.78	27.53
	Delayed	65.73	47.63	93.71	24.37

In the cued-recall task, I found a significant difference in response accuracy between the bolding and the reading-only conditions ($z = 2.59$, $p = .01$, multiplicity-adjusted $p = .03$, 95% CI: 0.05, 0.99), but not between bolding + glossing and reading-only or between bolding + glossing and bolding. The odds of obtaining a correct answer for the collocations learned in the bolding condition were 1.69 times the odds of those in the reading-only condition (95% CI: 1.14, 2.50) (Table 9).

In the form-recognition task, there was a significant difference between the bolding and the reading-only conditions ($z = 3.08$, $p < .01$, multiplicity-adjusted $p < .01$, CI: 0.24, 1.77) but not between bolding + glossing and reading-only. The odds of obtaining a correct answer for the items learned in the bolding condition were 2.73 times the odds of those in the reading-only condition (95% CI: 1.44, 5.17) (Table 10).

Table 9: Response accuracy on cued-recall test (treatment group): fixed effects

Parameter	Estimate	*SE*	*z*	*p*	Odds ratios	95% CI
(Intercept)	0.56	0.41	1.35	.18		
Treatment=BO+GL	0.18	0.20	0.88	.38	1.19	0.81, 1.77
Treatment=BO	0.52	0.20	2.59	.01	1.69	1.14, 2.50
Session=Immediate	0.31	0.18	1.76	.08	1.37	0.96, 1.94
Prior.Know=1	1.05	0.35	2.99	< .01	2.85	1.43, 5.65

Notes. Intercept levels: Treatment=Reading only; Session=Delayed; Prior.Knowledge=0.
BO=bolding; *BO+GL*=bolding-plus-glossing.

Table 10: Response accuracy on the form-recognition test (treatment group): fixed effects

Parameter	Estimate	*SE*	*z*	*p*	Odds ratios	95% CI
(Intercept)	2.77	0.38	7.33	<.001		
Treatment=BO+GL	0.37	0.30	1.24	.21	1.44	0.81, 2.58
Treatment=BO	1.00	0.33	3.08	< .01	2.73	1.44, 5.17

Notes. Intercept level: Treatment=Reading-only.

It is worth noting that, in the cued-recall test (Table 9), the odds of participants responding correctly to a collocation reported as known prior to the treatment was 2.85 time the odds of the unknown items. The participants were also somewhat less accurate on the delayed than on the immediate cued-recall test, but the effect of session did not reach statistical significance. Prior knowledge, session and VST did not affect form-recognition accuracy and were, therefore, dropped from the final model (Table 10).

Overall, the results of the declarative-knowledge post-tests can be summarised as follows: Bolding > Bolding + Glossing = Reading-only.

3.9.2 Primed LDT: Preliminary analysis

Incorrect responses in the LDT (n = 17, an accuracy rate of 99%) were removed prior to the data analysis. A preliminary analysis was conducted on the control group data only. Three extreme outliers were removed after the data had been inverse-transformed. Experimental condition (collocations vs. control pairs) was the primary-interest predictor in this analysis. The results (see Table 11) showed no reliable priming (t = 0.78, p = .44), with RTs to the control pairs 7ms faster than RTs to the collocations (552ms vs. 559ms). The analysis also showed a small but reliable contribution of session to the model (t = 3.31, p < .01), with RTs on the delayed test about 66ms faster than on the immediate test. There was a 125-ms difference between the RTs of the

participants with the lowest and highest VST scores (t = -1.94, p = .07). Overall, the absence of priming in this analysis confirmed that participants in the control group did not have measurable procedural knowledge of the collocations.

Table 11: RTs of lexical decisions (control group only): fixed effects

Parameter	Estimate	*SE*	*df*	*t*	*p*
(Intercept)	-1.78	0.08	21.63	-22.82	< .001
Exp.Cond=Intact.Collocation	0.02	0.03	243.29	0.78	.44
Session=Immediate	0.19	0.06	13.36	3.31	< .01
VST(centered)	-0.01	0.00	12.84	-1.94	.07

Notes. Intercept level: Exp.Cond=Control.Pairs, Session=Delayed.

3.9.3 Primed LDT: Main analysis

Two extreme outliers were removed from the analysis, leaving 1093 observations. In the main analysis of the primed LDT, I looked for a priming effect for each of the treatments; therefore, RTs to the collocations learned in each of the treatments were compared with the RTs to the control pairs.

Table 12: RTs of lexical decisions (treatment group only): fixed effects

	Estimate	*SE*	*df*	*t*	*p*
(Intercept)	-1.79	0.05	47.68	-34.39	< .001
Treatment=Reading-only	-0.06	0.02	980.13	-2.62	.01
Treatment=BO	0.00	0.02	978.03	-0.18	.86
Treatment=BO+GL	-0.01	0.02	978.25	-0.53	.60
Session=Immediate	0.10	0.03	30.20	3.79	< .01
Prior.Knowledge(centered)	-0.69	0.28	44.60	-2.44	.02

Notes. Intercept level: Treatment=Control.Pairs, Session=Delayed.

Priming was present for items learned in the reading-only condition (t = -2.62, p < .01, multiplicity-adjusted p = .026, 95% CI: -0.12 -0.01), but not in the bolding + glossing or bolding condition (Table 12). For the collocations learned in the reading-only condition, RTs to the terminal word were significantly faster (by 18ms) in the collocations than in the control pairs (538ms vs 556ms, respectively). The RTs were, on average, 32ms faster in the delayed than in the immediate test, and about 110ms faster for participants who reported knowing more collocations.

Descriptive statistics (see Table 13) suggest that collocation priming for the reading-only treatment was numerically greater in the immediate post-test than in the delayed post-test, but there was no statistically reliable interaction between priming and session in the data analysis; in other words, priming occurred in both the immediate and delayed post-tests.

Table 13: Mean RTs and standard deviations (in milliseconds) for collocations and control pairs in immediate and delayed LDTs

		Control pairs		Collocation		Priming CrtColl.-Ctrl.
Group/Condition	Session	*M*	*SD*	*M*	*SD*	
Control group	Immediate	686.53	252.87	688.44	251.55	1.90
	Delayed	615.48	273.90	621.97	224.44	6.49
Reading-only	Immediate			602.94	173.14	-32.65
	Delayed			584.02	179.79	-1.33
Bolding	Immediate	635.59	212.06	640.42	234.23	4.83
	Delayed	585.35	200.29	604.18	239.15	18.84
Bolding+Glossing	Immediate			623.92	176.85	-11.67
	Delayed			607.93	168.25	22.59

Prior knowledge of the medical collocations played a role in the retention of declarative knowledge in the cued-recall tests (both immediate and delayed), possibly indicating that partial knowledge of the collocations helped participants in the initial acquisition and the long-term recall of the items. However, no relationship was found between prior knowledge and retention in the form-recognition tests. In the primed LDTs, there was no interaction between prior knowledge and priming, suggesting that prior knowledge played no role in the development of procedural collocational knowledge.

Lexical proficiency, operationalised as VST scores, played no role in the retention of technical collocations in the declarative-knowledge tests, both immediate and delayed. Lexical proficiency played a role in the speed of access to the terminal word of the collocations and control phrases in both the immediate and delayed primed LDTs only for the control group, but not for the experimental group.

3.10 Discussion

In this discussion section, I firstly present a summary of the main findings from the first experiment. Then I evaluate the research questions and the hypotheses of the study in light of the results. Finally, I discuss the study's limitations.

3.10.1 Summary of findings

Nine exposures to 14 low-frequency, technical, lexical collocations in a spaced repetition regime in this experiment produced off-line, declarative knowledge in all three treatment conditions, as measured in tests of controlled cued-recall and form-recognition. Typographic enhancement (bolding) of collocations led to declarative knowledge more effectively than no typographic enhancement, while enhancement with marginal glosses was no more effective than no typographic enhancement. In a primed LDT, nine encounters with the typographically unenhanced collocations improved the fluency of the online processing of the collocations—in other words, produced procedural knowledge in the reading-only treatments. On the other hand, typographic enhancement—with and without glossing—had no significant effect on the processing fluency of the collocations.

3.10.1.1 Replication of Sonbul and Schmitt's findings for declarative knowledge

This experiment has replicated the findings of Sonbul and Schmitt's (2013) second experiment for the development of declarative knowledge of technical collocations. Both Sonbul and Schmitt's experiment and my experiment found, firstly, that participants gained substantial declarative knowledge (as measured in the cued-recall and form-recognition tests) through repeated exposure to the collocations in supportive contexts in incidental learning conditions—both typographically enhanced and unenhanced—when compared with the control group's results. The second replicated finding is that the presence of typographic enhancement of collocations led to greater retention than the absence of typographic enhancement. Thirdly, a small amount of collocational knowledge was lost between post-test sessions. This indicates that participants largely retained the declarative knowledge of the collocations gained from reading, regardless of the treatment, after two weeks. Fourthly, similar to Sonbul and Schmitt and learning studies in general (e.g., Pellicer-Sanchez, 2017), the participants'

ability to recognise the collocations was superior to their ability to retrieve them (in my study, 95% vs. 67% respectively). Overall, the results show that repetition of technical collocations in supportive contexts is effective at producing declarative knowledge of collocational form, especially when the collocations are typographically enhanced.

3.10.1.2 Extension of Sonbul and Schmitt's study for procedural knowledge

This experiment extended Sonbul and Schmitt's (2013) research into the development of procedural knowledge of technical collocations. Sonbul and Schmitt (2013, p. 153) point out that their "short, massed treatment … might have put the … enriched condition at a disadvantage because unenhanced exposure is said to aid development in knowledge gradually over time". They predicted that a longer treatment period with a larger number of repetitions of the medical collocations would produce long-term implicit (procedural) knowledge of the collocations, where three repetitions in one session in their study did not. My study partly fulfils this prediction by finding a priming effect for the reading-only condition (nine repetitions of the collocations, typographically unenhanced), in both the immediate post-test and the delayed post-test, confirming that procedural knowledge of collocations can develop over time. The priming effect observed in the present experiment, where Sonbul and Schmitt found none, is likely due to a larger number of exposures to the collocations (nine vs. three), more learning sessions (three vs. one), and the distribution of the learning sessions over two days.

3.10.2 The research questions

The study's research questions are now evaluated in light of the post-test results.

Research question 1: Does repeated exposure to technical lexical collocations lead to the development of declarative knowledge and/or procedural knowledge?

The answers to this research question are "yes" for declarative knowledge and a partial "yes" for procedural knowledge. Participants in the present study gained declarative knowledge of the medical collocations after nine contextual exposures across three sessions over two days. The learning gains among the treatment-group participants in the immediate and delayed declarative-knowledge tests were much greater than the control-group participants' gains. Although the control group made learning gains (for

an explanation, see below), it can be argued that gains by the treatment-group participants above the response accuracy rates of 5.61% in the delayed cued-recall test and 23.47% in the delayed form-recognition test were the result of the treatment (Sonbul & Schmitt, 2013).

The fact that the learning conditions, incorporating repeated occurrences of the collocations, resulted in a significant amount of learning in both the immediate and delayed declarative-knowledge post-tests can be explained to a large extent by the frequency effect. According to this model, repeated occurrences of a vocabulary item are more likely to produce learning of the item (Ellis, 2002; Ellis, Frey, & Jalkanen, 2009). While little research has been conducted on the learning effect of the contextual repetition of MWUs, gains in declarative knowledge of the collocations were observed in the two incidental learning conditions in Sonbul and Schmitt's (2013) second experiment: enriched (repeated occurrences not highlighted) and enhanced (repeated occurrences highlighted). As in the current study, Sonbul and Schmitt's non-native participants remembered many of the collocations in the immediate test: 35% and 52% of the answers were correct in the enriched and enhanced learning conditions respectively in the immediate cued-recall post-test. The comparable response accuracy scores in my study were 63% and 74% for the reading-only and reading-plus-bolding conditions respectively. Sonbul and Schmitt's participants recorded scores of 75% and 87% for the enriched and enhanced conditions in the immediate form-recognition post-test compared with 89% and 96% respectively in the current study.

A small amount of attrition was observed in the delayed declarative post-tests of both experiments. The attrition rates for the reading-only and bolding conditions in the cued-recall post-tests in the present experiment were slightly larger than those in Sonbul and Schmitt's experiment: approximately 1 percentage point and 6 percentage points respectively, compared with, rather surprisingly, a gain of 2 percentage points and a loss of 4 percentage points in the equivalent post-tests of Sonbul and Schmitt. In the form-recognition post-tests of the present experiment, the attrition rate was minimal: about 1 percentage point for both the reading-only and the bolding condition. These rates were smaller than those for the same conditions in Sonbul and Schmitt's experiment: 7 percentage points and 9 percentage points respectively.

The control group's learning gains were likely caused by one or more of the following three factors (the first two of which produced a testing effect): brief exposure

in the primed LDT to half of the collocations, which the participants saw twice (once in the immediate post-tests and once in the delayed post-tests); exposure to the correct answers (along with the distractors) in the immediate form-recognition (multiple-choice) test; and prior knowledge of some of the collocations, which was activated in either or both of the primed LDTs and/or in the immediate form-recognition test.

Participants showed gains in procedural knowledge in one of the three treatments: reading only (see Table 6). A significant difference in RTs for collocations and control pairs for the reading-only condition was found in both the immediate and delayed primed LDTs.

While priming was present for the items learned in the reading-only condition, no significant procedural learning was recorded for collocations in the bolding and bolding-plus-glossing conditions. The lack of priming for bolding and bolding plus glossing suggests that the procedural collocational knowledge gained was still fragile and that more than nine encounters over a longer period of time may be needed for such procedural knowledge to develop fully.

No significant difference was found between the RTs for the collocations and the control pairs for the control group. This result was expected since both the collocations and the control pairs were largely unfamiliar to the control-group participants.

Research question 2: Does typographic enhancement lead to the development of declarative and/or procedural knowledge of lexical collocations (compared to reading only)?

The answers to research question 2 are “yes” for declarative knowledge and “no” for procedural knowledge. Bolding of the collocations in the reading texts led to superior declarative knowledge in comparison with the reading-only condition. This learning advantage was found in both the cued-recall and form-recognition tests in both the immediate and the delayed sessions. This result corroborates Sonbul and Schmitt’s (2013) finding that highlighting of collocations produced better declarative knowledge of the collocations in their immediate post-tests than no highlighting. The finding of the present experiment indicates that typographic enhancement is more effective than no typographic enhancement at producing a long-term advantage in the recall of (mostly) previously unseen technical, lexical collocations. More generally, my results support claims that typographic enhancement of vocabulary items—in this case, collocations—

can promote learning of the forms of the items (Barcroft, 2002; Nation, 2013; Rott, 2007).

On the matter of the development of procedural knowledge, the reading-only treatment produced significant collocational priming in the immediate post-test session, whereas the bolding treatment did not. The role of reading without typographic enhancement in the development of procedural knowledge was weak in the delayed primed LDT. The reading-only treatment, involving fluent reading, promoted the establishment of procedural knowledge more quickly than the bolding condition, thus producing an advantage in the procedural-knowledge tests (primed LDTs) over the other two conditions.

It is likely that no measurable procedural knowledge was developed in the bolding condition than in the reading-only condition because the bolding treatment interrupted fluent reading somewhat by drawing extra attention to the forms of the collocations. While this action promoted declarative noticing of the items, which resulted in an advantage in the declarative-knowledge tests, the focus on the form of the items reduced the efficacy of their acquisition as procedural knowledge because they were internalised more as declarative knowledge (Paradis, 1994, p. 395).

Research question 3: Does typographic enhancement combined with glossing lead to the development of declarative and/or procedural knowledge of lexical collocations (compared to reading only)?

The answer to this question is “no”. Both the reading-only condition and the bolding-plus-glossing condition were approximately as effective as each other at fostering declarative collocational knowledge after nine exposures to each collocation. This finding contradicts the argument that the further level of support (i.e., meaning-focused elaboration) induced by glossing would result in better learning (Hulstijn et al., 1996). It is possible that when the participants took their eyes away from the text to look at the marginal glosses their online contextual processing of the collocations was interrupted and the forms of the collocations were not sufficiently encoded. A learning advantage for the glossing condition might have been predicted from the results of previous studies which have found an advantage for marginal glosses in the retention of lexical items (e.g., Abraham, 2008; Hulstijn et al., 1996; Jacobs, Dufon, & Fong, 1994). According to Barcroft’s (2002) type of processing-resource allocation model (TOPRA), semantic

elaboration—such as glossing—increases learning of the *semantic* properties of target vocabulary items while structural elaboration increases learning of the *structural*, or formal, properties. However, semantic elaboration leads to reduced learning rates of structural properties of target items when processing demands are high. Applying the TOPRA model to the contextual learning of L2 collocations, the semantic elaboration involved in processing the glosses appears to have decreased the learning rates of the structural properties of the collocations in the declarative-knowledge tests (and also the primed LDT).

Like the bolding condition, the bolding-plus-glossing condition did not promote the development of any measurable procedural learning. The lack of a priming effect for bolding plus glossing may be explained by three possible factors. Firstly, as already suggested in the explanation for research question 3 connected with declarative knowledge, the extra elaboration provided by the glosses may have reduced the efficiency of processing. Secondly, each medical collocation was explicitly defined in the first occurrence in which the participants encountered it in the texts. For example, in the first text that all participants read, the *stone heart* condition was defined as occurring "when heart muscles become stiff, which usually leads to the death of the patient." This definition was then supplemented in one of the learning conditions by the marginal gloss, "one side of the heart not beating," which elaborated on the definition. It may have been that once a participant established an initial link between the form of the collocation and its meaning in the text, the glossed definitions added no, or little, information to participants' knowledge of the collocation. A third possible factor accounting for no priming effect in the bolding-plus-glossing condition is the nature of the collocations and the nature of the texts. Even without the additional aids of bolding or bolding plus glossing, the medical collocations would have been easily explicitly identified by the advanced non-native participants for at least three reasons. One is that, as mentioned above, each collocation was defined at its first mention in a treatment text. A second reason is that the use of nine repetitions of each collocation meant that participants became familiar with the forms of the collocations, as attested to by the declarative cued-recall test results. A third reason is that vocabulary levels of the texts were relatively low (at least 97% of the individual words in each text were in the most frequent 3000 words of the BNC), while the vocabulary levels of the participants were advanced (estimated mean vocabulary size of 9200 word families, as measured in the

VST). This meant that very few words and phrases were likely to be previously unknown to the participants; this in turn made it easier for the participants to identify the novel collocations, with or without the aid of bolding or glossing.

In the present experiment, exposure to nine repetitions of mostly unknown technical lexical collocations in diverse contexts in multiple short texts produced substantial declarative knowledge of the forms and meanings of the collocations and the beginnings of procedural knowledge. However, the three learning conditions differently promoted the development of both types of knowledge. The reading-only condition—repetition alone of the collocations (without bolding or glossing)—was more effective at developing procedural knowledge of form than the other conditions but was less effective at promoting declarative knowledge than the bolding condition. Typographic enhancement (bolding), which made the collocations' forms more salient and promoted conscious noticing of them, was the most effective condition for the acquisition and retention of declarative collocational knowledge. It was less effective at promoting procedural knowledge, however: perhaps the action of noticing brought about by TE interrupted the normal reading process and the development of procedural knowledge. TE plus glossing was overall as effective as the reading-only condition at producing declarative collocational knowledge and the least effective at promoting procedural knowledge. Its lack of promotion of the development of procedural knowledge in particular is possibly because marginal glosses took the participants' attention away from the text, disrupting the natural processing of the collocations to a greater extent than TE.

The development of declarative and procedural collocational knowledge at different rates under various learning conditions supports Sonbul and Schmitt's (2013) suggestion that there is a dissociation between the development of the two types of knowledge. Procedural lexical knowledge appears to start developing before declarative lexical knowledge, as evidenced in research using ERPs (McLaughlin, Osterhaut, & Kim, 2004). However, the findings of both Sonbul and Schmitt (2013) and my experiment show that the LDT may not be a sensitive enough measure to detect the weak, initial traces of procedural knowledge.

3.10.3 Limitations

One limitation of this experiment was the small number of participants and items. I used the same number of medical collocations (15) as Sonbul and Schmitt's second experiment (2013). In order to address Sonbul and Schmitt's limitation of having only 43 participants, I recruited a larger number of participants (62 for the immediate post-tests). However, this limited number of items was still problematic and more robust results could have been observed with a larger number of collocations, especially on the procedural knowledge measure. Further, the number of participants who took part in the voluntary delayed test session (46) was similar to the number of Sonbul and Schmitt's participants; a greater number of participants in the delayed post-test could have been helpful in finding possible differences between the knowledge measured in the immediate and delayed post-tests.

I also acknowledge that the experimental treatments in this study were somewhat artificial. First, real texts are unlikely to have the same high density of repeated occurrence of collocations as the present study did. Typically, the three occurrences of each collocation within a certain text occurred within three consecutive sentences. Multiple repetitions of the collocations in short texts may have impeded the coherence and flow of the texts and reduced their naturalness. (In contrast, repeated occurrences of lexical collocations could be more naturally inserted into longer texts such as graded readers). Second, it is unusual to see repeated typographic enhancement, such as bolding, of key vocabulary items in authentic texts. In a typical university textbook, typographic enhancement of a technical term usually occurs only once (e.g., Clark & Randall, 2004; Field, 2009). Third, glosses were used in the present experiment in a somewhat unusual way: instead of supporting initial comprehension of the collocations, they were used as a form of repeated meaning elaboration. This may have contributed to the weaker-than-expected effect of the bolding + glossing treatment on learning. The first instance of each collocation in the texts was also accompanied by its embedded explanation, so it is possible that the glosses added little further useful information. Finally, even though the participants had been instructed to read all the glosses, some may have not followed the instructions, as reported by one participant in a post-test debriefing.

The experiment has several limitations connected with testing. First, some of the retention of the collocations recorded in the delayed post-tests will have been the result of not just the treatment condition and/or prior knowledge of the collocations but also of the testing effect (e.g., Carpenter & DeLosh, 2006). Evidence of this comes from the significant gain in learning experienced by the control-group participants; they were exposed briefly to half of the collocations in the immediate primed LDT test and to all of the collocations (among three distractors for each correct collocation) in the immediate form-recognition test. In the delayed primed LDT, participants again saw the same medical collocations that they had seen in the immediate version of the LDT, and, finally, saw the correct answers once more in the delayed form-recognition test. A second test-related limitation is a flaw in the design of the LDT. In the delayed primed LDT test, each participant saw the same set of collocations and control pairs that they saw in the immediate LDT post-test. For example, if a participant saw the collocation *cloud baby* and the control pair *rock heart* in the immediate LDT, they saw the same phrases in the delayed LDT test. However, this meant that they had unnecessary exposure to the phrases and, having seen these phrases twice, a memory trace of each would have been established, reducing the likelihood of finding a priming effect in the delayed LDT. A third limitation is that I did not test depth of knowledge of meaning to check whether the marginal glosses promoted deeper semantic knowledge of the collocations.

The non-random sampling of the participants in the study was another possible limitation. The large majority of the participants were ESL students at university. Thus, while the results of the experiments may be generalisable to ESL university students, caution should be exercised in extrapolating beyond that population.

Another limitation is lack of norming of the collocations. I discovered that a number of the medical collocations were familiar to the participants, most notably *gene therapy,* recognised by 21 of the 62 participants. Norming for familiarity with a comparable group of non-native speakers could have helped identify these more familiar collocations earlier, allowing a chance to replace them with less familiar ones.

Having found evidence of substantial declarative knowledge and some procedural knowledge of technical collocations in Experiment 1, I now began, in Experiment 2, to examine non-technical collocations. What role would typographic enhancement and lack of enhancement play in the development of knowledge of lexical and grammatical collocations?

Chapter 4: Experiment 2

This chapter presents Experiment 2. Firstly, I explain the aim and research questions of the study; then I describe the method, including the participants, materials (which includes the selection process for the target collocations), learning conditions, measures, and procedure. This is followed by the data analysis and discussion of the findings.

4.1 Aim and research questions of Experiment 2

Experiment 1 corroborated Sonbul and Schmitt's finding that repetition of low-frequency medical collocations in written input produces declarative knowledge of the collocations and that typographic enhancement of the collocations produces more declarative knowledge than no typographic enhancement. Experiment 1 also found that a larger number of contextual exposures to the collocations without typographic enhancement over a longer time period resulted in the development of some procedural knowledge in the reading-only condition, as measured in a primed LDT.

In Experiment 2, I focused on non-technical collocations, which are more commonly encountered by learners than technical collocations. Non-technical semi-transparent lexical and grammatical collocations such as *strong tea* and *by habit* may be quite easily understood but at the same time may be more easily forgotten. The aim of the experiment was to check whether learners' repeated exposure to such collocations in written contexts using typographic enhancement and no typographic enhancement would result in increased declarative and procedural knowledge of the collocations (when compared to a control group), and, if so, which treatment was superior for each type of knowledge. I also investigated differences in the types of knowledge gained of lexical (V+N) collocations versus grammatical (prep+N) collocations. (Factors that might make verbs in collocations containing nouns more difficult to encode than prepositions in collocations containing nouns include their length—they usually contain more letters than prepositions—and their frequency: while the great majority of single-word prepositions in English are in the first thousand most-common words, many of the nouns are of a lower frequency.)

The study addressed the following research questions:

1. Does repeated exposure to V+N and prep+N collocations in written input lead to the development of declarative and/or procedural knowledge?

2. Does typographic enhancement lead to greater development of declarative knowledge of V+N and prep+N collocations than no typographic enhancement?

3. Does lack of typographic enhancement lead to greater development of procedural knowledge of V+N and prep+N collocations than typographic enhancement?

4. Do repeated exposure and typographic enhancement differentially affect the development of declarative and/or procedural knowledge of V+N and prep+N collocations?

4.2 Participants

Eighty-three native speakers of Chinese—later reduced to 78 in the analysis as a result of researcher error—took part in the experiment. The 78 participants (56 females and 22 males) ranged in age from 17 to 35, with a mean age of 24.19 (SD 4.43). They reported that they had spent an average of approximately 9½ months living in New Zealand and other English-speaking countries; the mean age at which they had first been exposed to the English language was 9½[10] (see Appendix G).

Participants with English proficiency levels of IELTS 5.0 to 6.5 or equivalent (intermediate- and upper intermediate-levels) were recruited[11]. The lower limit of IELTS 5.0 was imposed so that the participants could be assumed to have a minimum receptive mastery or near-mastery of the great majority of the most common two thousand English word families, a level that was required for easy reading of the treatment texts. The upper limit of IELTS 6.5 was imposed as a proficiency ceiling to

[10] Two participants' data on their age of first exposure to English were missing. One did not supply an answer, while a second wrote "primary school".

[11] Three potential participants who had no recent evidence of their English proficiency level (i.e., in the previous two years) were accepted into the study. Two had scores from China's CET-6 (College English Test) which indicate they were at a suitable level. The third potential participant had no external English test scores. She was one of only two participants to gain perfect scores on the Vocabulary Levels Test at the 2k, 3k and 5k levels (0/90). On reflection, I should have excluded her results from the analysis.

restrict the number of collocations known by participants since higher proficiency learners would have been likely to know more of the collocations. The participants' mean score on the 2k (2000-word-family) level of the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001) was 27.62 out of 30 (SD 2.65) (range = 21 to 30). The vision of all the participants was normal or corrected-to-normal.

Participants were recruited through advertising posters at Victoria University of Wellington, mass emails in two schools at the university, the Chinese networking application WeChat, university Facebook pages and personal contacts. Interested potential participants were emailed a screening questionnaire. Each treatment-group participant was rewarded with a $30 supermarket voucher, while each control-group participant was given a $15 voucher.

4.3 Materials

In this section, I describe the materials used in this experiment. Firstly, I outline the selection process for the experiment's target collocations, including the trialling process. I then describe the selection process for the control phrases and, finally, the creation of the treatment texts.

4.3.1 Collocations

In the description of this experiment, the term *collocation* is used to refer to a word sequence which meets some researchers' criteria for a collocation. The terms *combination* and *phrase* are used as synonyms of *collocation* but also have a wider meaning. *Phrase* in particular is more often used here to denote a combination of words which is not a collocation (e.g., an unconventional phrase such as *learn knowledge*).

I planned to find 48 target collocations for the experiment: 24 pairs of V+N combinations and prep+N combinations in which the noun was the same for each pair. The purpose of having a reasonably large number of collocations was to have many data points for analysis, thus increasing the study's statistical reliability. The purpose of having matching pairs of combinations was to serve as a point of comparison between lexical and grammatical collocations. I employed a mixed-method approach to item selection, using a combination of statistical information to identify potential target items, and human intervention, to make judgements about their pedagogical value (Ackerman & Chen, 2013).

The process of selecting collocations comprised four stages. These stages are summarised in Appendix H. However, for the purpose of concision, I will describe only the final stage of the selection process.

4.3.1.1 Selection of collocations

The second edition of the *OCD* (McIntosh et al., 2009) was the source of nearly all the collocations. The dictionary gives few details about the selection process for its collocations. However, the editors explain that they took a pragmatic, rather than theoretical, approach to the inclusion of collocations, inserting a phrase into the dictionary if it was typical and if a learner of English would want to express the idea. "The aim was to give the full range of collocation – from the fairly weak (*see a movie* …) … to the strongest and most restricted (*see reason* …) … for around 9,000 headwords (McIntosh et al., 2009, p. v)." Free combinations and full idioms were excluded from the dictionary.

Apart from the introduction in the second edition of the dictionary, no articles appear to have been published about the compilation of that edition (Lea, personal communication). However, one of the managing editors of the first edition, the version on which the second edition was based, Diana Lea (2007), states that decisions on the inclusion of collocations in the first edition were based on frequency, statistical significance and editorial judgement about the collocations' usefulness to learners. In the first edition of the dictionary, the editors' working definition of collocation was a compromise between Sinclair's (1966) statistically-based concept (the frequency-based approach) and Hausmann's "more restricted concept based on semantic unpredictability and interdependency" (1984, as cited in Klotz, 2003, p. 58) (the phraseological approach). Corpora, mainly the Oxford English Corpus, were used to find "statistically salient" combinations, which were then assessed for their usefulness to the learner (p. 58).

For collocations to be included on my list in this approach, they met nine requirements:

1. For each collocation, one or more possible *parallel collocations* were able to be identified. In this study, I define parallel collocations of V+N collocations as prep+N combinations containing the same noun; the converse is true for parallel collocations of prep+N phrases. The purpose of the parallel collocations was to allow comparisons to

be made between the learning of selected V+N and prep+N combinations (e.g., *in the knowledge (that)* and *to the knowledge (of)* for *acquire knowledge*). This was particularly important for the measure of procedural knowledge as reading times on the terminal word of the V+N and prep+N collocations were compared.

2. At least one collocation in each V+N and prep+N pair appears in the *OCD* (MacIntosh et al., 2009). (In fact, every collocation except one, *on review*, can be found in the *OCD*.)

3. Each verb was to appear with only one noun in the treatment, since exposure to the same noun in two different V+N or prep+N phrases in the treatment texts and/or the tests could have produced an unwanted priming effect.

4. The preposition in each prep+N combination and the verb in each equivalent V+N combination are in the K1 frequency band of the BNC-COCA-25 list (Cobb, n.d., Nation, 2012)[12].

5. The collocations have meanings that are not transparent or not easily guessed, as assessed by two raters. For example, I avoided collocations containing the noun *emotion* because the two prepositional collocations listed for it are *with emotion* and *without emotion* (e.g., *She spoke with deep emotion*); both phrases contain fairly transparent prepositions which might have been easily guessed in the gap-fill test by participants even if they had not encountered the phrases previously.

6. The prepositional collocations do not commonly incorporate parts of speech other than articles (e.g., adjectives, pronouns—as in *on your mind*) since checking all the forms manually in COCA would have been unnecessarily time-consuming. For example, *by luck* was excluded since the example given in the *OCD* is "By sheer ~ we managed to get out in time" (McIntosh et al., 2009, p. 495). Each collocation comprises two or three words. Although the focus was on V+N and prep+N combinations, many combinations contain one or more extra words between the verb or preposition and the noun (Howarth, 1998a). For simplicity's sake, the only three-word collocations chosen were those containing articles (*a*, *an* or *the*). This meant the elimination of a number of potential target combinations containing adjectives and pronouns (e.g., *at this stage*).

7. The prep+N combinations do not have synonyms with which they might be interchangeable (e.g. *in disgust* and *with disgust*; *at a distance* and *from a distance*).

[12] I switched to searching the BNC-COCA-25 list (Cobb, n.d.; Nation, 2012) instead of the BNC in this phase when an academic pointed out the existence of this updated version.

8. The prep+N combinations are not commonly followed by another specific preposition (most frequently, *of*—e.g., *at the mention of*).

9. Both singular and plural forms of the countable nouns were listed in COCA.

Once I had a large number of collocations on the list, I went through the list removing collocations on three grounds:

1. High raw frequency: V+N collocations and prep+N collocations with raw frequencies of 1000 or more in COCA were removed. Although this was an arbitrarily chosen figure, it seemed reasonably likely that these more common combinations would be known by the participants. In many cases, the deletion from the list of one phrase also necessitated the deletion of its parallel phrase or phrases.

2. Low MI score: I decided that, for consistency, both prep+N and V+N collocations on the final list should have MI scores equal to or higher than the commonly accepted threshold for collocations: 3 (Hunston, 2002). Thus, all prep+N phrases with MI scores of less than 3—in other words, the constituent words were highly substitutable—were excluded.

3. High transparency: The V+N and prep+N combinations that were highly transparent (i.e., they were free combinations, e.g., *provide an explanation*, *without difficulty*) were eliminated. The collocations left were those that were less transparent, with the verbs and prepositions having a figurative meaning (e.g., *set a target*, *over the target*; *form character*; *in character*).

Because the number of suitable collocations remaining on the list after the above-mentioned removal process was under 24 pairs, I returned to COCA and started adding K2 verbs to the list of V+N phrases. This meant that there would not be equivalence of frequency bands between the prepositions, since all of the prepositions were K1. However, all verbs and prepositions would still be in the second-most common thousand words in English, and, as such, would be highly likely to be known by the intermediate- and upper-intermediate-level participants.

As I was examining potential target collocations in this stage, I noticed that, in the prep+N collocations with MI scores above 3, certain prepositions were over-represented on the list while others were under-represented. Of the 30 prep+N collocations whose MI scores I had checked in a previous item-selection approach, 11 contained *under* and six contained *beyond*. In comparison, shorter prepositions were

under-represented on the list: *on* appeared only once on the list of phrases with MI scores over 3 and *in* appeared not at all (although *in* was in five prep+N combinations with low MI scores). Many prep+N combinations listed in the *Oxford collocations dictionary* have low MIs and would thus not be considered collocations according to that MI-threshold-of-3 definition of collocations.

This bias in favour of certain prepositions and against other prepositions leading to a lack of variety of prepositions in high-MI target collocations, as well as the small number of prepositional combinations which reached the MI threshold of 3, were two factors which led me to question the usefulness of selecting prep+N combinations partly on the basis of MI scores. A second reason to question the criterion of high MIs for prep+N combinations was that many of those listed in the *OCD* (e.g., *off the agenda*, *with confidence*, *over the limit*, *within reason*) received low MI scores in COCA and thus did not qualify for inclusion on my list. As I examined this issue more, I discovered that the MI threshold of 3 favours lexical collocations (Stubbs, 1995). MI scores are higher for lower-frequency items which tend to be "exclusive" to each other, whereas phrases containing high-frequency items such as prepositions receive low MI scores (Gablasova, Brezina, & McEnery, 2016; Schmitt, 2010).

For the reasons just given, I abandoned the use of MI scores as a criterion for the selection of collocations. However, the majority of target V+N phrases on the final list (16/24) had MI scores of 3 or greater. In contrast, only 3 of the 24 target prep+N phrases had MI scores greater than 3 (see tables I1 and I2 in Appendix I).

4.3.1.2 Trials

As part of target-collocation selection, I conducted two trials with non-native speakers of English to test their knowledge of the collocations. One trial was conducted in the first approach and one in the fourth approach. I report here only the procedure for the trial for the fourth approach.

1. The first step was to create written gapfill sentences. For each V+N combination, one, two or three (and, in one case, four) letters of the missing verb were presented as clues. Each missing letter was represented by a short space, for example, *The temperature should r_ _ _ _ a maximum of 20 degrees today* [*reach*]. For sentences containing prep+N collocations, one longer space was supplied to represent each preposition. Some of the spaces for prepositions included a letter clue at the start or the

end of the space. This is an example of a prep+N gapfill sentence: *Good news: we're* _____ *target to get our sales goals* [*on*]. See Appendix J for the earliest version of the trial test.

2. Next, I trialled both V+N and prep+N combinations with native speakers of English. Over different stages of the trial, I modified the number of letters given to the participants as clues. I attempted to make the combinations easy enough for native speakers to guess while at the same time difficult for lower-level Chinese native speakers in order to increase the amount of measurable learning by non-natives. The aim was to make the trial test ecologically valid and fair. It could be argued that a test in which non-native speakers achieved very low scores was unfair if the results for native speakers were also low.

I wrote three or four versions of the test text in order to refine the sentence contexts. In particular, I experimented with the letter clues for the participants. For the verbs in the V+N phrases, I usually started by supplying the first letter of the verb as a clue. An exception to this was the three-letter verbs *set* and *put*, most of the clues for which I supplied the second and third letter, as giving the first letter might have made the process too easy for the trialists. When it became clear from the results of the first few native-English-speaking participants that one letter was not enough of a clue to lead to correct answers, I added a second letter as a clue, and, in seven cases, a third letter clue. One verb, *stage*, from *stage a protest*, needed four of the five letters to be given before all the native speakers could correctly complete the verb in the gap-fill. I found in an early stage of the trial that none of the three native-speaker participants were able to guess the missing verb in the collocation *face difficulties* as it was presented in the following sentence: *__a__ __ difficulties*. Subsequently, I changed the presentation of the phrase to *f__ __ __ difficulties*, and, after two participants had difficulty filling in the correct verb, *f__ __e difficulties*. (Ironically, two native-speaker participants in the trial commented that the existence of the letter clues in the answers hindered rather than helped their attempts to guess the answers.)

Some of the results of my gapfill test trial may have been confounded by the repetitious use of a number of verbs and many prepositions. For example, I included the verb *draw* in four collocations in the trial sentence: *draw strength*, *draw boundaries*, *draw a comparison* and *draw an audience*. It is possible or even likely that the correct identification by a triallist of *draw* in a gap an earlier sentence made it easier for them to

identify it in another phrase in a later sentence. Repetition of a limited number of prepositions was inevitably common in the trials with the prep+N phrases.

3. The latest version of the sentences was trialled with two advanced native speakers of Chinese and two advanced native speakers of other languages. As a result of this, I deleted several V+N combinations in which the same verb was used in more than two combinations.

4. Then I trialled the updated version of the same sentences with 13 native speakers of Chinese at upper-intermediate level. In each half of the trial test, five easy collocations were included (e.g., *use a computer*, *at work*) for affective reasons: to lessen discouragement that the trialists might have felt at not knowing many of the other answers. In this version of the test, no verb was used more than twice in the V+N combinations. I found that only four pairs of combinations (*close the deal/strike a deal* and *under the deal*; *stage a protest* and *under protest*; *draw strength* and *under strength*; *put a strain (on)* and *under the strain*) had correctness scores for each combination as low as 0 and 1 (maximum score = 13). The other collocations in the trial sentences had scores of 2 or higher. Other collocations had high scores. *Give an explanation* was correctly answered by all 13 participants; another, *on purpose*, was correctly answered by 12. This meant that the learning potential of the collocations that were already basically known would have been minimal. I realised that supplying two or more letters as clues might have given the participants too much information and that they might have been making random guesses about the verbs without in fact knowing the collocations. For example, the test item *pe__f__ __ __ a review*, which was correctly guessed by 10 out of 13 participants, may have provided too many clues about the identity of the target verb *perform*. (The item was later changed to *p__ __ __ __ __ __ a review.*)

5. In the final step of the trial, I reduced the number of letter clues in the gapfill words in an effort to minimise guessing by participants. The number of letters for the target verbs was restricted to one, with two types of exceptions. For three-letter verbs, no letter clues were given, while an extra letter clue was added to one five-letter verb (*stage*) for the purpose of disambiguation (*stage/start a protest*). For all but one of the prepositions, all letter clues were removed. However, the final letter of one preposition (*before*) was inserted to disambiguate it (*The band played _____e an audience of 5,000 people*).

Overall, I found that the collocations were not identified by the majority of intermediate- and upper-intermediate-level learners of English in trial testing and thus seemed to provide learning opportunities for students at that level.

Before I chose the final list of collocations, I gave a simple yes/no test in the form of a type of checklist test (Meara & Buxton, 1987) to two classes of intermediate students of English attended by 13 Chinese speakers and 13 non-Chinese speakers in order to determine basic receptive knowledge of the K3 nouns in the potential target collocations. The classes contained students the majority of whom had 80+% receptive knowledge of the second thousand most frequent words in English as measured by the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001). The students were given a list of 14 K3 words which were in the potential target collocations and asked to place a tick next to a word if they knew its meaning and a cross if they did not. Most words were reported as known by the great majority of students but three—*strain*, *obligation* and *dispute*—were known by only a minority, so the combinations containing those nouns were eliminated from the draft list. The four nouns that were placed on the final list of collocations were *target*, *belief*, *decline* and *audience. Target* was self-reported as known by 24 out of the 26 learners in the classes, *belief* and *decline* by 25 learners, and *audience* by 26 learners.

After making adjustments to the test, I gave it to another five advanced Chinese speakers of English for trialling. In the process of trialling I came to the conclusion that it was not possible to create gapfill sentences the results of which clearly discriminated between the results of native speakers (which I had hoped would be mostly correct) and the results of intermediate-to-upper-intermediate non-native speakers (mostly incorrect). When I implemented the minimalist gapfill approach with very few letters provided, many NS triallists were unable to supply the missing words in the gaps, instead providing legitimate alternatives. While this result did not meet my initial aim of creating a test which clearly separated native speakers and non-native speakers' results, I decided that the minimalist approach was adequate since most important purpose of the test was as a prototype for the gapfill post-test in the experiment, which would be a measure of form-recall knowledge after exposure to the collocations in the treatment.

4.3.1.3 Final selection criteria for collocations

The following were my criteria for the selection of target collocations.

1. Each collocation was paired with a *parallel collocation*. In this study, I define parallel collocations of V+N collocations as prep+N combinations containing the same noun; the converse is true for parallel collocations of prep+N phrases. The purpose of the parallel collocations was to allow comparisons to be made between the learning of V+N and prep+N combinations (e.g., *in the knowledge (that)* and *to the knowledge (of)* for *acquire knowledge*).

2. At least one collocation in each V+N and prep+N pair was found in the *Oxford collocations dictionary for students of English* (2nd Edition) (McIntosh et al., 2009) apart from one, *on review,* a parallel collocation for *perform a review* which was found instead in COCA (Davies, 2010).

3. No verbs were in more than one V+N combination on the collocation list. No prepositions were in more than three prep+N combinations[13].

4. The collocations contained only single-word verbs and single-word prepositions.

5. The collocations comprised two or three words only[14]. If a collocation contained three words, the extra word was an article (*a*, *an* or *the*).

6. The constituent words of the collocations were high-frequency words, the large majority appearing in the first and second most frequent thousand words of the BNC-COCA-25 list (Cobb, n.d.; Nation, 2012). However, five of the 24 target nouns at K3 level, which had been trialled with intermediate-level students and found to be known by them, were included in target collocations.

7. The collocations (both prep+N and V+N) occurred in COCA (Davies, 2010) no more than 2000 times.

8. The prepositional combinations did not have any synonyms (or near-synonyms) in which two or more prepositions co-occurred with the same noun (e.g., *at a distance* and *from a distance*).

---

[13] Following this criterion reduced the possibility of creating a repetition effect in the SPRT (Keating & Jegerski, 2015). However, it was impossible to avoid repeated use of prepositions in the 24 prep+N phrases because of the limited number of prepositions.

[14] 23 of the 48 V+N and prep+N combinations chosen contain one article between the verb and noun or the preposition and noun (e.g., *stage a protest, against the odds*).

9. The prep+N combinations were not commonly followed by another preposition (most frequently, *of*—e.g., *at the mention of*).

10. Both singular and plural forms of the countable nouns were listed in COCA.

11. The choice of the grammatical form of a collocation containing the singular form of a noun (e.g., *shed (a) tear*) versus the form containing the plural form of the noun (e.g., *shed tears*) was made largely according to the ability of the form of the collocation to fit as naturally as possible into the sentences and paragraphs in the treatment texts.

12. The collocations were not highly transparent, as assessed by me and another rater who was an applied linguist and a near-native English speaker. The verbs in the V+N collocations and the prepositions in the prep+N collocations are figurative, making the collocations semi-transparent (e.g., *hold the belief*, *under protest*). In a few cases, both words in the collocation carry a figurative meaning (e.g., *suffer a decline*; *at a loss*), making them fairly opaque[15].

4.3.2 Control phrases

In order to create control phrases for use in the SPRT (self-paced reading task), I looked for verbs and prepositions to combine with the nouns of the target collocations. I checked in the Oxford Dictionary thesaurus online (https://en.oxforddictionaries.com) and the Macmillan Dictionary online (https://www.macmillandictionary.com). I selected *non*-synonyms of the verbs and prepositions in the target phrases for inclusion in the control phrases rather than synonyms (Cruse, 1986); that is, the words were not synonyms in the sense of being verbs or prepositions which appeared in the online Oxford Dictionaries thesaurus or the online Macmillan Dictionary in the same entry and the same sub-entry of another verb or preposition. Nine of the study's 48 control phrases were created in this way. Eight of the verb non-synonyms and one of the preposition non-synonyms in the control phrases appeared in the same thesaurus or dictionary entry but in a different sub-entry from the verb or preposition from the

[15] Most of the V+N combinations are (restricted) collocations in the phraseological sense (Howarth, 1998a, Nesselhauf, 2003); furthermore, the noun is unrestricted but the verb is restricted in the sense that it can collocate with only a small set of nouns (e.g., *draw a comparison*, *raise the issue*). Six target V+N combinations were judged to be figurative idioms, containing two words in which both words have figurative meanings (e.g., *beat the odds, meet targets*). Most of the prep+N phrases are also collocations—that is, the prepositions are used in a figurative, non-spatial sense while the noun senses are literal (e.g., *in credit*, *beyond doubt*) although six phrases are fully figurative (e.g., *off target*, *against the odds*).

target collocation. For example, for the collocation *mark a turn*, the meaning sense of the verb *mark* matches most closely the following sentence and synonyms, taken from a sub-entry in the online Oxford Dictionaries thesaurus:

> 5 *'two great sea battles marked a new epoch in naval history'*
> SYNONYMS: represent, signify, be an indication of, be a sign of, indicate, herald

However, I took the verb *show* for the control phrase *show a turn* from another meaning sense of *mark*, found in another sub-entry[16]:

> 3 *'I've marked the relevant passages'*
> SYNONYMS: indicate, label, flag, tab, tick, show the position of, show, identify, designate, delineate, denote

I decided against including synonyms of verbs and prepositions in target collocations for control phrases. While there might have been enough suitable synonyms for the verbs in the V+N combinations, the choice of synonyms for any given preposition is limited, and I wanted consistency in the types of verbs and prepositions in the control phrases. For example, the figurative use of *under* has no synonyms suitable for the purpose of this study in the online Oxford Dictionaries thesaurus: it has only two-word synonyms, none of which are prepositions. No synonyms are given for *for*.

For the remaining 39 collocations, the online sources did not supply a control verb or preposition for each of the nouns which met my criteria (see below), so I thought of a control word myself. For example, the verb *gain* from *gain targets*, the control phrase for *meet targets*, did not come from a dictionary or thesaurus but from me.

Control words were selected according to four main criteria:

1. The control phrases sound as natural and plausible as possible as judged by myself and my second rater. Naturalness and plausibility were important since it meant that the control phrases could be seen by participants as valid options in the self-paced reading task.

2. The number of letters in each control word is no more than one letter different from the length of the word in the collocation. This is because word length of the

[16] Although *mark* and *show* have some semantic overlap, they are non-synonyms (Cruse, 1986).

preceding word affects the reading time on the following word in the SPRT measure, and thus the first word of the collocation and the control phrase had to be matched. For example, the target collocation *raise doubts* contains a verb comprising five letters; in the matching control phrase, *lift doubts*, the verb comprises four letters. The parallel prep+N collocation is *beyond doubt*, *beyond* containing six letters. The control phrase for that combination is *outside doubt*; *outside* comprises seven letters. However, there are two exceptions to the plus-or-minus-one-letter rule: the control phrase *with turns* for the collocation *by turns* and *over issue* for *at issue*. The differing lengths of these two pairs of prepositions was unavoidable because of the limited number of prepositions which would fit naturally into the control phrases.

3. The control words are in either the K1 or K2 frequency band of the BNC and COCA, as are the equivalent words in the collocations.

4. The maximum raw frequency of each control phrase as a semantically-connected *exact phrase* in COCA was 1. None of the V+N phrases appeared as semantically-connected phrases. Nineteen of the 24 prep+N phrases did not appear in COCA as semantically-connected phrases—the other five appeared once.

As I looked at the *exact phrase* raw frequency figures of the prep+N combinations in COCA, I noticed that a number of them had surprisingly high counts. For example, *of a loss* occurred 141 times in the corpus and *at credit* 72 times. On further inspection, I found that while the component words co-occurred, they were not collocations in a semantic sense. The preposition in each of these non-phrases was linguistically bound not to the following target noun but to another lexical word (usually prior) in the example sentence. For example, the words *at* and *warning* in *at warning* are not semantically linked even though the two words co-occur in the corpus in the following sentence segment from COCA: "…aiming its anti-drug efforts at warning women not to take drugs." In this sentence *at* is bound to *effort*, not *warning*.

Although it would have been ideal to have had no occurrences (incidental or otherwise) of any of the control phrases, this proved to be impossible in a corpus of 520 million words. However, the fact that the five phrases used in the study occurred only once in the corpus as semantically-linked phrases shows that they are highly improbable constructions (Gyllstad & Wolter, 2016). The final list of collocations and their matched control phrases is presented in Appendix L along with relevant statistical information.

4.3.3 Treatment texts

I created twelve 170-word treatment texts on various, mostly business- and employment-related, topics (e.g., a company's annual report, university staff cuts, unemployment) based around the target collocations (see an example text in Appendix K). Each text contained six target V+N combinations and six target prep+N combinations. Three of each combination type were bolded and three were unbolded. Both or all three words of each target collocation in my treatment texts appeared together on the same line (and were not split up in the sense that one or two words were at the end of one line and one or two words were at the start of the next line). Visually, such a split would have broken up a collocation and possibly made it less memorable.

The texts were screened for the frequency bands of individual words. The words in the texts which were assumed to be known by the students were K1 and K2 words in the BNC/COCA list (Nation, 2012), the K3 words which had been found to be known by intermediate student in a yes/no test trial, and well-known place names (e.g., Wellington). Thus, each treatment-text paragraph had, for the participants, a coverage score of 98%+, the generally-accepted threshold of vocabulary knowledge at which a reader has adequate text comprehension (Hu & Nation, 2000; Schmitt, Jiang, & Grabe, 2011).

The treatment texts contained a combination of varied contexts (Bolger, Balass, Landen, & Perfetti, 2008) and verbatim repetition (Durrant & Schmitt, 2010) of the target collocations. In the texts, participants would be exposed to a total of six occurrences of each collocation: three occurrences in varied contexts on the first day of treatment, all repeated once on the second day of treatment (3 x 2 = 6 occurrences). The varied contexts of the texts within each day's session would provide variety of content, aimed at maintaining the participants' interest in the texts (Bolger et al., 2008). The repetition of the texts on the second day of treatment was intended to strengthen the memory traces of the collocations by reducing the cognitive burden on participants and promoting fluency of form of the collocations (Durrant & Schmitt, 2010). To ensure that the participants were continuing to pay attention to the meaning of the texts, the comprehension questions following the repeated versions of the texts were different from the questions for the initial versions of the texts.

I allocated the collocations to alternating learning conditions (bolding vs. no-bolding) according to the order of the collocations' occurrences in the treatment texts. For example, the first V+N combination in the first text (*stage a protest*) was allocated to the bolding condition for treatment condition 1 and the second V+N collocation in the first text (*reach an audience*) to the no-bolding condition. Each prep+N collocation was placed in the same learning condition as its V+N equivalent and vice-versa to allow a comparison of equivalent V+N and prep+N combinations in the same treatment condition. For example, both the V+N collocation *stage a protest* and its prep+N equivalent *under protest* were in the bolding condition for treatment condition 1 and the non-bolding condition for treatment condition 2.

4.4 Measures

Two tests measuring learning for the treatment phase were used in the experiment: a SPRT and a pen-and-paper gapfill test. While the gapfill test was the same type of measure as that used in Experiment 1, the online measure in that experiment, the primed LDT, was replaced with a SPRT. Like the primed LDT, the SPRT forced the participants to respond under time pressure. However, it had the advantage of being a more naturalistic measure of the reading of collocations than the LDT was.

4.4.1 Self-paced reading task

The first test was the self-paced reading task (SPRT), using the software programme E-Prime. The non-cumulative moving window format was used (Juffs & Harrington, 1995; Just et al., 1982). Participants read short sentences onscreen: each sentence was presented one word at a time with the participant advancing the text at the push of a keyboard key (e.g., Fraundorf, Watson, & Benjamin, 2010; Kember, Choi, & Cutler, 2016; Lee & Fraundorf, 2019; Millar, 2011; see Jegerski's 2013 review). I chose this method over the less-frequently-used cumulative moving window method, in which all the words of the sentence remain onscreen after they appear and participants can re-read them; this is a more natural technique of reading. However, even though the non-cumulative window format promotes less natural reading, it has two major advantages which make it a better measure of online word processing: it provides a reading time for each word, and it restricts the use of reading strategies (Rayner, 1998).

The development of procedural knowledge of the collocations was operationalised as faster RTs for the nouns in the collocations than for the nouns in the control phrases. The comparison of plausible phrase targets with control-phrase targets is standard practice in L2 SPR studies (e.g., Roberts & Felser, 2011). Operationalising this comparison as measuring the development of procedural knowledge is based on the widely-accepted assumption in L2 experiments that SPR measures implicit knowledge (Marsden et al., 2018; Marinis, 2010). SPR fulfils most or all of R. Ellis's (2005) criteria for implicit knowledge, such as response according to feel, time pressure and a primary focus on meaning. Marsden et al. (2018), in their survey of 52 studies which employed SPR, state:

> We found no challenges to the notion that SPRs in L2 research are a measure of implicit knowledge, and no discussion of a potential role for awareness or attention. When explicit knowledge was mentioned, it was in relation to SPR reducing access to it or in relation to the other measures being used in the same study to elicit a different type of knowledge to the SPR. This perhaps reflects a consensus that reactions in SPRTs are deemed to operate below the level of consciousness, though empirical validation of this would be useful." (p. 10)

4.4.1.1 SPRT stimulus sentences

Nine-word stimulus sentences containing the target collocations were created for the SPRT. Each stimulus sentence for each of the V+N and prep+N combinations contained three words before the collocations. In the follow stimulus sentences, bolding has been added to the collocations and the target nouns have been underlined, for the purposes of clarifying the design (in the actual SPRT used in Experiment 2 no typographic enhancements were used): *He will really* ***push the limits*** *if he cheats; She has borrowed* ***over the limit*** *she can afford.* I created control sentences by replacing the collocations in the target sentences with the synonymous control phrases which were not collocations (e.g., *He will really* ***touch the limits*** *if he cheats; She has borrowed* ***after the limit*** *she can afford.*). This lexical matching produced strong internal consistency for each item (Keating & Jegerski, 2015).

4.4.1.2 SPRT post-stimulus comprehension questions

The comprehension questions following the stimulus sentences in the SPRT were simple questions requiring yes/no responses (e.g., *Will there be more work for jobless people?).* Their purpose was to focus the participants' attention on the meanings of the sentences and to keep the purpose of the task hidden from the participants (Keating & Jegerski, 2015; Marinis, 2010). I ensured that these post-stimulus questions did not contain either of the words in each target collocation or control pair—in order to avoid repetition of the form of the item. I also made sure that the sentences did not otherwise draw attention to the semantic content of the item (Keating & Jegerski, 2015). For example, for the stimulus sentence, "Today he is off work because he is sick," in which *off work* was the target collocation, the post-stimulus question was "Is he feeling sick?" (correct answer: yes). To access the answer to this question the participant would not need to recall the collocation *off work*; thus, attention would not be drawn to it. For many of the sentences it was very difficult, if not impossible, to create questions which tested the meanings of the whole sentences but which did not draw attention to the target collocations. That was because the core idea of each sentence was often found at least partly in the collocation, especially if it was a V+N combination. For example, for the stimulus sentence, "Rising prices will present a threat to the economy," the collocation *present a threat* is at the heart of the main idea of the sentence. A draft post-stimulus question that I had written for that sentence, "Will rising prices be good for the economy?" tested understanding of the main idea of the target sentence but also drew attention to the target collocation. Thus, I replaced it with another: "Is the sentence about prices that are increasing?"

Such post-stimulus sentences focusing on the target items were commonly used in one of the few studies that has used a SPRT to test knowledge of multiword units, Tremblay et al. (2011). Eight of the researchers' 20 questions containing target lexical bundles and non-lexical bundles were general topic questions (e.g., For the stimulus sentence, *Indeed, whatever you do about it feels weird*, the post-stimulus question was *Is the sentence about feelings?*) or questions asking participants to identify a word from the sentence (e.g. *Is there a man named Ron in the sentence?*). Twelve of their 20 questions tested understanding of the sentence, and at least nine of these 12 questions required the participants to focus on the target lexical bundle or non-lexical bundle—

e.g.: Stimulus sentence: *His friend's got nothing to do next Friday*; post-stimulus question: *Does his friend have something to do?*

Originally more than half of my draft post-stimulus questions started with "Is the sentence about…." However, since too many such sentences could have conditioned participants to look for the general topic in a sentence more often than for specific information, I reduced the number of such questions. For example, for the stimulus sentence, "The government will create work for many jobless people," I replaced the question "Is the sentence about doctors?" with the more specific question, "Are doctors going to do something for jobless people?" (correct answer: no) In the end, fewer than half (20/48) of my post-stimulus questions for the collocations in List A started with "Is the sentence about…?" Once the filler sentences were included with the collocations, the proportion of "Is the sentence about…?" questions was greatly diluted: only one quarter (20/80) started with that phrase.

4.4.1.3 SPRT presentation lists

For the SPRT I developed two presentation lists of trials (sentences), one for each experimental condition (see example in Appendix L). Each list of trials comprised 24 target sentences—each of which contained a target combination (12 V+N and 12 prep+N)—24 control sentences (containing a control phrase inserted in the other 24 target sentences) and 32 filler sentences. Below is an example of a target sentence and its matching control sentence:

> Target sentence: They have to reach an audience to sell insurance.
> Control sentence: They have to stir an audience to sell insurance.

Each participant saw only either the target sentence or the control sentence of each of the matching pairs of sentences, since seeing matching pairs could have produced repetition effects, with participants responding to items unnaturally (Keating & Jegerski, 2015). Both lists contained the same 32 filler sentences (see next section). The total number of trials on each SPRT testing list was 80.

4.4.1.4 SPRT filler sentences

Thirty-two filler sentences (e.g., *The government will soon introduce a new financial law*; *He is returning to his home town on Monday*) were included on both SPRT lists to hide the target items in the target sentences and the control phrases in the control sentences from the participants and to distract the participants from understanding the purpose of the SPRT, which was to measure RTs for the target collocations and the control phrases (Brisard, Frisson, & Sandra, 2017; Brothers, Swaab, & Traxler, 2017; Keating & Jegerski, 2015; Millar, 2011). I placed two filler sentences at the start of each block of trial sentences. I also inserted fillers between target sentences and control sentences where there were more than two target sentences or control sentences in consecutive order.

The ratio of target items to filler items in the present experiment (48 to 32) was 3:2. This was a proportionally higher number of filler items than that used by Millar (2011), who employed 32 target items and 16 filler sentences (2:1). According to Keating and Jegerski (2015), the greater the proportion of filler items, the better, to reduce the predictability of the types of sentences. They argue that at least 50% filler sentences is "highly desirable" (p. 17). In my experiment, however, practical considerations, particularly the possibility of the fatigue effect or boredom at the participants' reading a larger number of sentences in total, also needed to be considered, so I restricted the number of filler sentences in order to keep the total number of trials to 80 (Keating & Jegerski, 2015; Nation & Webb, 2011).

4.4.1.5 Ordering of sentences/sequencing of SPRT trials

For ordering the trials in my SPRT, I considered a number of options. Randomization across participants (e.g., Sonbul & Schmitt, 2013; Wolter & Gyllstad, 2011) was one option. However, as reported in Chapter 3, this method fails to prevent clustering (i.e., having three or four or more of the same type of phrase occur consecutively on a list), which can produce an unwanted repetition effect. I chose a type of pseudo-randomization, which has been used in SPRTs investigating MWUs (Millar, 2011) and LDTs investigating individual words (Elgort & Warren, 2014). This was the approach I followed in Experiment 1 and involves initial randomizing of the order of the trials and then manual intervention to avoid clustering of items. Using this approach meant that

each of the target collocations and their matching control phrases would appear in the same order on the two presentation lists (one for each testing group) and could be directly compared. Lack of task familiarity and the fatigue effect were minimised by my checking that there was a similar number of target sentences vs. control sentences at the start and the end of either of the lists.

I pseudorandomised the two testing lists. First, I arranged the initial Excel file (list A) into five blocks of sentences (12 target V+N, 12 target prep+N, 12 control V+N, 12 control prep+N, 24 fillers) (see Table 42). I then applied the randomizing function in Excel to the list. Next, I made manual changes to the list, breaking up clusters of three or more trials containing the three types of phrase (target, control and filler) and checking that not too many target sentences were at the beginning or end of either list (compared with control sentences). I then created list B by matching the items to those on list A. For example, the first control item on list B (e.g. *at arrangement*) was in the same order in the sentences (e.g., fourth on the list) as its target item equivalent on list A (*by arrangement*). In the SPRT, I gave testing group A list A and testing group B list B.

Table 14: Design of self-paced reading task

SPRT (48 target items)	
Test Group A's sentences	**Test Group B's sentences**
12 target V+N collocations (set 1)	12 target V+N collocations (set 2)
12 control V+N phrases (set 2)	12 control V+N phrases (set 1)
12 target prep+N collocations (set 2)	12 target prep+N collocations (set 1)
12 control prep+N phrases (set 2)	12 control prep+N phrases (set 1)
24 filler phrases	24 filler phrases
TOTAL: 72 test sentences, 24 target items + 24 control items	
12 practice sentences	12 practice sentences
TOTAL: 84 test sentences + practice sentences	

4.4.2 Gapfill test

The pen-and-paper gapfill test, comprising short definition prompts, was designed to measure cued recall of declarative collocational knowledge of form (see Appendix M). I created 48 single sentences (one for each target collocation) adapted from sentences in the treatment texts. As was the case in Experiment 1, the gapfill sentences in the present experiment were similar to those that the participants had already read in the treatment texts. The first word of each collocation (whether a verb or a preposition) was removed, and no letter clues were provided. Participants were instructed to supply the first word by writing in the gap. This test measured participants' ability to retrieve

the form of part of the target collocations from a context in which meaning and some form (the remaining word or words of each combination) were provided. Using the randomizing function in Microsoft Excel, I randomised the order of the items in the gapfill test within two blocks (V+N and prep+N) across participants (i.e., the order of test sentences was the same for all participants).

4.4.3 Order of measures

The SPRT was run before the gapfill test. The SPRT briefly exposed participants once to half of the target collocations (as well as to the control phrases) and likely had a testing effect, strengthening participants' memory traces of the collocations before the gapfill test (Rowland, 2014; Sonbul & Schmitt, 2013). However, placing the gapfill test before the SPRT would have produced a larger testing effect: it would have drawn attention over a longer period to all of the target collocations in the SPRT (which was to follow), possibly overpowering the effects of the incidental learning in the treatments and obscuring differences between the TE and non-TE conditions.

4.4.4 Rationale for no pre-test and no delayed post-tests

As with Experiment 1, in Experiment 2 I included neither a pre-test nor a delayed post-test session. The reasons for having no pre-test include the possibility of unintentional learning of target items and alerting participants to the specific purpose of the study. A control group was used as a baseline for participant knowledge of the target collocations. A two-tailed *t*-test found no significant differences between the scores for the 2k, 3k and 5k levels of the VLT for the treatment-group participants and the control-group participants ($p = .97$).

Unlike in Experiment 1, in Experiment 2, no delayed post-tests were administered. One reason for this was that the purpose of the tests in Experiment 2 was to measure learning only, not long-term retention. In such a situation, Hulstijn (2001) sees no need for delayed post-tests:

> If the aim of the investigation is to assess incidental learning taking place during initial exposure, an immediate post-test is all one needs. … Delayed post-tests, therefore, are appropriate only if the research question focuses on what happens

with information after initial exposure under various conditions of rehearsal or re-exposure (p. 274).

A second reason for running no delayed post-test session was practical: the two experimental sessions (including the immediate post-test session) lasted 2½ hours in total, and a number of potential participants might have been reluctant to have attended a study comprising three sessions, as was the case with Experiment 1. In fact, delayed post-tests appear to be uncommon in experiments on MWUs (and in experiments that measure RTs). An informal survey of more than a dozen studies on the learning of MWUs since 2006 shows that the great majority did not include a delayed post-test (e.g. Siyanova & Schmitt, 2008; Wolter & Gyllstad, 2011; Yamashita & Jiang, 2010). Many of the studies included RT tests, and only three contained delayed post-tests (Boers et al., 2014; Sonbul & Schmitt, 2013; Webb, Newton, & Chang, 2013).

4.4.5 Vocabulary Levels Test

For Experiment 2, I replaced the VST with levels 2, 3 and 5 from the Vocabulary Levels Test (VLT) (Schmitt, Schmitt, & Clapham, 2001). I wanted to use the VLT as a diagnostic tool. Because the English proficiency level of the learners in this study was lower than the level of those in the previous study (the participants in Experiment 2 were at the intermediate to upper-intermediate proficiency levels), it was important that I check that they had receptive recognition knowledge of the most common 2000 word families in English, which was the frequency level of the words in the treatment texts and in the tests. At the level of the second-thousand word family (K2), the mean score of the participants (N = 78) was 27.62 out of 30 (SD 2.65), with the minimum and maximum scores 21 and 30 respectively. From these scores it can be interpreted that the participants had fairly strong receptive knowledge of K2 words, meaning that the vocabulary level of the texts was at a suitable level for the participants. A two-tailed *t*-test found that there was no statistical difference between the control group's VLT scores for K2, K3 and K5 and the treatment group's scores ($p = .97$). The mean VLT score of the 25 control-group participants was 68.00 and the score of the 53 treatment-group participants was 68.15.

4.4.6 Yes/no test

An added measure of target-collocation knowledge employed in the present experiment was self-reporting by both the treatment-group and control-group participants. At the conclusion of their experimental sessions the participants were asked in a modified version of the yes/no test (Meara & Buxton, 1987) whether they had encountered the collocations before their involvement in the study (see Appendix N). Participants were asked to place a tick against each of the 48 collocations that they had seen or heard before. In this way, participants indicated whether they knew (or thought they knew) the forms of the collocations. This test assessed only the most rudimentary level of collocational knowledge: self-reported familiarity with the combinations' form. While the results of this test are instructive, they must be viewed cautiously because its self-report nature potentially compromises its reliability. As with Experiment 1, no pseudo-words were included in the test as the participants had already been exposed to each collocation six times in the treatment phase and were assumed to have some developed some knowledge of the collocations by that stage.

4.5 Procedure

Each participant was assigned to one of two treatment conditions after their initial expression of interest to participate in the study (learning condition 1: n = 27, learning condition 2: n = 26) (as well as one of two testing groups) or the control group (n = 25) in the order in which they volunteered for the study. If a participant withdrew from the study (before the beginning of the data collection), he or she was replaced by the next person who volunteered.

The experiment used a counterbalanced learning condition and a between- and within-participants design. Two variables, each with two levels, were used: bolding and no-bolding; and V+N and prep+N. Participants twice read 12 texts containing 48 non-technical collocations of low to moderate frequency. Each participant read: 12 bolded V+N combinations; 12 unbolded V+N combinations; 12 bolded prep+N combinations; 12 unbolded prep+N combinations. Each text contained six occurrences of V+N combinations and six occurrences of prep+N combinations (see Appendix O).

I ran two experimental sessions on consecutive days. As in Experiment 1, this approach was intended to produce a repetition regime (Elgort, 2011; Karpicke &

Roediger, 2007) and to facilitate overnight consolidation of the memory of the collocations (Wang, Savage, Gaskell, Paulin, Robidoux, & Castles, 2017; Lindsay & Gaskell, 2010). In the first session, the treatment participants read 12 different texts containing three occurrences of each collocation; in the second session on the second day, participants read the same texts again in a randomly pre–determined order. Thus, participants had a total of six contextual encounters with the same collocations in all texts over the whole learning period.

Before the start of the first treatment session, the participants were instructed to focus on the meanings of the texts in order to answer one comprehension question after each text. They were given a written sheet containing the instructions, which I read aloud, with participants following the text silently, before the first reading session. No mention was made of the bolded collocations in the texts.

The participants' knowledge of the target collocations was measured at the end of the learning period on the second day of sessions in the two post-tests, the self-paced reading task and the gap-fill task. These tasks measured participants' ability to retrieve collocational knowledge in an automatic manner and a controlled manner respectively. The test session was conducted in a language-learning laboratory with no more than eight participants at a time.

In the first measure, the SPRT, participants were seated at alternate desktop computers (or further apart) in order to reduce possible distraction. Button boxes were not used in this experiment to measure RTs; I wanted to run experimental sessions with groups of four or more participants at a time for the sake of time efficiency, but only two button boxes were available in the department. Therefore, keyboards, which have been previously employed in published research to measure RTs in SPRTs (e.g., Kember, Choi, & Cutler, 2016; Lee & Fraundorf, 2019; Schoot, Reijntes, & van Lieshart, 2012), were used instead of button boxes as input devices. After doing eight practice trials, participants were exposed to the 80 trials of the SPRT proper—divided into one block of 40 trials containing the V+N collocations and a second block containing the prep+N collocations[17]. The second test, the gapfill task, was a pen-and paper task.

Two intervening tasks were used in the experiment. Between the learning and the testing sessions I gave the participants five minutes to count backwards, subtracting

[17] Several participants, particularly those in the early sessions, continued on to the real test without stopping after the practice trials.

by 7s from 300 and writing their answers on a piece of paper (Kroll & Kellicutt, 1972). This task was a cognitively demanding numerical exercise which made the participants think about something different from the treatment. Between the two tests, participants completed a background information questionnaire, which also acted as an intervening task between the two sessions.

I administered two tasks after the participants had completed the post-tests: the Vocabulary Levels Test and the yes/no self-report recognition test.

As in Experiment 1, the control-group participants completed the same tests as the experimental group in order to control for the possibility of prior knowledge of the target collocations. The control-group participants attended one session in which they took the two post-tests, the VLT and the yes/no recognition test. They also completed the participant background questionnaire.

4.6 Data analysis

I conducted the analyses of the two post-tests (the gapfill test and the SPRT) separately using mixed-effects modelling in R, version 3.2.3 (The R Foundation for Statistical Computing, 2015). Participants and items were included in the model as crossed random effects.

The treatment type predictor comprised two levels: no-bolding and bolding. The collocation type predictor had two levels: V+N and prep+N. The initial alpha level was set to .05. Self-reported prior knowledge of the target collocations was included in the analysis as a fixed effect and a potential co-variate, while the participants and collocations were entered in the models as crossed random effects. Vocabulary Levels Test (VLT) scores were not included in the analysis as the VST had been for the analysis in Experiment 1. The VLT had been administered as a diagnostic test in order to confirm that the participants' proficiency levels were adequate to cope with the level of the treatment texts. However, since the VLT (in particular, the 2k, 3k and 5k levels that I used) sampled only three levels of frequency (compared with the 14 levels of the VST version used in Experiment 1), it was inadequate to be used as a proficiency covariate in this study.

For both the gapfill test and the SPRT, the accuracy of responses to the collocations encountered in the bolding treatment was compared with the accuracy of the responses to the collocations encountered in the no-bolding condition. The accuracy

of responses to the prep+N combinations was compared to that of the V+N combinations.

I employed a backwards stepwise approach to the data analysis in Experiment 2, starting with a full model, containing all theoretically and empirically motivated predictors. In the stepwise model selection procedure, I tested model fit after removing each predictor, one at a time, by comparing the Akaike Information Criterion (AIC) for the resulting and previous model, using the "anova" command in R. The best model fit in each case was the model with the lowest AIC value.

The analyses of the gapfill test and the SPRT each involved two steps. In the first step, the control group was compared with the treatment group to check that the control group had relatively little knowledge (declarative or procedural) of the collocations (categorised according to collocation type: V+N and prep+N). In the second step, the control group was excluded, and the two treatment conditions were compared with each other. In this step, the no-bolding treatment was the intercept.

4.6.1 Gapfill test analysis

In the gapfill test analysis, the response accuracy rates were analysed using mixed-effects logit modelling. The independent variables (IVs) were *collocation type* (two levels: V+N and prep+N) and *typographic enhancement* (two levels: bolding and non-bolding).

For both steps in both analyses, a model with accuracy as the dependent variable (DV) was fitted and included the IVs of collocation type and typographic enhancement. An interaction between collocation type and typographic enhancement was first tested to check if the effect of typographic enhancement was the same for both the V+N and prep+N collocation types. A significant interaction would signal that typographic enhancement had a different effect on the accuracy levels for the two collocation types. If the interaction were not significant, this would mean that the effect of typographic enhancement was the same for both collocation types.

4.6.2 Self-paced reading task

The IVs in the SPRT were collocation type and typographic enhancement and the DV was RTs on the noun (terminal) component of the collocations. As with the analysis of the primed LDT, I used inverse-transformed response times (-1000/RT) because of the

positive skew of the untransformed RTs' distribution. I compared RTs on the terminal noun in the collocations learned in the two treatments with those on the terminal noun in the control phrases. Minimum a priori outlier removal was performed (i.e., only extreme outliers were removed).

4.6.3 Exclusion of five collocations from analysis

In Experiment 1, one medical collocation, *gene therapy*, was excluded from analysis because one-third (21 out of 62) of the participants reported that it had already been familiar to them, thus providing little opportunity for learning. In contrast, the collocations in Experiment 2 were more frequent non-technical phrases, and as such were more likely to be familiar to non-native speakers. This assumption was borne out: many of the collocations were reported by more than half of the participants as being already familiar to them (see Appendix P). The mean number of collocations reported by all 78 participants as previously known was 51.07% (SD 0.16); a mean of 50.18% of the V+N collocations (SD 0.18) and 51.95% of the prep+N collocations (SD 0.13) were reported as previously known[18]. In this study, therefore, a higher threshold of prior knowledge for the exclusion of any target items from analysis seemed to be required.

Prior knowledge of the target collocations was assessed in this study from the results of two tests: the yes/no test (see section 5.4.6.) and the control-group participants' gapfill post-test. Since participants had been assigned to the two treatment groups and the control group alternately in the order in which they had volunteered for the study, I assumed that the control group was a representative sample of the same population that the treatment group participants were drawn from. That being the case, the control group's results for the gapfill test may be interpreted as being equivalent to all participants' prior knowledge of the collocations (Sonbul & Schmitt, 2013). (The reliability of the gapfill test is complicated by the fact that many of the test items have alternative answers. For example, an acceptable alternative to the collocation *draw a comparison* is *make a comparison*.) The control group had not been exposed to the

[18] The figures given here are for 44 of the total 48 collocations presented to participants. As is explained later in this chapter, four collocations were excluded from analysis because of experimenter errors in the treatment texts. The 44 collocations cited here include *make arrangements*, which I later excluded because of its high level of prior knowledge by participants. When *make arrangements* is also excluded from the calculation, the mean percentage of the remaining 43 collocations reported as previously known was 50.16% (SD 0.15); 48.29% (SD 0.16) of the 21 V+N collocations were known and 51.95% (SD 0.13) of the 22 prep+N collocations were known.

collocations in the treatment texts, although each participant had seen half of the collocations once each in the sentences in the SPRT (which was administered just before the gapfill test). However, as the task was performed under time pressure, there was little time to take in the collocations. Any new memory traces of the collocations that were established were likely to have been weak, possibly not strong enough to lead to correct answers in the gapfill test of cued recall. Consulting both measures of prior knowledge provides greater reliability than reference to scores from one test alone.

In response to the results of the two tests of prior knowledge, I excluded one collocation—*make arrangements*—from the analysis on the grounds that it was already known to too many participants: 56% of the control-group participants in the gapfill test and 90% of all participants according to their self-report. I also removed four further collocations from the analysis after I discovered experimenter errors in the treatment texts. In three instances I had inadvertently bolded one occurrence of a collocation that was in the unbolded condition; in a fourth instance, one occurrence of a collocation was unbolded where it should have been bolded. As a result, I excluded these four collocations: *shed tears*, *push the limits*, *by profession* and *under protest*. In all, five collocations were excluded from analysis (and the descriptives), meaning that a total of 43 of the original 48 collocations (21 V+N combinations and 22 prep+N combinations) were subjected to analysis.

4.7 Results

In this results section, I present the descriptive statistics and the analysis of the gap-fill test (measuring declarative knowledge) and the self-paced reading task (SPRT) (measuring procedural knowledge) of Experiment 2. As noted in a previous section, five collocations were excluded from the analysis, leaving 43 for which descriptives and analysis are presented.

4.7.1 Gapfill test descriptives

Table 15 shows the percentage scores for participants' accuracy rates for the target collocations in the gapfill test. I calculated percentage scores for scores in six categories: control group, treatment group, no-bolding, bolding, V+N collocations and prep+N collocations. The accuracy rate for the treatment group for both types of collocations was 20.67%. Within the treatment group, the bolding condition recorded a higher

accuracy rate than the no-bolding condition. The accuracy rates for prep+N collocations were higher than that for V+N collocations.

Table 15: Mean percentage scores for responses in gapfill post-test under both conditions

Condition	Mean (%)	SD (%)
Control group (*n* = 25)	5.02	5.44
Treatment group (both treatments combined) (*n* = 53)	20.75	14.57
No-bolding	16.73	13.57
Bolding	25.05	18.37
V+N collocations (*n* = 21)	16.71	14.73
Prep+N collocations (*n* = 22)	24.61	17.09
V+N no-bolding	15.80	14.71
V+N bolding	17.91	18.07
Prep+N no-bolding	17.96	16.09
Prep+N bolding	30.91	19.77

4.7.2 Gapfill test analysis

In the preliminary analysis of the gapfill test, the scores of the treatment group and the control group were compared to confirm that there was a difference between the groups' scores. Table 16 shows the best-fit model for the fixed effects of the gap-fill test (control group vs. treatment group). The odds of obtaining a correct answer in the test for those in the treatment group were 6 times the odds of obtaining a correct answer in the control group. In other words, as expected, treatment-group participants were significantly more likely than control-group participants to supply the correct answer.

On average, the 78 participants in this study reported in the yes/no test that they had previously seen or heard 21.73 of the 43 collocations retained for analysis—in other words, half of the collocations. This score is much higher than the approximately 10% of medical collocations reported as previously known in Experiment 1, but this was expected to some extent since the collocations examined in Experiment 2 are more widely used in everyday situations. In the treatment-group-versus-control-group comparison model, prior knowledge (as self-reported in the yes/no test) had a

statistically significant effect (at the 5% significance level) on gapfill test scores (z = 8.83, p <.001). The odds ratio can be converted to an effect size (d) by finding the natural log of the odds ratio and then dividing by 1.81 (Chinn, 2000). The odds ratio of 2.90 converts to a medium effect size of 0.59. In other words, familiarity with the target combinations moderately increased the odds of participants in all conditions (including the control group) gaining correct answers in the test.

Table 16: Response accuracy of gapfill test (control group vs. treatment group): fixed effects summary

Parameter	*Estimate*	*SE*	*z*	*p*	*Odds ratio*	*95% CI*
Intercept	-4.18	0.31	-13.39	<.001	0.02	0.008, 0.03
Group = Treatment	1.83	0.30	6.14	<.001	6.23	3.48, 11.17
Prior Knowledge = 1	1.07	0.12	8.83	<.001	2.90	2.29, 3.68

Notes. Intercept level: Group = Control; Prior Knowledge = 0.

Having established that there was a large difference in odds between the treatment group and the control group in the gapfill test, I analysed the data for the experimental group. In the best-fit model, I found a significant interaction between learning condition and collocation type (z = 3.08, p = .002), meaning that typographic enhancement had a different effect on the accuracy levels of the two collocation types (see Table 17). The odds ratio of 2.08 equates to an effect size of 0.40. Specifically, bolding had a larger role in the retention of prep+N collocations than on the retention of V+N collocations. Prior knowledge had a statistically significant role in the accuracy of the treatment group's answers; the odds ratio of 2.72 for prior knowledge converts to a medium effect size of 0.55.

Table 17: Response accuracy of gapfill test (two treatment conditions): fixed effects summary

Parameter	*Estimate*	*SE*	*z*	*p*	*Odds ratio*	*95% CI*
Intercept	-2.62	0.25	-10.36	<.001	0.07	0.04, 0.12
Treatment = Bolding	0.13	0.18	0.71	.48	1.13	0.80, 1.61
Collocation Type = Prep+N	0.07	0.18	0.37	.71	1.07	0.75, 1.52
Prior Knowledge = 1	1.00	0.13	7.61	<.001	2.72	2.10, 3.51
Treatment:CollocationType	0.73	0.24	3.08	.002	2.08	1.30, 3.31

Notes. Intercept level: Treatment = NoBolding; CollocationType = V+N; Prior Knowledge = 0.

Table 18 shows the predicted probabilities ("fit") (fixed effects only) of a correct response for each combination of learning condition and collocation type along with 95% pointwise confidence intervals. The predicted probability of a correct answer for a V+N collocation in the bolding condition (13%) was almost the same as the probability for a V+N collocation in the no-bolding condition (11%). The predicted probability of a

correct answer for a prep+N collocation in the bolding condition (25%) was approximately double that for a prep+N collocation in the no-bolding condition (12%) and for a correct V+N answer in the bolding condition (13%). The predicted probability of a correct answer for a prep+N collocation in the no-bolding condition was almost the same as that for a correct V+N collocation answer in the bolding condition. Figure 5 presents the same information in a chart.

Table 18: Predicted probabilities of correct response for each combination of learning condition and collocation type: fixed effects summary

Learning condition	Colloc type	*Fit*	*SE*	*Lower*	*Upper*
Bolding	Verb+Noun	0.13	0.24	0.09	0.19
No-bolding	Verb+Noun	0.11	0.24	0.07	0.17
Bolding	Prep+Noun	0.25	0.23	0.17	0.34
No-bolding	Prep+Noun	0.12	0.24	0.08	0.18

Figure 5: Predicted probabilities of gapfill test accuracy (with confidence intervals)

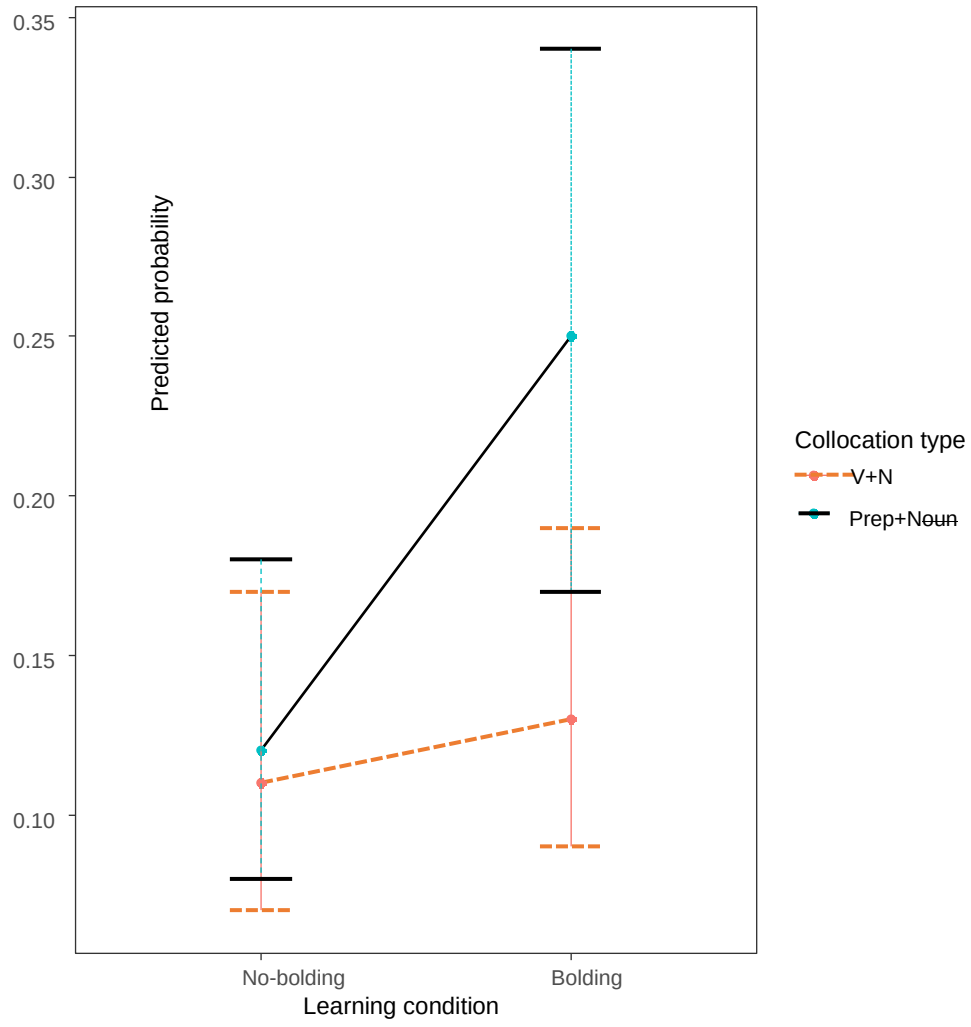


Table 19 shows the pairwise comparisons of the learning conditions and the collocation types. The scores for the prep+N combinations in the bolding condition were significantly higher than the scores for the prep+N combinations in the no-bolding condition (p <.001). However, no difference was found for the V+N collocations in the no-bolding versus the bolding conditions. The scores for the V+N collocations were significantly higher than the scores for the prep+N collocations in the bolding condition than in the no-bolding condition (p <.001). No difference was found between the scores for the V+N collocations in the no-bolding condition versus the prep+N collocations in the no-bolding condition.

Table 19: Multiple comparisons of means: user-defined contrasts

	Estimate	*Std Error*	*z*	*Pr(>|z|)*
NoBold V+N vs. Bold V+N	-0.15	0.17	-0.86	.79
NoBold Prep+N vs. Bold Prep+N	-0.90	0.16	-5.74	<.001
NoBold V+N vs. NoBold Prep+N	-0.04	0.18	-0.24	.99
Bold V+N vs. Bold Prep+N	-0.79	0.17	-4.78	<.001

4.7.3 Self-paced reading task descriptives

Table 20 shows the mean accuracy rates for the post-stimulus comprehension questions in the SPRT by participant. The treatment group recorded an accuracy rate of 92.08% for the questions containing target collocations and a rate of 91.71% for the questions containing control phrases. Control-group participants' mean accuracy rates for target-collocation sentences and control-phrase sentences were a little lower.

Table 20: Mean accuracy rates for post-stimulus comprehension questions in SPRT by participant

Condition	Target-collocation sentences		Control-phrase sentences	
	Mean (%)	SD (%)	Mean (%)	SD (%)
Control group (*n* = 25)	87.69	7.74	88.64	7.46
V+N collocations (*n* = 21)	86.51	10.20	88.44	10.95
Prep+N collocations (*n* = 22)	88.73	8.81	88.73	10.59
Treatment group (both treatments combined) (*n* = 53)	92.17	6.92	91.71	7.46
No-bolding	93.69	8.07		
Bolding	91.07	9.84		
V+N collocations	89.06	10.85	91.17	9.64
Prep+N collocations	95.03	6.80	92.11	9.27
V+N no-bolding	90.57	12.60		
V+N bolding	89.06	15.05		
Prep+N no-bolding	96.98	7.82		
Prep+N bolding	93.21	1.02		

Table 21 shows an accuracy rate of 85.26% for questions following stimulus sentences that contained target V+N collocations, while 95.48% was recorded for the questions after stimulus sentences containing target prep+N collocations.

Table 21: Mean accuracy rates for post-stimulus comprehension questions in SPRT by item

Condition	Target-collocation sentences		Control-phrase sentences	
	Mean (%)	SD (%)	Mean (%)	SD (%)
Control group (*n* = 25)	84.34	27.48	87.67	22.47
V+N collocations (*n* =21)	84.90	25.71	83.08	29.07
Prep+N collocations (*n* = 22)	83.78	29.49	91.97	12.76
Treatment group (both treatments combined) (*n* = 53)	92.68	9.40	92.11	8.77
No-bolding	93.52	9.72		
Bolding	91.29	10.34		
V+N collocations	89.35	11.23	91.16	9.60
Prep+N collocations	95.48	6.54	92.94	8.09
V+N no-bolding	89.63	11.74		
V+N bolding	89.04	12.18		
Prep+N no-bolding	96.93	5.91		
Prep+N bolding	93.34	8.04		

I performed a two-stage analysis of the incorrect answers to the post-trial comprehension questions. First I conducted an item analysis. I checked whether there were any trials for which more than 50% of the comprehension questions were incorrectly answered[19]. The lowest accuracy score for any trial was 58/78 (74.36%), for *offer a reply / spare a reply*. Since the accuracy rates were adequately high, I did not exclude any target collocations. In the second step of the first stage, I performed a participant analysis, checking for whether any participants had an accuracy rate of 50% or less[20]. However, no participants had any scores below 50%: the lowest accuracy rate of any participant was 33/43 (76.44%) (two participants). Therefore, no participant's data was excluded.

In the second stage of the post-stimulus question analysis, I needed to determine how to treat incorrect responses to the questions. There were several possible approaches. One was to exclude all the trials in which the participants responded

[19] I would have excluded the trials for those questions from the accuracy count but would have kept the trials in the SPRT analysis.

[20] I would have excluded any participants with accuracy rates below 50%.

incorrectly. This is standard practice, especially with native speakers, the assumption being that incorrect responses are due to nonstandard behaviour such as lack of attention (Juffs, 2001; Keating & Jegerski, 2015; e.g., Tremblay et al., 2011). Another possibility, commonly used with data from second-language learners, was to run two analyses—one with all trials and the other excluding trials with incorrect comprehension questions—and check for differences (ibid.). A third approach, which was consistent with mixed-effects modelling, was the one I took: I included trial comprehension question correctness in the analysis as a dummy variable (covariate). This approach took account of the post-trial question accuracy rate as a potential variable affecting the reading speeds on the nouns and did not require the exclusion of any data.

Table 22 shows the mean RTs in the SPRT for the 43 target collocations for all 78 participants. These are the raw RTs after the extreme outliers were trimmed but before the RT data were inverse transformed. The difference in mean RTs between the target-collocation nouns and the control-phrase nouns for the control group was +21.93 ms, while the difference for the treatment group (both conditions) was +13.09 ms. The treatment condition with the largest mean RT difference was the no-bolding condition, with +21.51; the bolding condition had a mean RT difference of +4.78 ms. prep+N collocations and V+N collocations had similar mean RT differences respectively.

Table 22: SPRT: Mean RTs in milliseconds for collocations and control phrases under all conditions (extreme outliers excluded)

Condition	Target collocation nouns: RT mean	SE	Control phrase nouns: RT mean	SE	Difference in RT mean: control phrases – target collocations
Control group	**589.16**	17.49	**611.09**	18.13	**+21.93**
Treatment group	**474.63**	8.46	**487.72**	8.56	**+13.09**
Treatment group					
No-bolding	**466.21**	12.04	**487.72**	8.56	**+21.51**
Bolding	**482.94**	11.89			**+4.78**
Prep+N	**442.99**	10.71	**453.31**	11.62	**+10.32**
V+N	**507.58**	13.05	**523.49**	12.45	**+15.91**
Prep+N NoBold	**431.75**	15.50	**453.31**	11.62	**+22.00**
Prep+N Bold	**454.15**	14.79			**-0.84**
V+N NoBold	**502.31**	18.33	**523.49**	12.45	**+21.18**
V+N Bold	**512.76**	18.60			**+10.73**

4.7.4 Self-paced reading task inferential analysis

Before the reading time (RT) analysis for the SPRT, I visually checked the RT distributions and trimmed outliers from the tails. Only times between 125 ms and 2500 ms were retained for analysis. The total number of data points before trimming was 3354; 24 data points, or 0.7% were excluding, leaving a total of 3332 data points for the analysis. Cut-off points in SPR studies at the fast end are usually set at between 100 ms and 200 ms; RTs shorter than these times are generally a result of unintentional button presses (Jegerski, 2013). Longer absolute cut-off times can vary from 2000 to 6000 ms (Fine & Jaeger, 2016; Jegerski, 2013; Pliatsikas & Marinis, 2013). I chose a shorter longer cut-off time as per Marinis, Roberts, Felser, & Clahsen (2005). In total, 0.66% of the RTs were excluded. Next, I applied an inverse transformation to the RTs to achieve normal distribution (Tabachnick & Fidell, 2013).

The preliminary analysis of the SPRT involved comparing the inverse-transformed RTs for the target-collocation nouns and the control-phrase nouns for the treatment group and the control group. Table 23 shows best-fit model estimates of the differences in RTs between the treatment group and the control group. I found no interaction between the groups in the experimental condition. This means that the treatment and control groups experienced the same trend in the SPRT. Specifically, the difference between the collocations and the control phrases was similar for the participants in the treatment and control groups. That is, participants in both groups reacted more quickly to the target-collocation nouns than to the control-phrase nouns (t = -2.06, p = .04). I had hoped for an interaction, in which the control group would record similar RTs for both target-combination and control-phrase nouns while treatment-group participants would record faster times for target-combination nouns than for control-phrase nouns. The likely explanation for the no-interaction finding is that some collocations were already familiar to the participants; this was confirmed by the result that the covariate prior knowledge was a significant factor (t = -2.09, p = .04). In other words, self-reported familiarity with specific collocations led to faster RTs for those collocations for both the treatment-group participants and the control-group participants. The inverse-transformed RTs for the first words of the combinations were a significant factor in the RTs of the nouns. That is, the speed with which a participant responded to the first word of the combination—whether a verb or a preposition—was connected to the time taken to respond to the noun.

When fitting models, I reviewed the collinearity of Word1 RTs and target noun RTs, checking for high correlations between Word1.RT and noun RT. The correlation between the treatment group and the control group was fairly low, at 0.28, while the correlation for the two conditions within the treatment group was reasonably high: 0.49. I centred the RTs to reduce collinearity.

Table 23: SPRT RTs (inverse transformed): control group vs. treatment group: fixed effects summary

Parameter	*Estimate*	*SE*	*df*	*t*	*p*
Intercept	-2.37	0.17	86.00	-14.36	<.001
Group = Treatment	-0.25	0.19	72.00	-1.29	.20
Exp Cond = Target Collocs	-0.06	0.03	3221.00	-2.06	.04
Prior Knowledge = 1	-0.07	0.03	3274.00	-2.09	.04
Word 1 RT (inv trans)	0.22	0.02	3324.00	13.00	<.001

Notes. Intercept level: Group = Control; Exp Cond = Control Phrases; Prior Knowledge = 0.

Table 24 shows the main analysis: of the inverse-transformed RTs of the treatment group to target-collocation nouns versus control-phrase nouns. There was a significant difference between the RTs for V+N collocation nouns and their controls and the RTs for prep+N collocation nouns and their controls, with the RTs for prep+N collocation nouns being faster (t = -2.13, p = .03). There was an interaction between learning condition and collocation type (t = -2.13, p = .03). The inverse-transformed RTs for the first words of the collocations were a significant factor in the RTs of the nouns. In other words, the speed with which a participant responded to the first word of the collocation—whether a verb or a preposition—was connected to the time taken to respond to the noun. Prior knowledge of collocations was not a significant factor in the RTs for the treatment group (t = -1.61, p = .11).

Table 24: SPRT RTs (inverse transformed) for treatment group: fixed effects summary

Parameter	*Estimate*	*SE*	*df*	*t*	*p*
Intercept	-2.53	0.13	93.80	-19.91	<.001
Treatment = NoBolding	-0.08	0.08	2179.60	-1.03	.31
Exp Cond = Control Phrases	0.08	0.07	2179.60	1.22	.22
Colloc Type = Prep+N	-0.16	0.08	2200.50	-2.13	.03
Word 1 RT (inv transformed)	0.20	0.02	2256.50	9.90	<.001
Prior Knowledge = 1	-0.07	0.04	2221.90	-1.61	.11
Treatment(NoBold):CollocType (pN)	-0.13	0.11	2179.00	-1.19	.24
Treatment(ControlPhr):CollocType (pN)	-0.19	0.09	2178.80	-2.13	.03

Notes. Intercept level: Treatment = Bolding; Exp Cond = Target Collocations; CollocationType = V+N; Prior Knowledge = 0.

Table 25 shows pairwise comparisons of the learning conditions and collocation types for the treatment-group participants (taking into account Word 1 as a covariate). These are comparisons between the inverse-transformed RTs of the target collocations only: matched control phrases are not included in the comparisons. There was a significant difference between the prep+N collocations in the no-bolding condition and the prep+N collocations in the bolding condition (z = -2.75, p = .04), with RTs being faster in the no-bolding (reading only) than bolding condition. The V+N combinations in the no-bolding condition were faster than the control phrases at a level that was (arguably) marginally significant (z = -2.40, p = .10) (see Pritschet, Powell, & Horne, 2016, for a discussion on psychology researchers' descriptions of p-values over .05 as marginally significant). This finding suggests, when V+N collocations were encountered in the reading only treatment, the development of procedural knowledge may have been triggered, which was not the case for the bolding treatment.

The inverse-transformed RTs of the target-collocation nouns for the prep+N collocations in the no-bolding condition were significantly faster than the times for the nouns in the V+N collocations in the no-bolding condition (z = 3.75, p = .001).

Table 25: Multiple comparisons of means of inverse-transformed RTs: fixed effects summary

Parameter	*Estimate*	*SE*	*z*	*Pr(>|z|)*
prep+N				
NoBold Prep+N vs Bold Prep+N	-0.20	0.07	-2.75	.04
NoBold Prep+N vs Control Phrases	-0.09	0.06	-1.39	.63
Bold Prep+N vs Control Phrases	0.11	0.06	1.79	.36
V+N				
NoBold V+N vs Bold V+N	-0.08	0.08	-1.03	.85
NoBold V+N vs Control Phrases	-0.16	0.07	-2.40	.10
Bold V+N vs Control Phrases	-0.08	0.07	-1.22	.74
V+N vs prep+N				
NoBold V+N vs NoBold Prep+N	0.29	0.08	3.75	.001
Bold V+N vs Bold Prep+N	0.16	0.08	2.13	.19

Figure 6 shows the predicted probabilities of the RTs for the two learning conditions in inverse-transformed units. For this chart, the inverse RTs were back-transformed to the original scale (in milliseconds).

Figure 6: Predicted mean (back-transformed from inverse) RTs in two treatment conditions for treatment group (with confidence intervals)

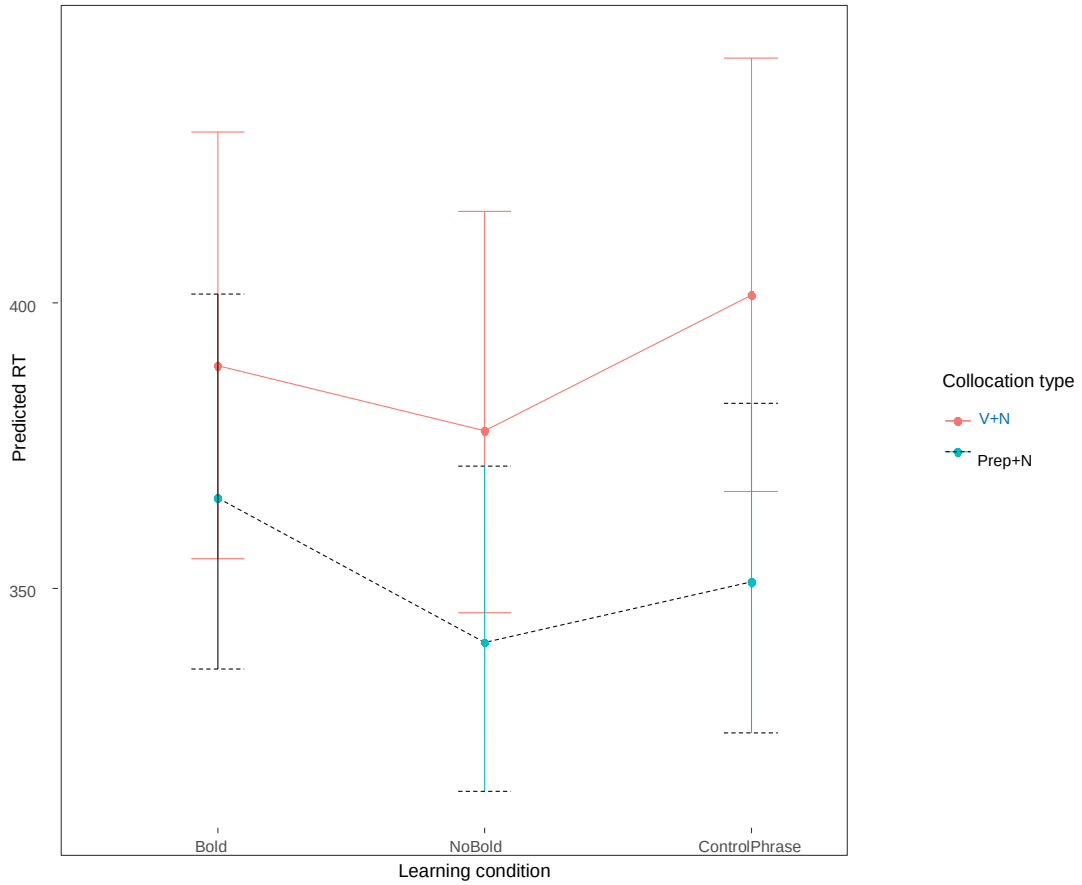


Table 26 shows inverse-transformed RTs for the target-collocation nouns and control-phrase nouns for the control group. There was no statistical difference between the RTs of the target-collocation nouns and the RTs of the control-phrase nouns (t = -1.43, p = .15). This result was as expected (as was the case in Experiment 1) since control-group participants had been previously exposed in the experiment to neither the collocations nor the control phrases. The table also shows that the RTs for the target prep+N collocations were faster than the RTs for the V+N collocations (t = -4.94, p <.001).

Table 26: SPRT RTs (inverse transformed): control group: fixed effects summary

Parameter	*Estimate*	*SE*	*df*	*t*	*p*
Intercept	-2.17	0.22	29.20	-9.93	<.001
Exp Cond = Target Collocs	-0.09	0.06	1022.30	-1.43	.15
CollocType = Prep+N	-0.31	0.06	886.00	-4.94	<.001
Prior Knowledge = 1	-0.10	0.06	1040.90	-1.56	.12

Notes. Intercept level: Exp Cond = Control Phrases; Prior Knowledge = 0.

4.8 Discussion

In this section, I present a summary of the main findings from Experiment 2. Then I evaluate the study's research questions in light of the results. I assess the role of prior collocational knowledge and then discuss the dissociation between the development of declarative and procedural collocational knowledge. Finally, I discuss the experiment's limitations.

4.8.1 Summary of findings

Six exposures to 43 non-technical lexical and grammatical collocations in diverse and verbatim contexts in multiple short written texts in this experiment facilitated an increase in declarative knowledge in both treatment conditions, as measured in a gapfill (cued-recall) test and when compared to control-group test scores. Typographic enhancement (bolding) led to a larger amount of declarative knowledge of prep+N collocations than no typographic enhancement; however, typographic enhancement resulted in no extra declarative knowledge of V+N collocations. Significantly more learning occurred of prep+N collocations than of V+N collocations in the bolding condition; however, in the no-bolding condition, no difference was found between the amount of declarative knowledge of the two collocation types. Self-reported familiarity with (prior knowledge of) the collocations played a significant role in the accuracy rates in the gapfill test.

The positive effect of typographic enhancement on the learning of prep+N collocations which was recorded in the gap-fill test was not found in the SPRT. In contrast, the *absence* of typographic enhancement of prep+N collocations conferred a processing advantage over typographic enhancement of prep+N collocations. However, robust procedural knowledge was not found for either the bolded or the unbolded prep+N collocations or V+N collocations, as measured in the comparison of the RTs of the target-collocation nouns with the RTs of the control-phrase nouns (although a marginal advantage of V+N collocations over control phrases was observed for the non-bolded treatment, see Table 53). Nevertheless, the predicted probabilities (see Figure 5) indicate a possible pattern of faster RTs for unbolded collocations (both prep+N and V+N) than for control phrases. The RTs for the nouns in the unbolded prep+N collocations were faster than those for the unbolded V+N collocations, but there was no difference for the two collocation types in the bolded condition. Prior knowledge of the target collocations unexpectedly affected the RTs of both treatment-group and control-group participants in the same way: both groups read the nouns in the collocations more quickly than the nouns in the matched control phrases. This result may have been because, on average, half of the collocations were reported as being familiar by treatment-group and control-group participants, while the matched control phrases were unfamiliar. (For details of the number of target collocations previously known, refer to section 5.9.3.). Tables 27 and 28 show key results of the post-tests.

Table 27: Key gapfill test results: declarative knowledge gained

prep+N	Bolding > No-bolding
V+N	Bolding ≈ No-bolding

Table 28: Key SPRT results: procedural knowledge gained (collocations vs. control phrases)

prep+N	No-bolding ≈ Control Phrases	Bolding ≈ Control Phrases
V+N	No-bolding ≈ Control Phrases	Bolding ≈ Control Phrases

Note. This summary table takes a conservative approach to the findings, setting aside any marginally significant comparisons.

As in the cued-recall tests of Experiment 1, prior knowledge of the target collocations—operationalised as participants reporting having seen or heard the collocations before—played a role in the gaining of declarative knowledge in the gapfill test of Experiment 2.

On the other hand, prior knowledge appeared to play no role in the development of procedural knowledge of the collocations in Experiment 2, as measured in the SPRT—a result similar to that found in the primed LDTs in Experiment 1.

4.8.2 The research questions

The experiment's research questions will now be answered in light of the post-test results.

Research question 1: Does repeated exposure to V+N and prep+N collocations in written input lead to the development of declarative and/or procedural knowledge?

The answers to research question 1 are "yes" for declarative knowledge and "no" for procedural knowledge. Six exposures to both V+N and prep+N target collocations resulted in gains in declarative knowledge of both types of collocations as seen in the difference in the gap-fill test scores of the treatment group and the control group. Gains were found for the collocations presented in both the bolding and no-bolding learning conditions. These findings are similar to those of Experiment 1 and Sonbul and Schmitt (2013), in which nine and three exposures respectively to each of the technical collocations produced declarative knowledge of the collocations. However, the amount of declarative collocational knowledge gained in Experiment 1 was substantially more than the amount gained in Experiment 2. I will discuss the reasons for this difference in Chapter 5.

Both the treatment-group participants and the control-group participants recorded faster times for the target-collocation nouns than for the matched control-phrase nouns; therefore, it cannot be claimed that repeated exposure to the collocations in the experiment produced procedural knowledge overall. The finding for the control-group participants—who were not exposed to the collocations in the treatment session—was unexpected as I had assumed that their RTs for the target-collocation nouns and control-phrase nouns would be approximately the same: such a result would have shown that the control-group participants had little knowledge of the collocations. The actual result seems likely to have been due to the control-group (as well as the treatment-group) participants already having partial knowledge of approximately half of the target collocations, leading to faster RTs for the nouns of the collocations in the SPRT. (A mean 52.51% of the prep+N combinations and 48.47% of the V+N

combinations were reported by the 78 participants to have been previously familiar.) It appears that, even though the control-group participants' cued-recall (declarative) knowledge of the collocations (as measured in the gap-fill test) was relatively low, they had a higher level of pre-existing knowledge—perhaps at the level of recognition—of a number of the collocations[21].

Research question 2: Does typographic enhancement lead to greater development of declarative knowledge of V+N and prep+N collocations than no typographic enhancement?

The answer to this question is a partial "yes". Typographic enhancement led to greater declarative (cued-recall) knowledge of prep+N collocations than no typographic enhancement. This result was expected since bolded collocations in Experiment 1 and Sonbul and Schmitt (2013) were learned better than unbolded collocations. What seems likely is that bolding of prep+N combinations made them more perceptually salient to the learners, drawing increased attention to the combinations and helping the learners detect formal and semantic links between the prepositions and the nouns (Boers et al., 2016; Choi, 2017; Sonbul & Schmitt, 2013).

Unexpectedly, no difference was found between the amount of declarative knowledge resulting from repeated exposure to V+N collocations in the typographic enhancement and the no-typographic enhancement conditions. It seems that, while bolding increased the participants' noticing (relative to the no-bolding condition) of the prep+N combinations, it made no difference (relative to the no-bolding condition) to the noticing of the *V*+N combinations. The result for the non-technical lexical (V+N) collocations in this experiment contrasts with that found for the technical lexical (adjective + noun and noun + noun) collocations in Experiment 1, in which the typographically-enhanced collocations had an advantage over the unenhanced collocations. Typographic enhancement of the V+N collocations in Experiment 2 appears not to have made the combinations any more perceptually salient to the

[21] It should be noted that, in the gapfill test, control-group and treatment-group participants supplied many verb and preposition answers which were acceptable alternatives but were marked as incorrect because they did not complete the target collocations (e.g., a participant answered *make a comparison* instead of *draw a comparison*). While control-group participants had a 5% accuracy rate in terms of the "official" answers, when answers that seemed acceptable alternatives were added, the control group's correctness rate was 18%. In comparison, the accuracy rate of the official answers for treatment group-participants was 21%, but, taking into account acceptable alternative answers, the rate increased to 31%.

participants than lack of typographic enhancement. It seems that the V+N combinations were just as salient (or non-salient) to the learners whether bolded or unbolded, and that bolding was just as effective (or ineffective) as no bolding in alerting the learners to the fact that the verbs collocated with the nouns. (The declarative learning rate for both learning conditions, as measured in the gapfill test, was fairly low anyway. As reported above, the predicted probability of obtaining a correct answer for a V+N collocation in the bolding condition was 0.13 compared with 0.11 for a V+N collocation in the no-bolding condition.) Further, it appears that once the prep+N collocations were *noticed* it was more likely they would be remembered, leading to greater declarative knowledge. Possible reasons for the difference in memorability of prep+N and V+N collocations will be examined in the discussion of research question 4.

Research question 3: Does lack of typographic enhancement lead to greater development of procedural knowledge of V+N and prep+N collocations than typographic enhancement?

The answer to research question 3 is a tentative, partial "yes". The nouns in the V+N collocations in the no-typographic enhancement condition were read faster than the control-phrase nouns at a marginally-significant level (see Tables 22 and 25); this may indicate the initial stages of development of procedural collocational knowledge. However, there was no difference between these two conditions for the prep+N collocations. Meanwhile, the times for the two conditions of the prep+N collocations indicate a clear processing advantage for the no-bolding condition (Table 25); however, this advantage does not equate to the development of procedural knowledge, which is operationalised as faster RTs for target collocation nouns than for control-phrase nouns. In fact, the RTs for the nouns in the bolded prep+N collocations were slower than the RTs for the control-phrase nouns (see Table 25 and Figure 6). This seems to indicate that the presence of bolding somehow hindered the processing of the bolded collocations. Perhaps bolding promoted conscious attention to (noticing of) the collocations, thus slowing the reading of the collocations and also slowing the development of procedural collocational knowledge. Generally speaking, it may be inferred from the SPRT results for both the V+N collocations and the prep+N collocations that the participants' repeated exposure to unbolded combinations conferred a processing advantage over repeated exposure to bolded combinations.

The finding that *lack* of typographic enhancement of prep+N collocations resulted in a processing advantage over typographic enhancement of prep+N collocations and that lack of typographic enhancement of V+N collocations may have promoted the establishment of some procedural knowledge is the reverse trend of what occurred in the gapfill test, although it is consistent with the results of the equivalent tests for the medical collocations in Experiment 1. In Experiment 1, as in Experiment 2, RTs for the second words in the target collocations were faster in the no-bolding (reading-only) condition (compared with control pairs) than in the bolding condition. As suggested in Chapter 3, a possible reason for this finding is that typographic enhancement may interrupt the normal reading process through noticing caused by bolding, leading to slower development of procedural knowledge (Paradis, 1994). This issue will be discussed in Chapter 5. It may be that bolding leads to conscious noticing more effectively than no-bolding and thus greater declarative knowledge, but *lack* of bolding promotes attention without awareness to collocational form and leads to faster processing; this process may promote the development of procedural knowledge more quickly than it does the development of declarative knowledge.

Research question 4: Do repeated exposure and typographic enhancement differentially affect the development of declarative and/or procedural knowledge of V+N and prep+N collocations?

Repeated exposure produced greater declarative knowledge of prep+N collocations than of V+N collocations in the gapfill test, but only in the bolding condition. There was no difference in the amount of declarative knowledge of prep+N combinations and V+N combinations gained in the no-bolding condition. As noted above, typographic enhancement seems to have made the prep+N collocations (but not the V+N collocations) more perceptually salient to the participants Prepositions may be often overlooked by readers for several reasons. Firstly, they tend to contain fewer letters than verbs. Of the verbs and prepositions in the 43 target collocations included in the analyses in this study, the mean number of letters in the verbs was 4.81 whereas the mean number of letters in the prepositions was 3.77. Secondly, prepositions tend to be more frequent than verbs in natural language[22], thus being seen as very well known

[22] To gain an idea of the frequency levels of prepositions and verbs in natural language, I checked the number of total occurrences of all the prepositions and verbs in the target collocations in COCA (Davies,

(Boers et al., 2014). Thirdly, prepositions may seem very familiar because they are highly polysemous. The prepositions in the target collocations are more polysemous than the verbs in the collocations. The 13 prepositions in the 22 prep+N combinations which underwent analysis each had an average of 14.15 meaning senses in the online Macmillan Dictionary. By comparison, the 21 verbs in the V+N collocations subjected to analysis had a mean 8.33 meaning senses each. If readers generally ignore prepositions to some extent, typographic enhancement may have a stronger effect in increasing the salience of prepositions than it does the salience of verbs. Boers et al. (2014) argue that many collocations lack novelty, and thus salience, because the constituent content words tend to be frequent and the constituent function words even more frequent.

I make no claims about the development of procedural knowledge of V+N and prep+N collocations since the RTs for neither combination types were significantly faster than their matched control phrases (see Table 51). However, the prep+N collocations did have a processing advantage over the V+N collocations, but only for those combinations seen in the no-typographic enhancement learning condition. There was no statistical difference between the times for the two collocation types in the bolding condition. The processing advantage for the unbolded prep+N collocations is the opposite of the result in the gapfill test, in which superior declarative knowledge was found in the bolding condition, but is consistent with findings from Experiment 1. A significant factor in the RTs of the nouns in the collocations was the RTs for the first words of the collocations (the verbs and the prepositions). When considered alongside faster RTs for unbolded prep+N combinations than for unbolded V+N combinations, this means that the RTs to the prepositions were faster than the RTs to the verbs.

There are several possible reasons why higher declarative-knowledge test scores and faster processing times were recorded for the prep+N collocations than for the V+N collocations by the treatment-group participants. One reason is related to the frequency effect, with participants likely having been exposed—both outside and within the

2010). Prepositions were initially counted as many times as they occurred in the target collocations. For example, the preposition *by* was in two target collocations so was included in the tally twice. All inflections of verbs were counted (e.g., *meet*, *meets*, *met*, *meeting*); this meant counting verb forms which were also other words forms (e.g., *meeting* is a noun as well as a past participle). The mean number of occurrences of each verb in COCA was 211,716.1; the mean number of occurrences of each preposition was more than ten times larger: 2,635,782.7. In a second calculation of the prepositions' frequencies, I counted each of the 14 prepositions in the 22 target prep+N collocations only once; the mean was 1,950,653.

experiment—to a larger number of the target prepositions or prep+N combinations than the target verbs or V+N combinations respectively. This greater frequency likely led to greater familiarity with the items. The issue of the frequency effect can be divided into several subfactors. First, as already mentioned, prepositions in general—and the target prepositions in this experiment specifically—have a higher frequency (possibly much higher) in real language than verbs[23]. Second, the prep+N combinations in this study have a much higher frequency in natural language than the V+N combinations[24]. Third, the repeated use of several prepositions in more than one collocation in the study may have produced a repetition effect in the SPRT, contributing to the faster RTs for the prep+N collocations than for the V+N collocations. For example, the preposition *on* occurred in the target collocations *on the decline* and in *on review*; *off* occurred in *off target*, *off work* and *off course*.

Other reasons for higher test scores and faster RTs for prep+N collocations than V+N collocations are connected to other properties of prepositions and verbs. One reason (already noted) is that prepositions are in general shorter than verbs, so they may be recognised more quickly. Another possible reason (especially related to gapfill test scores) is that there is a limited number of prepositions in English to choose from compared with a larger number of verbs. As evidence of this, Saint-Dizier's (2006, p. 2) "fairly complete" list of one-word English prepositions contains 65 prepositions. In contrast, there are thousands of verbs in English (Bonk, 2001; Simpson & Wiener, 1989). If a participant was unsure which first word of a collocation to write in a blank in the gapfill test, it would have been easier to correctly guess a preposition than a verb.

[23] All of the prepositions in the 22 prep+N collocations retained in the analysis are at K1 frequency in the BNC/COCA, whereas 18 of the 21 verbs in the V+N collocations are in the K1, with three in the K2. Controlling for equivalent frequency bands for the verbs and prepositions in the target collocations proved to be difficult, if not impossible. However, they are all high-frequency words which the intermediate and upper-intermediate learners would likely have been familiar with.

[24] The mean raw frequency in COCA of the prep+N collocations retained for analysis was 363.55, while the raw frequency for V+N collocations was much lower, at 47.10. This difference in frequency may have been a factor in the participants in the study reporting a slightly larger amount of prior knowledge of the prep+N combinations. In the yes/no recognition test, the participants reported previously knowing 52.51% of the prep+N combinations subjected to analysis compared with 48.47% of the V+N combinations. However, it is arguable how reliable this test is a measure of previous exposure to the collocations. If participants had seen certain prep+N collocations before but not remembered them, for the reasons outlined above, their scores on the yes/no test would be too low.

4.8.3 Dissociation between development of declarative and procedural knowledge

As in Experiment 1, declarative and procedural collocational knowledge developed at different rates under various learning conditions in Experiment 2. This finding supports Sonbul and Schmitt's (2013) suggestion of a dissociation between the development of both knowledge types. This issue will be discussed further in Chapter 5.

4.8.4 Limitations

One limitation of this experiment was that while the selection of non-technical collocations had the advantage of being phrases that were in general usage, many were already recognised by the participants. This was the likely cause of the control-group participants recording faster RTs for the target-collocation nouns than for the control-group nouns where normally similar times would be expected. Probably the main reason for the participants' familiarity with the collocations was that the collocations are more frequent in natural language than the medical collocations of Experiment 1. A likely related reason was the relatively high upper limit of English proficiency (IELTS 6.5) for participants. However, the high limit did mean that a larger number of participants were able to be recruited. It must also be remembered that prior knowledge was accounted for in the statistical model.

Another potential limitation was the lack of naturalness of some of the control phrases. While every effort was made to find controls that were as natural as possible, in some cases the resulting phrases may seem less natural. This was particularly true of some prep+N control phrases (e.g., *around reason*, *down work*) for which the choice of prepositions was limited. Ideally, I would have normed the control phrases for naturalness and predictability and then included these factors as covariates in the mixed-effects modelling in order to gauge their effect on the measures of declarative and procedural knowledge.

There were several other limitations of Experiment 2. One is that, as already noted, it was necessary for several prepositions to be included in more than one (but no more than three) target collocations, producing a possible repetition effect in the SPRT. Another limitation, already mentioned, is the possibility of a testing effect (e.g., Carpenter & DeLosh, 2006): exposure to the collocations in the SPRT may have had some effect on the accuracy rates of the gapfill test. A third limitation is that the

reliability of the gapfill test may have been lessened by the fact that many of the test items have alternative answers. For example, an acceptable alternative to the collocation *draw a comparison* is *make a comparison*. However, when participants supplied *make a comparison*, it was marked as incorrect since it was not the target collocation. While this practice may have been justifiable for treatment participants, it might be seen as unfair to control-group participants, who had not seen half the collocations in the experiment and who had seen the other half only briefly in the SPRT.

In the next chapter, I will bring together the results and discussion of both my experiments and make overall claims about what I have found in this thesis.

Chapter 5: General discussion and conclusion

In this chapter, I combine general discussion of my two studies with a conclusion. I firstly briefly review the background literature relevant to my thesis and then answer the overarching research questions of the thesis, outline the thesis's most important findings and compare the key findings of the two studies. Next, I discuss the theoretical implications of my findings with respect to the literature. I then outline the studies' limitations, offer recommendations for future research, discuss pedagogical implications of my research, and, finally, make concluding remarks.

5.1 Overarching research questions

The overarching research question for my thesis has been:

Which types of knowledge of collocations are developed through different input enhancement techniques in written input?

It appears that different incidental learning techniques do have different effects on the learning of second-language collocations in written input. However, to answer this question adequately, I split it into two further overarching research questions, one for each of the two experiments. I will briefly answer each of these questions in this section by giving firstly a summary of the method employed in each study and then a summary of the results.

Experiment 1: *Which types of knowledge of technical lexical collocations are developed through different input enhancement techniques in written input?*

Experiment 1 was a conceptual replication and extension of Sonbul and Schmitt (2013). I attempted to corroborate their results by finding evidence of declarative collocational knowledge; I also hoped, by increasing the number of repetitions of the collocations and the number of treatment sessions, and by adding an extra day to the treatment, to find evidence of procedural collocational knowledge where Sonbul and Schmitt had found none. In a counter-balanced learning and experimental condition, 62 advanced adult ESL speakers of various first languages were exposed to nine occurrences of 15 low-frequency technical collocations (e.g., *cloud baby*, *regional control*) in 500-word written texts in three sessions on two consecutive days. Three incidental learning treatments

were implemented: reading-only (typographically unenhanced), bolding-only and bolding-plus-glossing. Declarative collocational knowledge was assessed in a cued-recall (gapfill) test and a form-recognition (multiple-choice) test in both immediate and delayed sessions. Procedural collocational knowledge of form was operationalised as a priming effect in a primed LDT.

The results of the experiment show that declarative knowledge of technical lexical collocations was developed through input flooding both with typographic enhancement (plus or minus glossing)—which drew attention to the collocations' forms—and without it. A larger amount of declarative knowledge was gained through typographic enhancement and typographic enhancement plus glossing than through no typographic enhancement. Procedural knowledge was developed through input flooding without typographic enhancement but there were no indications of procedural knowledge through input flooding with typographic enhancement—with or without glossing.

Experiment 2: *Which types of knowledge of non-technical lexical and grammatical collocations are developed through different input enhancement techniques in written input?*

Experiment 2 focused on non-technical lexical (V+N) collocations and grammatical (prep+N) collocations. For this experiment, I changed a number of the elements of the design of Experiment 1. Seventy-eight intermediate-to-upper-intermediate-level adult native speakers of Chinese were exposed to six occurrences of each of 48 non-technical lexical (V+N) and grammatical (prep+N) English collocations (e.g., *suffer a decline*, *on the decline*) in two incidental learning conditions in a counter-balanced, within-participants study design. The participants read twelve 170-word treatment texts twice on two consecutive days. The immediate post-test session comprised a gapfill test, for measuring declarative knowledge, and a self-paced reading task (SPRT), for measuring procedural knowledge. No delayed post-test session was held.

The results of Experiment 2 show that declarative knowledge of non-technical lexical and grammatical collocations was developed through input flooding, both with and without typographic enhancement. Input flooding with typographic enhancement facilitated more declarative learning of grammatical collocations than input flooding without typographic enhancement. Typographic enhancement and no typographic

enhancement of the lexical collocations led to the same amount of declarative knowledge. Grammatical collocations were processed more quickly through input flooding without typographic enhancement than through input flooding with typographic enhancement, but there was no evidence of the development of procedural knowledge. No robust procedural knowledge of the lexical collocations was recorded through either learning approach, although there was marginally significant evidence of the development of procedural knowledge in the V+N collocations seen in the unbolded condition.

5.2 Comparison of both studies

In this section, I compare the key findings of both experiments, particularly the differences between them. I first discuss the techniques of input enhancement used in the studies and the types of knowledge—declarative and procedural—gained; then I discuss possible reasons for the large differences in the overall gains of declarative knowledge between the two experiments; finally, I examine the role of the co-variate prior knowledge.

5.2.1 Input enhancement and types of knowledge

Both experiments in this thesis employed two types of input enhancement of target collocations in written text: input flooding alone and input flooding plus typographic enhancement[25]. In both experiments, input flooding alone facilitated gains in declarative knowledge of the collocations, and, in some cases, procedural knowledge. Such gains were expected. (The theoretical implications of input flooding are discussed in subsection 5.5.5.)

In both experiments, learner exposure to input flooding alone had a different effect on the development of declarative and procedural knowledge than exposure to input flooding combined with typographic enhancement—and glossing. In Experiment 1 only, typographic enhancement and glossing, in conjunction with input flooding, promoted primarily conscious, declarative knowledge (as measured in gap-fill cued-recall and multiple-choice form-recognition tests). In both experiments, larger amounts

[25] Another type of input enhancement, input flooding plus typographic enhancement plus marginal glossing, was used only in Experiment 1, so it will not be a focus of discussion in this section.

of declarative knowledge developed through typographic enhancement than through no typographic enhancement, while some procedural knowledge developed, or may have developed, through input flooding in the absence of typographic enhancement but none through typographic enhancement; this pattern held true for both the technical lexical (medical) collocations in Experiment 1 and for the non-technical grammatical (prep+N) combinations in Experiment 2. However, unexpectedly, this was not the pattern for the non-technical lexical (V+N) collocations, for which no difference was found between the amount of declarative knowledge gained in the two treatment conditions, although some initial development of procedural knowledge may have been triggered by input flooding without typographic enhancement.

One possible explanation for the findings for declarative knowledge is that an advantage for typographic enhancement was only observed when typographic enhancement made the collocations more salient to, and thus noticed more by, the participants. The low-frequency medical collocations in Experiment 1 were largely unknown to the participants (although the high-frequency words comprising the collocations were most likely known to them). Typographic enhancement increased the collocations' salience, facilitating increased noticing of the word combinations as units with particular meanings in medical contexts, thus leading to greater declarative knowledge of the collocations. The advantage for typographic enhancement was also found for the higher-frequency, non-technical prep+N combinations in Experiment 2. As with the medical collocations, bolding of the grammatical collocations appears to have made them more perceptually salient, leading to greater noticing and recall than no bolding. That result may have been because the very high frequencies and short lengths (among other possible factors) of the prepositions in the prep+N combinations made the prepositions less salient and therefore more easily overlooked and less easily recalled in a gapfill test when unenhanced than when enhanced. In contrast, the non-technical V+N collocations contain verbs, which, while frequent, are less frequent in naturally-occurring language than prepositions, which are highly frequent (Gledhill, 2000). (For example, *meet*, from the V+N collocation *meet targets*, appeared about 79,000 times in COCA, while *off*, from the parallel prep+N combination *off target*, appeared about 410,000 times in the corpus.) As a result, the verbs, while likely known by the participants, may have been less familiar and consequently more salient to the learners than the often-overlooked prepositions (Durrant, 2009). Such greater salience

possibly meant that bolding did not lead to greater noticing of the verbs in the V+N combinations than no bolding as it apparently did with the prepositions in the prep+N combinations.

The findings from the test of procedural knowledge in Experiment 2, the self-paced reading task, which appear to show some advantage for the absence of typographic enhancement, partially corroborate the findings from the test of procedural knowledge in Experiment 1, the primed LDT. In both tests, collocations in the reading-only/no-bolding condition were processed more quickly than collocations in the bolding condition. Absence of typographic enhancement facilitated the development of procedural knowledge of medical collocations in Experiment 1 and there was a marginally significant indication of the development of procedural knowledge of V+N collocations in Experiment 2. There was no evidence of procedural collocational knowledge for the bolding condition in either the primed LDT in Experiment 1 or the SPRT in Experiment 2.

In comparing the results for the two types of lexical collocations in the two experiments—the medical (adjective + noun and noun + noun) collocations in Experiment 1 and the non-technical (V+N noun) collocations in Experiment 2—more procedural knowledge was established of the medical collocations. There are a number of possible explanations for this. Firstly, the medical collocations were presented more frequently: nine times each compared with six times each for the non-technical collocations. Secondly, the medical collocations were more highly contextualised than the non-technical collocations in Experiment 2. (This point is discussed in the next subsection.) The third possibility is that, in the SPRT, the RTs were measured on the terminal noun only, but not on the word following it. In some studies, "spillover effects" are found on words before or after the critical region (Marsden et al., 2018); in this experiment, however, I did not conduct such extra analysis due to lack of time.

5.2.2 Differences in gains in declarative knowledge between studies

In both experiments, six to nine exposures to target collocations in written texts led to a significant amount of declarative knowledge. Although the differences in the results of the two studies were not statistically analysed, it is clear that the total amount of declarative knowledge of the medical collocations measured in the cued-recall test in Experiment 1 was substantially more than the amount of the non-technical collocations

in the equivalent test in Experiment 2. In Experiment 1, all treatment-group participants scored 68.54% in the immediate cued-recall (gapfill) post-test (compared with a score of 0.48% for the control group). This compares with 20.67% for all treatment-group participants on the gapfill test in Experiment 2 (and the control-group accuracy score of 5.02%). The large difference between the treatment groups can be explained by several possible factors.

Firstly, the linguistic proficiency of the learners in Experiment 1 (advanced) was higher than that of the learners in Experiment 2 (intermediate to upper-intermediate). This may have contributed to the higher learning rate of new lexical items in Experiment 1 (Daneman & Green, 1986).

Secondly, the collocations in each experiment were of different types. The collocations in Experiment 1 were low-frequency technical collocations with restricted meanings; in Experiment 2, the collocations were non-technical combinations, which are in wider use in natural language. Although the non-technical collocations of Experiment 2 were more familiar to the participants than the technical collocations of Experiment 1 were to the participants of that study—as shown in the yes-no self-report test results—the lower scores in the declarative-knowledge tests could have been due to the collocations in Experiment 2 lacking novelty and so being less perceptually salient to the participants, particularly the unbolded prep+N collocations (Boers et al., 2014).

Thirdly, even though both experiments contained lexical collocations, the V+N collocations in the Experiment 2 may have been more difficult to learn than the noun + noun and adjective + noun combinations in Experiment 1. As noted in the literature review, V+N combinations can be more arbitrary, more abstract and more polysemous (Altenberg & Grander, 2001; Boers et al., 2014; Crossley et al., 2013; Gitsaki, 1999; Huo, 2014).

Fourthly, as already noted in the previous subsection, there were fewer repetitions of the collocations in Experiment 2 (6) than there were in Experiment 1 (9). This meant that the participants likely gave less attention to the target items in Experiment 2.

Fifthly, in Experiment 1, the thematic focus within paragraphs in the treatment texts was on the medical collocations, each of which was accompanied by not only an explicit contextual definition but also by three massed repetitions of the collocation

within the space of usually three sentences. In Experiment 2, the occurrences of collocations were spaced throughout six separate mini-texts, and were not the thematic focus of attention in the mini-texts. The latter spacing may have reduced the amount of learner focus on the target items.

Sixthly, it may be that the more supportive, or constraining, a text was of meaning inferences of the target collocations the greater the amount of declarative knowledge developed. While the technical (medical) collocations of Experiment 1 had high contextual support in the treatment texts through the provision of explicit definitions (Bolger et al., 2008; Daneman & Green, 1986), as well as through the provision of marginal glosses for a third of the collocations (as determined by learning condition), the non-technical collocations in Experiment 2 were not defined and so likely attracted less learner attention and promoted less elaboration of meaning. It is probable that the non-technical collocations did not require definitions to be mostly understandable; nevertheless, the lack of definitions likely meant that learners spent less time processing semantic information related to each occurrence of each collocation in Experiment 1 than they did processing such information for each medical collocation in Experiment 1.

The part-paragraphs below, taken from the treatment texts in experiments 1 and 2, illustrate the fifth and sixth factors described above. As can be seen in the following extract from Experiment 1, there is massed repetition of the technical term *regional control*, which occurs in three consecutive sentences. Note also the explicit definition of the term ("stopping the spread of cancer from its starting place to other locations"), which promotes elaboration of meaning:

> Cancer is the leading cause of death in New Zealand; it caused 8891 deaths, or nearly a third of deaths in the country, in 2011. Most deaths from cancer occur after the cancer has spread from its original location to other organs. Because of this, **regional control** is important. **Regional control**, or stopping the spread of the cancer from its starting place to other locations, can be done through the use of surgery, drugs or radiation. The use of **regional control** can have a positive effect on the treatment's success. [Note: I have added underlining to the text.]

In contrast, in the following part-paragraph from Experiment 2, each non-technical collocation occurs only once—the repetition of the collocations is spaced throughout

the other treatment texts—and is not defined—although some meaning of the collocations may be derived from the surrounding context:

> Applying for jobs is stressful, especially when you first **enter the profession** you have chosen and you have little work experience. Getting a good job can seem **against the odds**. If you are accepted for a job interview, you may need to make arrangements to be off work from your present job for a few hours to attend the interview. During that first interview, it is best not to raise the issue of pay. [Note: I have underlined two target collocations in this text in order to identify them.]

Finally, the gapfill test in Experiment 1, like the treatment texts, contained contextual semantic clues as well as definitions in the margin, thus facilitating further semantic elaboration of the collocations. In contrast, the gapfill test text in Experiment 2, also similar to the treatment texts, contained neither of these semantic clues.

5.3 Theoretical implications of thesis

In this section, I discuss the findings of the studies in my thesis within the wider context of the research literature. I relate the findings to a number of theoretical frameworks and models, namely: the interface debate, declarative/explicit and procedural/implicit knowledge, usage-based learning models; depth-of-processing and TOPRA models of vocabulary learning; attention and noticing frameworks, and claims about the comparative learnability of lexical and grammatical collocations.

5.3.1 Measures, the interface debate and declarative and procedural knowledge

Two types of tests employed in this thesis—cued recall (gapfill) and form recognition (multiple choice)—are assumed to have measured declarative knowledge of collocations; the other two types of measures—lexical decision and self-paced reading—are assumed to have quantified procedural collocational knowledge (Schmitt, 2000). Two points regarding these types of knowledge and the measures need to be re-iterated. Firstly, the distinction between declarative knowledge and the more commonly-referenced term, explicit knowledge—along with the distinction between their opposite terms, procedural knowledge and implicit knowledge—has not been

settled in the research literature. At issue in particular is whether all procedural knowledge is implicit or whether some of it is also sped-up explicit knowledge (Suzuki & DeKeyser, 2017). Thus, it is beyond the scope of this thesis to determine whether and how implicit collocational knowledge was gained in the experiments. While faster processing times were found in Experiment 1 for the targets of the typographically-unenhanced conditions than for the targets of the control phrases, it is unclear whether the unbolded collocations were learned (solely) implicitly since it is not known whether the faster times were the result of attention without awareness or attention with some awareness. That being the case, no evidence was found to support the weak interface position, which, of the three interface positions, seems to be the best supported by neurophysiological research (Hulstijn, 2002, 2015; Paradis, 1994; 2009). The weak-interface position would hold that explicit knowledge of collocations—as promoted by bolding, without or without glossing—indirectly facilitates the development of implicit knowledge. Although I make no claims about finding evidence of implicit collocational knowledge, I do contend that the learners were better able to access procedural knowledge of the target collocations online when collocations were presented in the unenhanced input flood treatment (Hulstijn, 2015).

The second point to make about the measures in this thesis is that although I have labelled them as measures of either declarative or procedural knowledge, cued-recall and form-recognition tests (declarative-knowledge measures) and lexical decision and self-paced reading (procedural-knowledge measures) may not respectively measure these types of knowledge exclusively. Although the element of time pressure and the priming manipulations are believed to limit the ability of participants to access declarative knowledge in primed LDTs and SPRTs (DeKeyser, 2003), the possibility that participants also access declarative knowledge cannot be completely excluded (R. Ellis, 2004). Further, participants may (and often do) also use procedural knowledge in pen-and-paper tests. Nevertheless, it is safe to assume that the cued-recall and form-recognition tests could not have been completed without a major contribution of declarative knowledge, and that primed lexical decisions and self-paced reading could not have been completed without a major contribution of procedural knowledge.

5.3.2 Development of declarative and procedural collocational knowledge

The results of this thesis support the claim that knowledge of MWUs, like knowledge of individual words, is both declarative (explicit) and procedural (implicit) (Doczi & Kormos, 2016; R. Ellis, 2004; Paradis, 1994; Graf & Schacter, 1989). Declarative collocational knowledge was found in untimed pen-and-paper tests in both experiments, while the existence of procedural knowledge was found in the primed LDT in Experiment 1 and indications of some procedural knowledge in the SPRT in Experiment 2—both timed measures (Sonbul & Schmitt, 2013).

My findings also corroborate Sonbul and Schmitt's (2013) claim that declarative collocational knowledge and procedural collocational knowledge are dissociated (Graf & Schachter, 1989). One aspect of this claim, supported by my results, is that implicit (procedural) collocational knowledge seems to take longer to establish than explicit (declarative) collocational knowledge. These findings appear to conflict with those of studies using ERPs (McLaughlin, Osterhout, & Kim, 2004; Mestres-Missé, Rodriguez-Fornells, & Münte, 2007) showing that procedural lexical knowledge apparently starts developing before declarative lexical knowledge. However, the apparent contradiction may be explained by the relative insensitivity of measures such as primed LDTs and SPRTs which record RTs. These are less direct measures than ERPs, which sense electrical brain activity. It is also possible that the development of collocational knowledge proceeds differently from that of individual word knowledge, tested in the neurolinguistics studies cited above. Clarifying this is an import area of future research.

5.3.3 Input flooding and usage-based and instance-based learning models

As expected, in both my experiments input flooding (with and without typographic enhancement and glossing) of the target collocations resulted in significant amounts of declarative knowledge as measured in the immediate and delayed cued-recall (gap-fill) and form-recognition post-tests (when compared with the control-condition results). These results corroborate the findings of a number of previous experiments (e.g., Sonbul & Schmitt, 2013; Webb et al., 2013), which report significant gains in declarative collocational knowledge through input flooding of collocations in written text. Input flooding alone (i.e., without typographic enhancement) also promoted the development of some procedural collocational knowledge. The results also corroborate Sonbul and

Schmitt's (2013) suggestion that recycling is needed for the development of procedural knowledge more than for the development of declarative knowledge.

All these results can be interpreted according to usage-based perspectives of language learning, particularly with regard to the theories' principal input property for the acquisition of formulaic constructions: frequency of the construction in the input (Wulff, 2018). In my experiments, the participants encountered the collocations multiple times in diverse contexts in written input. In the process, according to usage-based theories, the learners' language systems implicitly tallied the statistics of the collocations as they compared each instance of each collocation with memories of previous encounters with the collocation (N. Ellis, 2002, 2007; Ellis et al., 2016b; see also Eysenck, 1982; A. Reber, 1976; P. Reber, 2013); as a result, the learners gained significant declarative, and some procedural, collocational knowledge. While frequency is the most important factor in usage-based approaches relevant to my experiments, another factor that is pertinent is salience. According to usage-based theories, "low salience cues" are learned more slowly than "high salience cues" (Wulff, 2018, p. 24). The results of the experiments in this thesis partly support this claim. My findings indicate that typographic enhancement of collocations enhanced their salience, producing greater declarative knowledge of form. This was especially true for the prep+N collocations in Experiment 2, combinations that were likely less salient to the participants than the lexical collocations (Durrant, 2009). However, the greater salience of typographically enhanced items did not produce faster processing times than for the unenhanced items. It seems that salience promotes awareness of items, which leads to faster gains in declarative knowledge of form but slower (if any) gains in procedural knowledge of form.

My findings related to input flooding can also be accounted for by Reichle and Perfetti's (2003) and Bolger et al.'s (2008) instance-based word-learning framework, according to which each encounter with a word—or, in this case, I would argue, a collocation—produces a new episodic memory trace. In my experiments, each encounter with a collocation in its surrounding context can be interpreted as having produced a trace of the collocation in the long-term memory; the traces were strengthened through each new encounter with the collocation. Form-meaning mapping and abstraction of the collocations' meanings from the specific contexts in which they had been seen occurred as new encounters with the collocations activated memories of

meaning features from previous encounters. Another important factor in contextual learning is quality of context (more vs. less supportive). In Experiment 1, the contexts in which target collocations were placed were more supportive than those in Experiment 2; this could explain the more robust knowledge established for the target lexical collocations in Experiment 1 than for those in Experiment 2.

The results of my research corroborate those of the few other studies which have been conducted on the learning effect of the contextual repetition of MWUs in incidental learning conditions and which have found gains in declarative knowledge of the MWUs (e.g., Durrant & Schmitt, 2010; Sonbul & Schmitt, 2013; Webb et al., 2013). A number of studies which have employed tests of procedural knowledge (e.g., Arnon & Snider, 2010; Conklin & Schmitt, 2008; Kim & Kim, 2012; Tremblay et al., 2011) have found faster response times for higher frequency MWUs than for low-frequency MWUs and non-MWUs (although these were processing not learning studies). Such results, indicating the presence of a frequency effect and memory traces of the collocations, were repeated in both my studies. Importantly, however, to my knowledge my experiments are the only studies which indicate the development of the initial stages of procedural collocational knowledge as a result of learning treatments. This discovery suggests that procedural knowledge of collocations can develop over time (Sonbul & Schmitt, 2013).

5.3.4 Glossing and depth-of-processing and TOPRA models of vocabulary learning

According to depth-of-processing and related frameworks (e.g., Anderson & Reder, 1979; Craik & Lockhart, 1972; Hulstijn & Laufer, 2001), deep processing, involving semantic elaboration, facilitates stronger retention of vocabulary items than shallow processing, as does a larger number of elaborations, or more attention or involvement from the learner. However, the evidence from Experiment 1 does not appear to support this for the learning of collocations, or, at the least, shows that the amount of semantic elaboration that is useful for retention may be limited. A single elaboration for each collocation facilitated by a contextual definition within the text, followed by eight more repetitions of each collocation, was a feature of each of the learning conditions—reading-only, bolding-only and bolding-plus-glossing—and was adequate to produce a high level of retention. However, the further elaboration induced by the glosses in the bolding-plus-glossing condition did not promote increased learning, as measured in the

post-tests. In fact, the learning condition which facilitated the most semantic elaboration of the technical collocations—bolding plus glossing—led to the development of less declarative knowledge than the reading-only and bolding conditions in the cued-recall and form-recognition tests. In the primed LDT, there was also a lack of advantage of the extended semantic elaboration facilitated by the marginal glosses: extended elaboration led to no measurable procedural knowledge (priming) whereas less elaboration (along with lack of bolding) in the reading-only condition did. One possible explanation for these results is that when the participants looked away from the texts to the glosses their processing of the collocations during reading may have been disrupted, counteracting any advantage of gloss-induced meaning elaboration on learning. Another factor that should be noted is that the declarative-knowledge and procedural-knowledge post-tests in the experiments were tests of both form and meaning; a more dedicated test of meaning, such as a meaning-generation task, may have produced a different result, with possibly higher scores for collocations seen in the bolding-plus-glossing condition than for collocations seen in the other two conditions.

Barcroft's (2002) TOPRA model provides a more adequate explanation of the gloss-related findings of Experiment 1. His argument that semantic elaboration of lexical items can decrease learning of the formal properties appears to account for what occurred in my experiment(s). The learning condition which facilitated extended semantic elaboration—bolding plus glossing—led to the development of no measurable procedural knowledge whereas the least semantically-elaborate condition, reading-only, did. It appears that the semantic elaboration induced by glossing to some extent inhibited the encoding of the formal properties of the medical collocations. It is possible that bolding in some cases promoted explicit structural elaboration, for example, as learners noticed and then explicitly considered one or more formal aspects of the collocations. However, I am unaware of any evidence that bolding of collocations routinely facilitates structural elaboration, which Barcroft (2002) seems to interpret as an extended activity involving conscious effort, such as counting the number of letters in each word. In an eye-tracking study, Choi (2017) found that bolded target collocations were gazed at 24% longer than unbolded collocations. However, it is doubtful that participants would have engaged in structural elaboration in that extra

reading time since a structurally-elaborate activity such as letter-counting or rehearsing spelling would take much longer.[26]

5.3.5 Typographic enhancement, attention and types of knowledge

It is commonly argued that more attention to—or noticing of—particular linguistic items leads to better learning of those items (e.g., Schmidt, 1990). Learners often fail to notice MWUs because the constituent words are usually frequent and because many of the MWUs are common and lack novelty (Bishop, 2004; Boers et al., 2014). As a result, techniques which promote noticing of MWUs such as typographic enhancement have been found to lead to gains in knowledge (Boers et al., 2016; Peters, 2012; Szudarski & Carter, 2014). However, as my results indicate, the picture appears to be more complicated than that. The *type* of attention given to the target collocations in my two experiments seemed to determine the type of knowledge gained. Although I did not set out to specifically test for it, repetition of target collocations in the input, with and without typographic enhancement, facilitated the development of declarative knowledge of the items, as measured in the cued-recall and form-recognition tests—a finding which previous research has found to be robust (e.g., Sonbul & Schmitt, 2013). In both experiments, typographic enhancement led to a larger amount of declarative knowledge than no typographic enhancement. On the other hand, input flooding *without* typographic enhancement in general led to faster processing times and/or the development of more procedural knowledge than input flooding with typographic enhancement. These findings supported Sonbul and Schmitt's (2013) that procedural knowledge is more difficult to enhance than declarative knowledge. A possible explanation for the difference in the types of knowledge gained in this thesis is that the attention paid to the repeated occurrences of the typographically-unenhanced collocations did not reach the threshold of noticing—in other words, it was attention without awareness of learning (Robinson, 1995; Tomlin & Villa, 1994), leading to the collocations being internalised mostly as procedural knowledge. In contrast, the attention paid to repeated typographically-enhanced collocations did reach the threshold of noticing (awareness), leading to their internalisation mostly as declarative

[26] Take one of Choi's (2017) target collocations, *moderate exercise,* as an example. Timing myself, I read the collocation in less than a second, but counting the number of letters in it took me about 3½ seconds—about four times longer than it took to read the collocation. Rehearsing the spelling of the collocation took even longer: about 4½ seconds.

knowledge. It is likely that processing times were slower and that less procedural knowledge was developed in the bolding condition than in the no-bolding (reading-only) condition because the bolding treatment interrupted fluent reading somewhat by drawing extra attention to the forms of the collocations (see Choi, 2017). While this action promoted declarative noticing of the items, which resulted in an advantage in the declarative-knowledge tests, the focus on the form of the items reduced the efficacy of their acquisition as procedural knowledge (Paradis, 1994, p. 395).

The findings in both experiments appear to support Tomlin and Villa's (1994) concept of the dissociation of detection and awareness; in other words, language-learning can take place without awareness. Although I did not directly investigate the detection and awareness of collocations, I would argue that such dissociation can justifiably be inferred from the results of my experiments. Both input flooding with typographic enhancement (plus or minus glossing) and without typographic enhancement likely promoted detection of the collocations. However, typographic enhancement promoted greater (conscious) awareness of the target items, as attested to by the higher scores for the bolded items in the declarative-knowledge tests. These findings are also consistent with those of Godfroid et al.'s (2013) and Choi's (2017) eye-tracking studies, in which the greater the amount of attention that was focused on lexical items, the higher the scores on tests of declarative knowledge; further, attention to—and awareness of—collocations in Choi's study was increased through typographic enhancement. The converse effect is true for the absence of typographic enhancement: in both my experiments, lack of bolding promoted detection with less collocational awareness, as evidenced in the lower declarative-knowledge test scores for collocations seen in the no-typographic-enhancement conditions compared with the collocations seen in the typographic-enhancement conditions. In general, in the measures of procedural knowledge, which allow little participant awareness of the experimental stimuli (McDonough & Trofimovich, 2008; Marsden, Thompson, & Plonsky, 2018), RTs for typographically-unenhanced collocations were faster compared to the control phrases; however, this was not the case for the typographically-enhanced collocations. In summary, I propose that some encoding of the collocations in long-term memory occurred without awareness, although it is unlikely that no encoding occurred without learner attention to the collocations (Gass, Svetics, & Lemelin, 2003; Robinson, 1995, 2003; Schmidt, 1995; Tomlin & Villa, 1994).

5.3.6 Development of lexical and grammatical collocational knowledge

Both verbs and prepositions are difficult to learn since they are both polysemous and have figurative meanings (e.g., Crossley et al., 2013; Tyler et al., 2011), and their combinations with nouns cause problems for L2 learners (e.g., Gitsaki, 1999; Huo, 2014; Hemchua & Schmitt, 2006). Nevertheless, the findings of this thesis with respect to the learnability of lexical and grammatical collocations overall support the claims of researchers such as Howarth (1998b) and Ackerman and Chen (2013) that lexical collocations are more difficult than grammatical collocations to master and the claim that grammatical collocations are more easily internalised because they are more predictable (Ackerman & Chen, 2013). My findings also partially corroborate Gitsaki's finding of lower scores in a gap-fill task for V+N collocations encountered previously than for prep+N collocations encountered previously. In Experiment 2 of this thesis, a larger amount of declarative knowledge of prep+N collocations than of V+N collocations was gained when the combinations were made more perceptually salient through typographic enhancement. However, no difference in gains for the two phrase types was recorded in the no-bolding condition, meaning that in that condition they were equally difficult to learn. Since, to my knowledge, no other studies have compared the learning of typographically enhanced and unenhanced L2 lexical and grammatical collocations, the reasons for this lack of difference in the reading-only condition are unclear. Also supporting the claim that grammatical collocations are easier to master than lexical collocations is the clear processing advantage in the SPRT for the terminal word in the unenhanced prep+N phrases over the unenhanced V+N phrases, meaning that the prep+N collocations were recognised more quickly. Even though there was no significant difference between the RTs of the bolded prep+N and the bolded V+N collocations, it does seem that a general pattern of faster times for the prep+N phrases existed (see Figure 5). Again, the reason is not clear, although one possible reason is the greater frequency in natural language of the prepositions than the nouns and the prep+N collocations than the V+N collocations. Alternatively, it is also possible that the RTs on the terminal word in the prep+N collocations were faster than in the V+N collocations because the preceding word in the prep+N collocations was shorter (had fewer letters) than in the V+N collocations.

5.4 Recommendations for future research

Although my experiments built on Sonbul and Schmitt's (2013) study and produced some results of interest, the findings related to procedural knowledge should be seen as initial findings only: a great deal of further work needs to be done in this area to confirm the link between multiple repetitions in written input and the development of procedural collocational knowledge in incidental learning conditions. Future research could flood texts with more repetitions of typographically enhanced and unenhanced MWUs in a larger number of experimental reading sessions over a longer period of time to determine whether that could lead to the development of stronger procedural knowledge. Repetitions of single words in graded readers or other simplified materials have been found to produce gains in declarative receptive knowledge of the items (Waring & Takaki, 2003), enhance their retention (Horst, Cobb, & Meara, 1998) and increase their processing speeds (Elgort et al., 2017; Pellicer-Sánchez, 2015). However, further research is needed into the effectiveness of the insertion of repeated occurrences of MWUs into longer texts, with and without typographic enhancement, for the purpose of producing both declarative and procedural knowledge of the MWUs. It is possible that if participants were exposed to a larger number of repetitions of the collocations in my experiments the initial evidence of procedural knowledge resulting from 6-9 exposures, particularly in the reading-only/no-bolding condition, would produce a stronger result. It is also possible that, with participant exposure to more repetitions, evidence of the presence of procedural knowledge would extend to the collocations seen in the bolding and bolding-plus-glossing conditions.

For future experiments in this area, changing the experimental conditions might produce informative results. In my experiments, the treatment conditions were "all-or-nothing" treatments: in the no-bolding/reading-only conditions, all the collocations were unbolded in the context; in the bolding conditions, all the collocations were bolded; and in the bolding-plus-glossing condition of Experiment 1, all the collocations were both bolded and glossed. It would be interesting to blend one or two conditions to more accurately reflect real-life learning situations. In particular, target collocations in one condition could be typographically enhanced on the first occurrence and be unenhanced for the remainder of the occurrences. This procedure would imitate the approach frequently used in university textbooks to introduce technical terms (e.g., Field, 2013). In one study which employed this technique, Martinez-Fernandez (2008)

inserted repetitions of each target word in her treatment texts but bolded, glossed or deleted each target word (according to the learning condition) only once. This method is also consistent with N. Ellis's (2007) claim that focused attention may only need to be applied once to a non-salient linguistic item. After the item has been noticed the first time, "[t]he cue does not have to be repeatedly noticed thereafter; once consolidated, mere use in processing for meaning is enough for implicit tallying" (p. 30). Similarly, an item in a bolding-plus-glossing condition could be bolded and glossed on the first instance only, with the expectation that the process of implicit tallying will take place on the second and subsequent occurrences of the item in the text. Such an experimental approach, in which MWUs are repeated in the input and in which the first occurrence of the MWU is typographically enhanced while the remaining occurrences are unenhanced, might facilitate a more evenly balanced development of declarative and procedural knowledge of the target items.

Another suggestion for future research is to compare non-technical lexical collocations in two conditions in which the collocations are spaced differently within the texts—the massed repetition technique used in Experiment 1 and the spaced repetition technique used in Experiment 2. If texts are to be manipulated by materials designers to include repetitions of target MWUs, it would be useful to know which technique of repetition, massed or spaced, is more useful for learning.

More research is also needed into the role that glossing—and the explicit attention that it brings to the target items—can play in the retention of collocations in written texts. Very little research has explored this area. However, future studies should avoid the overuse of definitions in incidental conditions. Based on the results of Experiment 1, I would recommend the placement of either one contextual definition or one gloss in a treatment text for each target MWU but not both.

5.5 Pedagogical implications

The overall pedagogical implication of my two studies is that second-language learners need to be exposed to repeated occurrences of collocations in order to facilitate the development of both declarative and procedural knowledge of the collocations. Learners' encounters with collocations and other MWUs in natural contexts are generally infrequent, meaning that much of the knowledge gained in the first encounter may be forgotten by the time of the second encounter, leaving little to build on.

Specifically, I recommend the manipulation of longer written texts, such as textbooks, graded readers and electronic readers (Eldridge & Neufeld, 2009), to include multiple instances of MWUs, including collocations. It is true that manipulating texts to include multiple repetitions of vocabulary items is labour-intensive and time-consuming, requiring the materials developer to be resourceful (Boers et al., 2016). As such, creating such materials may not often be feasible for teachers, but it may be for course designers, and for publishers, who already insert multiple instances of single words in graded readers.

Whether to typographically enhance target MWUs is an issue that needs to be carefully considered by materials designers as well as to be the subject of further research. Materials designers intending to promote retention of certain MWUs need to decide between typographically enhancing MWUs, leaving the MWUs unenhanced, and some combination of both. As shown in my experiments, repeated typographic enhancement of collocations appears to have an advantage over the absence of typographic enhancement for the development of declarative knowledge, while the converse seems to be true for procedural knowledge.

Going beyond the manipulation of texts, creative, considered approaches to the use of the texts containing useful collocations may need to be applied by teachers. Choi (2017), who claims that there is a trade-off between learning enhanced collocations and recalling unenhanced text, suggests caution in the use of typographic enhancement when the main focus is on understanding and remembering the text.

5.6 Concluding remarks

The purpose of this thesis has been to investigate the development of declarative and procedural knowledge of second-language collocations through repeated exposure to the collocations and through different learning techniques. In my experiments, I have explored a gap in the literature, namely, the question of whether typographic enhancement (with or without glossing) or no typographic enhancement is more effective at promoting the development of declarative and procedural knowledge of lexical and grammatical collocations in incidental reading.

The thesis contributes to second-language vocabulary-acquisition research with its finding that different incidental learning techniques do appear to have different effects on the gaining of different types of knowledge of collocations. In particular,

repeated exposure combined with typographic enhancement is more effective at promoting declarative collocational knowledge of both lexical and grammatical collocations than repeated exposure without typographic enhancement. Conversely, repeated exposure with no typographic enhancement is in general more effective at promoting faster processing and the establishment of procedural collocational knowledge. The findings of these experiments have also corroborated previous claims that the development of declarative and procedural knowledge of MWUs are dissociated.

References

Abraham, L. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*(3), 199-226. doi:10.1080/09588220802090246

Ackermann, K., & Chen, Y-H. (2013). Developing the Academic Collocation List (ACL)—A corpus-driven and expert-judged approach. *Journal for Academic Purposes*, *12*, 235-247. doi:10.1016/j/jeap.2013.08.002

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken: New Jersey: John Wiley & Sons, Inc.

Ahn, J. (2014). Attention, awareness, and noting in SLA: A methodological review. *MSU Working Papers in SLS*, Vol. 5.

Alali, F., & Schmitt, N. (2012). Teaching formulaic sequences: The same as or different from teaching single words? *TESOL Journal*, *3*(2). doi:10.1002/tesj.13

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, *22*(2), 173-195. doi:10.1093/applin/22.2.173

Anderson, J. (1983). *The architecture of cognition.* Cambridge: Massachusetts: Harvard University Press.

Anderson, J. (1995). *Learning and memory: An integrated approach.* New York: John Wiley and Sons, Inc.

Anderson, J., & Reder, L. (1979). An elaborative processing explanation of depth of processing. In Cermak, L., & Craik, F. (Eds.). *Levels of processing in human memory.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*, 67-82. doi:10.1016/ j.jml.2009.09.05

Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*, Cambridge: Cambridge University Press.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. doi:10.1016/j.jml.2007.005

Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, *47*(1), 56-63. doi: 10.1093/elt/47.1.56

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, *21*(1), 101-114. doi:10.1016/0346-251X(93)90010-E

Balling, L. (2008). A brief introduction to regression designs and mixed-effects modelling by a recent convert. *Copenhagen Studies in Language*, *36*, 175-192.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning, 52*(2), 323-363. doi:10.1111/0023-8333.00186

Barcroft, J. (2003). Distinctiveness and bidirectional effects in input enhancement for vocabulary learning. *Applied Language Learning*, *13*(2), 47-73.

Barfield, A., & Gyllstad, H. (2009). Introduction: Researching L2 collocation knowledge and development. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: multiple interpretations* (pp. 1-18). Basingstoke, UK: Palgrave Macmillan.

Beck, I., McKeown, M., & McCaslin, E. (1983). Vocabulary development: All contexts are not created equal. *The Elementary School Journal*, *83*, 177-181. doi:10.1086/461307

Benson, M., Benson, E., & Ilson, R. (1986). *The BBI dictionary of English word combinations* (first edition). Amsterdam: John Benjamins Publishing Company.

Benson, M., Benson, E., & Ilson, R. (1997). *The BBI dictionary of English word combinations* (revised edition). Amsterdam: John Benjamins Publishing Company.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3), 275-311. doi:10.1075/ijcl.14.3.08bib

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*(3), 263-286. doi:10.1016/j.esp.2006.08.003

Biber, D., Johansson, S., Leech, G., Conrad, S., & Leech, G. (1999). *Longman grammar of spoken and written English.* Edinburgh Gate, Harlow: Pearson Education Limited.

Bishop, H. (2004). Noticing formulaic sequences—a problem of measuring the subjective. *LSO Working Papers in Linguistics, 4*, 15-19. Madison, Wisconsin: University of Wisconsin-Madison Department of Linguistics.

Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review*, *23*(3), 291-320. doi.org/10.1515/TLR.2006.012

Boers, F., & Demecheleer, M. (1998). A cognitive approach to teaching prepositions. *ELT Journal*, *52*(3), 197-204. doi: 10.1093/elt/52.3.197

Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2016). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, 26(2), 448-469. doi:10.1111/ijal.1214

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a Lexical Approach to the test. *Language Teaching Research, 10*(3), 1-17. doi:10.1191/1362168806lr195oa

Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition.* New York: Palgrave Macmillan.

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014). Some explanations for the slow acquisition of L2 collocations. *Vigo International Journal of Applied Linguistics, 11*, 41-62. doi:10.1080/09658416.2013.774008

Bolger, D., Balass, M., Landen, E., & Perfetti, C. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes*, *45*, 122-159. doi:10.1080/01638530701792826

Bonk, W. (2001). Testing ESL learners' knowledge of collocations. In T. Hudson & J. Brown (Eds.), *A focus on language test development: Expanding the language proficiency construct across a variety of tests* (Technical Report #21, pp. 113-142). Honolulu: University of Hawai'i, Second language Teaching and Curriculum Center.

Bramki, D., & Williams, R. (1984). Lexical familiarization in economics text, and its pedagogic implications in reading comprehension. *Reading in a Foreign Language*, *2*(1), 169-181.

Brisard, F., Frisson, S., & Sandra, D. (2001). Processing unfamiliar metaphors in a self-paced reading task. *Metaphor and Symbol*, *16*(102), 87-108. doi:10.1080/10926488.2001.9678888

Brothers, T., Swaab, T., & Taxler, M. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203-216. doi:10.1016/j.jml.2016.10.002

Brown, D. (2014). Knowledge of collocations. In J. Milton & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (pp. 123-139). London: Palgrave Macmillan.

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

Bruton, A., Lopez, M., & Mesa, R. (2011). Incidental L2 vocabulary learning: An impracticable term. *TESOL Quarterly*, *45*(4), 759-768. doi:10.5054/tq.2011.268061

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1). Doi: 10.5334/joc.10

Byrd, P., & Coxhead, A. (2010). *On the other hand*: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, *5*, 31-64.

Carnine, D., Kameenui, E., & Coyle, G. (1984). Utilization of contextual information in determining the meaning of unfamiliar words. *Reading Research Quarterly*, *19*(2), pp. 188-204. doi:10.2307/747362

Carpenter, S., & DeLosh, E. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory and Cognition*, *34*(2), 268-276. doi:10.3758/BF03193409

Carpenter, S., Sachs, R., Martin, B., Schmidt, K., & Looft, R. (2012). Learning new vocabulary in German: The effects of inferring word meanings, type of feedback and time of test. *Psychon Bull Rev*, *19*, 81-86. doi:10.3758/s13423-011-1085-7

Carrol, G., Conklin, K., & Gyllstad, H. (2016). The influence of the L1 on the reading of idioms in a L2. *Studies in Second Language Acquisition*, *38*, 403-443. doi:10.1017/S0272263115000492

Chan, T., & Liou, H. (2005). Effects of web-based concordancing instruction of EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning, 18*(3), 231-251. doi:10.1080/09588220500185769

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med*, *19*(22): 3127-3131. doi:10.1002/1097-0258(20001130)

Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing*, *27*(3), 419-436. doi:10.1177/0265532210364391

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, *21*(3), 403-426. doi: 10.1177/1362168816653271

Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, *22*(2), 115-144. doi: 10.1191/0267658306sr263oa

Cintrón-Valentín, M., & Ellis, N. (2015). Exploring the interface: Explicit focus-on-form instruction and learned attentional biases in L2 Latin. *Studies in Second Language Acquisition*, *37*, 197-235. doi:1017/S0272263115000029

Clark, M., & Randal, J. (2004). *A first course in applied statistics—with applications in biology, business, and the social sciences.* Auckland: Pearson Education New Zealand.

Cobb, T. (n.d.). Compleat Lexical Tutor. Available at https://www.lextutor.ca

Cosme, C., & Gilquin, G. (2008). Free and bound prepositions in a contrastive perspective: The case of *with* and *avec*. In S. Granger & F. Meunier (Eds.), *Phraseology: an interdisciplinary perspective* (pp. 259-272). Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Cowie, A. (1998). Introduction. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and Applications* (pp. 1-20). Oxford: Clarendon Press.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 672-684. doi:10.1016/ S0022-5371(72)80001-X

Crossley, S., Subtirelu, N., & Salsbury, T. (2013). Frequency effects or context effects in second language word learning: What predicts early lexical production? *Studies in Second Language Acquisition*, *35*(4), 727-755. doi:10.1017/ S0272263113000375

Cruse, D. (1986). *Lexical semantics.* Cambridge: Cambridge University Press.

Cumming, G. (2014). The new statistics: why and how. *Psychological Science*, *25*(1), 7-29. doi: 10.1177/0956797613504666

Dahlmeier, D., Ng, H., & Wu, S. (2013). Building a learner annotated corpus of learner English: the NUS Corpus of Learner English. *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational applications* (pp. 22-31). Atlanta, Georgia, USA: The Association for Computational Linguistics.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*, 1-18. doi: 10.1016/0749-596X(86)90018-5

Davies, M. (2004-) *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Available online at https://corpus.byu.edu/bnc/.

Davies, M. (2008-) *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Available online at https://corpus.byu.edu/coca/.

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348), Malden, MA: Blackwell Publishing.

Doczi, B., & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization.* Oxford: Oxford University Press.

Dunnings, I. (2012). An overview of mixed-effects statistics models for second language researchers. *Second Language Research*, *28*(3), 369-382. doi: 10.1177/027658312443651

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes, 28*(3), 157-169. doi:10.1016/j.esp.2009.02.002

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL, 47*, 157-177. doi:10.1515/iral.2009.007

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*, 163-188. doi:10.1177/0267658309349431

Eldridge, J., & Neufeld, S. (2009). The graded reader is dead, long live the electronic reader. *The Reading Matrix*, *9*(20), 224-244.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367-413. doi:10.111/j.1467-9922.2010.00613.x

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2017). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, 2017, 1-26. doi:10.1017/S0272263117000109

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit knowledge and the role of learner and item variables. *Language Learning, 64*(2), 365-414. doi:10.111/lang.12052

Ellis, N. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In Ellis, N. (Ed.), *Implicit and explicit learning of languages.* London: Academic Press.

Ellis, N. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition, *SSLA, 24*, 143-188. doi:10.1017/S0272263102002024

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge, *SSLA*, *27*, 305-352. doi:10.1017/S027226310505014X

Ellis, N. (2007). The weak interface, consciousness, and form-focus instruction: Mind the doors. In S. Fotos & H. Nassaji (Eds.), Form-focused instruction and teacher education: Studies in honour of Rod Ellis (pp. 17-34). Oxford: Oxford University Press.

Ellis, N. (2011). Implicit and explicit SLA and their interface. In C. Sanz and R. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism.* Washington, DC: Georgetown University Press.

Ellis. N. (2015). Implicit AND explicit language learning: Their dynamic interface and complexity. In Rebuschat, P. (Ed.), *Implicit and explicit learning of languages* (pp. 3-23). Amsterdam: John Benjamins.

Ellis, N., Frey, E., & Jalkanen, I. (2009). The psycholinguistic reality of collocation and semantic prosody. In U. Römer & R. Schulze (Eds.), *Exploring the lexis-grammar interface*, pp. 89-114. Amsterdam: John Benjamins.

Ellis, N., Römer, U., & O'Donnell, M. (2016a). Construction and usage-based approaches to language acquisition. *Language Learning*, *66*(S1), 23-44. doi: 10.1111/ lang.1_12177

Ellis, N., Römer, U., & O'Donnell, M. (2016b). Determinants of construction learning. *Language Learning*, *66*(S1), 45-68. doi:10.1111/lang.1_12177

Ellis, R. (1993). The structural syllabus and second language acquisition. *TESOL Quarterly*, *27*(1), 91-113. doi:10.2307/3586953

Ellis, R. (1999). Factors in the incidental acquisition of second language vocabulary from oral input. In R. Ellis, *Learning a second language through interaction* (pp. 35-61). Amsterdam: John Benjamins.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition, 24*, 223-236. doi:10.1017/ S0272263102002073

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*(2), 227-275. doi:10.1111/j.1467-9922.2004.00255.x

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *SSLA*, *27*, 141-172. doi:10.1017/s0272263105050096.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, *40*(1), 83-107.

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In Ellis., R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H., *Implicit and explicit knowledge in second language learning and teaching*. (pp. 3-26). Bristol, U.K.: Multilingual Matters.

Ercetin, G., & Alptekin, C. (2013). The explicit/implicit knowledge distinction and working memory: Implications for second-language comprehension. *Applied Psycholinguistics*, *34*, 727-753. doi:10.1017/S0142716411000932

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, *20*(1), 29-62. doi:10.1515/text.1.2000.20.1.29

Eysenck, M. (1982). Incidental learning and orienting tasks. In C. Puff (Ed.), *Handbook of research methods in human memory and cognition.* New York: Academic Press.

Farghal, M., & Obeidat, H. (1995). *Collocations: A neglected variable in EFL. IRAL*, *34*(4), 315-331. doi:10.1515/iral.1995.33.4.315

Field, A. (2013). *Discovering statistics using IBM SPSS statistics (4th Edition)*. London: Sage.

Fine, A., & Jaeger, T. (2016). The role of verb repetition in cumulative structural priming in comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1362-1376. doi: 10.1037/xlm0000236

Firth, J. (1962). A synopsis of linguistic theory 1930-1955. In Philological Society (Ed.), *Studies in linguistic analysis.* Oxford: Blackwell.

Folse, K. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, *40*(2), 273-293. doi:10.2307/40264523

Foster, P. (2001). Rules and routines: A consideration of their roles in the task-based language production of native and non-native speakers. In Bygate, M., Skehan, P., & Swain, M., *Researching pedagogic tasks: Second language learning, teaching and testing.* Edinburgh Gate, Harlow, UK: Pearson Education Limited.

Fraundorf, S., Watson, D., & Benjamin, A. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *J Mem Lang*, *63*(3), 367-386. doi.org/10.1016/j.jml.2010.06.004

Gass, S., Svetics, I., & Lemelin, S. (2003). Differential effects of attention. *Language Learning*, *53*(3), 497-545. doi:10.1111/1467-9922.00233

Gablasova, D., Brezina, V., McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(51), 1-25. doi:10.1111/lang.12225

Gitsaki, C. (1999). *Second language lexical acquisition: A study of the development of collocational knowledge.* San Francisco: International Scholars Publications.

Gledhill, C. (2000). *Collocations in science writing*. Language in Performance Series No. 22. Tubingen: Gunter Narr Verlag.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*, 483-517. doi:10.1017/ S0272263113000119

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice.* Cambridge: Cambridge University Press.

Graf, P., & Schacter, D. (1989). Unitization and grouping mediate dissociations in memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 930-940. doi:10.1037/0278-7393.15.5.930

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (ed.), *Phraseology: theory, analysis, and applications* (pp. 145-160). Oxford: Clarendon Press.

Grant, L., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics*, *25*(1), 38-61. doi:10.1093/applin/25.1.38

Gui, S., & and Yang, H. (2003). Chinese Learners English Corpus. Shanghai: Shanghai Foreign Language Education Press.

Gyllstad, H. (2014). Grammatical collocation. In C. Chappelle (Ed.). *The encyclopaedia of applied linguistics*. Oxford: Wiley, Blackwell.

Gyllstad, H., & Wolter, B. (2016). Collocational processing and semantic transparency in light of the phraseological continuum: Does semantic transparency matter? *Language Learning*, *66*(2), 296-323. doi:10.111/lang.12143

Haberlandt, K. (1996). Methods in reading research. In Gernsbacher, M, (Ed.), *Handbook of psycholinguistics* (pp. 1-32). San Diego: Academic Press.

Han, Z., Park, E., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, *29*(4): 597-618. doi:10.1093/applin/amn010

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*(2), 237-260. doi:10.1111/j.1473-4192.1994.tb00065.x

Hemchua, S., & Schmitt, N. (2006). An analysis of lexical errors in the English compositions of Thai learners. *Prospect, 21*(3), 3-25.

Henriksen, B. (2012). Research on L2 learners' collocational competence and development – A progress report. In Bardel, C., Lindqvist, C., & Laufer, B. *EUROSLA Monographs Series 2, L2 vocabulary acquisition, knowledge and use* (pp. 29-56).

Hernandez, M., Costa, a., & Arnon, I. (2016). More than words: Multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience*, *31*(6), 785-800. doi: 0.1080/23273798.2016.1152389

Hill, J. (1999). Collocational competence. *English Teaching Professional*, *11*, 3-7.

Hoey, M. (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.

Hoey, M. (2005). *Lexical priming: a new theory of words and language*. London & NY: Routledge.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207-223. doi:10.4236/ojml.2014.45051

Howarth, P. (1998a). Phraseology and second language proficiency. *Applied Linguistics, 19*(10), 24-44. doi:10.1093/applin/19.1.24

Howarth, P. (1998b). The phraseology of learners' academic writing. In A. Cowie (Ed.), *Phraseology: theory, analysis and applications* (pp. 161-186). Oxford: Oxford University Press.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430.

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research, 18*, 193–223. doi:10.1191/0267658302sr207oa

Hulstijn, J. (2003). Incidental and intentional learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). Malden, MA: Blackwell Publishing.

Hulstijn, J. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 25-46). Amsterdam: John Benjamins Publishing Company.

Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal, 80*(3), 327-339. doi:10.1111/j.1540-4781.1996

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hunston, S., & Francis, G. (1999). *Pattern grammar: A corpus-driven approach to the lexical grammar of English.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Huo, Y. (2014). Analyzing collocation errors in EFL Chinese learners' writings based on corpus. *Higher Education of Social Science, 7*(1), 87-91. doi:10.3968/5182

IBM Corp. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Jacobs, G., Dufon, P., & Fong, C. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, *17*(1) (1994), 19-28. doi:10.1111/j.1467-9817.1994.tb00049.x

Jafarpour, A., & Koosha, M. (2006). Data-driven learning and teaching collocation of prepositions: the case of Iranian EFL students. *Research on Foreign Languages*, Journal of Faulty of Letters and Humanities, Year 49, *200*, 1-30.

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.). *Research methods in second language psycholinguistics* (pp. 20-49). New York: Routledge.

Jiang, N., & Nekrasova, M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal, 91*(3), 433-445. doi:10.1111/j.1540-4781.2007.00589.x

Jimenez Catalan, R. (1996). Frequency and variability in errors in the use of English prepositions. *Miscelanea: A Journal of English and American Studies, 17*, 171-187.

Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. (1995). Does textual enhancement promote noticing? A think-aloud protocol analysis. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 183-213). Honolulu: Second Language Teaching and Curriculum Center, University of Hawai'i at Manoa.

Juffs, A. (2001). Psycholinguistically oriented second language research. *Annual Review of Applied Linguistics*, *21*, 207-220. doi:10.1017/S0267190501000125

Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: Subject and object asymmetries in wh-extraction. *SSLA*, *17*, 483-516. doi: 10.1017/S027226310001442X

Just, M., Carpenter, P., & Woolley, J. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228-238. doi:10.1037/0096-3445.111

Karpicke, J., & Roediger III, H. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33*(4), 704-719. doi:10.1037/0278-7393.33.4.704

Kasahara, K. (2011). The effect of known-and-unknown word combinations on intentional vocabulary learning. *System*, *39*, 491-499. doi: 10.1016/j.system.2011.10.001

Keating, G., & Jegerski, J. (2015). Experimental design in sentence processing research. *Studies in Second Language Acquisition*, *37*, 1-32. doi:10.1017/S0272263114000187

Kember, H., Choi, J., & Cutler, A. (2016), Processing advantages for focused words in Korean, *Speech Prosody*, 31 May-3 Jun 2016, 702-705. doi: 10.21437/SpeechProsody

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627-633. doi:10.3758/BRM.42.3.627

Kim, S., & Kim, J. (2012). Frequency effects in L2 multiword unit processing: Evidence from self-paced reading. *TESOL Quarterly*, *46*(4), 831-841. doi: 10.1002/tesq.66

Kim, Y. (2006). Effects of input elaboration on vocabulary acquisition through reading by Korean learners of English as a foreign language. *TESOL Quarterly*, *40*(2), 341-373. doi:10.2307/40264526

Klotz, M. (2003). Review of *Oxford Collocations Dictionary for Students of English*. *International Journal of Lexicography*, 16(1), 57-61.

Ko, M. (2005). Glosses, comprehension, and strategy use. *Reading in a Foreign Language*, *17*(2).

Krashen, S. (1981). *Second language acquisition and second language learning.* Oxford: Pergamon Press.

Krashen, S. (1982). *Principles and practice in second language acquisition.* New York: Pergamon Institute of English.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, *73*(4), 440-464. doi:10.1111/j.1540-4781.1989.tb05325.x

Kroll, N., & Kellicutt, M. (1972). Short-term recall as a function of covert rehearsal and of intervening task. *Journal of Verbal Learning and Verbal Behaviour*, *11*, 196-204. doi:10.1016/S0022-5371(72)80077-X

Laufer, B. (2005). Focus on form in second language vocabulary learning. In S. Foster-Cohen, M. del Pilar García Mayo and J. Cenoz (Eds.). *EUROSLA Yearbook*, *5* (pp. 223-250).

Laufer, B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, *24*(1), 29-49. doi:10.1093/ijl/ecq039

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399-436. doi:10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1-26. doi:10.1093/applin/22.1.1

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 35-61.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647-672. doi:10.111/j.1467-9922.2010.00621.x

Lea, D. (2007). Making a collocations dictionary. *ZAA*, *55*(3), 261-271. doi:10.1515/zaa.2007.55.3.261

Lee, E.-K., & Fraundorf, S. (2019). Native-like processing of prominence cues in L2 written discourse comprehension: Evidence from font emphasis. *Applied Psycholinguistics*, *40*(20), 373-398. doi:10.1017/S0142716418000619

Leech, G., Rayson, P., Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus.* Harlow: Longman.

Lehecka, T. (2015). Collocation and colligation. In Ostman, J.-O., & Verschueren, J. (Eds.), *Handbook of Pragmatics.* Amsterdam: John Benjamins Publishing Company.

Leo, R., Egi, T., Maria Nuevo, A., & Tsa, Y-C. (2003). The roles of textual enhancement and type of linguistic item in adult L2 learners' comprehension and intake. *Applied Language Learning*, *13*(2), 1-16.

Leow, R. (2013). Attention, noticing, and awareness in second language acquisition. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*, 1-7. doi:10.1002/9781405198431.wbeal0058.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, *18,* 85-102. doi: 10.106/j.jslw.2009.02.001

Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. In D. Wood (Ed.), *Perspectives on formulaic language: acquisition and communication* (pp. 22-46). New York: Continuum.

Lindsay, S., & Gaskell, M. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning*, *60*(s2), 45-63. doi:10.1111/j.1467-9922.2010.00600.x

Lindstromberg, S. (1998). *English prepositions explained.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Lindstromberg, S. (2001). Preposition entries in UK monolingual learners' dictionaries: Problems and possible solutions. *Applied Linguistics*, *22*(1), 79-103.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam: John Benjamins. doi:10.1074/sibil.2.07lon

Lowie, W., & Verspoor, M. (2004). Input versus transfer? The role of frequency and similarity in the acquisition of L2 prepositions. In M. Achard & S. Niemeier, (Eds.), *Cognitive linguistics, second language acquisition, and foreign language teaching* (pp. 77-94). Mouton de Gruyter: Berlin.

McDonough, K., & Trofimovich, P. (2008). Using priming methods in second language research. New York: Routledge.

McIntosh. C., Francis, B., & Poole, R. (Eds.) (2009). *Oxford collocations dictionary for students of English* (2nd Edition, 2009). Oxford: Oxford University Press.

McLaughlin, J., Osterhaut, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703-704. doi:10.1038/nn1264

McNamara, T. (2005). *Semantic priming: Perspectives from memory and word recognition.* Hove, East Sussex, UK: Psychology Press.

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In S. Unsworth & E. Blom (Eds.). *Experimental methods in language acquisition research* (pp. 139-162). Amsterdam: John Benjamins.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language processing. *Studies in Second Language Acquisition*, 27(1). doi:10.1017/S0272263105050035

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-Paced reading in second language research. *Applied Psycholinguistics*, *39*(5), 861-904. doi:10.1017/S0142716418000036

Martinez-Fernandez, A. (2008). Revisiting the Involvement Load Hypothesis: Awareness, type of task and type of item. In M. Bowles et al., *Selected Proceedings of the 2007 Second Language Research Forum*, pp. 210-228. Somerville, MA: Cascadilla Proceedings Project.

Martinez, R., & Murphy, V. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, *45*(2), 267-290. doi: 10.5054/tq.2011.247708

Matuschek, H., Kliegel, R., Vasisth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305-315. doi: 10.1016/j.jml.2017.01.001

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, *4*(2), 142-154. doi: 10.1177/026553228700400202

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*, 1856-1866. doi:10.1093/cercor/bhl094

Meyer, D., & Schvaneveldt, R. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227-234. doi:10.1037/h0031564

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, *32*(2), 129-148. doi:10.1093/applin/amq035

Millar, N. (2016). 'Lexical teddy bears' and the processability of learner language by nature speakers. *Studies in Foreign Language Education*, *38*, 17-28.

Mueller, C. (2011). English learners' knowledge of prepositions: Collocational knowledge or knowledge based on meaning? *System*, *39*, 480-490. doi:10.1016/j.system.2011.10.012

Nation, I. (2000). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*(1), 59-82. doi:10.3138/cmlr.63.1.59

Nation, I. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*(1), 59-82. doi:10.1353/cml.2006.0049

Nation, I. (2013). *Learning vocabulary in another language* (Second Edition). Cambridge: Cambridge University Press.

Nation, I., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9-13.

Nation, I., & Webb, S. (2011). *Researching and analysing vocabulary.* Boston: Heinle Cengage Learning.

Nation, P. (2012). The BNC/COCA word family lists. Retrieved from https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf.

National Heart, Lung, and Blood Institute (NIH) (n.d.). Heart attack. Available at https://www.nhlbi.nih.gov/health-topics/heart-attack

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*(2), 223-242. doi:10.1093/applin/24.2.223

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130-149. doi.org/10.1017/S0267190512000098

Paribakht, T. & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: a rationale for pedagogy.* Cambridge: Cambridge University Press.

Paribakht, T. & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, *21*, 195-224.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393-419). London: Academic Press.

Paradis, M. (2009). *Declarative and procedural determinants of second languages.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Pawley, A., & Sider, F. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191-226). London/New York: Longman.

Pellicer-Sánchez, A. (2015). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*, 97-130. doi:10.1017/S0272263115000224

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, *21*(3), 381-402. doi:10.1177/1362168815618428

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do *Things Fall Apart*? *Reading in a Foreign Language*, *22*(1), 31-55.

Peters, E. (2012). Learning German formulaic sequences: The effect of two attention-drawing techniques. *Language Learning Journal*, *40*, 65-79. doi:10.1080/09571736.2012.658224

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Testing Research*, *18*(1), 75-94. doi:10.1177/1362168813505384

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study, *Reading in a Foreign Language, 18*(1), 1-28.

Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, *16* (1), 167-182. doi:10.1017/S136672891200017X

Polio, C., & Gass, S. (1997). Replication and reporting: A commentary. *SSLA*, *19*, 499-508. doi:10.1017/S027226319700404X

Porte, G. (2013). Who needs replication? *Calico Journal*, *30*(1), 10-15. doi:10.1139/cj.30.1.10-15.

Pritschet, L., Powell, D., & Horne, Z. (2016). Marginally significant effects as evidence for hypotheses: changing attitudes over four decades. *Psychological Science*, *27*(7), 1036-1042. doi: 10.1177/0956797616645672

Prodromou, L. (1999). *Grammar and vocabulary for First Certificate.* Edinburgh Gate, Harlow: Pearson Education Limited.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, *24*, 146-161. doi:10.1017/S0267190504000078

Reber, A. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(1), 88-94.

Reber, P. (2008). Cognitive neuroscience of declarative and nondeclarative memory. *Advances in Psychology*, *139*, 113-123. doi:10.1016/S0166-4115(08)10010-3.

Reber, P. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia*, *51*, 2026-2042. doi:10.1016/j.neuropsychologia.2013.06.019

Reichle, E., & Perfetti, C. (2003). Morphology in word identification: A word-experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading*, *7*(3), 219-237. doi:10.2017/S1532799XSSR0703_2

Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds), *English corpus linguistics*. Longman: London/New York.

Rieder, A. (2003). Implicit and explicit learning incidental vocabulary acquisition. Vienna English Working Papers.

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*(2), 299-331. doi: 10.1017/S0142716410000421

Robinson, P. (1995). Aptitude, awareness, and the fundamental similarity of implicit and explicit second language learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 303-357). Honolulu: Second Language Teaching and Curriculum Center, University of Hawai'i at Manoa.

Robinson, P. (2003). Attention and memory during SLA. In Catherine J. Doughty & Michael H. Long (Eds.), *Handbook of second language acquisition* (pp. 630-678). Oxford: Blackwell.

Roediger, H. (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*(9), 1043-1056. doi:10.1037/0003-066X.45.9.1043

Rott, S. (2007). The effect of frequency of input-enhancements on word learning and text comprehension. *Language Learning*, *57*(2), pp. 165-199. doi: 0.1111/j.1467-9922.2007.00406.x

Rowland, C. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432-1463. doi: 10.1037/a0037559

Rundell, M. (2010). *Macmillan collocations dictionary for learners of English*. Oxford: Macmillan Publishers Limited.

Saint-Dizier, P. (2006). *The linguistic dimensions of prepositions and their use in NLP Applications.* Dordrecht, the Netherlands: Kluwer Academic.

Salazar Campillo, P. (2015). Effect of textual enhancement on idioms: An exploratory study with Spanish students. *Revista Española de Lingüística Aplicada*, *28*(1), 258-272. doi:10.1075/resla.28.12sal

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129-158. doi:10.1093/applin/11.2.129

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, *11*, 11-26. doi:10.4236/ojml.2013.32016

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Honolulu: Second Language Teaching and Curriculum Center, University of Hawai'i at Manoa.

Schmidt, R. (2001). Attention. In Robinson, P. (Ed.), *Cognition and second language Instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmitt, D., & Schmitt, N. (2011). *Focus on vocabulary 2: Mastering the Academic Word List.* White Plains, New York: Pearson Longman.

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

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* Hampshire, England: Palgrave Macmillan.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, processing and use.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

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. doi: 10.1111/j.15404781.2011.01146.x

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:10.1177/026553220101800103

Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced task. In N. Schmitt (Ed.), *Formulaic sequences: Processing, acquisition and use* (pp. 173-189). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*, Psychology Software Tools, Inc.

Schoot, M., Reijntjes, A, & van Lieshout, E. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Read Writ, 25,* 1665–1690. doi:10.1007/s11145-011-9337-4

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. Kroll and A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371-388). New York: Oxford University Press.

Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, *7*(2), 118-132. doi:10.1177/026765839100700204

Shelton, J., & Martin, R. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology, Learning, Memory, and Cognition*, *18*(6), 1191-1210. doi:10.3758/BF03213769

Shin, D., & Nation, P. (2008). Beyond single words: the most frequent collocations in spoken English. *ELT Journal, 62*(4), 339-348. doi:10.1093/elt/ccm091

Sicherl, E. (2004). On the content of prepositions in prepositional collocations. *Studies in the English Language and Literature in Slovenia, 1*(1-2), 37-46. doi: 10.4312/elope

Simpson, E., & Weiner, J. (Eds.) (1989). *The Oxford English Dictionary.* Oxford: Oxford University Press.

Simpson-Vlach, R., & Ellis. N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512. doi:10.1093/applin/amp058

Sinclair, J. (1966). Beginning the study of lexis. In C. Bezell et al. (Eds.). In memory of J. R. Firth (pp. 410-430). London: Longman.

Sinclair, J. (1987). Collocation: A progress report. In R. Steel and T. Threadgold (Eds.), *Language topics: Essays in honour of Michael Halliday*. Amsterdam/Philadelphia: John Benjamins Publishing.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Siyanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language. *Corpus Linguistics and Ling. Theory*, *11*(2), 285-301. doi:10.1515/cllt-2014-0016

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review, 64*(3), 429-458. doi:10.3138/cmlr.64.3.429

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*(2), 251-272. doi:10.1177/0267658310382068

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 776-784. doi:10.1037/a0022531

Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, *36*(5), 549-569. doi:10.1093/applin/amt054

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge acquisition of collocations under different input conditions. *Language Learning, 63*(1), 121-159. doi:10.1111/j.1467-9922.2012.00730.x

Squire, L., & Zola, S. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24), 13515-13522.

Steinel, M., Hulstijn, J., & Steinel, W. (2007). Second language idiom learning in a paired-associated paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency, *SSLA, 29*, 449-484. doi:10.1017/S0272263107070271

Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2010). Does chunking foster chunk-uptake? *Applications of Cognitive Linguistics*, *17*, 99-120.

Supasiraprapa, S. (2019). Frequency effects on first and second language compositional phrase comprehension and production. *Applied Psycholinguistics, 40*, 987-1017. doi:10.1017/S0142716419000109

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747-790. doi:10.1111/lang.12241

Szudarski, P. & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, *26*(2). doi:10.1111/ijal.12092

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (Sixth Edition). Boston: Pearson Education.

Tabossi, P., Fanari, R., Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, *37*(4), 529-540.

Tidy, C. (2017). Heart attack: Myocardial infarction. Available at https://patient.info/health/heart-attack-myocardial-infarction-leaflet

Tiersky, E., & Tiersky, M. (1992). *The language of medicine in English* (Revised Edition). Englewood Cliffs, New Jersey: Prentice Hall Regents.

Tomlin, R., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*(2), 183-203. doi:10.1017/S0272263100012870

Tremblay, A., Derwing, B., Libben, G., Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*(2), 569-613. doi:10.1111/j.1467-9922.2010.0062.x

Truscott, J., & Sharwood Smith, M. (2011). Input, intake and consciousness: The quest for a theoretical foundation. *Studies in Second Language Acquisition*, *33*, 497-528. doi:10.1017/S0272263111000295

Tyler, A., Mueller, C., & Ho, V. (2011). Applying cognitive linguistics to learning the semantics of English to, for and at: An experimental investigation. *Vigo International Journal of Applied Linguistics*, *8*, 181-205. doi:10.1075/aila.23.03tyl

Ullman, M. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(1), 105-122. doi:10.1017/S1366728901000220

Ullman, M. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*, 231-270. doi:10.10616/j.cognition.2003.10.008

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences, in N. Schmitt (Ed.), *Formulaic sequences* (pp. 73-95). Amsterdam: John Benjamins.

Vilkaite, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1632-1642.

Wagenmakers, E., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*(3), 830-841. doi:10.1037/0033-295X.114.3.830

Wang, Y., & Shaw, P. (2008). Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME Journal*, *32*, 201–232.

Wang, H.-C., Savage, G., Gaskell, M., Paulin, T., Robidouz, S., & Castles, A. (2017). Bedding down new words: Sleep promotes the emergence of lexical competition in visual word recognition. *Psychonomic Bulletin & Review*, *24*(4), 1186-1193. doi: 10.3758/s13423-016-1182-7

Waring, R. & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 130-163.

Warren, P. (2013). *Introducing psycholinguistics.* Cambridge University Press: Cambridge.

Watanabe, Y. (1997). Input, intake, and retention: Effects of increasing processing on incidental learning of foreign language vocabulary. *SSLA*, *19*, 287-307.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46-65. doi:10.1093/applin/aml048

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, *43*(1), 55-77. doi:10.1002/j.1545-7249.2009.tb00227.x

Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*, *32*(3), 259-276. doi:10.1093/applin/amq051

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learnin*g, *63*(1), 91-120. doi:10.1111/j.1467-9922.2012.00729.x

Wen, Q., Wang, L., Liang, M. & Yan, X. (2008). Written English Corpus of Chinese Learners. Foreign Language Teaching and Research Press.

West, M. (Ed.) (1953). A general service list of English words: With semantic frequencies and a supplementary words-list for the writing of popular science and technology. Harlow, Essex: Longman.

Williams, R., & Morris, R. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*(1-2), 312-339. doi: 10.1080/09541440340000196

Winke, P. (2013). The effects of input enhancement on grammar learning and comprehension: A modified replication of Lee (2007) with eye-movement data. *Studies in Second Language Acquisition*, *15*, 323-352. doi:10.1017/ S0272263112000903

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

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, *35*, 451-482. doi:10.1017/S0272263113000107

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, *32*, 231-254. doi: 10.1017/S02671905200013X

Wulff, S. (2018). Acquisition of formulaic language from a usage-based perspective. In Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (Eds.), *Understanding multi-word expressions: A second language acquisition perspective*. London, New York: Routledge.

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*(4), 647-668. doi:10.5054/tq.2010.235998

Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, *40*, 831-856. doi:10.1017/S0272263118000141

Yun, Z., & Youmei, G. (2006). A CLEC-based study of collocation acquisition by Chinese English language learners. *CELEA Journal, 29*(4), 28-35.

Zahar, R., Cobb, T. & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review, 57*(4), 541-572. doi:10.3138/cmlr.57.4.541

Appendices

Appendix A

Differences in the study design between Sonbul & Schmitt (2013) and Experiment 1 in this thesis

Sonbul & Schmitt	**Experiment 1**
Materials	
15 medical collocations	15 collocations: 13 from S&S; 2 replaced
1 x 1000-word reading text	9 x 500-word reading texts (S&S's text adapted)
3 contextual occurrences of each collocation	9 contextual occurrences of each collocation
Treatments	
2 incidental learning conditions: reading-only & bolding	3 incidental learning conditions: reading-only, bolding & bolding + glossing
1 deliberate learning condition	no deliberate learning condition
Procedure	
1 learning session	3 learning sessions
learning on 1 day	learning over 2 days
Measures / Post-tests	
1 space for each missing letter in gapfill test	1 line for each missing word in gapfill test
participants' L2 proficiency measure: self-rated proficiency	participants' L2 proficiency measure: vocabulary size (VST)
participants' prior knowledge of the target collocations not checked	self-reported prior knowledge of the target collocations

Appendix B

Medical collocations and control pairs in Experiment 1

Medical collocations	**Control pairs**
regional control	domestic control
smooth diet	plain diet
gap periods	hole periods
chief complaint	prime complaint
pure absence	clean absence
fixed end	solid end
gene therapy	code therapy
shell shock	round shock
stone heart	rock heart
golden hour	grand hour
silent areas	gentle areas
cloud baby	steam baby
iron lung	metal lung
split hand	fork hand
partial response	limited response

Appendix C

Treatment materials for Experiment 1 (participant group 1)

Name: _____

Cancer and heart disease

Cancer and heart disease are together responsible for the deaths of more than half the New Zealanders who die each year. However, these diseases can often be successfully treated.

Cancer

Cancer is the leading cause of death in New Zealand; it caused 8891 deaths, or nearly a third of deaths in the country, in 2011. Most deaths from cancer occur after the cancer has spread from its original location to other organs. Because of this, **regional control** is important. **Regional control**, or stopping the spread of the cancer from its starting place to other locations, can be done through the use of surgery, drugs or radiation. The use of **regional control** can have a positive effect on the treatment's success. The use of special drugs to attack cancer cells is called chemotherapy. The drugs are often given to patients in cycles, which consist of drug treatments for several days each followed by gap periods. No drugs are given to patients during these gap periods, which allow normal cells to recover from any negative effects of the treatment. The gap periods are longer than the cancer treatment periods. When the treatment kills some of the cancer cells, this is called **partial response [1]**. More specifically, a **partial response** means that the size of the cancer has decreased by at least 50% but the cancer has not disappeared. As a result of a **partial response**, treatment may be stopped unless the cancer starts growing again.

[1] a reduction in a cancer because of treatment

Sometimes a cancer patient's body reacts with discomfort to cancer treatments, and the patient must follow a **smooth diet**. On a **smooth diet**, a person eats foods containing little fibre (the part of fruit, vegetables and grains that the body does not break down) – foods such as milk shakes and soft cheeses. Because a **smooth diet** is a major change in the diet, a patient should check with their doctor before making such a change. A doctor may recommend different variations of this diet, depending on the capability of the patient's body to process food.

Heart disease

Heart disease is the second most common cause of death in New Zealand, accounting for 24% of all deaths in 2011. A large amount of heart disease is caused by lifestyle choices, particularly smoking, following an unhealthy diet, being physically inactive, keeping an unhealthy weight and managing stress badly. Making changes to these areas of lifestyle is important in order to prevent or treat heart disease. If such changes are not effective, medicine may be needed. If the heart disease is severe, surgery may be required. The chances of surviving heart surgery are very high, with the survival rate of one type, heart valve surgery, being over 98%. However, during heart surgery, a rare serious complication that has occurred is the stone heart condition. Stone heart occurs when heart muscles become stiff, which usually leads to the death of the patient. However, these days stone heart is almost always avoided through the use of modern medical techniques.

END

Now answer two questions about the text. For each answer, circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

1. Radiation is used to treat both cancer and heart disease.

a. true b. false

2. According to the article, exposure to chemicals is one cause of heart disease.

a. true b. false

1.2 **Name: _____**

Injuries

Injuries kill more than 5 million people around the world every year and result in tens of millions of people being treated in hospital. Injuries are caused by a variety of events, including traffic accidents and wars.

Accident injuries

The most common cause of deaths from injuries is traffic accidents. Worldwide, in 2010, about 1 million people died from road traffic injuries. It has been said that the survival rate of victims of traffic, and other, injuries decreases more than 60 minutes after the injuries occur. This idea is called the **golden hour [1]**. In fact, the **golden hour** now appears to have little scientific basis. However, the principle of the **golden hour** has been extended to refer to the importance of rapid medical intervention after a severe injury. Rapid transport to hospital during this period is particularly important for patients with severe head injuries, which can become critical. Regions of the brain where injuries apparently cause no symptoms are labelled **silent areas.** Although injuries to **silent areas** cause no obvious signs, some researchers believe that damage to them can affect people's ability to plan and to function socially. According to recent research, however, **silent areas** may not exist.

[1] the time immediately after a serious accident

War injuries

Wars cause the deaths of, and injuries to, tens of thousands of people annually. One of the largest death tolls in a war was in the First World War, when 16 million people were killed and 20 million were wounded. However, the war also had psychological effects on soldiers. A phrase first used in the First World War was shell shock. Shell shock referred to the trauma reaction of some soldiers to war, and to symptoms such as panic, fear and the inability to sleep or talk. More than 250,000 British soldiers were thought to have suffered from shell shock during the war.

Fractures

A frequent injury, which does not generally cause death, is bone fractures (breaks or cracks), of which there are different types. A compression fracture occurs when strong pressure is applied from one end of the bone to the **fixed end [2]**. That is the end of the bone that is not moving, often because the end is attached to the rest of the body or is on the ground. One example of this compression is a person jumping onto the ground from a great height. Another type of break, a separation fracture, occurs when the other end of the bone is forced *away* from the **fixed end**. A third kind of fracture is caused by a bone being twisted; in this case the **fixed end** of the bone remains in one place while the other end turns.

[2] the end of a bone that does not move

Not an injury

A relatively rare condition which looks like it could result from an injury but in fact does not is **split hand** syndrome. A person with a **split hand** has one or more middle fingers on their hand missing and the hand resembles a claw. **Split hand** occurs at birth, immediately before birth or immediately after birth.

END

Now answer two questions about the text. For each answer, circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

1. Traffic accidents cause more deaths than wars.

a. true b. false

2. There is more than one type of bone fracture (break).

a. true b. false

1.3 **Name: _____**

Infectious diseases

Infectious diseases are still a common global problem. About a quarter of the 56 million deaths around the world each year are caused by infectious diseases. Millions more people suffer from the effects of these diseases.

One common infectious illness is measles, which killed more than 140,000 people in 2013. The main reason that sufferers of this illness seek medical help, or their chief complaint, is a high temperature. Another infectious illness, cholera, kills more than 100,000 people each year; the sufferers' chief complaint is frequently severe diarrhea. A third illness, malaria, kills more than 360,000 people each year; the patient's usual chief complaint is sudden coldness followed by a high temperature.

AIDS

An extremely serious infectious illness is AIDS, which leads to the deaths of more than 1 million people each year. It is most often caught through sexual contact or through the sharing of drug needles. There is no cure for AIDs, but some drugs can be used to control the virus. A potential cure for AIDS is **gene therapy [1]**, a new treatment. In **gene therapy**, a gene is inserted into a patient's cells to correct the faulty genes. In the future, **gene therapy** may help diseased tissues and organs work properly.

[1] putting a healthy person's genes into a sick person's cells

Polio

An infectious disease which leaves a very small number of infected people unable to move is polio. While there were about 350,000 cases of the illness in 1988, the number fell sharply to about 400 in 2013 because of immunization programmes. A large machine, invented in 1927, which helps polio victims breathe is called the **iron lung**. A patient lies inside the **iron lung**, which copies breathing actions and causes air to flow in and out of the patient's lungs. The **iron lung** is rarely used these days, however, because the illness has almost disappeared and because smaller, more advanced machines are now used to help patients breathe.

Epilepsy

Infections also commonly play a role in cases of epilepsy, an illness which causes more than 170,000 deaths per year. Most cases of the illness are caused by both environmental factors, such as infections, and genetic factors. One symptom is a pure absence seizure; this is a sudden attack, when a person loses consciousness for a few seconds and stares blankly during that time. Pure absence attacks are often not noticed by other people because it can look as if the person is not paying attention. A number of different treatments are available for these pure absence attacks, including medicine, diet therapy and surgery.

Spreading infectious diseases

While many people commonly spread infectious diseases, some do it more efficiently than others. These are called "super spreaders", and can be adults or children. A baby who quickly spreads infectious diseases into its immediate environment is known as a **cloud baby [2]**. A **cloud baby** transmits infection to many more other people than the average baby does, but has no obvious symptoms of the illness. The **cloud baby** was first described in a journal article in 1960. END

[2] a baby who looks healthy but who rapidly spreads illness

Now answer two questions about the text. For each answer, circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

1. Polio has now disappeared.

a. true b. false

2. Epilepsy is caused by both environmental and genetic factors.

a. true b. false

1.4 **Name:** _____

Common killers

The most common killers of human beings used to be infectious diseases. Today the leading causes of death are non-infectious diseases such as cancer, and injuries. Meanwhile, many infectious diseases which used to be major killers are now more easily controlled.

Cancer

Cancer, which causes about 8 million deaths around the world annually, is one of the most common serious illnesses in modern society. It occurs when cells in a person's body form harmful growths. Cancer can appear anywhere in the body, but common sites include the lungs, the breasts, the skin and the blood. Some of the most frequent symptoms of cancer are bleeding, a thickening in any area, difficulty swallowing and unexplained weight loss. Treatment, through surgery, drugs or radiation, may lead to a **partial response [1]**. When a **partial response** occurs, the cancer has reduced in size for at least a month. **Partial response** also means that the cancer can still be seen on scans but is not growing.

[1] a reduction in a cancer because of treatment

Injuries

Injuries are another major source of death in the modern world, causing about 5 million deaths around the world each year. More than 1 million deaths are caused by traffic related injuries. It is essential that critically injured victims of traffic accidents and other accidents receive medical care during the **golden hour [2]** in order to prevent their deaths. The **golden hour**, which is not a true 60 minutes, can continue for several minutes or several hours, depending on the nature of the victim's injury. (It is believed that the concept of the **golden hour** came from data from World War One soldiers.) Many doctors believe that serious head injuries can lead to serious symptoms except when they affect **silent areas** of the brain. However, current evidence suggests that there are, in fact, no **silent areas**. This evidence against the existence of **silent areas** comes from research showing that many brain regions are much more active than previously thought. Another type of serious injury, which often occurs in sports games, results from force to the **fixed end [3]** of a bone. For example, the inner structure of the knee can be damaged when significant force is applied to it while the lower leg acts as the **fixed end**. Force to the **fixed end** can also result in serious breaks.

[2] the time immediately after a serious accident

[3]

the end of a bone that does not move

Infectious diseases less problematic

While cancer and injuries kill millions of people each year, many infectious diseases which used to kill many people are no longer major problems. Polio, for example, once left many people unable to move, but today vaccines effectively protect against the illness. A device called the **iron lung** was used in the 1920s to help polio patients breathe. The **iron lung** has a very small place in modern therapy, however, since most patients who cannot breathe without help now use machines other than the **iron lung**. Another infectious lung disease, tuberculosis, was the main cause of death in the United Kingdom in the nineteenth century but can be prevented today through the use of immunization.

END

Now answer two questions about the text. For each answer, circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

1. Difficulty swallowing is a symptom of cancer.

a. true b. false

2. Polio is an infectious disease.

a. true b. false

1.5 **Name: _____**

Non-infectious killer diseases

In the 21st century, infectious diseases that used to be widespread killers can now be prevented or diagnosed early and cured. Today's major killer diseases are non-infectious diseases, particularly heart disease and cancer. The chief complaint of each of these types of illnesses can differ, but treatment can reduce the number of deaths from these diseases.

Heart disease

The most common serious illness in modern society is heart disease, which causes the deaths of more than 17 million people globally every year. A typical chief complaint of a person suffering from a heart attack (an event caused by heart disease) is chest pains. Another possible chief complaint is pain in other areas of the body such as both arms or between the shoulders. Heart disease is caused by many factors of modern life, such as stress, unhealthy diet, lack of physical activity and use of cigarettes. However, fast emergency treatment can save victims from sudden death and is sometimes used instead of surgery. Nevertheless, heart surgery must be performed at times. While the success rate of heart surgery is high, it can in very rare cases lead to complications such as the stone heart condition. Stone heart is a fatal condition which can occur during open-heart surgery. Analyses have found that the stone heart condition is linked with extensive damage to the heart muscle.

Cancer

Another illness that is a major killer is cancer, which leads to the deaths of 8 million people every year. Today, many types of cancer can be cured, especially if they are found early. A cancer can sometimes be treated and cured by surgery alone, but sometimes radiation or drugs are used together with surgery. Powerful cancer treatment drugs used in chemotherapy help destroy cancer cells and may cure the cancer. A mixture of drugs is normally given over several days followed by gap periods of a few weeks. Gap periods allow the body to recover from the side effects. These side effects can be severe for a number of days during the treatment, but people may be able to live normally during gap periods. Cancer often spreads from its place of origin to another organ. This situation requires higher doses of chemotherapy and/or radiation and means that the cancer is more difficult to cure. In the early stages of the spread of cancer, **regional control** is the main aim of the treatment plan. **Regional control** is achieved through chemotherapy as well as through radiation and surgery. **Regional control** of cancer can prevent the cancer from spreading to other organs and can increase the survival rate of the patient. Natural cancer treatments that do not use drugs or radiation include types of **gene therapy [1].** Using materials created by the body of the patient or in a laboratory, **gene therapy** causes the body's natural defences to fight cancer. For example, it may strengthen the body's immune system. It is believed that **gene therapy** may reduce or even stop the growth of cancer cells.

[1] putting a healthy person's genes into a sick person's cells

END

Now answer two questions about the text. For each answer, circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

1. According to the article, lack of physical activity is one cause of heart disease.

a. true b. false

2. Chemotherapy is the use of radiation to treat cancer.

a. true b. false

1.6 **Name: _____**

Common diseases and conditions

Modern medicine has advanced tools for fighting illness and injury, which produce better health outcomes than in the past. However, illnesses and injuries can cause negative psychological effects, which also need to be treated.

Fewer deaths

Thanks to modern medicine, many diseases kill fewer people today than they did in past years. One example is tuberculosis. This is the second most deadly infectious disease these days, causing the deaths of more than one million people globally each year. However, it killed many more people in the nineteenth century, when it caused 25% of all deaths in Europe. Today, to test for the illness, a simple skin test is given in a medical check-up. If the results are positive, the condition can then be controlled or cured by medicine. Epileptic people with tuberculosis should be treated carefully since treatment can involve drug interactions with epilepsy treatment drugs. Certain treatments can produce an increased risk of seizures of different types, including pure absence. These pure absence attacks can lead to weakness or the loss of physical feeling. If it is left untreated, a pure absence attack may progress into a more serious, complex attack.

Less suffering among children

Modern medicine has also reduced the suffering caused by diseases which are common in children. One common infectious childhood disease is whooping cough, whose symptoms include severe coughing. The illness kills nearly 300,000 people around the world each year, but immunization saves the lives of 500,000 people a year. Many other common infectious childhood illnesses do not usually kill although they do cause suffering. Infectious diseases are often transmitted from one child to another; however, a **cloud baby [1]** can infect many more children than is usual. A **cloud baby** appears to be healthy, yet it spreads illnesses more rapidly than other babies. The concept of the **cloud baby** was first introduced more than 50 years ago.

[1] a baby who looks healthy but who rapidly spreads

Psychological effects

Some illnesses result in not only physical problems but also psychological problems. For example, many soldiers in World War I were found to be suffering from a condition called shell shock. Common symptoms of shell shock include extreme tiredness and inability to sleep. Since the Second World War, the term shell shock has been replaced by "combat stress reaction". Another condition that can have serious psychological effects, such as anxiety and depression, is irritable bowel syndrome. This is characterized by changes in toilet habits and a general lack of comfort, and it requires special food considerations such as a **smooth diet**. A **smooth diet** avoids food that is hard for the stomach to break down. Food allowed in a **smooth diet** can include soft cheese, eggs and milk, but not brown rice or potato. Yet another condition with negative psychological effects is **split hand** syndrome. A person with a **split hand** has one or more central fingers missing on the hand; this can affect communication with the hands and lead to social rejection. **Split hand** affects 1 in 8,000 to 25,000 individuals.

END

Now answer two questions about the text. For each answer, circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

1. Tuberculosis causes fewer deaths today than it did in the 19th century.

a. true b. false

2. Whooping cough is usually controlled through surgery.

a. true b. false

1.7 **Name: _____**

Modern treatments

The treatment of illnesses and medical conditions has advanced greatly in the last few centuries. Many serious illnesses now have higher survival rates than they did in the past, but, at the same time, some of the less serious illnesses and conditions still have no cure.

Cancer treatments

One serious illness whose survival rate has increased in recent decades is cancer; about 32 million people worldwide suffered from it in 2012. Until the late 19th century, surgery was the only method of treating cancer. However, few people lived through surgery for cancer because hygiene problems usually occurred during surgery. After its discovery in the 1890s, radiation was used as the first effective non-surgical cancer treatment. Then, from the 1940s, chemotherapy (cancer treatment drugs) was used. Following several days of such drug treatments for cancer, gap periods are introduced. These gap periods allow the body to recover from the effects of the cancer drugs. The gap periods last from several days to several months. Today chemotherapy is often used in combination with radiation. One approach in which they are used together is in the **regional control** of cancer. By applying **regional control**, a cancer is prevented from spreading from its place of origin. **Regional control** greatly increases patients' chances of survival from cancer. When cancer treatment is successful, it may lead to **partial response [1]**. **Partial response** means that the amount of a cancer in the body has decreased as a result of the treatment. However, **partial response** implies that further cancer treatment may be needed in the future to remove the cancer.

[1] a reduction in a cancer because of treatment

Heart surgery

The heart disease death rate in developed countries has decreased since the 1970s due to changes in people's lifestyles and the introduction of new heart disease treatments such as heart surgery and drugs. However, a condition that occurred occasionally in heart surgery, particularly in the early decades of heart surgery in the mid-1900s, was stone heart condition. In stone heart condition, the heart stops and cannot be started again, which causes the patient to die. In the first years of heart surgery, the condition could not usually be prevented. However, scientists can now prevent stone heart and keep the patient alive using a technique which involves stopping the heart and cooling it.

No cures yet

Despite the progress in the treatment of many serious illnesses, modern science has failed to find cures for some less serious illnesses and conditions, although the symptoms can be treated. One such illness is the common cold, which is the most frequent infectious disease in the world. The average adult catches a cold two or three times a year. A less frequent condition that modern medicine has not yet found a cure for is **split hand** syndrome. **Split hand** is a condition which occurs around the time of birth and results in missing fingers. Sometimes reconstructive surgery is performed to address the condition. In the meantime, however, researchers are attempting to identify the gene responsible for the **split hand** disorder.

END

Now answer two questions about the text. For each answer, circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

1. Chemotherapy was the first known treatment method for cancer.

a. true b. false

2. A cure has been discovered for the common cold.

a. true b. false

1.8 **Name: _____**

New medical understanding

In medical science, ideas about illnesses, injuries and medical conditions constantly change as researchers continue to make new discoveries.

Cleanliness

An example of such an advance is Joseph Lister's idea in the 1860s of the need for cleanliness in the operating room. He developed ways to clean patients' wounds and surgery instruments, and these methods were adopted by surgeons. As a result, the death rate from surgical infection dropped dramatically from 60% to 4%.

Treatment of injuries

[1] the time immediately after a serious accident

Another advance in medical knowledge is the changing definition of the **golden hour [1]**. In the 1950s a medical professor suggested that victims of serious injuries needed to be medically treated within 60 minutes of being injured in order to prevent their death. Today, however, the **golden hour** is viewed more generally as a period lasting from a few minutes to a few hours after an injury. The key idea of the **golden hour** still remains, though: the critically-injured patient needs treatment soon after being injured to prevent his or her death.

Brain responses

A discovery made recently by medical science is about parts of the brain called **silent areas**. These **silent areas** do not respond to injury. It used to be thought that this lack of response was because there is little brain activity. However, it is now known that, in fact, plenty of brain activity occurs in these **silent areas** and that these areas control a number of different activities such as making moral judgements and appreciating jokes.

Psychological reaction

Medical understanding about a condition which used to be called shell shock has also grown. In the First World War, many soldiers were found to be suffering from this condition. Soldiers who had shell shock felt helpless in some way, perhaps being unable to walk or to talk. At first many doctors thought that shell shock was due to physical damage to the brain, although others correctly saw it as a psychological reaction to war.

Diet

There has been recent medical discussion on the topic of the ideal amount of fibre that is needed in a healthy diet. In the past, many doctors have recommended that people with inflammatory bowel syndrome (whose symptoms include diarrhea and stomach pain) increase the amount of fibre in their diet by eating more foods such as cereal and rice. However, recent studies indicate that this type of diet may actually worsen the symptoms and that a **smooth diet** may be better. A **smooth diet** contains very little or no fibre. This **smooth diet** contains soft foods such as milk products.

Fractures

fixed end:

the end of a bone that

A final modern medical development is the study of human bodies to determine fractures in bones. There are several types of breaks. In one type, the **fixed end [2]** of the bone remains still while the other end bends. In another type of break, the **fixed end**

stays still while the other end is twisted. In a third type, the **fixed end** does not move while the other end is forced towards it.

END

Now answer two questions about the text. For each answer, circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

1. Joseph Lister's idea in the 1860s was to have clean procedures in the operating room.

a. true b. false

2. In the past, it was thought that people with inflammatory bowel syndrome should eat only a little fibre.

a. true b. false

1.9 **Name:** _____

Past and present infectious diseases

Some infectious illnesses which used to kill millions of people are now largely controlled, while other infectious diseases still require cures from medical science.

Polio

An infectious illness which used to be common but which has now been almost eliminated is polio. In 1952, more than 3000 people died of the illness and 21,000 in the United States were left unable to move to some extent. However, in 2013 there were only 413 cases of the illness globally, with an estimated fewer than 10 deaths. The illness affects mostly children. One device which was previously used to help patients breathe was the **iron lung**. Most patients spent no more than a week inside an **iron lung**; however, one woman in the United States who died recently at the age of 83 spent over 60 years in her **iron lung**.

The plague

Historically, an extremely serious infectious illness has been the plague. The chief complaint of this illness is swollen, painful glands. In the sixth century, the disease killed about 50 million people in the Roman Empire, and in the fourteenth century the illness (in this situation it is usually called the Black Death) killed 25 million in Europe, or a third of the population. Today, however, only about 200 people a year die of it.

Tuberculosis

The infectious illness which kills the second largest number of people today is tuberculosis. This disease is responsible for at least 1.2 million deaths around the world per year. A typical chief complaint of the illness is a cough that last for three weeks or more.

AIDS

These days the infectious illness which causes the largest number of deaths globally is AIDS, which killed 1.3 million people in 2013. A common chief complaint of AIDS is a high temperature. There is currently no cure for AIDS, but one possible future treatment for the illness that is being researched is **gene therapy [1]**. In the future, **gene therapy** may help the immune system gain control over the AIDS virus. Until now, however, there has been limited success in the development of such a **gene therapy**.

[1] putting a healthy person's genes into a sick person's cells

Epilepsy

Another common, but less serious, illness which has no current cure is epilepsy, which 65 million people globally suffer from. In this illness a sufferer has seizures, or sudden attacks. One such attack is a pure absence attack, during which a person stares at nothing for several seconds. Pure absence attacks can occur many times in one day without warning. People who suffer from pure absence attacks should not drive vehicles.

Spreading diseases

Infectious diseases are spread more easily by some people than by others. A baby who spreads infectious illnesses much faster than other babies is called a **cloud baby [2].** The term was first used by researchers in the 1960s, and since then the concept of the **cloud**

[2] a baby who looks healthy but who rapidly spreads illness

baby has been expanded. Medical professionals now recognise that an adult version of the **cloud baby** exists: that is, some adults spread infectious illness very quickly.

END

Now answer two questions about the text. For each answer, circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

1. The plague kills more people than AIDS each year.

a. true b. false

2. A person with epilepsy has seizures (sudden attacks).

a. true b. false

Appendix D

Analysis of Experiment 1 texts (1-9)

Measure	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Text 9
No. words in headings	7	9	9	8	6	12	9	13	13
Total tokens in sentences	493	491	491	492	494	488	491	487	487
Total tokens (MSW/CLT)	500/500	500/501	500/505	500/501	500/504	500/505	500/502	500/501	500/503
Total word types (CLT)	202	203	212	212	192	224	203	213	195
Type-token ratio (CLT)	40%	41%	42%	42%	38%	44%	40%	43%	38%
Tokens per type (CLT)	2.48	2.47	2.38	2.36	2.67	2.25	2.47	2.35	2.63
Total words K1-K3 (CLT)	98.40%	97.41%	97.03%	98.20%	98.60%	97.81%	97.80%	97.80%	97.03%
No. sentences (MSW)	27	26	28	27	28	30	29	26	28
Av. sent. lgth in wds (MSW)	18.2	18.8	17.5	18.2	17.6	16.2	16.9	18.7	17.2
Passive sents. (MSW)	29%	38%	39%	29%	35%	33%	34%	34%	28%
Key: MSW = Microsoft Word 2010; CLT = Compleat Lexical Tutor (Cobb, n.d.)									

Note. Online calculators used: Compleat Lexical Tutor (http://www.lextutor.ca); Microsoft Word 2010

Note: It can be seen in the table that Microsoft Word and the Compleat Lexical Tutor (Cobb, n.d.) give slightly different word counts for Texts 2-9. This is mostly accounted for by the differing methods used by the two programmes for counting numbers in documents. The most common difference occurred in the use of numbers containing a punctuation mark. For example, whereas MS Word counted *140,000* as one word, the Compleat Lexical Tutor counted it as two. *1.3* was counted by MS Word as one word and by LexTutor as two. Another difference is in the treatment of hyphenated words. For example, MS Word viewed *non-infectious* as one word while the Compleat Lexical Tutor treated it as two.

Appendix E

Cued-recall test for Experiment 1

Name: _____

The following passage summarizes some of the texts that you have read. There are a number of medical phrases in bold type which were in the original texts. The first word of each phrase is cut off. Please look at the context and the meaning provided, and then fill in each blank with the missing word.

Example:

0. reduced density of bones

Many elderly women suffer from **(0. __*porous*__ bone).**

Killer diseases and injuries

In the 21st century, infectious diseases that used to be widespread killers can now be prevented or cured. Today's major killer diseases are non-infectious diseases, particularly cancer and heart disease, while injuries also account for many deaths.

1. control of cancer in initial stages

2. periods when cancer treatments stop

3. a decrease in the size of a cancer after treatment

4. a diet containing little fibre

Cancer is the leading cause of death in New Zealand. Most deaths from cancer occur after the cancer has spread from its original location to other organs. Because of this, **(1. _____ control)** is important. The use of special drugs to attack cancer cells is called chemotherapy. The drugs are often given to patients in cycles, which consist of drug treatments for several days each followed by **(2. _____ periods).** If a **(3. _____ response)** occurs, treatment may be stopped. Sometimes a cancer patient's body reacts with discomfort to cancer treatments, and the patient must follow a (**4. _____ diet).**

5. the heart is not beating

Heart disease is the second-most common cause of death in New Zealand. If the heart disease is severe, surgery may be required. The chances of surviving heart surgery are very high. However, during heart surgery, a rare serious complication that has occurred is the **(5. _____ heart)** condition.

6. putting biological material into a patient's cells

7. a mechanical device for artificial breathing

8. the main symptom of an illness

9. a brief loss of consciousness

Although infectious diseases kill fewer people than they used to, they are still a common global problem. One extremely serious infectious illness is AIDS. There is no cure for AIDs; however, a potential cure for AIDS is **(6. _____ therapy)**, an experimental treatment which may help diseased tissues and organs work properly. An infectious disease which leaves a very small percentage of infected people unable to move is polio. The **(7. _____ lung)** can be used to help polio victims breathe. One common infectious illness is measles. The **(8. _____ complaint)** of sufferers of measles is a high temperature. Finally, infections also commonly play a role in cases of epilepsy. One symptom of epilepsy is a **(9. _____ absence)** seizure (or sudden attack).

10. a baby carrying a disease

While many people commonly spread infectious diseases, some do it more efficiently than others. One such type of person is a **(10. _____ baby)**.

11. the first hour after a serious accident

12. brain areas where injuries cause no definite symptoms

Injuries kill more than five million people around the world every year. The most common cause of injuries is traffic accidents. It is crucial for critically injured accident victims to receive medical care during the **(11. _____ hour)** in order to prevent their deaths. Rapid transport to hospital during this time is particularly important for patients who have severe head injuries. Some researchers believe that damage to **(12. _____ areas)** of the brain can affect people's ability to plan and to function socially.

13. a psychological condition after a shock

Wars lead to the deaths of, and injuries to, tens of thousands of people each year. However, the First World War had not just severe physical effects on soldiers: more than 250,000 British soldiers were thought to have suffered from **(13. _____ shock)** in the First World War.

14. the end of a bone that is not moving

15. a hand with an extended division between the fingers

A frequent injury, which does not usually cause death, is bone fractures. A compression fracture occurs when strong pressure is applied from one end of the bone to the **(14. _____ end).** A fairly rare condition which looks like it could be the result of an injury but in fact is not is **(15. _____ hand)** syndrome.

Appendix F

Language background questionnaire

1. What is your family name? _____

2. What is your given name (or names)? _____

3. What is your date of birth? _____

4. What is your first language? _____

5. For how many years have you been studying and/or speaking English? _____

6. How long have you been living in New Zealand? _____

7. Before you came to New Zealand, did you live in any other English-speaking countries?

Yes / No

8. If the answer to question 7 is yes, please state how long you lived in that English-

speaking country (or countries). _____ years and _____months

9. Do you fluently speak any languages other than your first language and English? Yes / No

10. If the answer to question 9 is yes, please name the other language(s). _____

_____

11. When was the last time you sat an IELTS, TOEFL, TOEIC or Cambridge Advanced English

exam? Month: _____ Year: _____

12. What was/were your score(s) in the exam?

IELTS: Overall: _____ Reading: _____ Writing: _____

Listening: _____ Speaking: _____

TOEFL: _____

TOEIC: _____

CAE: _____

13. Which level of study are you currently doing in New Zealand? Circle one answer.

a. Pre-degree (e.g. polytech diploma, English language course)
b. Bachelor's degree - first year
c. Bachelor's degree - second year
d. Bachelor's degree - third year
e. Honours OR post-graduate diploma
f. Master's degree
g. PhD

14. What is your major (subject)? _____

15. Have you studied health sciences or medicine at university or polytechnic? Yes / No

Appendix G

Characteristics of participants in Experiment 2

Characteristic	Mean	SD	Range
Age	24.19	4.43	18
Time in English-speaking countries	0.79	0.97	6.63
Age first exposed to English	9.53	3.09	15
VLT 2k score (max = 30)	27.62	2.65	9

Appendix H

Approaches to collocation selection for Experiment 2

Element	Approach 1	Approach 2	Approach 3	Approach 4 (final list)
Sources of unconventional phrases	Learner essays (WECCL, CLEC, EAP)	N/A	Learner essays (WECCL, CLEC, EAP)	N/A
Source(s) of target V+N phrases	English dictionaries	ACL (268 V+N)	ACL English dictionaries	OCD (COCA: 1 phrase)
Source of target prep+N phrases	English dictionaries	OCD	OCD	OCD
V+N phrases containing multiword verbs excluded?	X	√	√	√
Prep+N phrases containing multiword prepositions excluded?	X	√	√	√
Frequency range of component words	K1-K3 (BNC)	K1-K3 (BNC)	K1-K3 (BNC)	K1-K2, some K3 (BNC/COCA)
Raw frequency (in COCA) cutoff of V+N phrases	5000	561	1000	2000
Raw frequency (in COCA) cutoff of prep+N phrases	5000	1000	1000	2000
All verb inflexions checked in COCA	X	√	√	√
Noun singular & plural forms checked	X	X	√	√
MI cutoff for V+N phrases	3 (or in dictionary)	3	3	N/A
MI cutoff prep+N phrases	3 (or in dictionary)	N/A	N/A	N/A
Frequency bands of verbs and prepositions matched	X	X	X	√
High-transparency phrases excluded	X	√	√	√
Only 2-word phrases and 3-word phrases including *a*, *an* or *the*	√	√	√	√
Trialled with native Chinese speakers	√	X	X	√
Trialled with native English speakers	√	X	X	√
Gapfill test trial	√	X	X	√
Difficulty and congruence rated by triallists	√	X	X	X

Abbreviations: V+N: verb + noun; prep+N: preposition + noun; WECCL: Written English Corpus of Chinese Learners; CLEC: Chinese Learners English Corpus; EAP: English for Academic Purposes; ACL: Pearson's Academic Collocation List; OCD: Oxford Collocations Dictionary (2nd edition); K1-K3: the first one thousand to the third one thousand word families in the (BNC) corpus; BNC: British National Corpus; COCA: Corpus of Contemporary American English

Appendix I

Target collocations and control phrases in Experiment 2

Table I1: Final V+N collocations and verb + control phrases for Experiment 2

V+N target collocations						V+N control phrases					
	Verb freq band (BNC/ COCA) (K1-3)	Noun freq band (BNC/ COCA) (K1-3)	COCA phrse raw freq	COCA MI phrase	COCA raw freq exact phrse	Half-synonym/ control verb	Verb freq band (BNC/ COCA) (K1 etc)	COCA phrse raw freq	COCA MI phrase	COCA raw freq exact phrse	COCA raw freq exact phrse (real phrse)
meet targets	1	3	230	4.57	18	gain targets	1	2	0.03	1	0
make arrangements	1	1	1181	4.58	274	sort arrangements	1	5	0.11	0	0
reach an audience	1	3	322	4.33	31	stir an audience	2	2	3.02	0	0
hold the belief	1	3	263	4.15	10	state the belief	1	82	3.90	0	0
draw a comparison	1	2	56	3.49	12	spot a comparison	1	0	NP	0	0
set a course	1	1	86	NP	86	map a course	2	16	-0.69	0	0
extend credit	2	2	163	4.45	39	supply credit	2	16	1.29	0	0
strike a deal	1	1	556	3.98	135	frame a deal	2	0	NP	0	0
suffer a decline	1	3	108	4.94	9	receive a decline	2	0	NP	0	0
form expectations	1	1	27	1.31	8	start expectations	1	15	-0.94	0	0
raise the issue	1	1	1207	3.65	224	note the issue	1	44	0.31	0	0
push the limits	1	2	420	6.04	89	touch the limits	1	0	NP	0	0
take a loss	1	1	237	0.06	35	give a loss	1	38	-0.25	0	0
beat the odds	1	1	370	6.10	190	win the odds	1	0	NP	0	0
enter the profession	1	2	225	5.92	54	begin the profession	1	0	NP	0	0
stage a protest	1	2	93	7.79	11	effect a protest	2	3	-0.90	1	0
see reason	1	1	1246	1.41	53	know reason	1	522	0.61	2	0
offer a reply	1	1	18	1.69	0	spare a reply	2	2	4.19	0	0
perform a review	2	3	32	2.05	0	operate a review	2	0	NP	0	0
shed tears	2	1	316	7.61	131	lose tears	1	10	-0.83	0	0
present a threat	1	2	220	3.03	29	produce a threat	2	7	0.05	0	0
mark a turn	1	1	61	0.93	1	show a turn	1	58	-1.51	0	0
post a warning	1	2	56	4.67	3	list a warning	1	23	2.72	0	0
create work	2	1	330	0.57	41	develop work	2	191	-0.28	0	0

Note: NP = not provided in COCA

Note: The "COCA raw freq exact phrase (real phrase)" column describes the control phrases. It reports the raw frequency of "real" exact phrases: phrases whose two main component words have a semantic link[27].

Note: The collocates span for "COCA raw phrase frequency" was set at 3 to the left

Note: The column "COCA MI phrase" provides the MI scores for the two content words of each collocations (articles are excluded from the scores).

[27] I included these figures to show, particularly for the prep+N phrases (a few of which appeared several dozen times), that the number of occurrences in the corpus as a real (semantically-linked) phrase was negligible. For example, there are two instances of *know reason* (the control phrase for *see reason*) in COCA, but neither is a real phrase (in this case, verb + object). In one, "… if we didn't already **know reason** does not enter into it", *reason* is at the start of a subordinate clause (equivalent to "…know *that* reason…"); in the other sentence, "'Now you **know reason** why first dog name Captain,' Kwan adds," the omission of *the* before *reason* is clearly a non-native–speaker error, given the other grammatical errors in the sentence.

Table I2: Final prep+N collocations and prep+N control phrases for Experiment 2

prep+N target collocations						prep+N control phrases					
	Prep freq band (K1 etc)	COCA phrse raw freq	COCA MI phrase	COCA raw freq exact phrs		Half-synonym/ control prep	Prep freq band (K1 etc)	COCA phrse raw freq	COCA MI phrase	COCA raw freq exact phrse	COCA raw freq exact phrse (real phrse)
off target	1	118	0.73	81		near target	1	27	0.72	2	0
by arrangement	1	205	1.07	110		in arrangement	1	376	-0.35	5	1
before an audience	1	348	1.83	162		beside an audience	1	0	NP	0	0
beyond belief	1	431	5.69	401		outside belief	1	4	-1.68	0	0
for comparison	1	1348	1.55	876		with comparison	1	120	-1.59	9	1
off course	1	766	0.92	471		near course	1	29	-1.69	0	0
in credit	1	899	-1.40	281		at credit	1	284	-1.18	72	0
under the deal	1	241	0.82	65		upon the deal	1	16	-1.36	0	0
on the decline	1	399	0.05	285		at the decline	1	35	-2.99	13	0
against expectations	1	127	0.93	9		outside expectations	1	61	0.93	3	0
at issue	1	1125	0.38	1666		over issue	1	820	0.93	1	0
over the limit	1	275	1.74	79		after the limit	1	14	-2.42	3	1
at a loss	1	1579	1.12	1210		of a loss	1	5051	0.41	141	0
against the odds	1	827	5.44	167		outside the odds	1	0	NP	0	0
by profession	1	337	1.29	168		for profession	1	426	0.42	3	0
under protest	1	53	1.60	46		with protest	1	166	-0.65	38	0
within reason	1	174	1.07	168		around reason	1	26	-2.86	1	0
in reply	1	606	0.63	455		at reply	1	25	-2.09	0	0
on review	1	559	-0.74	43		at review	1	170	-1.99	4	1
through tears	1	384	2.62	130		around tears	1	16	-1.46	0	0
under threat	1	754	3.98	434		within threat	1	4	-2.95	1	0
by turns	1	323	-0.73	232		with turns	1	203	-2.27	7	1
in warning	1	373	-1.47	138		at warning	1	72	-1.96	3	0
off work	1	1425	0.68	535		down work	1	487	-1.23	28	0

One noticeable difference between Table H2 and Table H1 is that the numbers of the exact phrase counts for the prep+N phrases are higher than for the V+N phrases (e.g., there are 171 occurrences of *of a loss* compared with 0 occurrences for *give a loss*). Many instances in COCA of some phrases (e.g., *in credit*) are not "real", for example, "…why change in credit rules is…", "…included almost $10,000 in credit card charges." All the prep+N target collocations were initially found in the *OCD* (2009) except *on review*, which was found only in COCA.

Appendix J

Early trial sentences for Experiment 2

Trial sentences 1: verb + noun

Complete each sentence by writing one verb in each gap. Each single space [_] stands for one letter. For each verb, one or more letters have been given. If you have no idea of a word, leave the spaces blank.

1. We need to **me_ _ the obligations** that are written in law.

2. Our manager is going to **_e_ targets** for sales for the next 12 months.

3. At the conference the researchers will **p_ _ _ _ _t data** on child health.

4. The temperature should **r_ _ _ _ a maximum** of 20 degrees today.

5. By surviving a very serious illness she **b_ _ _ the odds**.

6. The managers will **pe_f_ _ _ a review** of the employees next month.

7. The New Zealand government hopes to **e_pa_ _ trade** with China and sell more New Zealand products there.

8. These worrying questions **r_ _ _e doubts** about whether he would be a good leader.

9. Because she was wearing a seat belt, she managed to **a_ _ _ _ injury** in the crash, and she walked away unhurt.

10. We all **f_ _m expectations** about how good leaders should behave.

11. We've been talking to the other company for weeks but now is the time to **cl_ _ _ the deal**.

12. The workers have stopped working and will **sta_e a protest** about their pay this afternoon.

13. I hope he'll finally **s_ _ reason** and start acting maturely.

14. Describing the characters in more detail will **_ _d depth** to your story.

15. If you don't have enough money to pay your bills, your bank may **e_t_ _d credit** to you.

16. I always respect what she says because I **p_ _ _e value** on her opinions.

17. I **_ _ _d the view** that the government should take care of poor people.

18. I've heard a thief is operating in the neighbourhood; we need to **po_ _ a warning** about that.

19. If sales are slow, the company may have to **t_ _ _ a loss** on this product.

20. Many people believe that having 8 as their house number will **b__ __ __ __ luck**.

21. It's impossible to **r__a__ __ perfection**, but we should do our best.

22. The company has begun to **s__ __fe__ a decrease** in its sales.

23. To get good marks for the essay, you should **f__ __l__ __ the guidelines**.

24. In this difficult time, they need to **d__ __ __ strength** from the support of their families.

25. By doing these exercises every day you will **b__ __ __d strength** in your body.

26. When she heard about the queen's death, she **s__ __ __ tears**.

27. That company's new product could **pr__ __e__ __ a threat** to sales of our own products.

28. When he leaves school, he wants to **f__l__ __ __ a trade** such as building or engineering.

29. The products were so popular that the manufacturer could not **m__ __ __ the demand** for them.

30. The American and Russian presidents are going to **__ __ __d discussions** about nuclear weapons next week.

31. You were absent yesterday; can you please **__i__ __ an explanation** why?

32. You should expect to **f__ __e difficulties** if you travel alone, but they may not be major.

33. Good financial advice will **__ __t a course** for a company's future financial success.

34. The loss of his job might **__ __t a strain** on their relationship.

35. When we think about this situation and one that happened last year, we can **d__a__ a comparison** between them.

36. Her actions have begun to **r__i__ __ suspicions** that she can't do her job well.

37. He is relaxed and is able to **h__ __ __ __e stress** easily.

38. You shouldn't be arrested if you **f__ __l__ __ the law**.

39. He exercises hard and always tries to **p__ __ __ the limits** of his body.

40. He plans to attend university and then **e__ __e__ a profession** such as law or accountancy.

41. They have been arguing about this issue for a long time, so I hope they can **se__ __ __ __ the dispute** soon.

42. The job of the student advisor is to **__i__ __ guidance** to students who need it.

43. The company wants to **st__ __ __e a deal** with the government to buy the national railway system for a cheap price.

44. If unemployment increases, the economy may **su__ __ __r a decline**.

45. Your workers will work more efficiently if you **b__ __ __d confidence** in them.

46. You'll achieve your academic goal if you **fo__ __ __ __ the course** of study that's described here.

47. The directors will soon **s__ __ the budget** for the year; less spending will be allowed this year.

48. This medicine should **r__ __i__ __e the pain**.

49. If we work faster we might be able to **b__ __ __ the clock** and finish early.

50. Tomorrow we need to **__ __ __e arrangements** for Grandma's funeral.

51. If we get a famous band to play the concert, that could **d__ __ __ an audience** of 10,000.

52. Social media are able to **r__ __ __h an audience** more effectively than some other types of advertising.

53. They **__o__d beliefs** that I strongly disagree with.

54. The two of us **s__ __ __ __ a belief** in democracy.

55. If that company doesn't work more efficiently, it will **l__ __ __ the contract** it has with the government.

56. She felt so angry that she couldn't **o__ __ er a reply** to his comment.

57. Even though many students hate them, tests do **s__ __ __e a purpose**.

58. Introducing lower taxes is a good idea because it would **s__ __ __en the impact** of the economic crisis on businesses.

59. We should **r__ __ __e the issue** at tomorrow's meeting because the issue is very important.

60. She needs to **t__ __ __ command** of the situation and show the workers that she is the boss.

61. The doctor will need to **p__r__ __r__ an examination** on your chest.

62. Lowering taxes could **m__r__ a turn** in New Zealand's economic situation.

63. The boss will often **c__ __ __ __e work** for us to do, but it's unnecessary work!

64. I'm not sure whether I should **e__ __ __r the debate** and give my opinion.

Trial sentences 2: preposition + noun

Complete each sentence by writing a preposition of <u>one</u> word in each gap. For some prepositions, one letter has been given. No clues are given about how many letters each preposition has.

Example: The fish are __*below*__ the surface of the lake.

Example: She jumped ___*onto*___ the table and danced on it.

1. The company made a loss last year and is now _____ **difficulty**.

2. We are _____ **an obligation** to pay our company tax in four weeks.

3. _____**n explanation** of his absence he told us that he had been visiting his dying mother in Australia.

4. Good news: we're _____ **target** to get our sales goals.

5. **O**_____ **consideration**, she has probably made a good decision.

6. She was married _____ **arrangement** to a friend of her family.

7. The band played **b**_____ **an audience** of 5,000 people.

8. What you did is **b**_____ **belief** – I can't understand why you did it.

9. This year we have limited our spending, so we are **w**_____ **budget**.

10. Their marriage broke apart _____ **the strain** of John's depression and anger.

11. To finish the project, we had to work _____**d the clock**: 24 hours a day!

12. She is a senior police officer, _____ **command** of a group of 10 other officers.

13. You should look at the results of both groups _____**r comparison.**

14. He is _____ **contract** to the government to do website design work for government departments.

15. The company is pleased that it is _____ **course** to make a large profit this year.

16. We are currently _____**n credit**: in other words, we don't owe the bank any money.

17. New Zealand can sell many of its products in Australia without trade tax _____ **the deal** with the Australian government.

18. Whether female soldiers should be allowed to fight in battle is the issue _____ **debate**.

19. New Zealand's economy is currently _____ **decline**.

20. I need to examine this issue _____ **depth** before I make a decision.

21. He is **b_____ doubt** an excellent manager; in fact, he's probably the best in the company.

22. **_____n examination**, his arm was found to be broken.

23. The sales figures for February were **b_____ expectations**, which is disappointing.

24. The top student failed half his exams, which was **_____t expectations**.

25. Workers must always wear safety equipment **_____ the guidelines**.

26. They died **_____ impact** in the car crash.

27. The professional football player ended his career **t_____ injury**.

28. The crash is currently **_____ investigation**.

29. The ideas **a_____ issue** need to be considered carefully.

30. Murder is obviously **_____ the law**.

31. He drank too much, was caught driving **_____ the limit** and lost his driver's licence.

32. I'm **_____ a loss** to explain why she did that -- it's hard to understand.

33. He crossed his fingers **_____ luck**.

34. We're **_____ luck**: I've managed to get two tickets to tomorrow's All Blacks game!

35. I believe employees should work **_____ the maximum** of their ability.

36. Because my parents paid for my studies, I'm **_____ an obligation** to them to get a good degree.

37. She survived cancer **_____ the odds**—the doctors said she was going to die.

38. The meal had been cooked **_____ perfection**.

39. She's a nurse **_____ profession**.

40. He did what he was told to do, but only **_____r protest**.

41. I don't think that was an accident: I think you did it **_____ purpose**.

42. I can do anything you like if it's **_____ reason**.

43. She told him, **_____n reply** to his question, that she was very unhappy.

44. The new payment system is **_____ review**.

45. The New Zealand Army is **_____r strength** by 5,000 soldiers, but the army is advertising for more soldiers next month.

47. When she heard of the singer's death, she was **_____ tears**.

48. The teaching of European languages in New Zealand may be **_____ threat** as fewer students want to learn them.

49. He's a builder **_____ trade**.

50. It's been a hard week at work, and I'm **_____ stress** at the moment.

51. Our house has increased **_____ value** by $30,000 in the last year.

52. Paintings by that painter are now **o_____ view** in the art shop.

53. When a stranger came close to the house, the dog barked **_____n warning**.

54. Because he has been very sick this week, he's been **o_____ work**.

55. He was, **b_____ turns**, sad and angry at the news.

Appendix K

Treatment materials in Experiment 2 (for one of the two experimental groups)

Name: _____

Text 1

Our company made a financial loss last year, a result that was **against expectations**.
On review, it is clear that one of the reasons it happened was that the economy began to suffer a decline. Another reason was that we had to **take a loss** on one of our major products. As a result, the company's yearly income was off target.

Yesterday our directors stood before an audience of share holders at the company's annual general meeting. I had heard that some share holders were going to **stage a protest** about the loss outside the building where the meeting was, but that did not happen. At the meeting, an elderly share holder was the first person to raise the issue of the loss of income. **In reply**, the directors said that they plan to **set a course** next year for the company which will mark a turn in the company's financial direction. They expect the company to make a profit next year and to be in credit with the bank.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Our company is expected to lose money next year.

a. True b. False

Name: _____

Text 2

In recent years it has been financially difficult for university departments, some of which have been under threat of being closed. Often courses and departments are closed after university directors **perform a review**. The directors may find that the department is failing to meet targets of student numbers and that some or many courses are losing money. Often the lecturers will say **in reply** to the directors that the subjects that they teach should be kept because they are needed by students. The staff may **post a warning** about the directors' plans and **stage a protest**. At that meeting, lecturers will stand before an audience of staff and students and ask the university to see reason and not to close the department or the courses. Occasionally academic staff and directors will **strike a deal** which includes saving one of the courses **against the odds**. But courses are usually closed, and staff lose their jobs; staff will often feel sad and angry by turns and may comfort each other through tears.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Some university courses are closed because they lose money.

a. True b. False

Name: _____

Text 3

New Zealand politicians make laws in Parliament's House of Representatives—commonly called the House. However, politicians also often argue there, and discussions about laws can go **off course** and onto other topics. Many politicians are lawyers **by profession** and good public speakers who try to reach an audience. They often **draw a comparison** between their own political party and opposing parties to make their party seem better. And, **in warning**, they often tell the public about the weaknesses of other parties and attack each other with words. Usually the other parties will **offer a reply** by attacking the first party. Sometimes a politician will **push the limits** of the House's rules and call another politician a bad name (e.g., a liar)—he may have to leave the House, often **under protest**. This behaviour is beyond belief for many members of the public (who visit the House by arrangement). Consequently, people's respect for politicians can suffer a decline. Every three years the government will **perform a review** of the House's rules.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Politicians in the New Zealand House of Representatives (Parliament) always speak politely to each other.

a. True b. False

Name: _____

Text 4

Applying for jobs is stressful, especially when you first **enter the profession** you have chosen and you have little work experience. Getting a good job can seem **against the odds**. If you are accepted for a job interview, you may need to make arrangements to be off work from your present job for a few hours to attend the interview. During that first interview, it is best not to raise the issue of pay. If an employer asks you what salary you would like, **offer a reply** that is general—you can discuss details in a later interview. Even then, do not **push the limits** by being greedy: you should ask for a salary that is within reason. You can check the salaries of similar jobs **for comparison**. When you are offered the job, that is the best time to discuss the pay you will receive **under the deal** with your new employer. After everything has been agreed, you may feel happy beyond belief and shed tears of joy.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

The writer suggests looking at the pay of similar jobs.

a. True b. False

Name: _____

Text 5

Many people **form expectations** of buying their own house, but plans to buy can go off target because of high house prices. To get enough money to buy their first house, most people need a bank to extend credit to them in the form of a loan. A first step is to talk to a bank staff member, by arrangement if you need to. The staff member will **perform a review** of your finances to see how much you can borrow. When looking for a loan, you should **draw a comparison** between the different banks' rates of payment—some will be lower than others. If you are allowed a loan, before you **strike a deal** with a bank, do not borrow **over the limit** that you can afford: borrow within reason. If the banks refuse to give you a loan, you may feel **at a loss** and shed tears. But if you succeed, congratulations! Take some time off work and celebrate: you are one step closer to buying your house.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

The payment rates for loans are the same for all banks.

a. True b. False

Name: _____

Text 6

Unemployment often increases when a country's economy is on the decline. Two periods in New Zealand history when many people lost jobs were after share market "crashes" (big falls in share prices). After the 1929 crash, thousands of workers became unemployed and were **at a loss** regarding where to find work. Workers got together to **stage a protest** in Wellington and were able to reach an audience of thousands. **In warning**, they told the government to listen to them. The government did not **offer a reply** at the meeting, but it did create work for the jobless.

Another period when many jobs were under threat was after the 1987 crash. The government sold some government departments to private companies. Those sales did **push the limits** of public acceptance. At issue was whether it was acceptable to sell publicly-owned organisations. Many people still hold the belief that it was not. **Under the deal** to sell one department, the railways, a private company ran the railways while the government owned the land.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Thousands of New Zealanders lost their jobs after the 1929 share market crash.

a. True b. False

Name: _____

Text 7

A general election in New Zealand is an important event. What is at issue is who the public should choose to govern the country.

On election night, the political party leaders stand before an audience of their supporters. If their party managed to **beat the odds** and succeed **against expectations**, the leaders may laugh and cry by turns and give a speech, sometimes through tears. In the speech they thank their supporters, but they do not usually **draw a comparison** between their own party and opposing parties. Their supporters clap and cheer **in reply**.

However, the party that gets the most votes on election day may not have enough votes to govern the country alone. They may have to **strike a deal** with another party to form a government.

After each election, a number of people **enter the profession** of politics for the first time. They also create work for their new office staff. However, they make arrangements to employ these workers the following week – after election celebrations have finished.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

The political party that gets the most votes in a New Zealand election automatically becomes the government.

a. True b. False

Name: _____

Text 8

You can own your own business even if you are not an accountant **by profession**. Starting your own business can be, by turns, frustrating and satisfying. Many businesses fail—about 50% fail within 5 years—but many **beat the odds** and succeed. Yours can succeed, too.

To have a successful business, make careful plans to **set a course** for the future. You may need to meet your bank manager by arrangement to ask her to extend credit to you. Another important thing to do is marketing: you must reach an audience who will buy your products or services. You should also make and meet targets for sales of your products and services. And pricing is also important: do not **take a loss** on sales. Finally, it will be useful to talk to other business owners about their businesses **for comparison**.

If your business is not on the decline after 5 years, congratulate yourself because you have done what is beyond belief for most people: owning a successful business.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Nearly all businesses fail within five years.

a. True b. False

Name: _____

Text 9

Problem gambling is a harmful addiction in which a person bets a lot of money often. He will try to **beat the odds** in order to win money but, instead, he will often **take a loss** on his bets – again and again. Such a person does not know when to stop gambling, and may take time off work to bet. The problem will present a threat to the person's finances, his job and his family life as he will often spend money **over the limit** that he can afford. His bank account is unlikely to be in credit. It may be difficult to raise the issue of his problem with him, and he might only accept help **under protest**. He may not see reason until he has lost his money, his possessions and his family, and, through tears, decides to change his life. **On review** of his actions, he may understand his problem better and decide to stop betting. Hopefully, this will mark a turn in his life.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Problem gamblers can easily stop their gambling.

a. True b. False

Name: _____

Text 10

Share markets are an important part of modern economies, but many people are **at a loss** to understand why. Sometimes share markets suddenly suffer a decline that seems to be **against the odds**. A large, dramatic drop in the prices of many shares—or "crash" —can present a threat to a country's economy and push it **off course**. It also leads to many job losses, causing people to shed tears. Many people hold the belief that economies should act within reason, but we need to **form expectations** of share markets that are realistic. Share prices are, in fact, often determined by emotion, not logic—they rise and fall according to people's feelings. Sometimes, when share prices are very high, experts (economists or stock brokers **by profession**) **post a warning** that the market might crash. A crash is a good time for a company to take over (buy) another company by buying most of its shares. **Under the deal**, the first company can replace the management of the second company.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

A share market crash causes many people to lose their jobs.

a. True b. False

Name: _____

Text 11

Balancing the budget—or keeping personal finances in order—can be difficult for many people, but especially for those who hold the belief that spending money is better than saving it. Their bank accounts may often be **over the limit**, which puts their financial plans **off course** and their finances under threat. If you are careless with money, it is a good idea to see reason and to **form expectations** of your finances that are realistic. First, you should find out what your total income is and exactly how you spend your money every week, month and year. If you think it is necessary, you may want to make arrangements with your bank to extend credit to you. You should also check the borrowing options of other banks **for comparison**. **In warning**, I should say that if you do borrow money, you need to make sure that can meet targets of payment; otherwise you may have to pay extra money to your bank, which you would probably do under protest.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

To keep control of your finances, you need to know what your income is.

a. True b. False

Name: _____

Text 12

If you **enter the profession** of economics, you could help your country's government manage the economy. When the economy has a surplus, that is similar to a bank account being in credit. But an economy that is on the decline and is in debt can present a threat to people's jobs. What is at issue for a government is the best way to encourage economic growth. A government will usually try to **set a course** for a failing economy that will increase its growth and mark a turn in the economy's fortunes.

Sometimes the economy improves or weakens **against expectations**. If economists believe that the economy is going to weaken, they often **post a warning** to businesses and the general public. If it does weaken, governments will make efforts to improve it. **On review**, we can see now that some plans are successful and some are more off target. One plan of the New Zealand government in the 1980s was to create work for the unemployed; this was partly successful.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Governments try to improve weak economies.

a. True b. False

Name: _____

Text 13

In recent years it has been financially difficult for university departments, some of which have been under threat of being closed. Often courses and departments are closed after university directors **perform a review**. The directors may find that the department is failing to meet targets of student numbers and that some or many courses are losing money. Often the lecturers will say **in reply** to the directors that the subjects that they teach should be kept because they are needed by students. The staff may **post a warning** about the directors' plans and **stage a protest**. At that meeting, lecturers will stand before an audience of staff and students and ask the university to see reason and not to close the department or the courses. Occasionally academic staff and directors will **strike a deal** which includes saving one of the courses **against the odds**. But courses are usually closed, and staff lose their jobs; staff will often feel sad and angry by turns and may comfort each other through tears.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Lecturers attend protests to support the closing of university departments.

a. True b. False

Name: _____

Text 14

Applying for jobs is stressful, especially when you first **enter the profession** you have chosen and you have little work experience. Getting a good job can seem **against the odds**. If you are accepted for a job interview, you may need to make arrangements to be off work from your present job for a few hours to attend the interview. During that first interview, it is best not to raise the issue of pay. If an employer asks you what salary you would like, **offer a reply** that is general -- you can discuss details in a later interview. Even then, do not **push the limits** by being greedy: you should ask for a salary that is within reason. You can check the salaries of similar jobs **for comparison**. When you are offered the job, that is the best time to discuss the pay you will receive **under the deal** with your new employer. After everything has been agreed, you may feel happy beyond belief and shed tears of joy.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

The writer recommends that you discuss the details of your pay in your first job interview.

a. True b. False

Name: _____

Text 15

Our company made a financial loss last year, a result that was **against expectations**.
On review, it is clear that one of the reasons it happened was that the economy began to suffer a decline. Another reason was that we had to **take a loss** on one of our major products. As a result, the company's yearly income was off target.

Yesterday our directors stood before an audience of share holders at the company's annual general meeting. I had heard that some share holders were going to **stage a protest** about the loss outside the building where the meeting was, but that did not happen. At the meeting, an elderly share holder was the first person to raise the issue of the loss of income. **In reply**, the directors said that they plan to **set a course** next year for the company which will mark a turn in the company's financial direction. They expect the company to make a profit next year and to be in credit with the bank.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

The company made a profit last year.

a. True b. False

Name: _____

Text 16

Unemployment often increases when a country's economy is on the decline. Two periods in New Zealand history when many people lost jobs were after share market "crashes" (big falls in share prices). After the 1929 crash, thousands of workers became unemployed and were **at a loss** regarding where to find work. Workers got together to **stage a protest** in Wellington and were able to reach an audience of thousands. **In warning**, they told the government to listen to them. The government did not **offer a reply** at the meeting, but it did create work for the jobless.

Another period when many jobs were under threat was after the 1987 crash. The government sold some government departments to private companies. Those sales did **push the limits** of public acceptance. At issue was whether it was acceptable to sell publicly-owned organisations. Many people still hold the belief that it was not. **Under the deal** to sell one department, the railways, a private company ran the railways while the government owned the land.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

After the 1987 share market crash, the government continued to own railways land while a private company ran the railways.

a. True b. False

Name: _____

Text 17

A general election in New Zealand is an important event. What is at issue is who the public should choose to govern the country.

On election night, the political party leaders stand before an audience of their supporters. If their party managed to **beat the odds** and succeed **against expectations**, the leaders may laugh and cry by turns and give a speech, sometimes through tears. In the speech they thank their supporters, but they do not usually **draw a comparison** between their own party and opposing parties. Their supporters clap and cheer
in reply.

However, the party that gets the most votes on election day may not have enough votes to govern the country alone. They may have to **strike a deal** with another party to form a government.

After each election, a number of people **enter the profession** of politics for the first time. They also create work for their new office staff. However, they make arrangements to employ these workers the following week – after election celebrations have finished.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Political party leaders in New Zealand thank their supporters on election night.

a. True b. False

Name: _____

Text 18

Balancing the budget—or keeping personal finances in order—can be difficult for many people, but especially for those who hold the belief that spending money is better than saving it. Their bank accounts may often be **over the limit**, which puts their financial plans **off course** and their finances under threat. If you are careless with money, it is a good idea to see reason and to **form expectations** of your finances that are realistic. First, you should find out what your total income is and exactly how you spend your money every week, month and year. If you think it is necessary, you may want to make arrangements with your bank to extend credit to you. You should also check the borrowing options of other banks **for comparison**. **In warning**, I should say that if you do borrow money, you need to make sure that can meet targets of payment; otherwise you may have to pay extra money to your bank, which you would probably do under protest.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

There is no need to compare the borrowing options of different banks.

a. True b. False

Name: _____

Text 19

Many people **form expectations** of buying their own house, but plans to buy can go off target because of high house prices. To get enough money to buy their first house, most people need a bank to extend credit to them in the form of a loan. A first step is to talk to a bank staff member, by arrangement if you need to. The staff member will **perform a review** of your finances to see how much you can borrow. When looking for a loan, you should **draw a comparison** between the different banks' rates of payment—some will be lower than others. If you are allowed a loan, before you **strike a deal** with a bank, do not borrow **over the limit** that you can afford: borrow within reason. If the banks refuse to give you a loan, you may feel **at a loss** and shed tears. But if you succeed, congratulations! Take some time off work and celebrate: you are one step closer to buying your house.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

The writer recommends that people who want a bank loan for a house first talk to a worker at a bank.

a. True b. False

Name: _____

Text 20

If you **enter the profession** of economics, you could help your country's government manage the economy. When the economy has a surplus, that is similar to a bank account being in credit. But an economy that is on the decline and is in debt can present a threat to people's jobs. What is at issue for a government is the best way to encourage economic growth. A government will usually try to **set a course** for a failing economy that will increase its growth and mark a turn in the economy's fortunes. Sometimes the economy improves or weakens **against expectations**. If economists believe that the economy is going to weaken, they often **post a warning** to businesses and the general public. If it does weaken, governments will make efforts to improve it. **On review**, we can see now that some plans are successful and some are more off target. One plan of the New Zealand government in the 1980s was to create work for the unemployed; this was partly successful.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

All government plans to improve economies are successful.

a. True b. False

Name: _____

Text 21

Problem gambling is a harmful addiction in which a person bets a lot of money often. He will try to **beat the odds** in order to win money but, instead, he will often **take a loss** on his bets—again and again. Such a person does not know when to stop gambling, and may take time off work to bet. The problem will present a threat to the person's finances, his job and his family life as he will often spend money **over the limit** that he can afford. His bank account is unlikely to be in credit. It may be difficult to raise the issue of his problem with him, and he might only accept help **under protest**. He may not see reason until he has lost his money, his possessions and his family, and, through tears, decides to change his life. **On review** of his actions, he may understand his problem better and decide to stop betting. Hopefully, this will mark a turn in his life.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Problem gamblers spend more money on gambling than they can afford.

a. True b. False

Name: _____

Text 22

You can own your own business even if you are not an accountant **by profession**. Starting your own business can be, by turns, frustrating and satisfying. Many businesses fail—about 50% fail within 5 years—but many **beat the odds** and succeed. Yours can succeed, too.

To have a successful business, make careful plans to **set a course** for the future. You may need to meet your bank manager by arrangement to ask her to extend credit to you. Another important thing to do is marketing: you must reach an audience who will buy your products or services. You should also make and meet targets for sales of your products and services. And pricing is also important: do not **take a loss** on sales. Finally, it will be useful to talk to other business owners about their businesses **for comparison**.

If your business is not on the decline after 5 years, congratulate yourself because you have done what is beyond belief for most people: owning a successful business.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Marketing is not usually needed for the success of a business.

a. True b. False

Name: _____

Text 23

Share markets are an important part of modern economies, but many people are **at a loss** to understand why. Sometimes share markets suddenly suffer a decline that seems to be **against the odds**. A large, dramatic drop in the prices of many shares—or "crash" —can present a threat to a country's economy and push it **off course**. It also leads to many job losses, causing people to shed tears. Many people hold the belief that economies should act within reason, but we need to **form expectations** of share markets that are realistic. Share prices are, in fact, often determined by emotion, not logic—they rise and fall according to people's feelings. Sometimes, when share prices are very high, experts (economists or stock brokers **by profession**) **post a warning** that the market might crash. A crash is a good time for a company to take over (buy) another company by buying most of its shares. **Under the deal**, the first company can replace the management of the second company.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Rises and falls in share prices are caused by people acting emotionally.

a. True b. False

Name: _____

Text 24

New Zealand politicians make laws in Parliament's House of Representatives—commonly called the House. However, politicians also often argue there, and discussions about laws can go **off course** and onto other topics. Many politicians are lawyers **by profession** and good public speakers who try to reach an audience. They often **draw a comparison** between their own political party and opposing parties to make their party seem better. And, **in warning**, they often tell the public about the weaknesses of other parties and attack each other with words. Usually the other parties will **offer a reply** by attacking the first party. Sometimes a politician will **push the limits** of the House's rules and call another politician a bad name (e.g., a liar)—he may have to leave the House, often **under protest**. This behaviour is beyond belief for many members of the public (who visit the House by arrangement). Consequently, people's respect for politicians can suffer a decline. Every three years the government will **perform a review** of the House's rules.

[The following information appears overleaf.]

Now answer a question about the text. Circle *a* (true) or *b* (false). Do <u>NOT</u> look back at the text on the previous page.

Sometimes New Zealand politicians call other politicians bad names in the House.

a. True b. False

Appendix L

Stimulus sentences and post-stimulus distractor questions for test group A in self-paced reading task

Sentence	Question
Block 1: V+N	
She is really fast at learning new English words.	Does she learn English vocabulary quickly?
He talked to her on the street last week.	Did he speak to her last week?
Rising prices will *produce a threat* to the economy.	Is the sentence about prices that are increasing?
Our company might **suffer a decline** this financial year.	Is the sentence about something that happened last year?
We have to **post a warning** about that man.	Is the sentence about a woman?
Her manager is quite bossy and unpleasant to her.	Is her manager nice to her?
Sue will definitely *beat the odds* if she tries.	If Sue tries to do something, will she fail?
We need to *sort arrangements* for Jill's big wedding.	Will Jill's wedding be large?
There is a strong wind in the city today.	Is it windy in the city?
The directors will *map a course* for their company.	Is the sentence about the government?
He is visiting his sick mother over the holidays.	Is he visiting his friend?
She wants to **enter the profession** of commercial law.	Does she wish to do something?
Bob could not *spare a reply* to Jill's question.	Did Jill ask a question?
He wants to find a better job very soon.	Does he want another job?
They have to *stir an audience* to sell insurance.	Are they trying to sell insurance?
The shop will **take a loss** on these bananas.	Is the sentence about a type of fruit?
Workers plan to *effect a protest* in town today.	Is the sentence about something that's happening tomorrow?
This year our company is doing very well financially.	Is our company now financially successful?
The government will *develop work* for many jobless people.	Is the sentence about jobs for people who already have jobs?
The sisters will *lose tears* after their mother's death.	Is the sisters' mother still alive?
Universities get a lot of funding from the government.	Do universities get money from the government?
The bank won't **extend credit** to Susan and John.	Is the sentence about a bank and two people?
We would like to have a baby next year.	Do we want a child this year?
He hopes they'll *know reason* and start acting wisely.	Are they acting wisely now?
He will really **push the limits** if he cheats.	Is it possible that he will cheat?
His new sales job does not pay very well.	Does he earn a lot of money in his job?
That job may **mark a turn** in his life.	Is the sentence about economic growth?
She's able to *spot a comparison* between the twins.	Is the sentence about something she can do?
Good managers usually **form expectations** that are quite high.	Is the sentence about something that good managers generally do?
She is looking for a teaching job in Wellington.	Does she want a nursing job?
Most people here **hold the belief** that war is bad.	Is the sentence about work?
Newspapers have not been selling well in recent years.	Have newspaper sales been good recently?
They will visit him at his house this evening.	Will they see him this evening?
We need to **meet targets** of strong financial growth.	Is the sentence about money?
That company may **strike a deal** with the government.	Is the sentence about what many companies might do?
Their sons play football every Saturday morning.	Do their sons play sport?

The managers will *operate a review* of the department.	Are students going to do something with the department?
She will probably **raise the issue** at the meeting.	Is the sentence about flying?
He's going to meet her at the restaurant later.	Will he see her at the library?
He was an engineer but is now a teacher.	Does he still work as an engineer?
Block 2: prep+N	
That's one of the best restaurants in the city.	Is that a good restaurant?
The meeting lasted for four and a half hours.	Was the meeting longer than four hours?
She said clearly *by reply* that she would visit.	Did she promise to visit?
Our jobs are *within threat* from the company's managers.	Is the sentence about our cars?
The economy is **on the decline** at this time.	Is the sentence about movies?
He is able to play the electric piano well.	Is he good at playing the electric piano?
Today she stood *beside an audience* and sang beautifully.	Did she sing badly?
She has borrowed **over the limit** she can afford.	Has she borrowed money?
She worked for 12 hours and felt very tired.	After work, did she have a lot of energy?
He won races **against expectations** and then started laughing.	Did he cry?
She has been working at the university since 2001.	Does she have a job at the city council?
What is **at issue** now is whether to study.	Is the sentence about travel?
Their finances were *near course* over the last year.	Is the sentence about money?
She's a doctor **by profession** but has no job.	Is she working at the moment?
The new medicine has reduced some of his pain.	Does he feel less pain now?
We're visiting him *in arrangement* at ten o'clock tomorrow.	Will we see him tomorrow?
She laughed loudly *around tears* and then stopped suddenly.	Did she slowly stop laughing?
They plan to sell their house later this month.	Will they want someone to buy their house?
He told her, **in warning**, to run very fast.	Was she told to run slowly?
He followed her with protest but then he disappeared.	After he followed her, did he leave her?
She works at a small printing company in Wellington.	Is the printing company in Auckland?
She is an extremely hard working high school student.	Is she a lazy student?
Tom's actions were **beyond belief** to all his friends.	Is the sentence about the things that Tom did?
The car always takes a long time to start.	Does the car start quickly?
It does seem, *at review*, that she was right.	Do we think that she was probably correct?
The government will soon introduce a new financial law.	Will the new law be introduced in a short time?
You should act *around reason* when you're at work.	Is the sentence about doing something at home?
The financial benefits **under the deal** were very large.	Were the benefits very big?
She used to be a well known pop singer.	Did many people know her?
His account is **in credit** at the present time.	Is the sentence about a situation that is true now?
She is often **at a loss** to understand John.	Is the sentence about reading a book?
The building was badly damaged in last night's storm.	Did the storm happen this afternoon?
Yesterday he was **by turns** confused and very pleased.	Was he happy?
She showed us, *with comparison*, the other two tables.	Did we see only one table?
He is returning to his home town on Monday.	Is he going back on Tuesday?
He finished first *outside the odds* and felt happy.	Did he feel disappointed?

Our plans are **off target** at the present moment.	Is the sentence about what our plans are like now?
Today he is *down work* because he is sick.	Is he feeling ill?
David's job is one of the most boring ever.	Does David have an interesting job?
He recently became one of the company's senior directors.	Is he now a senior company director?

Notes. Typographic enhancement has been added to this table to clarify the design. Target collocations are bolded and control phrases are italicised. Typographic enhancement was not used in the stimuli presentation in SPRT.

Appendix M

Gapfill test for Experiment 2

Name: _____

*The following sentences contain phrases (in **bold type**) that were in the reading texts that you read before. The first word of each phrase has been removed. Read each sentence and fill the space with the correct word. Write only ONE word in each space.*

1. Many people _____ **the belief** that the government should not sell government departments.

2. To get a bank loan, you will first have to _____ **arrangements** to meet a banker.

3. If you are a business owner, it is important that your business does not _____ **a loss** on sales.

4. Until he has lost all his money, a person who gambles too much may not _____ **reason** and stop gambling.

5. Some people are going to _____ **a protest** about the new law outside the Parliament buildings.

6. Politicians often _____ **a comparison** between their own political party and opposing parties to make their party seem better.

7. About 50% of businesses fail in their first few years, but many others _____ **the odds** and succeed.

8. For your business to be successful, you must _____ **an audience** who will buy your products.

9. Sometimes a politician will _____ **the limits** of Parliament's rules and call another politician a bad name.

10. A government will usually try to _____ **a course** for an economy that will increase the economy's growth.

11. When a person who gambles too much decides to stop gambling, that will _____ **a turn** in his life.

12. You could help your country's government manage the economy if you _____ **the profession** of economics.

13. When people lose their jobs, they sometimes _____ **tears**.

14. The political party that gets the most votes on election day may have to _____ **a deal** with another party to become the government.

15. People often lose their jobs when the economy begins to _____ **a decline**.

16. Sometimes university directors close a university department after they _____ **a review** of it.

17. One plan of the New Zealand government in the 1980s was to _____ **work** for the unemployed.

18. During the first interview for a job, it is best not to _____ **the issue** of pay.

19. If an employer asks you what salary you would like,_____ **a reply** that is general.

20. An economy that is in debt will often _____ **a threat** to people's jobs.

21. Businesses should make and then _____ **targets** for sales of their products and services.

22. Many people _____ **expectations** of buying their own house, but some are not able to afford one.

23. You will need to meet your bank manager if you want the bank to _____ **credit** to you.

24. Sometimes, when share prices are very high, experts _____ **a warning** that the market might crash.

25. When you find a job you would like, you should check the salaries of other similar jobs _____ **comparison**.

26. On election night, the political party leaders stand _____ **an audience** of their supporters and speak to them.

27. Many jobs were _____ **threat** after the 1987 share market crash.

28. People's plans to buy their own house can go _____ **target** when house prices increase quickly.

29. _____ **review** of his actions, he may understand his problem better and change his life.

30. Many politicians are lawyers _____ **profession**.

31. Members of the public can visit the Parliament building only _____ **arrangement**.

32. When a politician acts badly in Parliament, he may have to leave the House, and he often leaves _____ **protest**.

33. Many people think that share markets should act _____ **reason**, but in fact they often don't.

34. Sometimes the economy improves or weakens _____ **expectations**.

35. When you finally get a good job, you may feel happy _____ **belief**.

36. What is _____ **issue** for a government is the best way to encourage economic growth.

37. A person who gambles too much will often spend money _____ **the limit** that he can afford.

38. When people lose their jobs, they are often _____ **a loss** regarding where to find another job.

39. The directors expect the company to make a profit next year and to be _____ **credit** with the bank.

40. Politicians argue a lot in the New Zealand Parliament, so discussions about laws can sometimes go _____ **course**.

41. _____ **warning**, politicians often tell the public about the weaknesses of other politicians.

42. To attend a job interview, you may need to be _____ **work** for a few hours.

43. On election night, politicians thank their supporters, who clap and cheer _____ **reply**.

44. When people lose their jobs, they will often feel sad and angry _____ **turns**.

45. If you have little work experience, getting a good job can seem _____ **the odds**.

46. Unemployment often increases when a country's economy is _____ **the decline**.

47. The pay you'll receive in your new job, _____ **the deal** with your new employer, seems quite low.

48. People who lose their jobs may comfort each other _____ **tears**.

Appendix N

Yes/no test for target collocations in Experiment 2

Name: _____

Put a tick (√) next to each phrase that you had read (or heard) before you took part in this study. If you hadn't read (or heard) the phrase before, leave the space empty.

___ present a threat
___ under threat
___ suffer a decline
___ on the decline
___ post a warning
___ in warning
___ beat the odds
___ against the odds
___ make arrangements
___ by arrangement
___ set a course
___ off course
___ enter the profession
___ by profession
___ offer a reply
___ in reply
___ reach audiences
___ before an audience
___ take a loss
___ at a loss
___ stage a protest
___ under protest
___ create work
___ off work
___ shed tears
___ through tears
___ extend credit
___ in credit
___ see reason
___ within reason
___ push the limits
___ over the limit
___ mark a turn
___ by turns
___ hold the belief
___ beyond belief
___ meet targets
___ off target
___ strike a deal
___ under the deal
___ perform a review
___ on review
___ raise the issue
___ at issue
___ draw a comparison
___ for comparison
___ form expectations
___ against expectations

Appendix O

Collocations lists for treatment conditions

Treatment condition 1 collocations			**Treatment condition 2 collocations**	
V+N collocations bolded (12)			**V+N collocations unbolded (12)**	
stage	**a protest**		stage	a protest
offer	**a reply**		offer	a reply
push	**the limits**		push	the limits
perform	**a review**		perform	a review
post	**a warning**		post	a warning
strike	**a deal**		strike	a deal
take	**a loss**		take	a loss
set	**a course**		set	a course
form	**expectations**		form	expectations
draw	**a comparison**		draw	a comparison
enter	**the profession**		enter	the profession
beat	**the odds**		beat	the odds
V+N collocations unbolded (12)			**V+N collocations bolded (12)**	
reach	an audience		**reach**	**an audience**
create	work		**create**	**work**
hold	the belief		**hold**	**the belief**
meet	targets		**meet**	**targets**
see	reason		**see**	**reason**
suffer	a decline		**suffer**	**a decline**
raise	the issue		**raise**	**the issue**
mark	a turn		**mark**	**a turn**
extend	credit		**extend**	**credit**
shed	tears		**shed**	**tears**
make	arrangements		**make**	**arrangements**
present	a threat		**present**	**a threat**

prep+N collocations bolded (12)			**prep+N collocations unbolded (12)**	
under	**protest**		under	protest
in	**reply**		in	reply
over	**the limit**		over	the limit
on	**review**		on	review
in	**warning**		in	warning
under	**the deal**		under	the deal
at	**a loss**		at	a loss
off	**course**		off	course
against	**expectations**		against	expectations
for	**comparison**		for	comparison
by	**profession**		by	profession
against	**the odds**		against	the odds
prep+N collocations unbolded (12)			**prep+N collocations bolded (12)**	
before	an audience		**before**	**an audience**
off	work		**off**	**work**
beyond	belief		**beyond**	**belief**
off	target		**off**	**target**
beyond	reason		**beyond**	**reason**
on	the decline		**on**	**the decline**
at	issue		**at**	**issue**
by	turns		**by**	**turns**
in	credit		**in**	**credit**
through	tears		**through**	**tears**
by	arrangement		**by**	**arrangement**
under	threat		**under**	**threat**

Appendix P

Prior knowledge of target collocations in Experiment 2

Collocation	Gapfill (control group) /25	Previously read or heard /78	Overall score
make arrangements	0.56	0.90	1.46
hold the belief	0.40	0.72	1.12
over the limit	0.16	0.78	0.94
create work	0.16	0.71	0.87
under threat	0.28	0.58	0.86
in reply	0.12	0.72	0.84
meet targets	0.08	0.72	0.80
in credit	0.12	0.67	0.79
by arrangement	0.20	0.58	0.78
for comparison	0.16	0.58	0.74
suffer a decline	0.00	0.74	0.74
offer a reply	0.00	0.65	0.65
in warning	0.08	0.56	0.64
off work	0.04	0.60	0.64
raise the issue	0.00	0.63	0.63
within reason	0.00	0.60	0.60
set a course	0.00	0.60	0.60
on review	0.04	0.55	0.59
against expectations	0.04	0.55	0.59
on the decline	0.04	0.53	0.57
at a loss	0.04	0.53	0.57
draw a comparison	0.00	0.56	0.56
by turns	0.00	0.53	0.53
before an audience	0.04	0.49	0.53
off target	0.04	0.45	0.49
present a threat	0.00	0.49	0.49
extend credit	0.00	0.46	0.46
post a warning	0.00	0.46	0.46
take a loss	0.04	0.40	0.44
perform a review	0.00	0.42	0.42
through the tears	0.00	0.42	0.42
off course	0.04	0.37	0.41
beyond belief	0.00	0.41	0.41
strike a deal	0.00	0.40	0.40
beat the odds	0.04	0.33	0.37
reach an audience	0.00	0.37	0.37
against the odds	0.00	0.36	0.36
mark a turn	0.00	0.35	0.35
stage a protest	0.00	0.29	0.29
form expectations	0.00	0.29	0.29
at issue	0.00	0.29	0.29
under the deal	0.00	0.28	0.28
see reason	0.00	0.28	0.28
enter the profession	0.00	0.27	0.27

Appendix P shows the extent of participants' prior knowledge of the target items. The second column from the left shows the proportion of control-group participants (N = 25) who correctly completed each collocation in the gapfill test (max = 1.00). The third column shows the proportion of all participants (N = 78) who indicated in the yes/no test that they had previously read or heard a collocation. The column on the right is the sum total of the other two columns. The difference between the scores of the two measures can be explained at least partly by the nature of the knowledge measures. The "previously known" score is a test of recognition, which allows for more partial knowledge than the gapfill test, which, as a measure of recall, measures knowledge at a higher threshold.

The mean overall score for the V+N collocations is 0.59 (SD 0.18) and the equivalent score for prep+N collocations 0.52 (SD 0.22).

Note: Because *make arrangements* was so well-known (overall prior-knowledge score = 1.46), it was excluded from the analysis.

Appendix Q

Ethics approval

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

Phone 0-4-463 5676
Fax 0-4-463 5209
Email Allison.kirkman@vuw.ac.nz

MEMORANDUM

TO	Mark Toomer
COPY TO	Irina Elgort
FROM	Dr Allison Kirkman, Convener, Human Ethics Committee
DATE	13 March 2015
PAGES	1
SUBJECT	**Ethics Approval: 21595** The effects of various learning conditions on lexical and grammatical collocational knowledge

Thank you for your application for ethical approval, which has now been considered by the Standing Committee of the Human Ethics Committee.

Your application has been approved from the above date and this approval continues until 28 February 2017. If your data collection is not completed by this date you should apply to the Human Ethics Committee for an extension to this approval.

Best wishes with the research.

Allison Kirkman
Human Ethics Committee