The Acquisition and Processing of Formulaic Language: A Focus on Binomials and Collocations and the Effects of Congruency, Frequency and Learning Conditions

Abdulaziz Altamimi

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

May 2021

Abstract

Research in corpus linguistics, psycholinguistics, and second language acquisition indicates that formulaic sequences pervade language (e.g., Erman & Warren, 2000; Howarth, 1998). Thus, mastery of them is crucial for maintaining fluent communication and achieving a native-like competence of the target language (Pawley & Syder, 1983). Although extensive research has been carried out on formulaic language, it is still not clear how factors such as cross-language congruency, frequency of exposure, type of exposure, and transparency influence the learning of formulaic language in an L2. For this thesis, I conducted three studies combining offline and online measures to gain greater insight into how knowledge of binomials and collocations is developed, as well as how these factors influence learning and processing.

In the first study, I used self-paced reading and forced-choice tasks to investigate the role cross-language congruency and frequency of exposure play in the learning and processing of binomials. In this study, Arabic second language learners of English and English native speakers were presented with three types of binomial phrases: English-only binomials, Arabic-only binomials, and congruent binomials (occurring in both English and Arabic). The results revealed that both native and non-native speakers developed durable knowledge of the 'correct' order of binomials (i.e., *fish and chips,* not *chips and fish*) after only two exposures in a reading treatment. The results also showed that when a binomial phrase overlapped in the two languages, it was responded to more quickly and accurately. Frequency of exposure had a minimal effect on the learning outcomes, with no difference in performance between two and five exposures.

The second and third studies investigated transparent collocations (Study 2) and opaque collocations (Study 3). In both studies, eye-tracking and offline measures were used and had three aims. The first aim was to discover the most effective learning condition to

develop knowledge of transparent and opaque collocations for L2 learners and native speakers of English. The effectiveness of three learning conditions was evaluated: training-only (i.e., participants were explicitly asked to learn the target collocations), reading only (i.e., target collocations appeared four times in a text), and training plus reading (i.e., the collocations were presented before reading (a pre-reading exposure) and also occurred in reading). Participants took part in three phases (pre-reading, reading, and testing), and their performances were compared to a control group who only took part in the testing. The second aim was to investigate whether prior study of a collocation affected learners' processing when encountering the same collocation during reading. The third aim was to investigate if online processing predicted learning gains of collocations. Overall, the results of Study 2 and Study 3 were consistent for both language groups. Training plus reading was the most effective learning condition for developing knowledge of the form and meaning of transparent and opaque collocations, while the reading-only condition was more effective than the training-only condition. With regard to processing, prior study of a collocation led to a processing advantage when encountering it during reading. Fixation patterns during pre-reading were indicative of performance for form recognition in Study 2, as well as indicative of performance for form recall and meaning recognition in Study 3; fixation patterns during reading were indicative of performance for form recognition (Study 3).

The studies make a number of important contributions to the field. First, the findings show that congruency facilitates the learning and processing of formulaic language. Second, the results demonstrate that the type of exposure has the same effect for transparent and opaque collocations. Third, prior intentional study of a formulaic sequence results in substantial learning gains, when followed up with an incidental exposure. Finally, attention to a formulaic sequence during pre-reading contributes to learning.

Acknowledgements

I would like to express my deepest and most sincere gratitude to my primary supervisor, Kathy Conklin, for her invaluable guidance, continuous support, and patience throughout my doctoral studies. Her immense knowledge and guidance helped in all the time of research and writing of this thesis. My thanks also go to the members of the Psycholinguistics Research Group at the University of Nottingham for their assistance in various ways during my PhD, in particular Fabio Parente.

My special thanks go to my family for all the support they have shown me during my PhD. I am very much thankful to my wife, Amal, for her love, understanding, and continuing support to complete this research work. Her encouragement and love made me see the positive side even during the most challenging times. Also, I would like to express my special thanks to my parents, my brothers, my sisters, and my father-in-law for their sincere prayers and for their unconditional love.

Finally, I would like to express my gratitude for my friends and colleagues, Mahdi Al-Majdoa, Mubarak Al-fuhailah, and Mohammed Abo-Wahib, for their fruitful discussions and intellectual contribution.

Declaration

I declare that the work presented here is my own and was conducted during my time as a PhD student at the University of Nottingham.

Contents

List of Tables

List of Figures

Abbreviations

ADJ-N	Adjective + Noun
AIC	Akaike information criterion
BNC	British National Corpus
EFL	English as a Foreign Language
ECM	Entrenchment and Conventionalization Model
EEG	Electroencephalography
ERP	Event-Related Potentials
ESL	English as a Second Language
L1	First Language
L2	Second Language
LDT	Lexical Decision Task
LexTALE	Lexical Test for Advanced Learners of English
LME	Linear Mixed-Effects Model
M	Mean
MI	Mutual Information
ms	Milliseconds
NS(s)	Native Speaker(s)
NNS(s)	Non-native Speaker(s)
ROI	Region of Interest
RT(s)	Response Time(s)
SD	Standard Deviation
SE	Standard Error
VIF	Variance Inflation Factor

Chapter 1. Introduction

Although knowledge of grammar allows language users to generate novel utterances that have not been produced before, utterances tend to be produced in very predictable ways. This means that language use does not always require creativity and coming up with novel utterances using grammar rules; rather, most language use operates over a set of recurring word combinations (Pawely & Syder, 1983). For example, it is grammatically correct to say, *pepper and salt*, as well as *salt and pepper*. However, native speakers (NSs) of English generally say the latter, and they will have encountered it far more than the former. Similarly, while *powerful wind* and *give him a visit* both represent grammatical phrases, English NSs will generally say *strong wind* and *pay him a visit.* Along similar lines, Beckner et al., (2009) explain that English NSs prefer saying "*I want to marry you*", to "*I want marriage with you",* even though they both represent acceptable English sentences according to English grammar and they convey a very similar meaning. Sinclair (1991) explains that language users are not highly creative in their language production, in the sense that they can choose freely amongst acceptable formulations, but instead they tend to assess their language choices in terms of how conventionalized they sound.

The recurrent word combinations characterising language use are commonly referred to as formulaic sequences. A formulaic sequence is defined by Wray (2002) as: "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (p.9).[1] Notably, the co-occurrence of the components of formulaic sequences is too frequent to be accounted for merely by chance (Manning & Schutze, 1999). A range of different types of formulaic

[1] The terms formulaic language, formulaic sequences, and formulaic utterances will be used interchangeably in the present thesis.

sequences can be characterized by such descriptions: collocations (e.g., *strong coffee*), binomials (e.g., *fish and chips*), multiword verbs (e.g., *keep away from*), idioms (e.g., *face the music*), proverbs (e.g., *better late than never*), speech formulae (e.g., *what's up*), and lexical bundles (e.g., *in the middle of*).

Formulaic language serves many essential functions in language use. Wray (2002) emphasizes that use of formulaic language is less cognitively demanding than assembling single words together, which may help account for its ubiquity in language. This processing efficiency attributed to formulaic language helps language users overcome the limited capacity of working memory (Conklin & Schmitt, 2008, 2012), thus maintaining language fluency. Formulaic language also serves important social/pragmatic purposes. It can signal a speaker's identity/membership within a speech community (Wray, 2002). For example, different speech communities prefer using different formulaic expressions in the same situation to convey the same meaning; in one region the preference might be to say, "*I don't mind"*, while in another it might be "*I don't care to"*. Such expressions can serve as markers of speaker's identity within communities (Bardovi-Harlig, 2019, p. 97). Thus, formulaic language can be used to construct speakers' social identities. Formulaic language also plays a key role in social interactions. For example, sequences like "*nice weather today"* or *"oh, I see what you mean"* help maintain the flow of the conversation (Schmitt & Carter, 2004, p.10). Formulaic language is crucial in realising a number of social functions, such as making a request (e.g., "*would you [mind] X?"*) or expressing sympathy (e.g., "*I'm [very] sorry about/to hear [about] X"*) (Nattinger & DeCarrico, 1992, p. 62-63). Formulaic sequences help evoke the intended communitive purpose in a very reliable and effective way because of their familiarly to language speakers (Conklin & Schmitt, 2008).

The significant social/pragmatic utility of formulaic sequences helps to explain their ubiquity in language use. Wray and Perkins (2000) believe that formulaic sequences are the

"default setting", which enables interlocutors to focus on the meaning and function of utterances rather than on the analytical generation of these utterances. NSs' language production comprises a high proportion of formulaic sequences that depend on the distributional relations amongst words (Arnon & Christiansen, 2017; De Cock et al., 1998). Corpus-based research shows that between 30% to 50% of language use in written discourse is formulaic (e.g., Erman & Warren, 2000; Howarth, 1998) and it is also ubiquitous in spoken discourse (e.g., Kuiper, 2004; Sorhus, 1977).

While NSs appear to formulate a particular message in the same way or employ what we might think of as language patterns – which is commonly referred to as *formulaic language* – is the same true of non-native speakers (NNSs)? Pawely and Syder (1983) note that second language learners (L2) in the early stages of learning commonly produce utterances which are judged by NSs to be grammatical but unnatural. For their speech to sound more native-like, they not only need to adhere to the set of rules for what is acceptable and what is not in the L2, but additionally, they need to learn which of these acceptable utterances are marked and which are not marked based on NSs' usage (Pawely & Syder, 1983). In other words, NNSs' lack of formulaic language is what in part makes their language production marked and distinguishes them from NSs (Pawley & Syder, 1983; Wray, 2002).

Because of their ubiquity, knowing and understanding formulaic sequences has been acknowledged as a significant goal for language learners (Matthews & Bannard, 2010; Peters, 1983), and for achieving a native-like mastery of the language (Pawley & Syder, 1983). It also means that it is important to study how formulaic language is learned and processed in an L2 compared to a first language (L1). In the past two decades, there has been considerable research in this area on: collocations (e.g., Durrant & Schmitt, 2010; Sonbul & Schmitt, 2013; Wolter & Yamashita, 2017; Yamashita and Jiang, 2010), binomials (e.g., Siyanova-Chanturia, Conklin, & van Heuven, 2011), multiword verbs (Matlock & Heredia, 2002),

idioms (Carrol, Conklin, & Gyllstad, 2016; Conklin & Schmitt, 2008; Underwood, Schmitt, & Galpin, 2004), and lexical bundles (e.g., Arnon & Snider, 2010; Tremblay, Derwing, Libben, & Westbury, 2011). The majority of studies demonstrated that the entire formulaic utterance may have some level of representation in NSs' memories. This is generally evidenced by faster processing observed for formulaic language relative to matched non-formulaic language (e.g., Arnon & Snider, 2010; Tremblay, Derwing, Libben & Westbury, 2011). However, research findings for NNSs have been inconsistent, with some studies demonstrating fast processing as a result of phrase frequency effects (Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & van Heuven, 2011), while others have failed to find such a processing advantage (Underwood, Schmitt, & Galpin, 2004; Siyanova-Chanturia, Conklin, & Schmitt, 2011).

Since more than 50% of the world's population speaks a L2 (Grosjean, 2010), it is important to advance our understanding of how formulaic language is learned and processed in an L2, which has important practical implications for language teaching. There are a number of factors that have been shown to be important when learning formulaic language in an L2: the nature of learning (i.e., whether the learning is intentional or incidental), congruency (i.e., the amount of overlap between languages), transparency (i.e., whether the meaning of a formulaic sequence can be deduced from its constituents), frequency of the phrase, and L2 proficiency. Sonbul and Schmitt (2013) found that an intentional learning condition was more beneficial for learning formulaic language than an incidental learning condition for both NSs and NNSs. Effects of congruency were noted in the processing of formulaic language. Wolter and Gyllstad (2011) and Yamashita and Jiang (2010) found that formulaic sequences that overlap across languages have a processing advantage. In addition to congruency, transparency of the sequence plays an important role. The learning of less literal formulaic sequences is more difficult than their more literal counterparts (Irujo, 1986;

Macis & Schmitt 2017a, Siyanova & Schmitt, 2007). A study by Northbrook and Conklin (2019) demonstrated a clear effect of frequency in very early L2 learners. With regard to proficiency, Siyanova-Chanturia et al. (2011), demonstrated that only higher proficiency learners, but not lower proficiency learners, were able to process frequent compositional phrases in a similar way to native speakers.

Despite its pervasiveness and significance in language use, several questions in relation to the learning of formulaic language in an L2 remain unanswered. There is a general lack of research investigating the role of factors like congruency, the nature of learning (i.e., intentional versus (vs.) incidental), and frequency of exposure, as well as how the interaction of such factors impacts the learning and processing of formulaic language. Moreover, most of the studies on the learning of formulaic language have focused on assessing knowledge after exposure to a particular learning treatment (i.e., using offline tests targeting the learning of the form and meaning of a particular type of formulaic language). Thus, little is known about what occurs during real-time processing and how a particular treatment impacts learners' processing at the point of the treatment itself. Combining offline measures (i.e., assessing learning outcomes) with online measures (i.e., self-paced reading task or eye-tracking), can provide a more comprehensive understanding of the learning and processing of formulaic language, which can inform pedagogical practice.

The present thesis seeks to fill some of the gaps in the field by exploring the learning and processing of two types of formulaic language – binomials and collocations. The first study investigates how frequency of exposure and congruency influence English NSs and NNSs' learning and processing of binomial phrases. More precisely, the effects of congruency and frequency of exposure in the learning and processing of binomials will be investigated using self-paced reading and forced-choice tasks. The second and third studies explore the impact of different learning conditions (i.e., training only, reading only, training

plus reading) on the learning and processing of transparent and opaque collocations by NSs and NNSs. In these studies, eye-tracking and offline tasks are used to examine the learning and processing of collocations presented in the different learning conditions. Further, eye-movements are monitored to explore how exposure to a collocation phrase before reading influences learners' processing during reading, and how online processing can predict later learning outcomes.

The structure of the thesis is as follows. Following this introductory chapter, Chapter 2 provides a theoretical overview related to the acquisition of formulaic language. It concludes with a discussion of how formulaic language is acquired by both children in their L1 and adults in their L2. Following this, Chapter 3 provides a general background for all three studies – discussing features of formulaic language, defining binomials and collocations, summarising research on binomial and collocation processing in NSs and NNSs, highlighting congruency and other factors that might affect processing, and finally discussing the learning of formulaic language in an L2. In the subsequent three chapters – Chapters 4, 5, and 6 – the empirical studies are presented. Each study is a stand-alone study, that has a brief introduction presenting its objectives, followed by sections presenting its methodology, results and a discussion interpreting the findings. Chapter 4 presents the first study on how congruency affects the learning of binomials by NSs and NNSs. Chapter 5 presents the second study on how different learning conditions affect the learning of transparent collocations by NSs and NNSs. Chapter 6 presents the third and final study in this thesis, replicating Study 2 but with opaque collocations. Finally, Chapter 7 brings the findings together, and discusses their implications. It also identifies limitation of the studies presented in the thesis and concludes with some recommendation for future research.

Chapter 2. Formulaic Language within Language Acquisition Theories

2.1 Introduction

Pawely and Syder (1983) explain that native-like selection of lexical and structural choices is determined merely by convention which comes about through language use. This suggests that the repeated use of (formulaic) utterances is what makes them familiar and conventional to NSs. Sinclair (1991) highlights another fundamental property of formulaic language in his idiom principle, namely that while formulaic language can be broken into smaller chunks, it is treated as an unanalysed whole:

> "The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments" (p. 110).

Due to the pervasiveness of formulaic language in language use, this chapter is dedicated to explaining its nature and describing it in relation to linguistic theories of language acquisition. This chapter will conclude by comparing the differences in acquiring formulaic language in L1 and L2 learners.

2.2 Frequency Effects

Frequency is one of the key features distinguishing formulaic sequences from novel language. Experience with language plays an important role in the creation, entrenchment (i.e., representation in memory)[2] and processing of linguistic information. Frequency effects reflect implicit learning, such that representations are strengthened each time they are

[2] Langacker introduced the notion of entrenchment by explaining that "every use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence"(1987, p.59).

encountered (Ellis, 2002b). Evidence for the role of frequency (or usage) on memory representations can be drawn from speakers' sensitivity to statistical distribution of constructions, token frequencies (i.e., the frequency counts of a particular lexical item), and from the fact that high frequency language elements usually resist regularization (Bybee, 1995, 2001, 2007). In terms of regularization, it appears that high token frequency consolidates mental representations and the entrenchment of items in memory (Langacker, 2009), which makes them resistant to change.

While there is considerable evidence demonstrating the role for frequency effects on individual word recognition (Forster & Chambers, 1973; Scarborough, Cortese, & Scarborough, 1977; Whaley, 1978), there is less evidence for units that span more than a word. However, Bod (2000, 2001) demonstrated that frequent phrases, such as *I like it*, are recognized faster than infrequent control phrases such as *I keep it.* Further, both children and adults have been shown be sensitive to statistical patterns of language use (e.g., Bannard & Matthews, 2008; Hilpert, 2008; Saffran, Johnson, Aslin, & Newport, 1999; Saffran & Wilson, 2003). For example, Hilpert (2008) demonstrated that constructions bias the choice of one lexical item over another. He presented participants with the *make*-causative construction (in the carrier phrase *they made me*) and manipulated the initial sound of the following verb so that it could either be perceived as /trai/ or /krai/. Most participants perceived *cry*, which was the more frequent verb in the construction, despite the fact that *try*, in general, is more frequent than *cry*.

In what follows, we will look at accounts of frequency effects in some prominent models: Universal Grammar (UG) and usage-based approaches such as Construction Grammar (CG) and Entrenchment and Conventionalization (EC).

2.2.1 *Frequency Effects within UG Theory*

The words and rules approach (Pinker, 1991; Pinker & Ulman, 2002), which falls under the umbrella of UG, proposes that lexical items and grammar are subject to different mechanisms. It treats rules, which are applied to language forms, as the main property of language. Lexical items are stored and retrieved from the memory, while utterances, phrases and clauses are computed using grammar. That is, language emerges from two domains functioning differently: rules and words. Ullman et al. (2005) say that both domains are modulated by different cognitive mechanisms. Crucially, frequency effects *only* extend to lexical items, which are stored in memory, but not to the other language phenomena where generation and computation are at play. Because the meaning of idioms cannot be computed compositionally (i.e., *kick* + *the* + *bucket* ≠ 'die'), idioms are said to be anomalous phrases that are stored in memory and retrieved from the lexicon. However, fully compositional phrases like *salt and pepper* are *not* stored in memory, and therefore, according to the words and rules approach, should *not* exhibit frequency effects.

Evidence demonstrating frequency effects for larger sequences is incompatible with this theory. More specifically, if the more frequent phrase *salt and pepper* is processed more quickly than the grammatically acceptable and possible, but less frequent, *pepper and salt*, the words and rules approach (and UG more broadly) could not account for this.

2.2.2 *Frequency Effects with Usage-Based Models*

While frequency effects for formulaic strings are incongruous with the words and rules approach, evidence that language users process highly frequent items with ease would be compatible with usage-based theories. That is, in opposition to generative approaches which assume that the grammar and lexicon are two separate entities, a usage-based theory holds that the lexicon and grammar are intertwined, and that linguistic representations interact dynamically with each other (Bybee, 1998; Ellis, 2008). Usage-based models propose that all aspects of language are shaped by language use and experience (Bybee, 2006, 2010;

Goldberg, 2006; Jurafsky et al., 2001), and the frequency of language patterns influences their processing and their entrenchment in memory (Jurafsky et al., 2000). Jurafsky et al. (2001) confirm that language users are highly sensitive to frequency, not only at the word level, but also this effect extends to the combination of words and structures. In other words, language users are sensitive to frequency and this is evident in sound combinations, morphemes, single words, multiword sequences, constructions, etc. Crucially, this sensitivity influences how easily language users perceive and recognize elements of language.

In short, usage-based proponents advocate for the notion that all aspects of language are subject to frequency effects, and they do not see any clear distinction between syntax and the lexicon. This view does not treat syntax and the lexicon as separate, dichotomous entities and means that frequency effects are expected for repeated sequences of words (i.e., formulaic language). As a usage-based perspective is more compatible with formulaic language, in that it captures its features, like frequency for linguistic elements above a word, formulaic language will be investigated in this thesis in light of usage-based approaches.

Usage-based proponents account for sensitivity to frequency by explaining that any exposure to language input is registered in memory, and multiple exposures to a particular language unit can either strengthen its entrenchment or update that pre-recorded linguistic unit (e.g., Bod, 2006; Ellis, 2007; Ellis & Larsen-Freeman, 2009). In other words, language development involves a statistical accumulation of many usage events. According to this view, language users are exposed to a lot of language, with every language encounter leaving a trace in the memory. Language users can detect patterns between these memory traces, which allow them to conceptualize schema and prototypes from linguistic exemplars (Ellis & Robinson, 2008). Then, through mechanisms like association and analogy, they are able to draw generalizations between exemplars (Ellis & Larsen-Freeman, 2009). Notably, the linguistic knowledge emerging from experience can be at any level of granularity, and crucial

to the current research, it encompasses longer formulaic sequences. That is to say, language learners use the statistical distributions between sounds, words, and formulaic utterances, as a roadmap to help learn that language.

Ellis (2002a) holds that language learning reflects sensitivity to the statistical distributions governing the regularization among thousands of constructions.[3] This argument is based on the assumption that language users are attuned to both token frequencies and type frequency. Whereas type frequency refers to the number of different lexically filled items licensed in a particular structure or 'construction', token frequencies refer to the frequency counts of a particular lexical item or a phrase (Ellis, 2002a). Ellis shows that high type frequency promotes language productivity because it permits the insertion of many lexical elements into a particular construction frame. On the other hand, high token frequency, which is true of formulaic sequences, is less productive (Bybee, 2010). That is, repeated use leads some formulaic sequences to "lose analysability and compositionality" (Bybee, 2010, p. 95), and this may be the reason why they tend to be less productive, as their internal elements become less identifiable. According to Bybee (2010), high token frequency leads formulaic utterances to become autonomous, which distances them from other exemplars occupying the same construction. Based on this analysis, we can infer that formulaic sequences evolved from a general exemplar cluster (e.g., V-NP: *lend a hand*), and with repeated use they reach an autonomous status, characterized by a special semantic status, which distances them from other similar exemplars.

[3] Goldberg (2006) notes that constructions are built when a form-meaning is paired within a linguistic token. Examples of constructions could be as basic as single lexical items (words) or a combination of lexical items where direct form-meaning pairing is established. An example for the latter is found in the passive or ditransitive constructions (ditransitive: Subj +V +Obj1 +Obj2, as noted in Goldberg, 2003, and 2006), or in idioms (e.g., *break the ice*).

Along similar lines, Schmid (2015, 2017) proposed the Entrenchment and Conventionalization Model (ECM). Understanding how language patterns emerge from usage is of primary interest in this model. The ECM claims that language patterns are a by-product of the interaction between entrenchment and conventionalization operations that are connected through usage. Entrenchment is defined as, "the continuous routinization and re-organization of associations, depending on exposure to and frequency of identical or similar processing events, subject to the exigencies of the social environment" (Schmid, 2015, p.10). In other words, the repeated use of lexical items or language patterns in several social interactions increase the entrenchment of such lexical entries. Schmid argues that entrenchment operates over three cognitive processes: association, routinization, and schematization.

According to the ECM, different association types are triggered upon encountering linguistic stimuli. For example, symbolic association occurs when the form and meaning relationship of a language item is established. Syntagmatic associations, on the other hand, are activated when utterances are formed in a predictable sequential order. Schmid also reports that syntagmatic associations are responsible for facilitating meaning integration (in comprehension) and for arranging the sequential flow of utterances (in production). One main feature of syntagmatic association is the linear predictability of lexical items. The model postulates that repeated exposure to the syntagmatic form of fixed formulaic language means that activation of an initial word spreads activation to its subsequent components. Also, it can weaken the individual symbolic associative representation of the individual words, thus initiating a holistic association of the whole unit. What this means is that frequent processing of sequentially determined patterns leads to routinization of syntagmatic associations due to high contextual entrenchment. Building strong syntagmatic association leads to items being represented as a chunk, which reduces their compositional semantic status. As a result, the

effort that language users need to process some fixed phrases such as idioms, collocations, and irreversible binomials will be significantly reduced. For example, *fish and chips* is more syntagmatically associated and has stronger sequential ties than *chips and fish*; therefore, production and processing of the former will be more facilitated.

According to this model, using highly frequent conventional utterances maintains the routinization of associative types, resulting in reaching a high baseline activation level. This high level of activation primes their use in different communicative settings, resulting in greater conventionality than other items, and stabilizing their syntagmatic relations. For example, considerable repetition of language patterns (mainly formulaic sequences such as *salt and pepper* or *pay attention*), instantiates their entrenchment, thus biasing their further use due to their facilitated production and processing. Such repeated use reflects greater degrees of conventionalization for formulaic sequences in the speech community compared to non-formulaic sequences.

Overall, the usage-based models briefly reviewed all agree that frequency of use plays an essential role in language development and it is one of the main determining factors for the level of entrenchment in memory. However, while acknowledging and highlighting the role of frequency for language acquisition, we should not dismiss other factors such as noticing, attention, and salience (Beckner et al., 2009). For example, Ellis (2006) explains that the low salience of function words, despite their high frequency, may cause difficulty in language learning.

2.3 L1 and L2 Acquisition of Formulaic Language

As described above, usage-based theory says that exposure to language input is the key driver for language acquisition. Therefore, the acquisition of formulaic language by child L1 learners and adult L2 learners should be greatly influenced by the amount of exposure both populations receive.

2.3.1 *L1 Acquisition of Formulaic Language*

A study by Bannard and Matthews (2008) highlights the role of phrasal frequency in child language acquisition. Using a sentence repetition task, they found that two- and three-year-old children are sensitive to the phrasal frequency of a formulaic utterance, when the frequency of its component words is controlled for (e.g., *sit in your chair, you want to play* vs. *sit in your truck, you want to work*). They found that children's production accuracy was influenced by the frequency of the formulaic utterances in their input. Based on their findings, Bannard and Matthews argue that formulaic sequences are entrenched in memory to some degree, which makes their production easier. The findings also emphasize that child language output is relatively aligned to the input.

Matthews and Bannard (2010) investigated how patterns of usage affect child production of formulaic sequences. They explored some factors that influence children's (two- and three-year olds) production of a terminal word in lexically specified four-word sequences. One of the key factors was slot entropy (i.e., the probability that the final position accepts alternative words). It was predicted that children would tolerate modifications of the last word in a sequence if it had higher entropy. For example, the entropy of the frame *back in the X* (e.g., *back in the box*) was greater than that of *let's have a X* (e.g., *let's have a look*). Thus, it was expected that children would more easily produce unfamiliar forms of the former (e.g., *back in the town*) than the latter (e.g., *let's have a think*). Another important factor was semantic density (the probability of the final word being replaced by a semantically similar one). In a sequence repetition task, Matthews and Bannard found that better production accuracy of unfamiliar sequences was associated with higher entropy (e.g., *back in the X*). In addition, the more semantically similar the words which filled the final slot, the more likely children produced variants of the sequence correctly. The Matthews and Bannard (2010) study suggests that production of novel sequences is influenced by exposure and a sensitivity

to the statistical/semantic information of a sequence. That is, unfamiliar versions of a sequence which conform to the lexically specified pattern, in terms of its statistical distribution and semantic properties, will be produced with better accuracy. This study highlights the role of lexically specified patterns for productivity in child language production.

Formulaic sequences in child speech might come about either through undersegmentation or chunking (Arnon & Christiansen, 2017). Undersegmentation means that formulaic sequences are acquired as a single unit without analysing or segmenting the internal elements (Arnon & Christiansen, 2017; Wray, 2002). Infants perceive language units based on prosody, and because they lack a sense of word boundaries, they may segment at utterance level rather than at word level (Arnon & Christiansen, 2017). Thus, children acquire an utterance as a whole unit, akin to a complex morphological word, before they start analysing its internal components. It is only as their cognitive and linguistic ability grows that children start decomposing an utterance's constituents. Arnon and Christiansen (2017) explain that the considerable amount of repetition in child input helps children extract formulaic utterances during the segmentation process. Segmenting language patterns as meaningful chunks is followed by extracting smaller units from them, for example, learning *do you want any more* leads to extracting *any more* (MacWhinney, 2017). In addition to undersegmentation, another mechanism that helps children learn formulaic language is chunking. Chunking refers to "the development of permanent sets of associative connections in long-term storage" (Ellis, 1996, p. 107). The formation of formulaic sequences is a product of chunking processes (Bybee, 2010). In other words, repeatedly co-occurring sequences of words (i.e., formulaic utterance) are recorded as a single meaningful chunk. However, Arnon and Christiansen (2017) argue that this does not necessarily imply that formulaic sequences are stored holistically; rather, the formulaic sequence (e.g., *do not have to worry*) is

represented in memory alongside its internal parts. This means that the frequent co-occurrence of these chunks will result in their becoming represented in memory.

Similarly, in her Need-Only analysis of formulaic language, Wray (2002) argues that children begin by learning formulaic language and other language patterns as non-analysable wholes, and as their cognitive and linguistic abilities grow, they move to an analytical stage where they analyse the internal parts of a string to obtain grammatical information and to extract grammatical generalizations (Wray, 2002, 2008). However, according to this model, unless there is a *need* to analyse the internal structure of a formulaic utterance (i.e., to obtain grammatical information), some formulaic utterances may resist analyzation (i.e., mainly phrases where internal alternations would alter the intended message such as *thank you, good night, happy birthday*). That is, some formulaic sequences may remain unanalysed if there is no need to break them down. In contrast, L2 learners, according to the model, tend to analyse internal parts of formulaic sequences when it is not needed. This is thought to underpin L2 learner's difficulties with formulaic sequences. Further, this account argues that formulaic strings are stored and retrieved from long-term memory, which reduces the cognitive effort needed to compute them.

2.3.2 *L2 Acquisition of Formulaic Language*

Similar to child language development, Ellis (2002a) stresses the role of language input with regard to L2 learners' acquisition. According to him, for language learners to achieve native like fluency, they need to approximate a repertoire of exemplars of similar sizes to the ones of NSs and make accurate generalizations of those exemplars. Usage-based proponents agree that making use of encountered exemplars is what facilitates language learning (e.g., Bybee, 2006, 2008; Goldberg, 2006; Tomasello, 2003).

According to a usage-based approach, both child L1 and adult L2 learners utilize language input in a similar way to acquire, comprehend and produce formulaic sequences.

However, in the L2 literature, we see a number of studies documenting malformed formulaic forms. For example, Laufer and Waldman (2011) compared the use of English verb-noun collocations in the writings of Hebrew, L2 English learners and English NSs. They found that the language learners underused the collocations produced by English NSs. Siyanova and Schmitt (2008) investigated adjective noun collocations in the writing materials of Russian, L2 English learners. They found that only half of the collocations occurred frequently in the British National Corpus (BNC). Further, Yorio (1989) found that L2 learners' writing contained many malformed formulaic forms (e.g., *put more attention* or *made a great job*).

One explanation for adult L2 learners' performance is that they come to the L2 with knowledge about the units of language (e.g., words); this sensitivity to the units that are internal to formulaic language may impede the process of learning and producing formulaic language in a native-like way (Arnon & Christiansen, 2017). Another possibility is that adult L2 learners transfer L1 knowledge of formulaic language to the L2, resulting in formulaic sequences that are malformed in the L2. As evidence for this, some studies indicated that L1 formulaic sequences affect L2 formulaic processing (for a discussion, see Conklin & Carrol, 2019). However, this does not imply that chunking does not occur during L2 learning. It is simply that adult L2 learners bring additional knowledge to the task of language learning that L1 children do not have: literacy, knowledge about the components of language, and a repertoire of L1 formulaic language. This knowledge very likely impacts the acquisition, production and comprehension of formulaic language in the L2.

It is worthwhile mentioning that attending to words, rather than multiword utterances, might stem from the input L2 learners receive. It appears that formulaic sequences appear more frequently in the L1 input than the L2 input (Arnon & Christiansen, 2017). Thus, it may be that adult L2 acquisition and processing is the same as that of L1 learners, just to very different input. In a similar vein, language learning materials might encompass formulaic

language that is not representative of that of NSs. Northbrook and Conklin (2019) found that language teaching materials do not represent NSs' formulaic sequences, particularly at longer lengths (four or more words long). In addition to the quantity and quality of the input that L2 learners receive, a whole host of other factors have been demonstrated to impact L2 learning, such as L2 learning environment, feedback, age, aptitude, attitude, motivation, etc. (see Larsen-Freeman, 1997).

2.4 Conclusion

This chapter aimed at introducing formulaic language within the context of linguistic theories of language acquisition. Specifically, it explains formulaic language from the usage-based perspective, which highlights the role of frequency for language acquisition. It discusses the acquisition of formulaic language by L1 child and L2 adult learners. In summary, differences in L1 child and L2 adult formulaic sequences are likely in large part due to L2 learners' pre-existing linguistic knowledge and the L2 input they receive. However, once receiving sufficient L2 input, adult L2 learners' production of formulaic language will likely become more native-like. While this chapter provides a general overview covering key issues in the learning of formulaic language, the next one provides a more focused overview of the empirical investigations of formulaic language learning research that is more relevant to the empirical work conducted in this thesis.

Chapter 3. L2 Processing and Learning of Formulaic Language

The previous chapter summarises some theoretical concepts behind the acquisition of multiword units, and how child L1 and adult L2 learners may differ. This chapter summarises the experimental work that has been done on formulaic language, focusing mainly on binomials and collocations because they are the types of formulaic language under investigation in this thesis. This chapter will cover a range of topics that includes: characterizing formulaic language; NSs' and NNSs' processing; and the role of proficiency, frequency of occurrence and congruency on processing. This chapter concludes by discussing incidental learning for formulaic language, an area of research that has recently received increasing attention.

3.1 Introduction

Formulaic sequences differ in terms of their frequency, fixedness, semantic transparency, literal plausibility, and compositionality. Idioms represent a clear example of how multiword utterances may have features that vary across the scope of these criteria. For example, while some idioms are decomposable, such that the components contribute to the whole figurative meaning (e.g., *pop the question*), others are non-decomposable, such that the components do not contribute to the meaning (e.g., *kick the bucket*). Additionally, some idioms can be interpreted both literally and figuratively depending on the context. For example, *at the end of the day* depending on the context could either mean *eventually* or *in the evening* (Siyanova-Chanturia, Conklin & Schmitt, 2011).

In the literature on formulaic language processing, idioms have received by far the most attention in L1 research (e.g., Gibbs, 1980; Gibbs & Gonzales, 1985; Kyriacou, Conklin, & Thompson, 2019; Swinney & Cutler, 1979; Tabossi, Fanari, & Wolf, 2009; Titone, Holzman, & Levy, 2002) and in L2 research (e.g., Abel, 2003; Cieślicka, 2006; Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin & Schmitt, 2011; Underwood,

Schmitt, & Galpin, 2004). One important formulaic language type, which has been understudied, is binomials. Another type of formulaic language which has received increasing attention is collocations. Studying the learning and processing of binomials and collocations is the main goal of this thesis.

Binomials are "recurrent (frequent), familiar (conventional) expressions" consisting of two words from the same lexical class joined by a conjunction, where one word order is more frequent (e.g., *king and queen* not *queen and king, fish and chips* not *chips and fish*) (Siyanova-Chanturia, Conklin & Van Heuven, 2011, p.2). Binomials differ from idioms and other formulaic language in many ways. Lambrecht (1984) explains that binomials are structurally more predictable than idioms (e.g., N $_{Conj}$ N; Adj $_{Conj}$ Adj; Adv $_{Conj}$ Adv; or V $_{Conj}$ V vs. idioms that have various syntactic forms). Binomials are also more frequent than idioms. Further, binomials are mostly transparent such that the overall meaning can be computed from the parts, which makes them much more semantically transparent than idioms (e.g., *break the ice*). This is important because it has been said that idioms, in part, pose a challenge for L2 learners due to their opaque nature (Carrol, Conklin, & Gyllstad, 2016). However, for binomials, the challenge for L2 learners may be developing a sensitivity to the word order preference (e.g., *salt and pepper* not *pepper and salt*).

An important facet of binomials, in contrast to other types of formulaic sequences, is that a change of word order does not alter the meaning or result in syntactic irregularity. The flexibility in word order (*fish and chips*, *chips and fish*), without introducing confounds (e.g., unigram frequencies, syntactic and semantic properties), makes binomials a good choice for investigating the role of phrasal frequency – by manipulating the order of the content words what is primarily altered is phrasal frequency (Siyanova-Chanturia et al., 2011).

Another type of formulaic sequence that is relevant to this thesis is collocations. Unlike idioms, collocations are fluid in language use and reflect tendencies

rather than exclusiveness (Wray, 2002). For example, the component parts of one collocation (e.g., *strong feeling*) can also be used in other collocations (e.g., *strong coffee* and *strange feeling*). Collocations also lack the input salience associated with idioms (Durrant & Schmitt, 2008). Wray (2012) claimed that the exceptional form/meaning pairing (e.g., *kick the bucket* means 'die') may make idioms more noticeable. This (assumed) salience could result in stronger memory traces, and could facilitate recognition of idioms. In contrast, L2 learners may fail to detect collocations in written text due to lack of input salience.

The decision to study collocations was in large part based on the fact that they are, like binomials, pervasive in language use (Sinclair, 1991). Their ubiquity suggests that learning collocations would help language learners achieve more natural language use and native-like fluency. According to Ellis (1997), "Speaking natively is speaking idiomatically using frequent and familiar collocations, and the job of the language learner is to learn these familiar word sequences." (p. 9). In a similar vein, Millar (2011) believes that deviations to collocations disrupts communication. However, L2 learners seem to produce a large number of non-standard forms of collocations (Bahns & Eldaw, 1993; Foster, 2001; Gabrys-Biskup, 1992; Granger, 1998; Nesselhauf, 2003). Granger (1998) and Foster (2001) both demonstrated that non-native speakers overuse non-collocate word pairs and underuse frequent collocations. Durrant and Schmitt (2009) reported that non-native speakers tend to underuse highly associated collocations (i.e., collocations such as *densely populated* where their probability of co-occurrence is high).

The definition of collocation varies based on two main approaches: a frequency versus a phraseological approach. A frequency-based approach identifies the notion of collocation based on "the relationship a lexical item has with items that appear with greater than random probability in its (textual) context" (Hoey, 1991, p. 7). The frequency-based approach identifies collocations according to the statistical co-occurrence of words and

employs corpus frequency as a main identifying criterion (Carter, 1988; Hoey, 1991; Moon 1998; Sinclair, 1991). Several measures have been put forward to calculate the association between the two words of a collocation and to establish whether they co-occur more frequently than would be expected by chance alone (e.g., t-score, z-score, and log-likelihood, Mutual Information (MI) scores). For example, using the BNC search tools for nouns that collocate with the node adjective *deep* reveals that both *sea, roots, sigh, ocean, voice, breath, sleep* have high MI scores, indicating an above chance co-occurrence with *deep.* Typically, corpus linguists have adopted the minimum MI scores of 3.0 for identifying collocations (e.g., Hunston, 2002; Stubbs, 1995).

A phraseological approach sees a collocation as "a composite unit which permits the substitutability of items for at least one of its constituent elements (the sense of the other element, or elements, remaining constant)" (Cowie, 1981, p. 224). The approach favours natives' intuition over corpus frequency in identifying collocations (Greenbaum, 1988). In a phraseological approach, semantic properties (e.g., sense restriction) is taken into account (Nesselhauf, 2005; Gyllstad & Wolter, 2016). One main criterion of identifying collocations in a phraseological approach is based on the notion of 'restricted sense': the substitutability of either word parts of a collocation is restricted to only a limited set of items in an arbitrary way (Nesselhauf, 2003). Thus, this substitutability is not semantically motivated but is dictated by convention. For example, in the collocation *reach a decision,* the verb *reach* is restricted such that it can only be used with a limited set of nouns (e.g., "*conclusion, verdict, compromise,* or *goal"*); however, it cannot be used with *aim,* suggesting that this restriction to the sense of *reach* is arbitrary (Nesselhauf, 2003, p. 225). In contrast, the phrase *read a newspaper* is not treated as a collocation since the verb is unrestricted such that it can be used with any noun that is grammatically and semantically acceptable. Based on this notion, Nesselhauf (2003, p. 226) distinguished between three types of phrases:

Free combinations (e.g., *want a car*):

The senses in which the verb and the noun are used are both unrestricted, so they can be freely combined according to these senses.

Collocations (e.g., *take a picture*):

The sense in which the noun is used is unrestricted, but the sense of the verb is restricted, so that the verb in the sense in which it is used can only be combined with certain nouns (*take a picture*/*photograph*; but e.g. **take a film*/*movie*).

Idioms (e.g., *sweeten the pill*):

Both the verb and the noun are used in a restricted sense, so substitution is either not possible at all or only possible to an extremely limited degree.

In this thesis, collocations are identified based on the cooccurrence of the constituents in a corpus. This was thought to be the more appropriate choice, as an important focus of the thesis is on the impact of input (i.e., frequency) on L2 formulaic language acquisition. A disadvantage of a frequency-based approach is that it assumes that corpus-driven data approximates actual language input (Durrant & Schmitt, 2010), which might be particularly problematic in the case of NNSs (for a discussion see Conklin, 2019). This is addressed by complementing corpus frequency data with language users' intuition about the collocations (gathered through norming studies of potential materials).

3.2 Predictability and Frequency in Formulaic Language

Predictability is one of the factors that distinguish binomials and collocations from novel sequences. For example, upon encountering *salt and …* English NSs will most likely be able to predict the missing constituent (*pepper*). Ellis (2012) points out that relying merely on overall frequency effects is misleading as the salient phrase *once in a blue moon,* while of low overall frequency, has a higher than chance statistical coherence (i.e., the co-occurrence of the lexical elements of this phrase is unlikely to be due to chance). If recurring

combinations had some level of representation, this would help explain the predictability of final components upon seeing initial ones (Jurafsky, 1996). Jurafsky et al., (2001) found that highly predictable words (e.g., *kind of*) tend to be reduced in speech production, indicating that probabilistic relations are encoded in language processing. McDonald and Shillcock (2003a, 2003b) demonstrated that the increased bigram probabilities decreased reading times. Shorter fixation durations were recorded for collocations with higher transitional probabilities (e.g., *resolve disputes)* as compared with collocations with lower transitional probabilities (e.g., *resolve meanings*). Thus, such faster processing cannot be attributed merely to phrasal frequency for the collocations but also to probabilities of co-occurrence for the words making up the collocation (McDonald and Shillcock 2003a, 2003b, 2009).

From a probabilistic perspective, language users encode the statistical properties of linguistic input and the co-occurrence of lexical information (McDonald & Shillcock, 2003a). This can be explained by lexical priming theory (Hoey 2005, 2012), which says that word A will pass activation to word B if they co-occur. Thus, once language users encounter word A they are primed for (i.e., anticipate) word B. This sensitivity to the statistical distribution of language allows language users to predict the likelihood of (word $_{n+1}$) given (word $_{n}$). Durrant and Doherty (2010) and Wolter and Gyllstad (2011) both used a lexical decision task to examine priming effects for collocations. Both studies demonstrated that collocations exhibit priming effects for NSs, as evidenced in the processing advantage observed for collocations. In order for L2 learners to demonstrate collocational knowledge similar to that of NSs, they should demonstrate priming for the second constituent of a collocation (n+1) by the first (n).

As the above discussion highlights, there are potentially two issues at stake for formulaic sequences: entrenchment in memory of the formulaic form and encoding of the statistical co-occurrence of words (i.e., predictability). However, entrenchment does not necessarily entail a 'holistic' representation of formulaic language. More specifically, does

the processing advantage for *fish and chips* over *chips and fish* arise because: 1) the probability of having *chips* after *fish* is higher than the reverse, thus priming triggers an automatic spread to *chips*; or 2) the former phrase is more frequent which leads to entrenchment of the form in memory?

Siyanova-Chanturia, Conklin, and van Heuven (2011) addressed this question in an eye-tracking study. They found that binomials (e.g., *time and money*) exhibited shorter fixations than their reversed forms (e.g., *money and time*) for English NSs and higher proficiency NNSs. This processing advantage was not due to differences in the predictability of the binomial phrases and their reversed forms (assessed using completion scores of a cloze task). Siyanova-Chanturia et al., (2011) concluded that the binomial processing advantage may not be due solely to predictability, but to phrasal frequency and entrenchment in memory of the lexical pattern as well.

As well as predictability, frequency is another main determining factor of formulaic language which takes the form of the observed frequency counts of a phrase in a corpus. Frequency has been shown to be a key driver in shaping the lexicon and in word processing (Bod, Hay, & Jannedy, 2003; Froster, 1976). Ellis (2002a) says that frequency effects are ubiquitous in many aspects of language (e.g., phonology and phonotactics, reading and spelling, lexis and morphosyntax). However, Siyanova-Chanturia and Martinez (2014) note that studying frequency effects in multiword utterances has attracted less empirical attention than examining its effects for word level. Finding evidence for the role of frequency in larger lexical entries, such as formulaic language, is not straightforward. As noted by Jurafsky (2003), the frequency of multiword sequences is lower than that of individual words, making it challenging to separate effects arising from the constituents of the sequence and the sequence itself. It is important to note that investigating binomials and their reversed forms circumvents this issue, as the constituents of the frequent and less frequent forms are the

same. Crucially, many studies have demonstrated an impact of phrasal frequency on processing. A representative summary of this literature is provided in the next section.

3.3 Formulaic Language Processing in Native Speakers

Many studies have demonstrated, using different types of formulaic language and methodologies, that NSs' processing is facilitated for formulaic language compared to matched novel or less frequent language. Using a phrasal decision task, Bod (2001) demonstrated that English NSs needed less time to judge the acceptability of high frequency strings than low frequency strings (e.g., *I like it* vs. *I keep it*). Both Arnon and Snider (2010), using a phrasal decision task, and Tremblay, Derwing, Libben and Westbury (2011), using self-paced reading, provided evidence that high frequency phrases (*e.g., don't have to worry, in the middle of the*) yielded faster response time than low frequency phrases. In two follow-up word and sentence recall experiments, Tremblay et al., (2011) demonstrated that more frequent phrases (e.g., *in the middle of the*) had a better recall accuracy than less frequent ones (e.g., *in the front of the*).

Evidence for phrasal frequency effects can also be drawn from priming and self-monitoring studies. For example, using a lexical decision task, Durrant and Doherty (2010) investigated whether processing of the second word of a collocational pair varied as a function of its association with the first word. They found that the initial words in highly frequent collocations (e.g., *foreign debt*) and highly frequent and strongly associated collocations (e.g., *estate agent*) significantly primed the second words in the strategic priming experiment (i.e., the prime word was presented for 600 milliseconds (ms)). However, only the highly frequent and strongly associated collocations showed facilitative effects in the automatic priming experiment (i.e., the prime word was presented for 60 ms). Using a word monitoring task, where participants were asked to press a key when they heard *of,* Sosa and MacFarlane (2002) observed slower detection rates for *of* when it appeared in high frequency

collocations (e.g., *lot of, one of*) relative to low frequency collocations (e.g., *sense of*). Based on this finding, they concluded that high frequency collocations are represented as an autonomous unit, thus reducing the identifiability of their parts.

More recently, using eye-tracking, Carrol and Conklin (2020) investigated the effect of frequency (among other variables) on the processing of binomials, and collocations[4] for English NSs. The findings indicated that phrase frequency, cloze probability, and semantic association predicted the reading patterns for binomials (i.e., cloze probability was assessed in a norming study where participants provided completions of the phrases). As for collocations, the findings indicated that cloze probability and MI predicted the reading patterns. This suggests that while phrase frequency, among other factors, was associated with the processing advantage observed for binomials, predictability and frequency of co-occurrence was associated with the processing advantage observed for collocations.

Further evidence for phrasal frequency effects can also be drawn from electroencephalography (EEG) studies. For example, Siyanova-Chanturia, Conklin, Caffarra, Kaan, and Van Heuven's (2017) study pointed to a similar conclusion. Siyanova-Chanturia et al. (2017) used event-related potentials (ERP) to investigate binomial processing using frequent binomials (e.g., *knife and fork*), novel associated phrases (e.g., *spoon and fork*) and novel non-associated phrases (e.g., *theme and fork*). In a following experiment, they introduced the same stimuli without the conjunction *and* (e.g., *knife-fork, spoon-fork, theme-fork)*. Findings indicated that binomials yielded larger P300s and a reduced N400s compared to novel associated phrases and novel non-associated phrases. The larger P300s suggests an activation of something akin to a 'template' represented in memory for binomial phrases and the reduced N400s suggests a faster processing and easier semantic integration. Crucially,

[4] This study also investigated the processing of idioms. However, I concentrate here only on results for binomials and collocations as they are more relevant to the thesis.

when binomials were not introduced in their predictable way (without the conjunction), processing differences between phrases disappeared, indicating that processing advantage is only observed when there is a match between the stimuli being processed and the template represented in memory. This led to the conclusion that frequent lexical patterns, when presented in their conventionalised form, are characterized by faster processing, more facilitated semantic integration, and a "pre-activation of the mental template that uniquely matches the unfolding configuration" (Siyanova-Chanturia et al., 2017, p. 121).

As can be seen, these studies, which employed different measures – word monitoring task, eye tracking, ERPs, self-paced reading, word and sentence recall, phrasal and lexical decision tasks – all demonstrated a processing advantage for formulaic sequences, which was attributed to phrasal frequency. Thus, it appears that increased frequency leads to greater entrenchment in memory, which is then reflected in a faster processing. Further, such findings support usage-based models that assume that frequency effects extend single word *and* multiword sequences.

3.4 Formulaic Language Processing in Non-native Speakers

The literature on NSs demonstrates an effect of frequency, such that more frequent formulaic sequences are processed more quickly than matched novel or less frequent sequences. However, research on NNSs is still scarce (Siyanova-Chanturia & Van Lancker Sidtis, 2019). Whether NNSs are attuned to phrasal frequency is discussed in what follows.

An early study by Underwood, Schmitt, and Galpin (2004) used eye-tracking to investigate the processing of formulaic language (i.e., idioms) by NSs and NNSs. They integrated idioms into a reading passage and measured the number and duration of fixations on the last word of the idiom (e.g., formulaic: "*met the deadline by the skin of his teeth*" vs. non-formulaic: "*the dentist looked at his teeth"*). They found that while English NSs had fewer and shorter fixations on the final words of formulaic language, as compared with non-

formulaic control utterances, NNSs only had fewer fixations but not shorter fixation durations.

Attempting to replicate the study by Underwood et al. (2004), Conklin and Schmitt (2008), using a self-paced reading task with NSs and NNSs, measured the reading latency for idioms used figuratively and literally such as ("*everything but the kitchen sink"*) and non-formulaic utterances. They found that both NSs and NNSs had shorter reading times for idioms than matched non-formulaic utterances. Similarly, Siyanova-Chanturia, Conklin, and Schmitt (2011) monitored the eye movements of NSs and NNSs' reading idioms (e.g., *at the end of the day*) and matched control phrases (e.g., *at the end of the war*). Contrary to the Conklin and Schmitt (2008) study, they found that while NSs spent less time reading idioms compared to novel phrases, no processing advantage was detected among advanced L2 learners of English. Carrol and Conklin's (2014, 2017) studies pointed to a similar conclusion. Using a lexical decision task and eye-tracking, they found a faster processing for idioms (e.g., *spill the beans*) over control phrases by English NSs. However, this processing advantage did not extend to NNSs.

While idioms have been explored extensively, binomial processing has received less attention. An early investigation of binomial processing among NNSs was carried out by Siyanova-Chanturia, Conklin, and van Heuven (2011). Using eye tracking, they found that phrasal frequency (i.e., frequent vs. infrequent) alone did not explain binomial processing. It appeared that phrasal entrenchment (i.e., binomial vs. reversed) had an effect over and above that of frequency; once frequency is accounted for, there was still a difference between the entrenched phrase and the reversed form. This pattern was found for both NSs and higher proficiency NNSs, but not for lower proficiency NNSs. With increased proficiency there is likely increased exposure (i.e., greater frequency) to the L2, which leads to more entrenchment in memory for formulaic sequences. It should be noted, however, that

Siyanova-Chanturia et al. did not account for L1 congruency (L2 participants came from different L1 backgrounds), a variable, as we will see later, that has been shown to influence online formulaic language processing.

We can see that there is a clear advantage for formulaic language amongst NSs that is not always evident in NNSs. Why is this the case? Wray (2002) attributed the differences between the two groups to the different learning strategies they employ. Wray proposed that unlike NSs, L2 learners (or at least low-proficiency L2 learners) have a tendency to assemble the individual parts of the formulaic utterance compositionally. Focusing on the meaning of the multiword sequences as the sums of its part may obscure the idiomatic meaning of the whole phrase (where the parts do not yield the meaning of the whole). It is important to note that many studies have been unable to demonstrate Wray's analytical account. For example, Sonbul and Schmitt's (2013) study on collocations demonstrated that NSs do not have an advantage when it comes to learning collocations. Both NSs and NNSs learned more from direct input (i.e., decontextualized and enhanced) than from indirect input (i.e., enriched contextual exposure). Other evidence contradicting Wray's claim comes from Hernández, Costa, and Arnon's (2016) study, where intermediate and lower proficiency NNSs demonstrated sensitivity to multiword frequency.

It is worthwhile mentioning that many studies which have challenged Wray's account have said that if L2 learners do not show phrasal frequency effects, it might be due to insufficient L2 exposure rather than fundamentally different processing (e.g., Durrant & Schmitt, 2010; Northbrook & Conklin, 2019; Siyanova-Chanturia et al., 2011). More specifically, NNSs, particularly lower proficiency NNSs, might not have encountered formulaic phrases frequently enough for them to be sufficiently entrenched in memory for a processing advantage to emerge (Conklin, 2019). Increased L2 proficiency should lead to

more entrenchment of formulaic phrases in memory and then the typical processing advantage should be apparent.

Another explanation for the discrepancies between NSs and NNSs' performance might be due to the fact that researchers have relied on corpus frequency for identifying formulaic sequences and/or as a variable in their analyses. While reflecting an approximation of language use, corpus frequency may not reflect actual language input. Conklin (2019) claims that the mismatch between corpus frequency and actual language exposure should be more evident for L2 learners. Conklin (2019) went on to propose that subjective familiarity measures may be a better tool to assess frequency of exposure, in particular in the L2.

Northbrook and Conklin (2019) more directly explored the relationship between the input L2 learners receive and their performance on a phrasal judgement task for lexical bundles. Participants (L1 Japanese) were studying English as a foreign language (EFL). Northbrook and Conklin compared lexical bundles frequent in the participants' EFL textbooks (e.g., *do you play*) to matched lexical bundles unattested in textbooks (e.g., *do you hear*). The authors found that the L2 learners responded faster and more accurately to lexical bundles that appeared in their textbooks. They also found frequency effects in processing within these lexical bundles extracted from textbooks. This study indicates that, similar to child L1 learners, L2 learners are influenced by input. Thus, if L2 learners are tested on utterances extracted from their input, they exhibit a processing advantage for formulaic language and frequency effects. In other words, L2 learners may not produce or process formulaic language in a similar manner to NSs because they may have been exposed to input which is different in type and amount to that of L1 learners.

3.5 Factors Influencing Formulaic Language Processing for Non-native Speakers

3.5.1 *Proficiency*

The amount of exposure to a language, be that the L1 or L2, is at the heart of usage-based approaches to language acquisition (Bybee, 2006; Tomasello, 2003), whereby repetitive encounters of a lexical item strengthen the connections between its form and meaning. Thus, familiarity with recurrent patterns depends to a large extent on the number of encounters with such patterns. With increased proficiency it is likely that there has been greater exposure to a language, which ensures a greater entrenchment of formulaic utterances and a concomitant processing advantage. For example, Wolter and Yamashita (2017) demonstrated that with increased proficiency, NNSs gradually develop more sensitivity to collocation frequency than to word-level frequency of collocations.

Sonbul's (2015) study is one of the rare studies to investigate whether phrasal frequency effects are modulated by proficiency. The study showed that both NSs and NNSs exhibited frequency effects for collocations (e.g., *fatal mistake* vs. *awful mistake* vs. *extreme mistake*) in online (eye-tracking) and offline tasks (rating how typical the collocation is in English). Notably, Sonbul (2015) found a correlation between proficiency and performance in the typicality rating task, but no effect of proficiency in online reading.

More recently, Hernández, Costa, and Arnon (2016) examined NNSs' sensitivity to phrasal frequency along a continuum (e.g., *do not have to worry* vs. *do not have to wait*; *do not know how much* vs. *do not know how many*). They also investigated whether performance was modulated by different types of L2 exposure (classroom vs. immersion). On the phrasal decision task, both NSs and NNSs responded faster to high frequency phrases than low frequency ones, suggesting similar sensitivity to frequency for both groups. Crucially, both groups' sensitivity to frequency was continuous across a range of frequencies. Notably, type of exposure – naturalistic or immersion – did *not* modulate the frequency effects, indicating that a naturalistic learning context is not required to develop a sensitivity to the frequency of multiword utterances. Also, the proficiency level of the L2 learners did not influence their

response times.

Achieving a native-like competence of formulaic language almost certainly requires an advanced level of L2 proficiency and extensive exposure to L2 input. However, achieving high L2 proficiency does not necessarily imply that L2 learners have the same productive and receptive knowledge as NSs. As shown by Durrant and Schmitt (2009), even highly proficient NNSs may only use a limited set of collocations (i.e., in particular, collocations with high MI scores). In addition to productive knowledge, their intuitions of collocation frequency may be different. Siyanova-Chanturia and Schmitt (2008) showed that even when NNSs have a similar repertoire of collocations to that of NSs, their perception and intuition about the frequencies of them may not mirror NSs' intuitions.

It is important to bear in mind that it is difficult to determine what level of proficiency is needed for formulaic language processing and/or production to be native-like. And frequency effects for formulaic language may be apparent before native-like proficiency is achieved. For example, both Hernàndez, Costa, and Arnon (2016) and Siyanova-Chanturia, Conklin, and van Heuven (2011) found evidence for phrase-frequency effects in NNSs of different proficiency levels.

NSs and NNSs might also demonstrate differences in processing for formulaic sequences because L2 speakers bring their L1 knowledge to any processing task. Yamashita and Jiang (2010) and Wolter and Gyllastad (2013) stress that an L1 may facilitate or impede processing of formulaic language. The next section will discuss how congruency (L1-L2 overlap) influences the processing of formulaic sequences.

3.5.2 *Congruency Effects*

While proficiency is an important factor in L2 processing, it is not the only one. Another potential factor is the linguistic overlap between the L1 and L2. Unlike monolingual children learning their L1, adult L2 learners are equipped with a mature cognitive capacity

which has already been wired with a linguistic system – the L1 (MacWhinney, 2017). MacWhinney (2017) remarks that previously encoded L1 patterns may interfere with the new non-corresponding L2 patterns. For example, malformed formulaic utterances in an L2 may emerge due to being translated from L1 formulaic utterances, which are not matched in the L2. Evidence for this argument can be drawn from many corpora-based studies which demonstrated that L2 collocation pair choice may be altered by L1 collocational counterparts (e.g., Granger, 1998). On the other hand, where there is an overlap between the two languages, it may also facilitate processing (Wolter & Gyllastad, 2013; Yamashita & Jiang, 2010).

The two languages of bilinguals are thought to dynamically interact with each other (De Groot, 2011) and this could lead to congruency effects (i.e., effects due to the linguistic overlap between an L1 and L2). A plethora of research has demonstrated parallel activation of both L2 speakers' languages when using one (Bialystok, 2001). This co-activation has important implications for L2 processing. In particular, when there is orthographic, phonological and/or semantic similarity across languages, we have seen an influence of one language on the processing of the other (e.g., Cristoffanini, Kirsner, & Milech, 1986; Dijkstra & van Heuven, 1998; Kroll, Michael, Tokowicz, & Dufour, 2002; Dijkstra & van Heuven, 2002). Further, co-activation occurs when languages share a script and when their scripts differ (e.g., Allen & Conklin, 2013). Such co-activation can be explained by an automatic spread of activation to similar items which is blind to language identification (Dijkstra & Van Heuven, 2002).

While most studies exploring non-selective, co-activation across language have investigated items at the word level, fewer studies have attempted to explore co-activation, congruency effects for multiword sequences. The studies that have explored this are considered in what follows.

Congruency Effects in Formulaic Language. While some formulaic utterances exist in similar form or meanings across languages (i.e., congruent), other formulaic utterances do not necessarily correspond across languages (i.e., incongruent, e.g., some formulaic utterances are language specific). How formulaic language overlaps across languages spans a spectrum. For example, formulaic language can 1) overlap in meaning and form (e.g., some binomials such as *food and drink* exist in English and Arabic and with the same specific order preference; another example is the English collocation *warm reception* which has a word-for-word translation with the same meaning in an Arabic collocation); 2) overlap in meaning and partially overlap in form (e.g., binomials such as *day and night* in English and *night and day* in Arabic share the same words in the two languages but with a reversed order preference); 3) overlap in meaning but not in form (e.g., to describe something as expensive: an English speaker would say *costs an arm and leg* and an Arabic speaker might say *costs the blood of my heart*); 4) or have no formulaic equivalent in either language (e.g., some binomials such as *safe and sound* only exist in English and do not correspond literally to a binomial phrase in Arabic; some binomials such as *mosques and charges* only exist in Arabic).

Yamashita and Jiang (2010), following Jiang (2000), put forward a framework to explain the role of congruency in formulaic language processing. According to them, processing congruent collocations involves using ready-made links to L1 knowledge which assists comprehension, whereas processing incongruent collocations is either more compositional and/or requires contextual cues to infer meaning, which takes some effort. Yamashita and Jiang say that initially when learning collocations, L2 learners use links to L1 collocational counterparts to map the L2 collocational meanings. This process helps ensure faster recognition of L2 collocations, if they have an equivalent L1 collocation counterpart. Crucially, they believe that L2 learners need extensive exposure before they can overcome the effortful processing for incongruent items.

To explain the role of the congruency effects for idioms, Carrol, Conklin, and Gyllstad (2016) put forth a similar explanation, emphasizing the role of cross language priming. Specifically, when Swedish-L1 learners of English-L2 encounter an English word (e.g., *break*), which comprises an initial part of an idiom phrase, its L1 counterpart receives activation (e.g., the Swedish translated equivalent *bryta*). Assuming that the components of idioms are connected in the lexicon and/or represented, they propose that, activation of the L1 word would pass activation to the L1 idiom (*bryta isen*, literally 'break ice' and figuratively having the same meaning in the two languages). Activation of the L1 idiom activates its components (*bryta + isen*) as well as its English counterparts (*break + ice*) which speeds processing of the L2 idioms.

Considerable research has investigated the role of congruency in the processing of formulaic language (e.g., Carrol, Conklin, & Gyllstad, 2016; Wolter and Yamashita, 2017; Yamashita and Jiang, 2010). Most of the studies investigated congruency effects using idioms or collocations. For example, Carrol, Conklin, and Gyllstad (2016) used eye-tracking to look at the processing of three categories of idioms: English-only idioms, not found in Swedish; Swedish-only, translated idioms, not found in English; and congruent idioms, found in both languages. For English NSs, findings demonstrated the expected processing advantage for English-only and congruent idioms; however, that advantage was not found for unfamiliar Swedish-only items. For Swedish advanced learners of English, they found a processing advantage in congruent and Swedish-only idioms (i.e., with an advantage for idioms over controls), but not in the English-only idioms (i.e., no difference between idioms and controls).

Titone, Columbus, Whitford, Mercier, and Libben (2015) investigated the effect of congruency on processing English idioms that varied in their cross-language overlap with French. English-French bilinguals read English sentences containing idioms or matched

control phrases (e.g., *she lived*/*told a lie*) in a word-by-word presentation, where the final word was either in English or in French (e.g., intact condition: *he played with fire* vs. code-switched condition: *he played with feu*). Participants were asked to judge the meaningfulness of each sentence. They found that the code-switched condition was more disruptive for idioms than for control phrases, as indicated by judgment time. Importantly, congruency between languages reduced language switch cost, such that increased cross-language overlap reduced response times for idioms in the code-switched condition.

Beck and Weber (2016) did not find an effect of congruency. They presented German-English bilinguals and English monolinguals with audio sentences containing English idioms differing in their cross-language overlap: word-for-word translation equivalents vs. matching idiomatic concepts that could *not* be translated word-for-word. Presentation of idioms (e.g., *pull someone's leg)* was followed by a lexical decision task on English words which were either literally related to the idiom (e.g., *walk*), figuratively (e.g., *joke*), or an irrelevant control word. Both L1 and L2 participants showed a priming effect for literal and figurative targets compared to control words, with no differences in response times between literal and figurative targets. Notably, the study did not find any difference based on congruency, meaning that all idioms, regardless of their translatability, showed similar priming effects.

Using eye-tracking, Cieślicka and Heredia (2017) compared the processing of congruent and incongruent idioms by Spanish-English bilinguals. Participants read idioms that shared their form and meaning (similar idioms, e.g., English: *save your skin*, Spanish: *Salvar el pellejo*) or that differed in form but shared meaning (different idioms, e.g., English: *skate on thin ice*, Spanish: *Ir pisando huevos).* The study failed to demonstrate a benefit of congruency on processing. In fact, Cieślicka and Heredia found that similar idioms were processed more slowly than different ones. They suggested that the processing cost was due

to interlingual competition. However, the finding of a processing cost for congruent idioms is at odds with the wider literature that either finds no difference for congruent and incongruent idioms or a benefit for congruent ones.

In addition to idioms, evidence for congruency effects has also been observed in collocations. For instance, Yamashita and Jiang (2010) asked English NSs, Japanese studying English as a second language (ESL) participants, and Japanese EFL participants to judge if phrases were acceptable in English when they were congruent collocations, existing in both languages (e.g., *heavy stone*) or incongruent, existing only in English (e.g., *kill time*). English NSs showed no difference in response times and accuracy for congruent and incongruent items. For Japanese EFL learners, incongruent items had longer response times and resulted in more errors compared to congruent items. The ESL learners behaved similarly to EFL learners in that they had more errors for incongruent than congruent items; however, the ESL response times resembled the NS group – with no difference between the two item types. The results highlight the difficulty even advanced learners have with collocations which are not shared with their L1.

Support for the role of congruency effect can also be found in a study by Wolter and Yamashita (2017). They examined how knowledge of collocations in the L1 (Japanese) influences the processing of collocations in the L2 (English), using an acceptability judgment task.[5] They also investigated the interaction between proficiency and collocational frequency on processing. English NSs and NNSs (assigned to two groups based on proficiency levels: high vs. low) were instructed to judge if collocational pairs are acceptable in English. NNSs (of both proficiency groups) recognized congruent collocations (i.e., existing in both Japanese

[5] By employing an acceptability judgment task, instead of double lexical decision task, the researchers tried to overcome the limitations of their previous study (Wolter & Yamashita, 2015). They argued in favour of using an acceptability judgment task as they believe it encourages participants to focus on meaning rather than on form. They assume that a lexical decision task puts the focus on the form of a string.

and English, e.g., *strong wind*) faster than incongruent collocations (i.e., English-only, e.g., *low speed*), while NSs demonstrated equal performance across those two conditions. Notably, there was an interaction between proficiency and frequency on processing, such that NSs as well as higher proficiency NNSs showed more sensitivity to collocational frequency than did lower proficiency NNSs. The authors concluded that with increased proficiency learners gradually shift from attending to word-level frequency to attending to phrase-level frequency.

An earlier study by Wolter and Gyllstad (2013) had similar findings. Also using an acceptability judgment task, they demonstrated that congruency modulated L2 participants' processing, and that highly proficient L2 learners exhibited frequency effects of L2 collocations, regardless of whether an item was congruent. Other studies reached the same conclusion that congruent collocations experience a processing benefit compared to incongruent ones (e.g., Wolter & Gyllstad, 2011).

Using a different kind of task – cue recall – Pritchett, Vaid, and Tosun (2016) found an effect of congruency. In their study, Russian-English bilinguals were exposed to adjective-noun combinations of idiomatic meanings in English and/or Russian (e.g., English only: *blue blood*; English-Russian: *blue moon*; Russian only: *blue distances*). Following an exposure phase, participants were instructed to write down any item they could remember. The test indicated a better recall rate for idiomatic phrases that exist in both languages, than for phrases that only exist in one language.

Broadly speaking, studies have linked congruency with ease of processing for formulaic utterances (e.g., Carrol, Conklin, & Gyllstad, 2016; Wolter & Gyllstad, 2011; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2017; Yamashita & Jiang, 2010). However, some studies failed to confirm the role of congruency (e.g., Beck & Weber, 2016; Cieślicka & Heredia, 2017). Such discrepancy in results might be attributed to differences in how congruency is operationalised in these studies (i.e., they considered different degrees of

cross-language overlap when classifying congruent/incongruent items). Most studies which demonstrated a faciliatory effect of congruency on processing used incongruent items that had no conceptually and lexically matching translation equivalents in either language. Other studies (e.g., Beck & Weber, 2016; Cieślicka & Heredia, 2017) used incongruent items that share the same idiomatic concepts but could not be translated word-for-word.

Taken together, this section and the previous one indicate that proficiency and congruency are two key factors explaining L2 learners' performance on formulaic language. However, the discussion thus far has focused on L2 processing of formulaic language and not on its acquisition. It is important to consider how formulaic language is learned, and as has been highlighted elsewhere, there is still a great deal of uncertainty about this (Pellicer-Sánchez, 2017).

3.6 The Learning of L2 Formulaic Language

An important question is: how do we ensure that L2 learners acquire formulaic language? Two divergent learning approaches emerged: intentional learning and incidental learning. Intentional learning refers to the learning process that occurs when the learners' attention is directed to the learning process (Nation, 2001). Incidental learning refers to "the more or less 'unintentional', 'incidental' acquisition (or 'picking up') of language (grammar, vocabulary, orthography, pronunciation, etc.) during the performance of communicative tasks requiring attentional focus on the meaning and function of language rather than on its form" (Hulstijn, 2003, p. 373). Implicit learning is the learning mechanism involved in an incidental learning task.

Nation (2001) and Schmitt (2008) argue that intentional learning is more effective in L2 vocabulary learning than L1 learning. Research in favour of intentional learning – explicitly teaching – of formulaic language has exploited learning methods that promote noticing the target items, based on Schmidt's (1990, 2001) noticing hypothesis. This

hypothesis states that intentionally focused attention ('noticing') directed to linguistic features is essential for successful learning. It implies an explicit focus on forms and meanings of formulaic sequences. For example, Laufer and Girsai (2008) employed explicit contrastive and translation instruction (i.e., participants performed L2-L1 and L1-L2 translation tasks) and found this method to be effective for promoting collocational knowledge. Szudarski (2012) found that learners' knowledge of collocations was enhanced by the addition of explicit exercises (e.g., gap-filling exercises) that followed a reading activity. Chan and Liou (2005) explored the effect of intentional learning for collocational knowledge through the use of concordancer (i.e., teaching with bilingual concordance lines from a web-based bilingual concordancer). They found that explicit learning through an online concordancer triggered the learning of collocational knowledge. Laufer (2011) employed a dictionary task, presenting L2 learners with verb-noun collocations with a missing verb and asked them to use a dictionary to complete the missing collocations. The findings suggested that use of a dictionary improved collocational knowledge.

While explicit learning methods are effective in promoting collocational knowledge, they depend on tasks that require time to carry out. However, the amount of formulaic language more broadly, and collocations and binomials more specifically, is too large to be taught to L2 learners using traditional explicit methods in limited classroom time – or even outside of the class as 'homework'. Thus, we need to think of the less demanding and more effective ways of introducing formulaic sequences to learners. Recently, studies have started to explore whether formulaic sequences can be learnt incidentally (e.g., Conklin & Carrol, 2020; Durrant & Schmitt, 2010; Pellicer-Sánchez, 2017; Sonbul & Schmitt, 2013; Szudarski, 2012; Szudarski & Carter, 2016; Toomer & Elgort; Vilkaitė, 2017; Webb, Newton, & Chang, 2013).

Ellis (1999) points out that vocabulary is incidentally acquired when learners' attention is directed towards comprehending the meaning of the text rather than towards learning the words. Importantly, incidental learning requires less conscious effort. Research has demonstrated the incidental learning of single words in an L1 learning context (e.g., Jenkins et al., 1984; Nagy, Anderson, & Herman, 1987; Nagy, Herman, & Anderson, 1985; Shu, Anderson, & Zhang, 1995) and also in an L2 one (e.g., Day, Omura, & Hiramatsu, 1991; Dupuy & Krashen, 1993; Horst, 2005; Horst, Cobb, & Meara, 1998; Hulstijn, 1992; Kweon & Kim, 2008; Montero Perez, Peters, & Desmet, 2015; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Pitts, White, & Krashen, 1989; Van Zeeland & Schmitt, 2013). While the incidental learning of single word items has been studied extensively, less attention has been paid to incidental learning of multiword items. Incidental learning at the word level is different from the multiword level, which requires acquiring knowledge of the co-occurrences of words at a certain distance, and *if* a multiword item has a figurative meaning, it requires mapping a meaning that is not the sum of its parts. Studies investigating the incidental learning of formulaic language are discussed in what follows.

Vilkaitė (2017) examined whether collocations could be learned incidentally when they were adjacent (e.g., *designed experiments*) or non-adjacent (e.g., *design a number of different experiments*). L2 learners' collocational knowledge was tested using form recall and recognition tests. The findings revealed immediate collocational gains that lasted two weeks as a result of incidental learning. Interestingly, the adjacency of the collocations had no effect on the learning outcomes, suggesting that L2 learners can detect the distributional properties of collocations even if the components are not contiguous. The study also indicated that L2 proficiency promoted collocational knowledge.

Durrant and Schmitt (2010) examined the effect of frequency on learners' retention of collocations. In this study, highly proficient ESL learners were instructed to read aloud novel

collocations (e.g., *warm flat*) embedded in sentences. Collocations used in the study varied in number of occurrences (i.e., one vs. two) and in type of repetitions (i.e., verbatim repetition: the target collocation appeared twice in the same sentence vs. varied repetition: the target collocation appeared twice in different sentences). Post-test results (of a recall task) showed that L2 learners retained memory traces of collocations even after one exposure, with two exposures yielding higher recall compared to the single exposure condition. Interestingly, the varied repetition condition was more effective than the verbatim repetition condition. However, the explicit nature of the task used (read-aloud) limits the generalizability of the findings. Another limitation of the study is that it did not look at the durability of the learning gains, which was addressed in the following study by Sonbul and Schmitt (2013).

Sonbul and Schmitt (2013) investigated whether different learning conditions (i.e., enriched input, enhanced input, and decontextualized input) promoted explicit versus implicit knowledge of collocations (i.e., unknown medical collocations, e.g., *vanishing lung).* Target collocations were either presented: three times in context (reading-only condition), three times in context and made salient with bold font (reading-plus-enhanced condition), or for 10 seconds onscreen, and participants were asked to learn them (decontextualized condition). Explicit knowledge was tested using form recall and recognition tests, and implicit knowledge was tested with a priming paradigm in a lexical decision task. A delayed test (two weeks after exposure) showed that all exposure conditions resulted in gains for both NSs and highly proficient NNSs. However, the lexical decision task showed that neither condition promoted implicit knowledge of collocations for both groups. The finding also indicated that NNSs benefited more from enhanced input than from unenhanced input (i.e., reading-only condition), whereas NSs did not exhibit any differences. Notably, the Durrant and Schmitt (2010) and Sonbul and Schmitt (2013) studies are amongst the few to show that even limited exposure to collocations results in learning gains.

Sonbul and Schmitt's (2013) study was recently replicated by Toomer and Elgort (2019) who questioned whether increasing the number of encounters (nine instead of three as in Sonbul and Schmitt) would promote implicit knowledge, which was not observed in the original study. Further, Toomer and Elgort predicted that difficulty in learning formulaic sequences could be due to a lack of salience. To address this, they introduced another learning condition: bolding-plus-glossing. Toomer and Elgort replicated Sonbul and Schmitt's findings, demonstrating durable explicit knowledge of collocations. The superiority of reading-plus-enhanced condition over reading-only condition for L2 learners' accuracy responses was also replicated. Interestingly, the bolding-plus-glossing condition was no more advantageous than the reading-only condition. Importantly, Toomer and Elgort found evidence of implicit knowledge of collocations, although only in the reading-only condition, which was attributed to the increased in the number of repetitions.

The use of low-frequency, medical collocations in Sonbul and Schmitt (2013) and Toomer and Elgort (2019) helped control for participants' prior knowledge. However, the nature of technical, medical collocations limits their generalisability to collocations used in everyday language. First, medical collocations are often made up of words that are of low frequency (e.g., *vicarious haemorrhage*, with the first word occurring 95 times and the second occurring 323 times in the BNC). This means that both the form and the meaning of the individual words that make up a collocation may be unknown, which adds an extra burden when learning. In addition, the degree of transparency was not controlled for (i.e., *cloud baby, stone heart, smooth diet, and golden hour* seems less transparent than *regional control* or *gene therapy*, thus possibly resulting in participants paying increased attention to their forms). Concerns about these studies make it difficult to draw clear conclusions about learning gains for collocations in the different conditions. Any suggestions about the potential effect of the treatments explored in these studies need to be viewed with caution.

It is important to note that, Sonbul and Schmitt's (2013) and Toomer and Elgort's (2019) findings are contrary to those of Szudarski (2012). Szudarski (2012) explored the effect of different types of input (reading-only vs. reading-plus) on learning delexicalised verb-noun collocations (e.g., *make a profit, give birth*) which appeared six times in text. EFL participants were assigned to three groups: (reading-only, reading-plus, and a control group). While the reading-only group read stories containing the target collocations (i.e., incidental learning condition), the reading-plus group read the same stories but were also asked to complete some explicit activities targeting the collocations. Both productive and receptive tests revealed that the reading-plus condition led to greater collocational knowledge than reading-only. Crucially, the reading-only group was not different from the control group. Based on the findings, Szudarski concluded that the number of incidental encounters (i.e., reading-only condition) with the target collocations needs to be higher for considerable learning to occur. However, the target items used in this study were delexicalised verb noun collocations, which seems to be problematic for L2 learners (Altenberg & Granger, 2001). For example, since nouns carry most of the semantic weight in such collocations (Szudarski, 2012), it is possible that learners paid less attention to verbs while reading.

The Effect of Frequency of Exposure on Incidental Learning

Since incidental learning is influenced by repetition (Nation, 2001) and formulaic knowledge is influenced by repeated exposures (Hoey, 2005; Siyanova-Chanturia & Martinez, 2015), the role of the number of encounters (i.e. frequency) is an important area of research in formulaic language learning. As discussed above, input is thought to be one of the main drivers of the emergence of a formulaic language processing advantage. However, many questions remain about the most advantageous type of input, e.g., precise number of exposures needed, benefit of spaced versus massed input, etc. (Conklin, 2019). Furthermore,

would more encounters with a formulaic sequence during reading result in better learning gains.

Frequency of occurrence refers to the number of times linguistic elements, specifically formulaic sequences in this case, are encountered by language users. The frequency with which language users encounter L2 language patterns determines their degree of entrenchment in memory (Carrol et al., 2016). The role of repeated exposures on memory representations is nicely articulated by James (1890): "Objects once experienced together tend to become associated in the imagination, so that when any one of them is thought of, the others are likely to be thought of also" (p. 561). In this way, when we encounter the word *abject* we are likely to think of the word *poverty*.

Arguing that frequency of occurrence influences the way formulaic language is represented is in accord with usage-based models highlighting the role of language experience (Bybee, 2006; and Tomasello, 2003). Because formulaic language is experience driven (Ellis, 2001), one could argue that the more exposed an L2 learner is to L2 input, the more fluent they will be at recognizing the distributional properties of multiword units. As articulated by Durrant and Schmitt (2010), "adult learners who implicitly retained memories of the words that go together in their input ought to establish strengths of association that are appropriate to that input" (p. 168). Many researchers have emphasized the role of even one single exposure to a word combination. The first encounter with a language pattern forms an initial memory trace that is further consolidated by subsequent encounters (Logan, 1988). As stated by Sonbul and Schmitt (2013), each exposure to a word combination contributes to either storing a new word pair or to reinforcing/updating a previously encoded word-pair. This view was further substantiated by Goldberg (2007) who maintains that since repeated exposures to a language pattern have an effect on the way it is represented, then, it can be concluded that even one exposure to a linguistic unit may leave a memory trace. For

example, Gurevich, Johnson, and Goldberg (2010), found that NSs were able to significantly recall sentences even after a single exposure. They were also able to recall the sentences and reproduce them in a delayed test (six days after the initial test). Thus, a single exposure was enough to create an initial memory trace. In short, the more exposure to frequently paired words (i.e., binomials and collocations), the stronger their entrenchment in memory, such that the presence of one element should elicit activation of the other.

Frequency effects may be more profound for L2 learners than adult L1 users. Conklin (2019) points out that less frequent items benefit more from increased exposure. Items that are high frequency are generally at, or close to ceiling; thus, performance cannot be improved much by increased exposure. In contrast, increased exposure to low frequency items can make their recognition faster. Because L2 learners generally have less exposure to a language than L1 speakers, increased exposure will result in a greater processing benefit for L2 speakers (Conklin, 2019).

Webb, Newton, and Chang (2013) looked at effects of frequency (1, 5, 10, and 15 repetitions) on EFL collocation learning in a reading-while-listening task. Their findings indicated that the number of encounters significantly improved collocational knowledge (i.e., more than 1 encounter significantly increased performance on receptive test scores). Encountering collocations 15 times resulted in the best performance, and 5 occurrences resulted in better gains than the control group. The researchers highlighted the role of contextually based repeated exposure for incidental collocational learning. They concluded that five encounters seem to be the acceptable threshold for learning gains.

Pellicer-Sánchez (2017) investigated learning gains for six unknown adjective-pseudoword collocations (e.g., *loud twoser*) presented four or eight times in a story context based. One week later, learners' knowledge of the form and meaning of the collocations was assessed (at recall and recognition levels). Results demonstrated successful incidental

learning of these collocations. Interestingly, there was no difference in learning gains for four and eight exposures. However, we should be cautious when interpreting these results, as the saliency of the pseudowords may have attracted participants' attention, thus promoting better learning than would occur with known words. In addition, the pseudowords replaced actual words in existing collocations (*magic cambul:* 'magic ring'; *loud twoser:* 'loud noise'). Thus, participants may have mapped the words onto the pseudowords and existing collocations. In this way, collocational knowledge may have assisted learning.

The lack of a frequency effect observed in Pellicer-Sánchez (2017) was also noted by Szudarski and Carter (2016). They examined the learning of delexical verb-noun collocations which appeared either 6 or 12 times in different conditions (reading-only or reading with enhancement via underlining). Their results indicated that the reading with enhancement condition led to greater learning gains than the reading-only condition, irrespective of the number of encounters.

Taken together, some studies have indicated that learning collocations incidentally from input is possible (e.g., Sonbul & Schmitt, 2013; Toomer & Elgort, 2019), while others have failed to confirm incidental learning (e.g., Szudarski, 2012). Much of the current literature pays particular attention to the role of frequency of encounter on incidental learning. However, there has been little agreement. Durrant and Schmitt (2010) and Webb, Newton, and Chang (2013) found that increasing the number of encounters to a collocational phrase positively affect the learning outcome, while Pellicer-Sánchez's (2017) and Szudarski and Carter's (2016) studies did not demonstrate an advantage for increased exposures.

Across the various studies, the methodologies in terms of the items themselves (e.g., real words vs. pseudowords) and how they were presented and assessed differed, making it difficult to compare across studies. Thus, there are still uncertainties about the effectiveness of different learning conditions in the learning of formulaic language. Further, the interaction

between frequency of encounter and congruency is still poorly understood. In general, increased frequency is expected to benefit learners. Overlap may benefit learning when an L1 form and meaning 'points' L2 learners in the 'right direction' but can be harmful when it points them in the 'wrong direction' (e.g., in the L1 it is *pepper and salt* and in the L2 it is *salt and pepper*). These are open questions that the current thesis will address.

3.7 General Summary and Conclusion

Collectively, the studies reported in this chapter reveal a somewhat inconsistent pattern. Some studies found a processing advantage for formulaic language in an L2 (e.g., Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & van Heuven, 2011), while other studies failed to demonstrate one (e.g., Underwood, Schmitt, & Galpin, 2004; Siyanova-Chanturia, Conklin, & Schmitt, 2011). The processing advantage in Siyanova-Chanturia et al. (2011) was only attributed to higher proficiency participants, highlighting the role of proficiency for formulaic language processing. Differences in the L2 populations recruited and differences in the proficiency measures administered, could account for the different findings. This suggests an urgent need to understand how/why the sensitivity evident in the L1 to formulaic language is not always observed in L2 populations.

A general observation that seems to meet with very little opposition is that congruency has a direct role on formulaic language processing. However, it remains unclear whether such congruency effects may interact with L2 learning and proficiency. For example, although Jiang (2000) noted that increased proficiency should limit L1 congruency effects when comprehending L2 items, many studies observed congruency effects in advanced L2 learners (e.g., Wolter & Gyllstad, 2011, 2013; Yamashita & Jiang, 2010). Further, Wolter and Yamashita (2017) found faster response times for congruent collocations over incongruent ones in a high proficiency group. Conversely, Yamashita and Jiang's (2010) study found the same pattern but only with the participants with lower proficiency, while those with high

proficiency had similar response times for congruent and incongruent items. Reasons for such discrepancies may be the frequency of L2 input (e.g., some studies recruited mainly EFL learners) and lack of standardized tests for proficiency in some studies.

The majority of studies examining congruency have generally focused on idioms and collocations. As noted by Conklin and Carrol (2019), language specific items (incongruent) tend to have a figurative meaning; therefore, figurativeness is often confounded with congruency in many studies (mainly in studies on collocations). Thus, research needs to carefully consider and/or control transparency alongside congruency. Importantly, to the best of my knowledge, no study has investigated the role of congruency on binomial learning and processing.

Research addressing the learning of formulaic language has not come to a consensus. Exploring this topic would advance our understanding of two elements of vocabulary learning that remain insufficiently explored: binomials and collocations. Since classroom time is generally very limited, making it challenging to explicitly teach collocations and binomials, investigating other methods that take less effort, such as incidental learning becomes an increasingly important area in language learning research.

The research on incidental learning has made use of items that are very diverse, making it difficult to draw clear conclusions: delexical verb-noun collocations (e.g., Szudarski, 2012; Szudarski & Carter, 2016), pseudoword collocations (e.g., Pellicer-Sánchez, 2017), technical, medical collocations (e.g., Sonbul & Schmitt, 2013; Toomer & Elgort, 2019), high frequency collocations (e.g., Vilkaitė, 2017; Webb, Newton, & Chang 2013), or novel collocations (e.g., Durrant & Schmitt, 2010). Another limitation of research in the field is the focus on assessing knowledge after treatment and not looking at processing during reading treatment. Examining reading behaviour during treatment with methods like eye-

tracking would advance our understanding of how online processing during reading can predict later learning outcomes.

The current thesis seeks to contribute to our understanding of formulaic language learning and processing in relation to the points summarised above, by carrying out three studies: 1) the first study used self-paced reading and forced-choice tasks to examine the learning and processing of binomial phrases while manipulating congruency and frequency of encounter; 2) the second study used eye-tracking and offline tasks to examine incidental learning and processing of novel transparent collocations; 3) the third study used eye-tracking and offline tasks to examine incidental learning and processing of novel opaque collocations.

Chapter 4. Learning of Binomial Phrases

4.1 Introduction

As demonstrated in the previous chapter, NSs have a reduced processing cost for formulaic language that is not always evident in NNSs. This leads to the question of whether NNSs attend to formulaic language in a different way than NSs, potentially by processing it more compositionally. There are a range of factors that may explain why some researchers found a processing advantage for formulaic language, while others did not: proficiency of the participants, participants' L1 and the relationship between the L1 and L2 (degree of overlap or congruency), frequency of encounter, time of testing, etc.

The purpose of the current study was to examine how incidental learning of formulaic language interacts with congruency and frequency of encounters. Specifically, this research investigated the role of congruency and frequency of encounter for binomial learning and processing among English NSs and Arabic learners of English. In order to achieve this, the research looked at binomials which exist only in English, only in Arabic, and in both languages (congruent). In order to examine learning gains, participants engaged in four tasks: 1. pre-test; 2. exposure to the binomials; 3. immediate post-test; and 4. a delayed test. The exposure session involved presenting participants with binomial items in a reading passage. Frequency of encounter was manipulated by presenting binomials either two times or five times in a short paragraph. Durability was investigated by testing participants immediately after their exposure and then again one week later. Therefore, the current study aimed to address the following questions:

1) Does incidental exposure to binomials help learners acquire them and how durable is the knowledge?

2) What is the role of repeated exposures on learning and durability of knowledge?

3) What is the influence of congruency on learning and durability of knowledge?

4.2 Methods

4.2.1 *Participants*

Two groups were recruited for this study: Arabic learners of English (n = 47, all male) and British NSs of English (n = 18, female = 10, male = 9). Participants were undergraduate students who participated in exchange for course credit or for £20 for participating in all sessions. The proficiency level of both groups of participants was estimated using two objective tests: *Lexical Test for Advanced Learners of English* (LexTALE; Lemhöfer & Broersma, 2012) and a short multiple-choice vocabulary test (Carrol & Conklin, 2014, 2017; Carrol, Conklin, & Gyllstad, 2016) (i.e., 20 vocabulary items from different frequency bands). The multiple-choice vocabulary test was used to account for the possibility that the LexTALE might involve overestimation of vocabulary knowledge. The vocabulary test is presented in Appendix 4A. Participants also filled out a questionnaire where they provided information about their language background and were asked to self-rate their knowledge of English on a 7-point scale (1 = very low, 7=native-like) (see Appendix 4B). Table 4.1 summarises the demographic and language proficiency data.

Table 4.1 Means with Standard Deviations in Parentheses of Demographic Data and Self-ratings of Proficiency on a 7-Point Scale (1 = very low; 7 = native)

	NS	***NNS***
Age	19.83(1.25)	20.19 (2.15)
Age of English acquisition	0 (0.0)	11.46 (2.88)
Years lived in an English-speaking country	19.44 (1.53)	0.24 (0.92)
Short vocabulary test score (a 20-point scale)	18.77 (0.87)	9.29 (3.25)
LexTALE score (%)	92.29 (6.72)	56.84 (8.21)
Self-rating (1 to 7)		
Overall English proficiency	7 (0)	4.44 (1.47)

Proficiency in speaking	7 (0)	4.17 (0.91)
Proficiency in understanding	7 (0)	5.57 (0.94)
Proficiency in reading	7 (0)	5.10 (1.28)
Proficiency in writing	7 (0)	3.91 (1.50)

4.2.2 *Materials*

Item Selection. The items were all literal binomial phrases of the form *noun and noun* and were of three types: English-only binomials (occurring only in an English corpus such as *fish and chips*), Arabic-only binomials (occurring only in an Arabic corpus but translated in English such as *coffee and dates*), and congruent binomials (occurring in both corpora such as *food and drink*). The following procedure was followed for selecting items. First, English and Arabic binomials were extracted from the BNC (of 100 million English words) and the KACST Arabic corpus (of 1,182,515,633 Arabic words) respectively. Second, only binomial phrases whose forward (e.g., *food and drink)* phrasal frequency was at least three times larger than their backward (e.g., *silver and gold*) frequency were selected. Third, the Arabic binomials were translated into English. Translation accuracy was confirmed in a rating task, where four Arabic native speakers with an advanced knowledge of English were asked to evaluate the proposed translation from one to five (5 = 'excellent translation' and 1 = 'poor translation'). Only phrases that achieved a rating score of four or five were selected (M =4.66, SD= 0.51). Lastly, the binomial items were all reversible, meaning that both orders were semantically and logically possible. This left us with 14 experimental items per binomial type. All of the experimental items were introduced during a reading exposure session and tested three times (before reading, immediately after reading, and seven days following the reading session). An additional 14 English only items were included only in the testing sessions to ensure if any learning took place outside the exposure context. These are referred to as 'control items'. The target items are presented in Appendix 4C.

Items occurred in the middle of a sentence within short paragraphs, with the same binomial item presented either two times or five times in the paragraph. Paragraphs were either two sentences or five sentences long. Only one binomial occurred in a paragraph (e.g., only *fish and chips*). The number of repetitions for each binomial item was counterbalanced in two lists (e.g., if an item appeared twice in a paragraph in list 1, the same item would appear five times in list 2; an example is provided below). The order of presentation of the paragraphs was randomized. A complete list of the stimuli is presented in Appendix 4D.

Here is an example of the paragraphs for the item *fish and chips* when it occurred twice and five times:

- Winston Churchill once described fish and chips as 'the good companions'. England's popular dish of fish and chips first appeared in UK shops in 1860 and was sold by a Jewish immigrant.
- Winston Churchill once described fish and chips as 'the good companions'. England's popular dish of fish and chips first appeared in UK shops in 1860 and was sold by a Jewish immigrant. By the 19th century, the presence of fish and chips shops had increased significantly to meet the demands of a growing industrial population. It started as a popular take-away meal among the British working class, and then later fish and chips became popular among all social classes. Nowadays, the popularity of fish and chips has reached many other English-speaking countries around the world.

It was important that the texts were easy to understand by the participants. Therefore, the frequency profile of the texts was checked using the VocabProfile tool (Cobb, n.d). With the exclusion of proper nouns, 80.04% of the words in the texts came from the first thousand frequency band, 6.68% were from the second thousand frequency band, and 6.35% came from the Academic Word List. Therefore, given participants' proficiency level, it was

expected that that English words used are representative of the English words they would know.

All binomial items' frequencies were transformed using the Zipf scale (as adopted by van Heuven, Mandera, Keuleers, & Brysbaert, 2014). Table 4.2 summarises the Zipf transformed frequencies of all item categories. A set of *t-tests* based on BNC confirmed no significant differences between all the English based items: congruent, control and English-only items, all *p's* > .05. However, the Arabic-only items were significantly lower than all of the English based items, *p's* < .05. A set of *t-tests* based on the Arabic corpus confirmed no significant differences between Arabic-only and congruent items, $p > .05$, and between English-only and control items, $p > .05$. However, both Arabic-only and congruent items' frequencies were significantly higher than English-only and control items (*p's* < .05). In other words, the frequency analysis confirmed the item categories, in that English-only and congruent items were matched in their BNC frequencies, and Arabic-only and congruent items were matched in their Arabic corpus frequencies.

Table 4.2 Frequencies of Items from BNC Corpus and KACKAST Arabic Corpus

Item category	BNC frequencies	KACKAST corpus
Congruent	3.36	3.08
English-only	3.19	1.88
Control	3.11	1.48
Arabic-only	1.95	2.85

Since it has been claimed that frequency biases binomial ordering preference, such that the more frequent word often precede the less frequent word (Benor & Levy, 2006), I made sure that frequencies of the first word making up the binomial phrase was not more frequent than the second word. A set of *t-tests* confirmed no differences in frequencies between the nouns making up the binomial phrases across all item categories (all *p's* > .05).

Norming Studies. Familiarity ratings of the target items were collected from English NSs (n = 34) and Arabic NNSs (n = 82) in three norming studies. Six different groups of NSs and NNSs, who did not participate in the actual experiment, were recruited to complete the norming studies (no subject participated in more than one norming study). Table 4.3 summarises the percentage ratios of the correct responses in the norming studies.

Table 4.3 Average Correct Responses in the Norming Studies across the Item Types for NSs and NNSs

Item category	Judgement task		Forward completion		Backward completion	
	NS (n=13)	NNS (n=40)	NS (n=10)	NNS (n=32)	NS (n=10)	NNS (n=11)
Congruent	90.47%	68. 93%	94.16%	44.2%	73.57%	27.14%
English-only	100%	51.07%	61.69%	13.84%	52.85%	5.71%
Arabic-only	66.68%	60.90%	48.05%	21.88%	35.00%	7.14%

The first norming study was a forced-choice phrase judgment task, where participants were asked to judge which order of the binomial phrase sounded more natural/familiar to them (e.g., is it *gold and silver* or *silver and gold*). For the NSs, a set of *t-tests* showed that there was a difference in accuracy between congruent and Arabic-only ($p < .05$), and between English-only and Arabic-only ($p < .05$). However, no difference was found between English-only and congruent items ($p > .05$). For the NNSs, a set of *t-tests* indicated that there was a difference between congruent and Arabic-only, between English-only and Arabic-only, and between English-only and congruent ($p < .05$).

The second norming study was a forward completion task, where participants were given the first noun and the conjunction *and* from a binomial phrase and were asked to provide the first word that comes to mind (e.g., *gold and* _____). The third norming study

was a backward completion task, where participants were given the second component of a binomial phrase and provided the word they thought should go in the gap (e.g., ____ *and silver*). Several independent sample *t*-tests were carried out to compare differences in completion scores between forward and backward tests in both NS and NNS groups. For the NNSs, there was a significant difference between forward and backward Arabic phrases, $p < .05$. There was also a significant difference between forward and backward congruent phrases, $p < .05$. Additionally, the difference between forward and backward English phrases was significant, $p < .05$. For the NSs, on the other hand, there was no significant differences between forward and backward Arabic binomial phrases, $p > .05$. There was no significant difference between forward and backward English-only phrases, $p > .05$. There was, however, a significant difference between forward and backward congruent phrases, $p < .05$.

Norming data were considered as variables in the analyses.

4.2.3 *Procedure and Apparatus*

The experiment was carried out in accordance with the research ethics procedures at the University of Nottingham, which provided ethical approval. The experiment took place in a quiet room in Imam University, Saudi Arabia for Arabic NSs who were studying English as a foreign language and at the University of Nottingham, for English NSs. The experiment was carried out in three sessions. In the first session (pre-testing session), participants signed the consent form upon arrival. Then, they took part in a forced-choice judgment task in which they were instructed to choose which form of a binomial phrase sounded more natural (e.g., *king and queen* vs. *queen and king*). They were encouraged to respond as quickly and accurately as possible. Participants' response times and accuracy scores were recorded using Inquisit 5 from Millisecond Software (2015). At the end of this task, participants were asked to complete the two proficiency tests (LexTALE and the multiple-choice vocabulary test),

and the language background questionnaire. The pre-testing session took around 15 minutes. Participants took part in the second session 10 to 14 days after the pre-testing session.

In the second session (i.e., treatment + immediate post-test), participants were exposed to the target items embedded within different paragraphs. Participants engaged in a non-accumulating, region by region self-paced reading task using Inquisit 5. Self-paced reading allowed participants' reading time to be measured and non-accumulating text ensured that they did not revisit previously read items, which could have confounded the repetition manipulation. Each trial began by displaying rows of dashes and blank spaces across a monitor's screen. The dashes corresponded to all of the non-white-space characters in a paragraph. The binomials were always placed towards the centre of a line in one region. To reveal the first phrase-region, participants pressed the spacebar, causing the dashes corresponding to this region to be replaced by text. Each subsequent press of the spacebar caused the just read region to revert to dashes, while simultaneously revealing text of the next region. Reading times were collected for each phrase-region. Once participants reached the end of the paragraph, a simple yes-no question appeared on the screen to ensure that participants read for comprehension. Analysis of the questions revealed good comprehension of the text (NSs = 93.0% accuracy; NNSs = 79.5% accuracy). No further analysis of these questions was carried out. Once the reading task was completed, an unannounced forced-choice judgment task was carried out (i.e., the immediate post-test). The post-test was identical to that in the pre-testing session. The whole session (i.e., reading treatment + immediate post-test) took around 40 minutes.

Once participants were finished, they were invited to participate in the delayed post-test session, which occurred exactly seven days after the reading + immediate post-test session. In the delayed task, participants again did the same forced-choice judgment task.

4.3 Results

The data were analysed using mixed-effects modelling in *R* version 3.6.1 (R Core Team, 2019). For the accuracy in the forced choice task (a binary variable, correct = 1 or incorrect = 0), logistic mixed-effects models were fitted (Jaeger, 2008), while for response time data, linear mixed-effects models were fitted using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). The *p*-values were estimated using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2015), and interactions were inspected using the *emmeans* package (Lenth, 2019) and plotted with the *effects* package (Fox, 2009). *P*-values for all pairwise comparisons were adjusted using the Bonferroni correction.

Response time data (RT), from both forced-choice task and self-paced reading task, were log-transformed before the analysis to reduce skewness. All of the other continuous variables (e.g., proficiency scores and variables from norming data) were also log-transformed to ensure that variables were on the same scale. Data were trimmed by deleting data points that fell above or below 2.5 standard deviations for each condition (session, item type, repetitions, etc.) in each language group (NS or NNS) separately. This led to a loss of 1.91% of the RT data for the forced-choice task and 3.08% of RT for the self-paced reading task.[6] A few data points from the self-paced reading were removed due to technical issues related to the recording of RTs for these trials (0.25% of the data).

The best models (reported in this study) were chosen based on likelihood ratio tests and AIC scores for model comparisons. That is, models' predictors and interactions were only included if they significantly improved the fit of the final models. All reported models had low collinearity (i.e., all VIF values were < 2.0). Models always included random intercepts for subjects and for items. Random effect structures were not kept maximal

[6] Data trimming did not result in losing data unevenly from the conditions and groups.

because maximal random effect structure may usually result in uninterpretable models that are difficult to estimate (Bates et al., 2015; Matuschek et al., 2017).

Models included the item type (English-only, Arabic-only, congruent), participant group (NS or NNS), session (time of testing: pre-test, immediate post-test, and delayed test), repetitions (two or five),[7] The following variables were also added as covariates whenever they improved the model fit: proficiency scores (LexTALE scores and the short vocabulary test scores), length of the binomial phrase, frequency (individual word frequency and binomial phrase frequency), binomial forward/backward ratio (i.e., forward binomial frequency divided by the backward binomial frequency), and pre-test scores (i.e., accuracy scores in pre-test to account for the effect of previous knowledge of the items on the reading times in self-paced reading task). Familiarity ratings (i.e., norming data from phrase judgement tasks by NS and NNS) was also considered as covariates. The norming data from forward and backward completion tasks were not included as covariates because 1) their additions did not make an improvement to the models, and 2) they were highly correlated with each other (all r's > 0.57 and all p's $< .05$).

Proficiency is an important variable to consider. There were three measures of proficiency: participants' self-ratings; their scores on the multiple-choice vocabulary test; and their LexTALE scores. Participants' self-ratings were strongly correlated with the multiple-choice vocabulary scores ($r = 0.70$, $p < .05$) and with the LexTALE scores ($r = 0.53$, $p < .05$), which were thought to provide more objective measures. Both proficiency tests (the abbreviated vocabulary test and the LexTALE) were also strongly correlated with each other ($r = 0.76$, $p < .05$). In addition, likelihood ratio tests confirmed that the multiple-choice vocabulary test made a greater contribution to the models than the LexTALE, thus it was

[7] Since the repetition variable only took effects after treatment, it was always included in post-test and delayed test data.

considered in analyses. Since most proficiency scores for NS reached ceiling level, the proficiency analysis was limited to the NNS to avoid the possibility that proficiency effect might be partialled out by the language group effect.

4.3.1 *Analysis of Accuracy Data*

Table 4.4 presents the summary of NS's and NNS's accuracy in the different testing sessions (pre-test, immediate post-test test, delayed test) for the binomial types and the control items that only appeared in the three tests.

Table 4.4 Means and Standard deviations of Accuracy scores

Group	**Session**	**Item Type**							
		Arabic		Congruent		English		Control	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
NNS	Pre-test	8.11	2.10	10.80	1.80	8.74	2.06	9.26	1.68
NNS	Post-test	10.40	1.74	11.00	1.98	9.81	2.03	8.94	1.71
NNS	Delayed test	9.49	2.10	11.10	1.68	9.62	1.99	8.85	1.75
NS	Pre-test	8.11	1.29	13.20	1.12	13.30	1.05	11.70	1.25
NS	Post-test	10.60	1.17	13.20	0.91	13.60	0.67	11.50	1.54
NS	Delayed test	9.61	1.83	12.90	1.13	13.40	0.88	11.40	1.01

Note. The maximum score was 14.

Table 4.5 Model Outcome for Accuracy Scores

	Accuracy for experimental items			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	-3.44	2.17	-1.58	0.113
Group [NS]	0.02	0.14	0.15	0.878
Item Type [Congruent]	0.76	0.25	3.09	**0.002**
Item Type [English]	0.09	0.27	0.35	0.725
Session [Post-test]	0.85	0.11	7.98	**<0.001**
Session [Delayed test]	0.49	0.10	4.71	**<0.001**

Familiarity ratings - NNS	0.62	0.54	1.14	0.253
Familiarity ratings - NS	0.32	0.13	2.52	**0.012**
Group [NS] * Item Type [Congruent]	1.43	0.19	7.65	**<0.001**
Group [NS] * Item Type [English]	2.58	0.22	11.87	**<0.001**
Item Type [Congruent] * Session [Post-test]	-0.80	0.17	-4.76	**<0.001**
Item Type [English] * Session [Post-test]	-0.46	0.16	-2.89	**0.004**
Item Type [Congruent] * Session [Delayed test]	-0.41	0.16	-2.49	**0.013**
Item Type [English] * Session [Delayed test]	-0.18	0.16	-1.15	0.249
Random Effects				
σ^2	3.29			
τ_{00}	$0.14_{subject}$			
	0.27_{Item}			
ICC	0.11			

The mixed-effects logistic models for accuracy, summarised in Table 4.5, confirmed a significant improvement in both the immediate and delayed post-tests for the experimental items. The documented improvement after treatment that lasted in the delayed test was not attested for the control items (for the analysis of the control items see Appendix 4E), suggesting that learning did not occur because of the repeated testing. Therefore, analysis of the control items was not considered further. While the familiarity ratings provided by the NS group predicted accuracy scores, the familiarity ratings provided by the NNS was not a significant factor. As also reported in Table 4.5, group interacted with binomial type.

Pairwise test revealed while the NS group outperformed the NNS group in congruent (odds ratio = 0.23, $p < .05$) and English-only items (odds ratio = 0.07, $p < .05$), but no difference emerged between the groups in the Arabic-only items (odds ratio = 0.97, $p > .05$). Analysis of the interaction showed that while the NNS group exhibited no differences in accuracy across the binomial types, the NS group was more accurate in congruent and English-only items than they were in the Arabic-only items. There was also a significant interaction between binomial types and session. Pairwise tests of this interaction showed that congruent items were the only item type that did not exhibit any significant difference in accuracy across the three testing sessions (all *p's* > .05). However, there was a significant increase in the Arabic-only items and English-only items between the pre-test and immediate post-test sessions and between the pre-test and the delayed test (*p's* < .05*).* Arabic-only items had significant decrease in accuracy in the delayed test relative to the immediate post-test ($p < .05$); there was no significant differences in the English-only scores between the immediate post-test and delayed tests ($p > .05$). There was no interaction between group and session ($\chi^2 = 0.00$, $p > .05$), suggesting that both groups experienced the same improvement over time. However, when inspecting each group separately, post hoc analyses demonstrated an improvement for both groups at immediate post-test for Arabic-only items. Only the NNS group showed a statistically significant improvement in the English-only items, which is unsurprising since the NSs were already at ceiling in the pre-test for the English-only and congruent items.

In another model, looking at NNSs, proficiency was considered as a factor. The output from this model indicated that higher proficiency resulted in greater accuracy ($\beta = 0.46$, SE= 0.15, z = 3.17, $p < .05$). However, there was no interaction with testing session (immediate, delayed) ($\chi^2 = 0.25$, $p > 0.05$) nor with binomial type ($\chi^2 = 0.68$, $p > .05$), indicating that there was an overall increase in accuracy with increased proficiency scores regardless of the time of the test, or the binomial type.

In order to examine the effect of repetitions (two repetitions vs. five repetitions), I fitted a model for repetition for the immediate and delayed post-tests, where repetition was a factor. Summary of descriptive statistics and model output are presented in Appendix 4F. Analysis of repetition indicated that the effect of number of repetitions on accuracy scores was not significant (β = 0.08, SE= 0.07, z =1.13, p > .05). There was no interaction with testing sessions (immediate, delayed) nor with item type.

4.3.2 *Analysis of RT Data*

Forced-choice Task. Table 4.6 presents the descriptive statistics of NS's and NNS's response times in the different testing sessions (pre-test, immediate post-test test, delayed test) for the binomial types and the control items that only appeared in the three tests. A linear mixed-effects model was fitted in order to examine the effect of test session on the response times for both experimental and control items. The model output is presented in Appendix 4G. The output from this model indicated that there was a decrease in RTs over time (i.e., pre-test vs. immediate post-test; immediate post-test vs. delayed test). There was a significant interaction between test session and binomial types, indicating that the decrease in RTs over time was larger for experimental items than for control items.

Two additional linear mixed-effects models were fitted in order to further explore the effect of reading treatment on the response time for the experimental items. As can be seen in Tables 4.6 and 4.7, NSs required significantly less time to respond to binomial phrases in the post-tests than NNSs. There was a significant effect of test session, with both the immediate and delayed post-tests eliciting significantly less response time than pre-test (pairwise comparisons indicated that delayed test RTs were also significantly shorter than immediate post-test RTs). The significant interaction between group and testing session showed that while session had an effect on reducing groups' RTs, the magnitude of this effect was different across groups. In particular, the difference in RTs between pre-test and post-test was

larger for NNS than for NS. The significant interaction between group and item type indicated that the effect of item type on RT was different for the two groups. For the NNS group, pairwise comparisons showed that while congruent items elicited shorter response times than either English-only or Arabic-only items ($p < .05$), both English-only and Arabic-only had equivalent response times ($p > .05$). For the NS group, there was no difference in response time for the congruent and English-only items. However, Arabic-only phrases elicited longer response times than both congruent and English-only items for the NS group. Length was a significant predictor, with shorter phrases being processed faster.

Table 4.6 Means and Standard Deviations of RTs in milliseconds (ms) from the Forced-Choice Task

Group	**Session**	**Item Type**							
		Arabic		Congruent		English		Control	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
NNS	Pre-test	3787	1569	3108	1459	3640	1470	3758	1582
NNS	Post-test	2565	1048	2207	949	2627	1037	3064	1273
NNS	Delayed test	2541	1022	2076	972	2536	1097	2612	1139
NS	Pre-test	1889	813	1431	574	1375	487	1741	832
NS	Post-test	1564	750	1193	403	1124	317	1387	503
NS	Delayed test	1382	551	1117	356	1075	302	1267	474

Table 4.7 Model Outcome for RT from the Forced-Choice Task

	RTs for experimental items			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	7.19	0.17	41.71	**<0.001**
Group [NS]	-0.71	0.06	-11.64	**<0.001**
Session [Post-test]	-0.33	0.01	-25.75	**<0.001**

Session [Delayed test]	-0.39	0.01	-30.06	<**0.001**
Item Type [Congruent]	-0.17	0.03	-5.11	<**0.001**
Item Type [English]	-0.00	0.03	-0.10	0.922
Length [log]	0.37	0.06	5.82	<**0.001**
Group [NS] * Session [Post-test]	0.15	0.02	6.55	<**0.001**
Group [NS] * Session [Delayed test]	0.15	0.02	6.64	<**0.001**
Group [NS] * Item Type [Congruent]	-0.08	0.02	-3.42	**0.001**
Group [NS] * Item Type [English]	-0.28	0.02	-11.96	<**0.001**
σ^2	0.11			
τ_{00}	$0.04_{subject}$			
	0.01_{Item}			
ICC	0.30			

In order examine the effect of proficiency level on response times for the NNS group, proficiency was included in a model of only the NNS data. Proficiency scores significantly predicted response times, with higher proficiency scores eliciting faster response times **(**β =-0.18, *SE* = 0.07, *t* = -2.46, $p < .05$*).* There was a significant interaction between proficiency and testing session (χ^2= 20.24, $p < .05$*).* Analysis of the interaction showed that higher proficiency levels elicited shorter response times in the delayed test, however proficiency did not have an effect on the immediate post-test.

The influence of repetition (two vs. five) on response times was considered in a model of the immediate and delayed post-test data (summary of the model outcome is presented in Appendix 4H). There was no effect of repetition (β = -0.01, *t* = -1.34, $p > .05$) nor any interactions with repetition.

Self-paced Task. As can be seen in Tables 4.8 and 4.9, NSs' reading times were faster than NNSs'. There was a significant speed up with greater repetitions. Trial number was significant with all binomial types eliciting faster reading times as the experiment progressed. Further, shorter words were read faster. Pre-test scores was not significant, suggesting that previous knowledge of the binomials did not affect reading times.

Table 4.8 Means of Reading Times from the Self-paced Reading Task across Repetitions (Rep)

Group	Item Type	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5
NNS	Arabic-only	1687	1240	1168	1130	1002
NNS	Congruent	1407	1167	1079	1019	946
NNS	English-Only	1658	1293	1092	1127	1055
NS	Arabic-only	618	552	565	504	490
NS	Congruent	532	498	515	463	458
NS	English-only	602	500	491	486	487

The main variables were involved in interactions. There was a significant interaction between group and item type. Pairwise comparisons indicated that NSs read all binomial types at the same speed, but NNSs read congruent items significantly faster than English-only items. However, for NNSs, no difference was found between Arabic-only items and congruent items or between Arabic-only and English-only. There was a significant interaction between number of repetitions and group. Analysis of the interaction showed that while number of repetitions elicited faster RTs, this effect was modulated by group. That is, decrease in RTs as a function of repetitions affected both groups differently. In particular, the difference in RTs between each number of repetitions was larger and more noticeable for NNSs than for NSs. This was evident when conducting pairwise comparisons between groups. For example, for NNSs, each additional exposure to an item resulted in a significant acceleration in RTs (except when comparing the third and fourth exposure's RTs). However,

for NSs, it was only the second exposure that resulted in a significant RT acceleration (when compared to the first exposure). Lastly, there was a significant interaction between number of repetitions and item type. As can be seen in Table 4.10, the pairwise comparisons suggest that Arabic-only phrases had a steeper decrease in RTs (with increased repetitions) than the other binomial types (i.e., for Arabic-only phrases, almost each repetition resulted in a significant decrease in RTs).

Table 4.9 Model Outcome for RTs from Self-paced Reading Task

		RTs		
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	6.43	0.16	39.50	**<0.001**
Group [NS]	-0.99	0.06	-15.31	**<0.001**
Item Type [Congruent]	-0.12	0.03	-3.65	**<0.001**
Item Type [English]	0.02	0.03	0.69	0.492
Reps [2]	-0.26	0.02	-16.07	**<0.001**
Reps [3]	-0.33	0.02	-16.35	**<0.001**
Reps [4]	-0.37	0.02	-18.72	**<0.001**
Reps [5]	-0.47	0.02	-23.72	**<0.001**
Length [log]	0.51	0.06	8.61	**<0.001**
Trial number [log]	-0.07	0.00	-19.99	**<0.001**
Pre-test scores	-0.01	0.01	-1.60	0.109
Group [NS] * Item Type [Congruent]	0.03	0.02	1.39	0.165
Group [NS] * Item Type [English]	-0.06	0.02	-3.01	**0.003**
Group [NS] * Reps [2]	0.15	0.02	7.85	**<0.001**
Group [NS] * Reps [3]	0.23	0.02	9.73	**<0.001**
Group [NS] * Reps [4]	0.21	0.02	8.73	**<0.001**

Group [NS] * Reps [5]	0.27	0.02	11.13	**<0.001**
Item Type [Congruent] * Reps [2]	0.07	0.02	3.02	**0.003**
Item Type [English] * Reps [2]	0.01	0.02	0.37	0.714
Item Type [Congruent] * Reps [3]	0.07	0.03	2.46	**0.014**
Item Type [English] * Reps [3]	-0.06	0.03	-2.38	**0.017**
Item Type Item Type [Congruent] * Reps [4]	0.04	0.03	1.56	0.119
Item Type [English] * Reps [4]	0.00	0.03	0.15	0.877
Item Type [Congruent] * Reps [5]	0.08	0.03	3.09	**0.002**
Item Type [English] * Reps [5]	0.06	0.03	2.35	**0.019**

Random Effects

σ^2	0.10
$\tau_{00\ subject}$	0.05
$\tau_{00\ Item}$	0.01
ICC	0.35

Table 4.10 Summary of the Pairwise Comparison Results of the Interaction between Repetition and Item Type, where (+) Indicates that the Decrease in RTs with more Repetitions was Significant, and (–) Indicates that it was not Significant.

Item Type	Rep 1 vs. Rep 2	Rep 2 vs. Rep 3	Rep 3 vs. Rep 4	Rep 4 vs. Rep 5
Arabic-only	+	–	+	+
Congruent	+	–	+	–
English-only	+	+	–	–

4.3.3 *Summary of the Findings*

The main findings from the study can be summarised as follows. Accuracy improved for the items that occurred in the passages but not for the items that only occurred in the tests, highlighting the role of the reading treatment. The improvement in accuracy was still evident seven days after exposure in the delayed test. There was no interaction between group and session in accuracy results, suggesting that treatment exerted the same effect on both groups' accuracy scores. Judgment times reflected the same pattern as the accuracy; responses were faster in the immediate and delayed post-tests than in the pre-test for both groups. Further, responses were faster in the delayed than in the immediate post-test for both groups. Such a decrease in RTs over time was also detected for control stimuli. However, the interaction between test session and binomial types indicated that decrease in RTs over time was more pronounced for experimental items than for control items. This suggests that the reading treatment indeed affected the processing speed, and this effect was beyond the effect of repetition (i.e., practice effects).

Overall, NSs were more accurate and faster (in both self-paced reading and the forced-choice task) than NNSs. However, they were not more accurate on the Arabic-only items, where both NSs and NNSs had similar levels of accuracy. The NS group was more accurate in congruent and English-only items than they were in the Arabic-only items, unlike the NNS group who demonstrated no differences in accuracy scores across the item types. Interestingly, the interaction between binomial type and testing session indicated that improvement in accuracy was larger for Arabic-only phrases than all other item types, with congruent items showing no improvement across sessions for both groups. Subsequent analysis for groups separately showed that whereas both language groups recognized Arabic-only items better after treatment, it was only the NNS group who also recognized English-only items significantly better after treatment.

The interaction effect between groups and binomial types for judgment time pointed to somewhat similar directions. NSs responded to English-only and congruent items with the same speed; however, they required longer time to respond to Arabic-only items. NNSs judged congruent items faster than Arabic-only and English-only times, and no difference was found between Arabic-only and English-only judgment times. Similarly, RTs from self-paced reading data showed that while NSs read all item types with the same overall speed, NNSs read congruent items faster than English-only items (no difference was found in NNSs' reading time between Arabic-only items and English-only, or between Arabic-only items and congruent items).

Findings from the self-paced reading task indicated that the Arabic-only items decreased more steeply (across each additional exposure) than other binomial types. This pattern of results can be attributed to the lower phrasal frequency of the Arabic-only binomials. Since there are cognitive constraints conditioning how fast processing can be, as a function of frequency, larger frequency effects would be observed for lower frequency words or sequences of words (Conklin, 2019). Similarly, NNSs demonstrated a steeper decrease in RTs (over time of testing in the judgement task and over number of repetitions in the self-paced reading task) than NSs, suggesting clearer processing gains for the NNSs as a result of repeated exposures. Nevertheless, number of repetitions during treatment did not have an influence on learning gains. That is, it was not a significant factor in determining either processing time or correct responses in forced-choice task results.

4.4 Discussion

The goal of this study was to assess whether the word order preference characterizing binomial phrases (e.g., it is *fish and chips* and not *chips and fish*) could be learned during reading. To explore this question, participants read passages in which the target binomial phrases were embedded. Each target item was repeated either two or five times. A pre-test on

the items was administered 10 to 14 days before the main task. After the reading treatment, it was administered again (immediate post-test), and participants returned seven days later to take the test a final time (delayed test). The tests contained both experimental stimuli (included in the training session) and control stimuli (included only in the testing). The tests assessed participants' recognition of the 'correct' order of binomial phrases (*salt and pepper* vs. *pepper and salt*), recording their reaction times and accuracy.

4.4.1 *The Learning of Binomials by L2 Learners*

Accuracy and RTs for the forced-choice task indicated that binomial phrases can be learned by both NSs and NNSs when they were exposed to them in a reading task. The fact that accuracy was similar seven days after the reading treatment demonstrates that learning gains observed for both NS and NNS, while small, were also (somewhat) durable. The Arabic-only phrases were learned by both the NNSs and NSs, while the English-only phrases were only learned by the NNSs. The results demonstrate that NNSs retained some memory of binomial ordering preference; they had better performance in the immediate post-test than the pre-test for English-only and Arabic-only items (but not congruent items) from just a few input exposures. It could be that participants were relatively familiar with the congruent items in the pre-test, which might explain the small improvement observed for the congruent items. The results suggest that acquiring an intuition of the word ordering preference in binomials may not necessarily require an intentional strategy. This finding is consistent with the Sonbul and Schmitt (2013) study, although on another type of formulaic language, showing that both NSs and NNSs learnt collocations incidentally after three exposures.

The findings fail to support Wray's (2002) view that L2 learners attend to language by analysing its individual word components (a word-based approach). In other words, because L2 learners were able to retain some information about which binomial word order is more frequent without being explicitly asked to attend to the binomials, it can be concluded that L2

learners are able to notice formulaic sequences in natural input. However, such a conclusion must be interpreted with caution since only a specific aspect of form recognition knowledge has been measured.

The learning gains observed in this study cannot be characterized as intentional because the task instructions did not instruct participants to learn the binomial phrases, and participants were not forewarned that they would be tested on the target items; it cannot be characterized as *purely* incidental, either. Taking part in the pre-test might have clued participants on the study being about binomials. Thus, although participants engaged in a meaning-focused activity that directed their attention to understanding the text, they could have focused on the binomial phrases. A learning condition in which the primary goal is to understand the text while some features of the text (e.g., bolding) is manipulated such that it would attract learners' attention to particular formulaic sequences was described by Pellicer-Sánchez and Boers (2019) as semi-incidental. The reading task in this study can be characterised as semi-incidental because learners engaged in a meaning-focused activity in which the target items were made salient (i.e., repetitions and prior exposure in the pre-test) to attract more attention.

The current study supports evidence from previous research (e.g., Durrant & Schmitt, 2010; Sonbul & Schmitt, 2013; Webb, Newton, & Chang, 2013; Vilkaitė, 2017), showing that formulaic sequences can be learned from reading. The finding that reading-only exposures yielded learning gains for technical collocations (Sonbul & Schmitt, 2013), pseudoword collocations (Pellicer-Sánchez, 2017), and frequent collocations (Vilkaitė, 2017) was also noted for binomials in the present study. The present study expands previous findings on the learning of formulaic language, suggesting that learning gains for binomials can last seven days. It is important to note that comparing the current study to previous ones is not straightforward. For example, previous studies used other types of formulaic sequences

(i.e., mostly collocations vs. binomials in the current study). Although transparent binomial phrases might be similar to transparent collocations, many of the collocations used in some studies were less transparent than binomials. Furthermore, the studies employed different methodologies, such as using a between subject design or a control group instead of testing participants' prior knowledge (e.g., Durrant & Schmitt, 2010; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016; Webb, Newton, & Chang, 2013) or using a different treatment (e.g., a reading-while-listening treatment in Webb et al., 2013).

Overall, the ability to recognize the binomial order preference after having encountered binomial phrases (only twice) in one particular configuration (compared to control items) supports usage-based theories (e.g., Bybee, 1998; Goldberg, 2006; Tomasello, 2003), which highlight the role of language input in learning. These theories suggest that only minimum exposure to a language pattern is needed to leave a memory trace. This is supported by Goldberg (2007), who states that, because increased exposure to a language pattern influences its memory representation, it can be concluded that even one exposure may leave a memory trace. That memory trace will be reinforced by each exposure (Sonbul & Schmitt, 2013). Gurevich, Johnson, and Goldberg (2010) demonstrated that participants could recognize sentences even after a single exposure; such recognition had a durable effect (i.e., a six-day delay).

Effects of proficiency observed for the NNS group, such that increased proficiency results in overall better and faster recognition performance, highlights the role of language use. These patterns of results support the view that with increased proficiency, L2 learners experience more phrasal frequency effects (e.g., Fernández & Schmitt, 2015; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Wolter & Gyllstad, 2013). This view suggests a progression towards a more native-like processing for formulaic language with increased exposure. In this study, proficiency had a stronger influence on the delayed test than the

immediate post-test for RTs. This observation indicates that immediate effect of exposure helped mitigate the gap in performance between lower and higher proficiency groups in the post-test. The finding that increased proficiency may not necessarily lead to better performance after treatment suggests that L2 learners from different proficiency levels may benefit equally from the reading treatment.

4.4.2 *Repetition Effects*

Although the results confirm that both NSs and NNSs have durable learning gains after only two exposures, the study did not find an effect of repetition (i.e., two vs. five repetitions). In other words, increasing the number of occurrences from two to five did not translate into a better recognition, either in terms of accuracy or reaction times for either group. This finding is in line with Pellicer-Sánchez (2017) who found no effect of frequency of exposure (i.e., four to eight) on the acquisition of collocational knowledge (in both form recall and recognition post-tests). However, the results conflict with those of Webb et al. (2013) indicating that a greater number of exposures results in more learning gains (i.e., 15 exposures resulted in the highest gains, 10 exposures resulted in better performance than 5 or 1 exposures, and 5 exposures had better gains than 1 or no exposure). This was mirrored in Durrant and Schmitt's study (2010), where learners who were exposed to collocations twice outperformed those who were exposed once. The results of this study contrast with those of Peters' (2014) who found a positive durable effect of frequency for learning when comparing gains from one single exposure to five.

The discrepancy in results could be attributed to different types of treatments (i.e., input) used. Unlike the current study which employed a silent reading treatment, Webb et al.'s study used reading-while-listening, Durrant and Schmitt's study asked participants to read aloud, and Peters' (2014) study used explicit learning of decontextualized activities. Lin (2019) argues that knowledge of formulaic sequences can be more easily enhanced in spoken

than written input because formulaic sequences tend to be articulated as one intonation unit. While formulaic sequences lack salience in a reading-only condition, the prosodic salience of the intonation patterns of formulaic sequences in a reading-while-listening condition (or in spoken input in general) may assist the learning of formulaic sequences. Further, explicit learning activities may show clearer effects of frequency. Pellicer-Sánchez (2017) speculated that a reading-only treatment might require a greater range of occurrences in order for significant frequency effects to emerge. That is, a limited range of repetitions, two vs. five in this study, might have been too small for frequency effects to emerge.

The lack of a repetition effect could also be due to the fact that all repetitions of an item were in a single paragraph. Indeed, after the initial exposures to the item, participants may develop a prior expectation of the item that may result in recruiting lower attention resources when engaging with the item, thus possibly masking any greater advantage due to repetitions. It might be that spacing of repetitions (occurrences distributed in texts) would be more likely to produce frequency effects. However, research on the effect of spacing of repetitions on incidental learning is still lacking (Webb, 2014).

A few more possibilities might account for the lack of a repetition effect in the study. It might be that recognition of binomial word order is more easily acquired, requiring very little exposure for durable form recognition; thus, two exposures is sufficient and there are no further gains for the additional three exposures. As indicated by Pellicer-Sánchez and Schmitt (2010), different aspects in word knowledge may require a different number of encounters. Notably, in their metanalysis on the effect of repetitions on incidental vocabulary learning, Uchihara, Webb, and Yanagisawa (2019) found that frequency of occurrence was more significantly correlated with learning gains when tested with recall tests than when tested with recognition tests. Thus, since recognition knowledge is less difficult to acquire than recall knowledge (e.g., Laufer & Goldstein, 2004; Webb et al. 2013; Sonbul & Schmitt,

2013), a greater number of occurrences may not be needed to increase recognition knowledge of an item, but it could continue to increase the recall knowledge of that item (Uchihara, Webb, & Yanagisawa, 2019). Thus, more demanding tasks (i.e., recall) may show clearer frequency effects than less demanding ones (i.e., form recognition). The findings of Szudarski and Carter (2016) support such a view. They found that an increase in encounters (from 6 to 12) improved the form recall of collocations, while it did not reflect any better gains at the level of form recognition. In the current study, the recognition task (e.g., is it *fish and chips* or *chips and fish*) is arguably less demanding than typical form recall formats, thus possibly masking any repetition effects.

Finally, it is important to acknowledge that the role of frequency of encounter on learning is very complex because of many methodological variables (e.g., learner individual variability and items' distinctive characteristics) influencing the relationship between repetition effects and vocabulary learning (Uchihara, Webb, & Yanagisawa, 2019), and this is an important reason behind the inconsistency in the literature. Further research is needed to disambiguate these possibilities and to advance our understanding of the factors that determine the impact of frequency of encounter on formulaic language learning. While debate about how many repetitions are sufficient for acquisition to occur has not yet been resolved (Peters, 2014), it is expected that the number of contacts with a word combination should be a main determining factor for learning.

4.4.3 *Congruency Effects*

The study showed that when a binomial phrase overlapped in the languages, it was responded to more quickly and accurately. For example, NNSs' RTs for congruent phrases were faster than Arabic-only (in judgment time) and English-only (in judgment time and reading time). Further, the reading treatment was more effective for NNS for recognising Arabic-only and English-only items than congruent items. This might be because NNSs'

performance on the congruent items was already quite good at the outset of the study (77.36%) compared to the English-only (62.40%) and Arabic-only (57.90%) items. Thus, the congruent items had relatively little scope for improvement in an incidental reading task with only two or five exposures. The exposures from reading led to more gains for the English-only (70.06%) and Arabic-only items (74.62%), such that they reached an accuracy level similar to that of the congruent items. A similar pattern was found in NS data (i.e., English-only and congruent items did not improve with the reading treatment; only Arabic-only items did).

Overall, the advantage in recognizing the specific order preference of the congruent items is compatible with other studies which found confirmatory evidence for the role of congruency on processing or learning. For example, Wolter and Gyllstad's (2011, 2013) studies, using lexical decision and phrase judgment tasks respectively, showed a processing advantage for congruent items (i.e., L1 Swedish L2 English) over incongruent L2-only items. L2 learners in Yamashita and Jiang's (2010) phrasal judgment task also judged L1-L2 collocations with greater accuracy than L2-only collocations.

Notably, there was a clear influence of L1 knowledge on NNSs; however, this influence was limited to binomial phrases that were the same in L1 and L2, and not ones that had been translated from the L1 (Arabic-only items). Arabic NSs generally did not show any difference in performance between the Arabic-only and the English-only items. Thus, no evidence in the current study supports that L1 knowledge influences the recognition of translated L1 binomial phrases that are not registered in the L2 lexicon.

The current findings contrast with those of Carrol and Conklin (2014a, 2017), who showed an L1 facilitative effect when Chinese NNSs of English read Chinese idioms translated into English (equivalent to the Arabic-only condition in the current study). Similarly, Carrol, Conklin and Gyllstad (2016) found that while English NSs read Swedish-

based idioms with much difficulty, Swedish participants did not experience that disruption when reading the Swedish-only idioms.

Thus, the current study is incompatible with the cross-language priming account speculated by Carrol, Conklin, and Gyllstad (2016) and Carrol and Conklin (2017). This account predicts that when an item in L2 input is activated (e.g., the Arabic-only phrase *hunger*), automatic priming takes effect that extends activation to its L1 translated form, which then triggers activation to its collocates (*hunger and thirst),* thus resulting in facilitation in both processing and accuracy. However, this prediction was not met in our study, such that advantage was only found when the phrase overlaps in both L1 and L2.

The cross-priming account did not receive empirical support in other studies, as well. For example, Wolter and Yamashita's (2015) lexical decision task results demonstrated no evidence for an advantage accorded to L1 translated (Japanese) collocations neither in error rates nor in reaction times. Wolter and Yamashita (2017), using a phrasal acceptability task, found evidence for congruency in that L2 learners responded faster to congruent collocations than English-only collocations. No processing advantage was found for L1-only translated items (Japanese) over control items. A lack of L1 influence led them to conclude that processing advantage found for congruent items may not be attributed to a joint L1-L2 activation or automatic priming. Further, Pritchett, Vaid and Tosun's (2016) study indicated that congruent phrases (i.e., figurative meaning in both English L2 and Russian L1) were better remembered than phrases that existed in only one language (L1 only or L2 only).

These results point to the lack of L1 influence for L1-only formulaic sequences, suggesting that cross-language priming account, while it may explain an L1 influence in idioms, does not explain the congruency effects in more compositional or transparent formulaic language like the binomials in this study.

Wolter and Yamashita (2017) explain congruency effects with the mapping hypothesis (Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006; Monaghan & Ellis, 2002), which can account for the role of congruency and L1 knowledge such that the L1 influences the L2 only when there is L1-L2 correspondence. Wolter and Yamashita's account assumes that both L1-L2 and L1-only collocations are equally transferrable and utilised by L2 learners, but only in their initial stages of learning. However, as L2 learners become more proficient, they are more likely to encounter L1-L2 items than L1-only items in L2 input. This provides more reinforcement for L1-L2 items than L1-only items and results in further entrenching them in memory. Under this account, due to lack of exposure and reinforcement, L1-only items gradually becomes "a less prominent part of the network" until they eventually no longer trigger any collocational associations in L2 lexical network, thus eliciting more cognitive effort in processing than congruent items (Wolter & Yamashita, 2017, p. 16). Wolter and Yamashita's view is compatible with Wolter and Gyllstad's (2013) finding that once an item is registered in L2 lexicon, it is L2 frequency but not L1 frequency, which explains any processing advantage, and that L2 collocational frequency effects operate beyond congruency (with more frequent incongruent collocations being judged faster than less frequent congruent collocations).

The current findings expand Wolter and Yamashita's results (2017) to binomials. In relating their explanation to the current findings, L2 learners were more likely to recognise word order preference of congruent items (faster and better) than incongruent ones (whether Arabic-only or English-only) mainly because of repeated exposure. Congruent items are repeatedly encountered (in two input sources: L1 and L2), thus becoming more entrenched in memory. Based on Wolter and Gyllstad's (2013) finding that highlights the role of L2 frequency, it can also be predicted that as NNSs become highly proficient, they could recognise L2-only items faster and better than L1-only ones.

4.4.4 *Limitations and Future Directions*

It is important to note that the current study simply starts to explore the interaction between frequency of exposure and congruency in binomials and there are some limitations to bear in mind when considering the results. For example, although using real binomial phrases provides more ecological validity, participants' prior knowledge needed to be assessed (i.e., the pre-test). Efforts were made to eliminate an influence of the pre-test on any learning by separating the two (i.e., they were 10-14 days apart). However, pre-testing could have alerted the participants to the purpose of the study and the items that were under investigation. The study only assessed form recognition; future studies could investigate both receptive and productive knowledge. In particular, it might be that the number of repetitions plays a different role in recognition and recall tasks, as had been suggested in the literature (Pellicer-Sánchez & Schmitt, 2010; Uchihara, Webb, & Yanagisawa, 2019). Also, an influence of repetition might be more apparent when occurrences are spaced rather than masses as they were in the current research. Future research should explore these possibilities. Finally, although the study was not designed to test the effect of intentional learning, future studies should examine how receptive and productive knowledge interact with different treatment interventions (i.e., incidental vs. intentional). Further, it is unclear how incidental versus intentional tasks might impact learning of an aspect of formulaic language knowledge like word order (i.e., it is *salt and pepper* not *pepper and salt*).

4.5 Conclusion

Chapter 4 presented a study examining a relatively unexplored topic in the literature: the effect of congruency and repetition for the learning and processing of binomial phrases. Results showed that reading exposure benefited recognition of the preferred binomial form. Congruency played an influential role in recognising word order preference of binomial phrases. The effect of congruency was evident when a binomial phrase existed in the two

languages. Knowledge of binomial word order was not enhanced by repetition. However, this study is limited in that it examined the learning of formulaic language under only one learning condition (i.e., reading). To better inform pedagogical practice, studies need to evaluate the effect of different learning conditions for learning formulaic sequences. This is the main objective of the next two chapters.

Chapter 5. The Learning and Processing of Transparent Collocations

While the previous chapter discusses the effect of frequency and congruency on learning binomials incidentally, this chapter explores the effectiveness of different learning conditions on learning collocations. It begins by reviewing investigations of the effectiveness of different learning approaches – intentional learning, incidental learning, and semi-incidental learning – on the learning of transparent collocations. It then introduces the second study of this thesis by presenting its methodology and reporting its findings. This chapter concludes by interpreting the findings of the study.

5.1 Introduction

As noted in Chapter 3, many investigations in vocabulary learning (both on words and multiword sequences) have looked at the effectiveness of intentional and incidental learning. Intentional learning occurs when there is a deliberate effort to learn language items (Nation, 2001). Incidental learning refers to, "the picking up of new linguistic features while attention is focused on understanding the input" (Ellis, 2015, p. 147) and can emerge from communicative interactions where attention is not focused toward learning language items (Schmitt, 2010). Crucially, operationalising a clear distinction between intentional and incidental learning is difficult (Pellicer-Sánchez & Boers, 2019; Pellicer-Sánchez, Conklin, & Vilkaitė-Lozdienė, 2021). It can be difficult to ascertain whether learners' attention is on learning language items when engaging in an activity that is focused on content (e.g., reading a story). However, Pellicer-Sánchez and Boers (2019) suggest that it is possible to create a situation that does not elicit learners' deliberate efforts to learn language items.

The current study aims to compare incidental and intentional learning. Following Nation and Webb (2011), learning is considered 'incidental' if learners are not explicitly asked to pay attention to language items, but rather their attention is directed towards

understanding the content. A learning condition is considered ‘intentional’ when learners’ attention is directed toward learning language items.

5.2 Research on the Learning of Collocations

5.2.1 *Intentional Learning*

Studies have examined the effectiveness of intentional learning for L2 collocations using various tasks. For example, Szudarski (2012) evaluated the benefit of explicit post-reading exercises on learning collocations that had occurred in short stories. He compared learning gains for learners who read a story (reading-only), and learners who read the same story and afterwards completed some exercises (e.g., gap-filling) involving the collocations (reading-plus). The results showed that the reading-plus group had greater learning gains than the reading-only group, while the reading-only group results were no different from a control group that had not read the story.

Laufer and Girsai (2008) examined the effectiveness of different tasks on learning unknown collocations (i.e., as determined in a pre-test). After exposure to collocations in a reading passage, learners received different treatments. A meaning-focused instruction group engaged in content-focused tasks (e.g., comprehension questions). A form-focused instruction group did tasks to consolidate their knowledge of form (e.g., multiple choice and gap filling exercises). A contrastive analysis and translation group did L2-L1and L1-L2 translation exercises. The findings showed that tasks involving intentional learning (i.e., contrastive analysis and translation) yielded the greatest learning gains, while incidental tasks (meaning-focused) led to the smallest gains. However, it is important to note that the testing format used in the post-test involved translation tasks which closely resembled the tasks performed by learners under the translation condition. This might have provided learners under the translation condition with an advantage compared to learners tested under the other

conditions. Thus, it is very hard to draw firm conclusions from this study about the effects of the learning conditions examined.

Webb and Kagimoto (2009) also explored the effectiveness of different tasks (i.e., receptive and productive) on learning collocations. They presented EFL learners with collocations that were embedded in three sentences and which were glossed with L1 translations. They asked one group to read the collocations and the other group to copy the collocations presented in the gloss into three gapped sentences. Both groups showed substantial learning gains in both productive and the receptive conditions, with no difference between the two learning conditions. When learners were grouped based on their proficiency level, differences between the learning conditions were found. The productive condition was more beneficial for higher proficiency learners and the receptive condition was more beneficial for lower proficiency learners.

Zhang (2017) presented EFL learners with unknown collocations (i.e., verified in a pre-test) in three different learning conditions: a receptive group (i.e., learners were asked to learn the collocations by reading them four times), a productive group (i.e., learners were asked to learn collocations by writing them in sentences four times), and a receptive/ productive group (i.e., a combination of both learning conditions). The results indicated that while all study conditions were superior to a control group that had not encountered the collocations, the combined study condition had the greatest learning gains.

Another line of research created an intentional learning condition by explicitly asking learners to identify formulaic sequences used in texts (e.g., 'text chunking' proposed by Lewis, 1997). Boers, Eyckmans, Kappel, Stengers, and Demecheleer (2006) adopted such an approach by engaging EFL learners in text chunking activities throughout the course, drawing learners' attention to the phraseological dimension of formulaic sequences. Text chunking activities involved asking learners to identify (e.g., highlight) formulaic language

used in texts and engaging learners in gap-filling exercises targeting formulaic language. Both, learners who engaged in text chunking (text chunking group) and a control group (whose attention was not directed to formulaic language) were exposed to the same in-class multi-mode input. At the end of the course, participants' knowledge of formulaic sequences was assessed in an interview by counting the number of formulaic sequences used in a conversation. The interview showed that the text chunking group used more formulaic sequences than the control group. However, such an advantage did not reflect the development of a larger overall formulaic language repertoire as a result of the text chunking activities provided throughout the course. This is because most of the formulaic sequences used by the text chunking group were not encountered from the texts explored in class; they were rather recycled from the prompt text which was given to students beforehand in preparation for the interview. It needs to be noted that Boers et al. (2006) had two limitations: 1) they did not conduct a pre-test to better compare learners' uptake of formulaic sequences relative to their baseline; 2) one interview at the end of the course may underestimate learners' uptake of formulaic sequences.

Peters (2009) addressed a limitation of the Boers et al. (2006) study by conducting a more controlled experiment. The Peters' study used a pre-test/post-test design (i.e., a translation task) that targeted the items used in a text. After taking the pre-test, Peters assigned EFL learners to one of two groups: one group was asked to devote their attention to unfamiliar vocabulary when reading a text, another to devote their attention to both unfamiliar vocabulary and collocations. Vocabulary and collocations were underlined and presented in marginal glosses with L1 translation for both groups. Both groups were forewarned that they would be tested on their knowledge of the target items afterwards. Post-test results showed that both groups' knowledge of the target items improved substantially. Notably, the two groups paid equal attention to collocations (i.e., there was no difference

between groups in the recall of collocations). Peters concluded that exposure to collocations in the pre-test (which occurred on the day of treatment), as well as the fact that collocations were presented in marginal glosses for both groups, may have drawn attention to collocations during the study.

These studies, using a range of explicit tasks, demonstrated learning gains with intentional learning. While the benefit of intentional learning is clear, a practical issue is that explicit tasks require time to carry out. However, classroom time is often limited, and there is a vast number of multiword sequences, which would make it impractical to explicitly teach them all in the classroom or even as 'homework'. Thus, an important question is whether it is possible to learn formulaic language, like collocations, incidentally from input. Beyond the benefit of learning new vocabulary, incidental learning activities (e.g., extensive reading) may also foster other skills such as reading fluency and reading comprehension (Nation & Webb, 2011).

5.2.2 *Incidental Learning*

As discussed in Chapter 3, a number of studies have shown that learners can acquire collocations incidentally (e.g., Durrant and Schmitt, 2010; Pellicer-Sánchez, 2017; Sonbul & Schmitt, 2013; Toomer & Elgort, 2019; Vilkaitė, 2017; Webb, Newton, & Chang, 2013). The number of occurrences has also been shown to play an important role in incidental learning. Durrant and Schmitt (2010) showed that two exposures to a collocation in a read-aloud task generated better performance in a recall task than a single exposure by ESL learners. Webb, Newton, and Chang (2013), who modified a short story by incorporating multiple instances of the same collocation into the text (i.e., from 1 to 15), found that more encounters resulted in better learning outcomes of collocations for their EFL learners. In sum, much of incidental learning research describes frequency of encounter as a powerful accelerator of incidental learning.

Thus, it appears that multiple occurrences are needed to learn collocations incidentally. Webb et al. (2013) suggest manipulating materials to allow for multiple occurrences of a collocation. However, Pellicer-Sánchez and Boers (2019) say that it is very unlikely that unmodified authentic materials would have frequent repetitions of a set of collocations and that it would be practically difficult for practitioners to manipulate authentic teaching materials in this way (e.g., Webb et al., 2013). Thus, L2 learners may not encounter sufficient instances of the same collocation in authentic materials for any gains to emerge incidentally. For example, Boers and Lindstromberg (2009) demonstrated that very few collocations occurred more than once in a 120-page novel. A further issue with authentic texts, raised by Boers and Lindstromberg (2012), is that distribution of formulaic sequences is likely to be sporadic. They point out that this may hamper learning in two ways: 1) any initial memory traces would be lost before they could be strengthened by subsequent exposure; and 2) learners may fail to identify a word combination as a recurring formula, making it difficult for them to develop knowledge of word partnerships and the syntagmatic behaviour of language. In other words, forming strong memory traces may be impeded by having few, sporadic encounters with a formulaic item.

5.2.3 *Semi-Incidental Learning: Attention-Drawing Techniques*

As we have just seen, the sporadic and few encounters with formulaic items in authentic materials pose a challenge to incidental learning. Further, it may be difficult for educators to add sufficient occurrences to materials. In addition, even with repeated exposures, formulaic language may not be salient to learners. Laufer and Girsai (2008) explain that because many L2 lexical items are characterized by low frequency and low salience, "they may go unnoticed unless attention is drawn to them" (p. 697). Further, L2 learners may fail to recognize formulaic sequences in written input and focus instead on single words. An additional problem is that learners may generally encounter formulaic

language in written materials; written input is more frequent than spoken input in EFL classroom materials, as well as in L2 learning in general (Lin, 2012, 2019). The word boundaries imposed by written input may hamper the identification of unknown formulaic sequences (Bishop, 2004). Prosodic patterns that signal a formulaic sequence as one intonation unit in spoken input (Lin, 2019) are also lacking in written input.

To overcome some of these challenges, researchers have explored different ways of making formulaic sequences more perceptually salient in written input. As previously mentioned, Choi (2017) suggests using attention-drawing techniques (e.g., input-enhancement like underlining, highlighting, or bolding). The purpose of attention-drawing techniques is to make formulaic sequences more noticeable for learners. The difficulty in recognizing whether a word combination is a recurring formula (as a result of formulaic sequences occurring infrequently in unmodified text) might be overcome by using input-enhancement (Boers & Lindstromberg, 2012). Further, Pellicer-Sánchez and Boers (2019) believe that input-enhancement has some practical advantage over input flooding (i.e., manipulating the input by creating multiple instances of the same item in an incidental learning context) as it is much easier to carry out. Pellicer-Sánchez and Boers (2019) classify such a learning condition as 'semi-incidental' because it encourages learners' uptake by directing their attention to the phraseological dimension of formulaic sequences in a text, while keeping the primary focus on the content of the text (i.e., reading is still simply being done for comprehension).

Attention-drawing techniques are thought to be beneficial for vocabulary and formulaic sequence learning because they draw learners' attention to the target items. In other words, vocabulary learning benefits from tasks that induce higher degrees of attention to and engagement with the target items (Schmitt, 2008). The role of attention in language learning was highlighted by Schmidt (2001) in his Noticing Hypothesis: "people learn about the things

they attend to and do not learn much about the things they do not attend to" (p. 30). In other words, attention to language features is conducive to successful language learning. Attention-drawing techniques presumably stimulate depth of initial processing and higher engagement with the target items by attracting more attention to them. Deeper processing involves "a greater degree of semantic or cognitive analysis", which leads to stronger memory traces (Craik and Lockhart, 1972, p. 675). This means that tasks eliciting higher engagement levels with the target items would most benefit learners.

Little research has been done on the effect of attention-drawing techniques on the learning of formulaic sequences. However, Boers et al. (2017) examined the effectiveness of input-enhancement techniques and found that formulaic sequences were better remembered when they were underlined than when they had no typographic enhancement. With a focus on collocations, Szudarski and Carter (2016) found that input enhancement (i.e., underlining) and input flooding (i.e., manipulating the number of exposures, 6 vs. 12 exposures) led to greater overall learning gains than input flooding alone. Importantly, Szudarski and Carter concluded that the number of exposures necessary for learning gains can be reduced when collocations are enhanced. Sonbul and Schmitt (2013) compared the outcomes of intentional, incidental, and semi-incidental learning of collocations in a within-subject design. The target collocations were presented in an enhanced input condition (i.e., semi-incidental: target collocations were bolded), an unenhanced input condition (i.e., incidental: target collocations were presented without typographical enhancement), and a decontextualized input condition (i.e., an intentional learning condition). For the NSs, the decontextualized condition led to greater learning gains than the unenhanced condition, with no difference in the learning between the enhanced and unenhanced conditions. For the NNSs, the enhanced condition led to greater learning gains than the unenhanced condition, with no difference between these two conditions and the decontextualized condition. Toomer and Elgort (2019) replicated

Sonbul and Schmitt's (2013) findings, confirming that an enhanced condition led to greater learning gains than an unenhanced condition, even with an increased number of encounters (nine instead of three as in Sonbul and Schmitt). Taken together, these studies suggest that repeated exposures in combination with attention-drawing techniques, benefit learning collocations for L2 learners (Toomer & Elgort, 2019; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016).

These studies used offline post-tests to measure learning gains, whereas eye-tracking may offer a more nuanced understanding of the effects of attention-drawing techniques during online reading. Choi (2017) asked EFL learners to read one of two versions of an authentic text while their eye movements were monitored. In one version, collocations were boldfaced (enhanced); in another version, the same items were unenhanced. Enhanced collocations generated more and longer fixations than unenhanced collocations, indicating that input-enhanced items attract more attention. This effect was more pronounced in collocations which were judged as unfamiliar by participants in a pre-test. Crucially, a trade-off emerged in a free recall test: the group who read the enhanced version had greater recall of target collocations but less recall of other unenhanced segments in the text than the group who read the unenhanced version.

Overall, attention-drawing techniques in formulaic language learning research have primarily involved input enhancement. Another way of making the written input more salient – although it has received far less attention in the literature – is by introducing target items in a pre-reading exposure. In other words, explicit exposure to lexical items is initiated prior to the reading stage. A pre-reading stage often occurs in the classroom in order to prepare students by setting reading goals, generating their interest in reading, enabling their understanding of the main ideas of the text, and clarifying some information (e.g., vocabulary) to facilitate comprehension (Grabe & Stoller, 2019). In a pre-reading exposure,

form-meaning knowledge of target lexical items can initially be established and then reinforced during reading. Nation (2001) advocates direct teaching of vocabulary (at least for high frequency words) in an L2 as it can contribute to the incidental learning of these words, attract learners' attention during reading, and assist in comprehension. It is conceived of as a "consciousness-raising" process that should thus not consume much classroom time (Nation, 2001, p. 252). Many studies have examined pre-reading exposure in relation to its role in reading comprehension (e.g., Alemi & Ebadi, 2010; Floyd & Carrell, 1987; Jahangard et al., 2012; Taglieber, Johnson, & Yarbrough, 1988; Tudor, 1990), leaving open the question of the role of pre-reading exposure in vocabulary learning. In one of the few studies to investigate the benefit of pre-reading exposure on L2 learning, File and Adams (2010) found that words which were taught prior to or during reading led to larger learning gains than words which were simply encountered in reading (i.e., incidental learning).

More relevant to the current study, but with a focus on single words, Pellicer-Sánchez et al. (2021) investigated the effect of pre-reading vocabulary instruction on learning pseudowords. In their study, NSs and NNSs participants were either 1) explicitly taught the target items before reading, then read a text in which the target items were repeated eight times (pre-reading instruction group); 2) asked to read the same text but without engaging in any pre-reading activity (reading-only); 3) asked to read the same text but with pseudowords replaced by known English words (reading-baseline group); or 4) explicitly taught the target items and read an irrelevant text (instruction-only). While participants read the text, their eye movements were monitored. Post-tests assessing form and meaning knowledge, showed that the pre-reading instruction condition led to the greatest lexical gains, while the reading-only condition generated the least. The superiority of a pre-reading instruction condition was also evident in the eye-tracking data. In initial encounters, both L1 and L2 participants, as expected, spent more time processing pseudowords (regardless of whether they were pre-

taught) than known words in the reading-baseline. Notably, by the third encounter for L1 and the eighth encounter for L2 participants, pseudowords in the pre-reading instruction were read in a similar way to known words. Pseudowords, which were not pre-taught (i.e., reading-only) continued to exhibit longer processing times than known words even in the eighth encounter (for the L2 group). The authors suggested that pre-reading instruction did not increase attention during reading, which contrasts with Choi's (2017) finding that input enhancement techniques attracted more attention. The authors speculated that increased attention occurs when an item is made salient during online processing (i.e., textual enhancement) but not when attention is drawn to it before reading.

In sum, most of the studies discussed thus far have had similar limitations. First, while all studies reported compared learning outcomes of different learning conditions using offline learning scores, very few studies examined the learning process by examining real-time processing using eye-tracking. Second, almost all of the studies reported (with the exception of Sonbul & Schmitt, 2013; Toomer & Elgort, 2019) assigned different learning conditions to each group. Thus, different learners' profiles and possible between-group variability may be driving the pattern of results rather than the conditions themselves.

The discussion of the literature touched on the effectiveness of different learning conditions in the learning of collocations. Intentional learning of collocations was described as necessary, very effective (Laufer & Girsai, 2008; Webb & Kagimoto, 2009), and potentially able to promote greater L2 vocabulary gains than incidental learning (Laufer, 2003). However, direct teaching of formulaic sequences is problematic due to constraints on limited classroom time (Boer & Lindstromberg, 2012). Alongside this, learning gains reported from incidental exposures to collocations are small (Pellicer-Sánchez & Boers, 2019), and for any sizable gains to occur, reading materials would need be designed such that they ensure sufficient exposure to collocations in a short period of time (Pellicer-Sánchez &

Boers, 2019; Webb et al., 2013). Insufficient encounters in authentic texts and a lack of perceptual salience in written input seem to hamper incidental learning of collocations. A semi-incidental approach might help overcome some of these issues.

Semi-incidental learning techniques, involving attention-drawing devices, have some practical advantages over incidental and intentional learning techniques. Input enhancement is easier to implement in teaching materials compared to input flooding (Pellicer-Sánchez & Boers, 2019). Attention-drawing devices depend on less effortful methods than direct teaching and should thus not consume much classroom time. Attention-drawing techniques like input-enhancement and pre-reading instruction have promise. Input-enhancement was found to be as beneficial as intentional learning and more beneficial than incidental learning (reading-only) for L2 collocational learning gains (Sonbul & Schmitt, 2013). Learning gains from input-enhancement were attributed to the increased attention paid to an item (Choi, 2017). This supports Schmidt's (2001) *Noticing Hypothesis* in which intentionally-focused attention is likely to result in successful learning. Nevertheless, the benefit of pre-reading exposure (as a type of attention-drawing technique) on the learning of collocations remains largely unknown. One study has looked at the impact of pre-reading on the learning and processing of L2 words (Pellicer-Sánchez et al., 2021). Notably, pre-reading led to gains in both processing and vocabulary knowledge.

5.3 The Present Study

Little is known about the effect of pre-reading exposure on L2 vocabulary learning; to the best of my knowledge, no study has investigated its role on collocation learning and processing. The available studies reported reveal conflicting results on the different attention-drawing techniques used: Choi (2017) found that input-enhancement attracted more attention and benefitted the learning of collocations; Pellicer-Sánchez et al. (2021) found that pre-reading exposure replicated the attested learning gains on vocabulary learning, but no

evidence for its role in increased attention was detected. Thus, more research is needed to address these conflicting findings in order to better understand the role of attention in the learning of collocations and has important pedagogical implications.

In this study, eye-tracking and offline test measures were used to evaluate the effectiveness of incidental, intentional, and semi-incidental learning on the processing and learning of collocations. Linking fine-grained real time cognitive processes (captured in eye-tracking data) with learning outcomes (captured in offline form recognition and recall tests) offers a more comprehensive explanation of the effect of different learning conditions on collocation learning. This study aims to answer the following research questions:

1. Which condition leads to more learning gains in form recognition and recall: Training plus Reading, Reading only or Training only?
2. What is the effect of each learning condition (Training plus Reading, Reading-only) on online processing during reading?
3. What is the impact of reading times (i.e., amount of study) in the training phase on a) reading in the reading phase and b) performance in the testing phase?

5.4 Methods

5.4.1 *Participants*

Sixty-one participants took part in the eye-tracking study (NSs = 29, NNSs = 32; male = 22, female = 59). An additional 20 participants (a NS control group) took only part in the off-line tasks to provide a baseline against which learning gains could be compared. Participants received course credit or were paid £7 for their participation. The participants were mostly undergraduate students (75.30 %), but some were postgraduate students (24.70 %). The NNS group came from different L1 backgrounds (Chinese = 13, Spanish = 3, Arabic = 2, Vietnamese = 2, Bengali = 2, Greek = 2, Gujarati = 1, Hindi = 1, German = 1, Italian =

1, Kannada = 1, Kashmiri = 1, Malian = 1, Urdu = 1). The *Lexical Test for Advanced Learners of English* (LexTALE; Lemhöfer & Broersma, 2012) was used to assess participants' English proficiency. Participants were asked to complete a short language background questionnaire providing information about their age, the age they began studying English, their length of stay in an English-speaking country, estimates of their proficiency in reading, writing, listening, speaking and understanding on a 7-point scale (1 = very low; 7 = native). Table 5.1 summarises the demographic and language proficiency data.

Table 5.1 Means and Standard Deviations of Demographic Data and Self-ratings of Proficiency on a 7-Point Scale (1 = very low; 7 = native)

	NS	**NNS**	**NS (control group)**
Age	20.10 (3.38)	22.81 (3.84)	18.40 (0.63)
Age of English acquisition	0.00 (0.0)	4.99 (2.70)	0.00 (0.0)
Time lived in an English-speaking country (in years)	18.93 (5.02)	3.74 (5.72)	18.13 (1.06)
LexTALE score (%)	93.82 (5.61)	75.57 (11.64)	90.93 (5.91)
Self-rating (1 to 7)			
Overall English proficiency	7.00 (0.00)	5.68 (0.89)	6.93 (0.25)
Proficiency in speaking	7.00 (0.00)	5.87 (0.94)	7.00 (0.00)
Proficiency in understanding	6.96 (0.18)	6.00 (0.67)	7.00 (0.00)
Proficiency in reading	6.93 (0.25)	6.21 (0.83)	6.93 (0.25)
Proficiency in writing	6.93 (0.25)	5.59 (1.07)	7.00 (0.0)

5.4.2 *Materials*

Forty-eight novel adjective-noun (ADJ-N) collocations were created based on several criteria. First, none of the collocations were phrases in English, as confirmed by the BNC (range = 0.00 to 0.03 occurrences per million; M = 0.004 per million; SD = 0.007). Second, the items were all transparent, that is, the meaning of a collocation was deducible from its constituents. Third, the items were controlled for their backward transitional probability (i.e.,

the probability that a noun would be preceded by a particular adjective), such that none of the items had a noun that was predicted by an adjective in an existing collocation. For example, the word *opening* was not a candidate for selection because it has a high backward transitional probability with the word *official* in the collocation *official opening* (phrase frequency > 50; MI > 3.0). Fourth, as confirmed by the University of South Florida database for association norms (Nelson et al., 1998), none of the word pairs in the novel collocations were semantic associates in either the forward or backward directions.

The forward transitional probability (the probability that an adjective predicts a set of nouns) was considered and manipulated. For example, the sequence *deep X* has a high type frequency since the noun slot can be filled with many variants (e.g., *sea, ocean, smile, voice*). In contrast, *abject X* has a low type frequency since the noun slot is more or less restricted to *poverty*. In the current study, novel collocations were classified as high type frequency (HT) if the node word (the adjective) predicted a number of noun collocates (e.g., *nuclear* predicts *weapons* and *war* with the novel collocation being *nuclear battery*) or low type frequency (LT) if the node word did not predict any noun collocates (e.g., *tempting* does not predict a particular noun, yielding the novel collocation *tempting pastries*). A novel collocation was considered HT if the adjective occurred in an AJD-N collocation with a frequency greater than 50 and had MI values greater than 3.0. A complete list of the items used in the study is presented in Appendix 5A.

To ensure that the items were indeed novel, a norming study was carried out to rate the items for their subjective familiarity. Fourteen NS participants (who did not take part in the main study) rated the novel collocations on a 5-point scale (from 5 = I have very frequently heard/used this phrase to 1= I have never heard/used this phrase). A set of existing English collocations of varying frequencies were included for comparison and to ensure that the full scale was used. An independent sample *t*-test confirmed a significant difference

between existing collocations (M = 4.66) and the novel ones (M = 2.66), t = 15.04, p < .001. To account for any minor differences in familiarity, the ratings were considered as a covariate in the analyses.

5.4.3 *Treatment Conditions*

A within-subject design was adopted in which each participant encountered all the three treatment conditions (except the control group). The target items appeared in one of the following three treatment conditions: (a) Training only (T-only) (i.e., an intentional learning condition in which each item appeared, and participants were explicitly asked to learn the items); (b) Reading only (R-only) (i.e., an incidental learning condition where each item occurred four times in different sentences, with each sentence appearing in a different block); (c) Training plus Reading (TR) (i.e., items were trained and occurred in the reading). Items were counterbalanced across three lists, such that an item appeared in a different treatment condition in each list. For example, the item *cynical viewers* appeared in the T-only condition in List 1, in the TR condition in List 2, and in the R-only condition in List 3. Each condition had 16 items, half HT and the other half LT.

In the training phase, a combination of T-only and TR items were displayed on two screens in a random order (16 items per screen). Participants were asked to familiarise themselves with the items as they would later occur in sentences. In an attempt to mimic an authentic study context, no time limit was imposed for the training phase, which is in contrast to Sonbul and Schmitt's (2013) study where participants were shown collocations for 10 seconds and asked to remember them. However, in the current study, eye-movements were monitored in the training phase, establishing the 'learning' time for each item. In the reading phase, a combination of R-only and TR items occurred in sentences presented in four blocks (32 sentences per block). Each block had a single exposure to an item. The materials used in the reading phase are available in Appendix 5B.

5.4.4 *Measurement Instruments*

Online Measures. While Study 1 used a self-paced reading task to measure reading and judgement times, the second study used eye-tracking to measure online processing/attention. The underlying assumption in eye-tracking is that the amount of time spent fixating an item during reading represents the degree of cognitive effort expended in processing it. Eye-tracking is an informative tool for measuring online reading processes and has several advantages over self-paced reading (and other techniques traditionally used to record response times). For example, eye-tracking does not require engaging in a secondary task (e.g., pressing a key) simultaneously while reading, thus allowing for a more natural form of reading that is closer to normal reading than that experienced via self-paced reading. Second, eye-tracking recordings offer a rich data source by capturing aspects of moment-to-moment cognitive processes (Conklin & Pellicer-Sánchez, 2016; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989; Rayner, 1998). While self-paced reading considers only one reading time measure (i.e., total reading time from when stimuli are first encountered), eye-tracking data considers a wide range of measures, capturing all processing activities of each fixation made during reading. These measures provide several millisecond reports of all fixations made on pre-defined regions of interest (ROI). Carrol and Conklin (2014b) recommended adopting a hybrid method of analysis when analysing formulaic sequences using eye-tracking, whereby individual words and the whole phrase are both taken into account in the analysis. Thus, in this study, the ROI was defined to examine novel collocations at both the word and phrase levels.

In eye-tracking, a number of early and late measures can be analysed. Indeed, one of the great advantages of eye-tracking is that it disentangles the effects of early and late stages of processing, tapping into different cognitive processes. Early measures are sensitive to early lexical access processes that are highly automatic (e.g., word recognition), while late

measures are sensitive to later, more conscious, and effortful processes (e.g., re-reading and re-analysis) which are associated with semantic and lexical integration (i.e., post-lexical processing) (Conklin & Pellicer-Sánchez, 2016). Broadly speaking, early measures reflect how easily a word's form is initially activated; late measures reflect how easily that word is integrated into a sentence (or wider discourse). The current study explored early and late measures that are commonly used in formulaic language research. A description of each measure is provided in Table 5.2.

Table 5.2 Eye-tracking Measures explored in the Current Study, with Descriptions and Stage of Processing for Each Unit of Analyses Examined

Stage of processing	Type of measure	Description
	Word level	
Early	Likelihood of skipping	The probability that a word is skipped the first time it is encountered.
Early	First fixation duration	The duration of the first fixation made on a word.
Early	First pass reading time	The sum of all fixation durations made on a word the first time it is encountered (i.e., before exiting to the right or left).
Late	Total reading time	The sum of all fixation durations made on a word.
Late	Total fixation count	The total number of fixations made on a word.
	Phrase level	
Late	Total reading time	The sum of all fixation durations made on a phrase.
Late	Total fixation count	The total number of fixations made on a phrase.

Offline Measures. Participants' knowledge of the novel adjective-noun collocations was assessed in two offline tasks: form recall and form recognition. In the form recall task, participants were given the adjective from each phrase and were asked to provide a one-word

completion (e.g., cynical _____). A response was considered as correct if it matched the noun completion for the novel collocation. The goal was to assess collocational knowledge rather than word form knowledge, thus minor misspellings were disregarded if the noun was recognizable (e.g., for the phrase *cynical viewers*, the misspelling *veiwers* was considered 'correct'). The form recognition task also provided the adjective as the prompt. Participants were given five choices to select from: the correct answer; three distractors (e.g., for *cynical*, the choices were *observers*, *viewers*, *audience*, and *readers*); and "I do not know" to minimize guessing. The distractors were selected if 1) they were not collocations in English; 2) they were semantically associated with the correct answer. The two tasks are included in Appendix 5C.

5.4.5 *Procedure*

The study was carried out in accordance with the research ethics procedures at the University of Nottingham, which provided ethical approval. Participants were tested individually in a quiet lab at the University of Nottingham. First, the study was explained in general terms and participants signed an informed consent form. Participants were randomly assigned to one of the three lists for the experimental group or to the control group. The experiment had three phases: training, reading, and testing. The three phases of the experiment were completed in one session lasting approximately 55 minutes. Depending on their group, participants took part in only one phase (control group: testing only), or in all the three phases (experimental group). Eye movements were monitored in the training and reading phases using an EyeLink 1000 Plus desktop-mounted eye-tracker (SR Research, Canada), with data sampled at 1000Hz. Participants were seated in front of a computer monitor with a chin- and forehead-rest to minimise their head movements. The text was displayed double-spaced in black Courtier New font (size 14) on a white background. Participants initially performed a nine-point calibration, and this was repeated before every

reading block. Each screen was followed by a drift correction and recalibration was carried out as needed.

In the training phase, participants were told that they would see a set of phrases on two screens that they may not have encountered before. They were asked to familiarise themselves with the phrases, as they would see them later used in sentences. Participants were informed that there was no time limit and to press the space bar once they were done looking at the phrases. Then, 16 phrases appeared split into two columns. When the space bar was pressed the second set of 16 phrases appeared also split into two columns. Participants' eye-movements were monitored during the training phase to determine how long each collocation was 'studied'.

Before beginning the main reading phase, there were four practice sentences followed by comprehension questions. The reading phase was divided into four blocks with 32 sentences in each – one occurrence of a novel collocation in a sentence per block. Sentences were presented one at a time on the screen, and participants pressed the spacebar to move forward. A quarter of the sentences were followed by yes-no comprehension questions to ensure that participants were reading for comprehension. Accuracy on the comprehension questions confirmed participants' understanding of the sentences (NSs: 92.25% accuracy; NNSs: 85.0% accuracy). At the end of the reading phase, participants filled out the biographical questionnaire and did the LexTALE task.

In the testing phase, all of the participants completed the collocation recognition and recall tasks.

5.4.6 *Analysis*

Eye tracking data from one NS were removed due to poor calibration. The data were cleaned using the Eyelink Data Viewer's four-stage procedure: 1.) very short fixations of < 40 ms were merged with neighbouring fixations within 0.5 degrees; 2.) those = 40 ms were

merged with those within1.25 degrees; 3.) any three consecutive fixations of < 140 ms were merged into one; and 4.) any fixations shorter than 80 ms or longer than 800 ms were deleted, as they "are assumed to represent, respectively, minor location errors rather than true fixations, and momentary losses of concentration"(Carrol & Conklin, 2020, p. 104). Each trial for each participant was visually inspected to remove any cases of trackloss or cases in which the entire collocation was skipped (leaving 9553 data points after removing 1.94% of the data).

Analyses were carried out on both the phrases (i.e., the whole adjective-noun phrase) and the final words (i.e., the noun). The analyses only considered phrase-level data for the training phase, while both individual word data and the whole phrase were considered for the reading phase. For the phrase-level analyses, two late eye-tracking measures were considered: total reading time and fixation count. For the word-level analyses, a range of measures were considered: early measures (likelihood of skipping, first fixation duration, and first pass reading time); late measures (total reading time and fixation count). Durational analyses were limited to the non-skipped items.

The experimental data was analysed by fitting a series of linear mixed effects models using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in R, version, 3.6.1, (R Core Team, 2019), with the *p*-values estimated by the *lmerTest* package (Kuznetsova *et al.* 2015). For the offline data (form recall and recognition task results), generalized linear mixed effects models were fitted with *binomial* distributions (Jaeger, 2008) (since the outcome variable was binary: 1 = correct, 0 = incorrect). For the likelihood of skipping, generalized logistic mixed effects models were fitted, and for fixation count generalized linear models with *poisson* distributions were fitted. Interactions and post-hoc comparisons were inspected using the *emmeans* package (Lenth, 2019) and plotted using the *effects* package (Fox, 2003). *P*-values for all post-hoc analyses were adjusted using Bonferroni

correction. Multicollinearity of each model constructed was assessed using the package *performance* (Lüdecke et al., 2020).[8]

For each model, the random slopes and intercepts of subjects and items were included as random effects (Barr, Levy, Scheepers, & Tily, 2013). Following the recommendation of Bates, Kliegl, Vasishth, and Baayen (2015) and Matuschek, Kliegl, Vasishth, Baayen and Bates (2017) random effect structures were not kept maximal because it may result in uninterpretable and overspecified models that are difficult to estimate.

Model parameters were added additively in a stepwise manner by consulting likelihood ratio tests with the package *afex* (Singmann et al., 2016). Parameters and interactions were only included in the final model if they significantly improved its fit. The reported models are the best fits with variables that made significant contributions (as indicated by likelihood-ratio chi-squared tests). Models included the following main predictors: Group (NS or NNS), Condition (treatment: T-only, R-only, or TR), Repetition (first, second, third or fourth) and Item Type (HT or LT). Since HT items had more frequent adjectives, making frequency a possible confounding factor with Item Type, a variable (Summed Frequency) was included that represented the sum of frequencies of the first word + second word of the phrase. Summed Frequency did not significantly improve the fit for any of the models. The following covariates were also considered: LexTALE score,[9] Phrase and Word Length (in letters), Trial Number (to control for practice or fatigue effects), Familiarity (from norming data), and Frequency (frequency of the second word).[10]

---

[8] VIF values of parameters in all models constructed indicated that the multicollinearity assumption was not violated.

[9] Because there was a correlation between LexTALE scores and the self-rated proficiency ($r = .35$, $p < .01$), the LexTALE scores were used in the analysis as it is a more objective measure of proficiency.

[10] In word-level analyses, only the frequency of the second word (target word) was considered. Phrasal frequency was not considered in phrase-level analyses as the phrases were novel and did not occur in the BNC.

All durational measure variables were log-transformed to reduce skewness. All of the other continuous predictors were also log-transformed to ensure that variables had the same scale. Word frequency was transformed into Zipf scale values (van Heuven, Mandera, Keuleers & Brysbaert, 2014).

5.5 Results

Results are reported in relation to each research question.

RQ1: Which condition leads to more learning gains in form recognition and recall: Training plus Reading, Reading only or Training only?

To address this question, I compared response accuracies on the form recall and form recognition tasks between the experimental groups and the control group (who only took part in the final testing task). As can be seen in Table 5.3 and the model summary in Table 5.4, there was a significant effect of training/exposure, such that both the NSs and NNSs who took part in the treatment conditions (T-only, R-only and TR) outperformed the control group in both form recognition and recall tests. Follow up post-hoc analysis showed no significant difference between NSs and NNSs in either task (*p's* > .05).

The three treatment conditions were also compared separately to the control group (i.e., the control group was considered a baseline condition). Form recall, following each of the treatment conditions (T-only, R-only, TR), had significantly more correct responses than the control group for both NSs and NNSs (*p's* > .05), with no differences between NSs and NNSs. Form recognition after the R-only and TR conditions yielded significantly greater accuracy than the control group for both NSs and NNSs (*p's* > .05), with no differences between NS and NNS. However, in the T-only condition, no difference was found between any of the groups (NS, NNS, and the control group).

Table 5.3 Mean Collocational Gains and Standard Deviation for the Trained/Exposed NNS and NS and the Untrained/Unexposed Control Group

	Form Recall		***Form Recognition***	
Predictors	Means	SD	Means	SD
Control Group	0.05	0.21	9.40	4.23
NNS Group	11.10	6.38	29.40	5.55
NS Group	8.93	5.51	29.37	5.43

Note. Maximum score is 48.

Table 5.4 Overall Effect of Treatment on Collocational Gains

	Form Recall				**Form Recognition**			
Predictors	*Odds Ratios*	*SE*	*CI*	*p*	*Odds Ratios*	*SE*	*CI*	*p*
(Intercept)	0.00	1.04	0.00 – 0.00	**< .001**	0.18	0.21	0.12 – 0.28	**< .001**
Group [NS]	546.28	1.05	70.44 – 4236.57	**< .001**	17.48	0.22	11.35 – 26.91	**< .001**
Group [NNS]	846.61	1.04	109.41 – 6550.95	**< .001**	17.86	0.22	11.68 – 27.31	**< .001**

Next, to evaluate the effectiveness of each treatment condition, the analysis only considered the NSs and NNSs in the treatment conditions. As shown in Tables 5.5 and 5.6, there were no overall differences between the groups in the form recall and recognition tasks. For the form recognition test, considering NSs and NNSs together, TR (M = 0.78, SD = 0.41) yielded significantly more learning gains than R-only (M = 0.75, SD = 0.43) and both of them more than T-only (M = 0.29, SD = 0.45). There was no interaction between Group and Condition. For form recall, the pattern was the same: TR (M = 0.33, SD = 0.47) yielded significantly more learning gains than R-only (M = 0.26, SD = 0.43) and both of them more than T-only (M = 0.03, SD = 0.19). For the form recall there was a significant interaction between Group and Condition, as illustrated in Figure 5.1. We can see that while there was no overall effect of Group, a significant difference between groups emerged in the R-only

condition, such that the NNSs significantly outperformed the NSs (odds ratio = 0.54, *SE* = 0.16, $p < .05$) (with no differences between groups in the remaining conditions).

In both form recall and recognition, Item Type did not improve the model fit, neither its interactions. Familiarity was a significant factor in form recall – phrases which had been judged as more familiar in the norming had higher form recall rates. However, familiarity did not make a significant improvement for the form recognition model.

Overall, in both form recall and recognition, both NS and NNS learned the novel collocations better when they were presented in the TR condition than when they were presented in either the R or the T condition. The most effective learning treatment can be summarised as: TR > R > T.

Table 5.5 Summary of the Mean Collocation Gains per Condition with Standard Deviations in Parentheses

	NS			**NNS**		
Type of knowledge	*T-only*	*R-only*	*TR*	*T-only*	*R-only*	*TR*
Form recall	0.75 (1.75)	3.57 (3.19)	4.64 (2.54)	0.47 (0.83)	4.66 (3.19)	5.94 (3.60)
Form recognition	4.89 (2.95)	12.00 (2.52)	12.60 (2.49)	4.62 (2.78)	12.20 (2.74)	12.60 (2.45)

Note. Maximum score is 16.

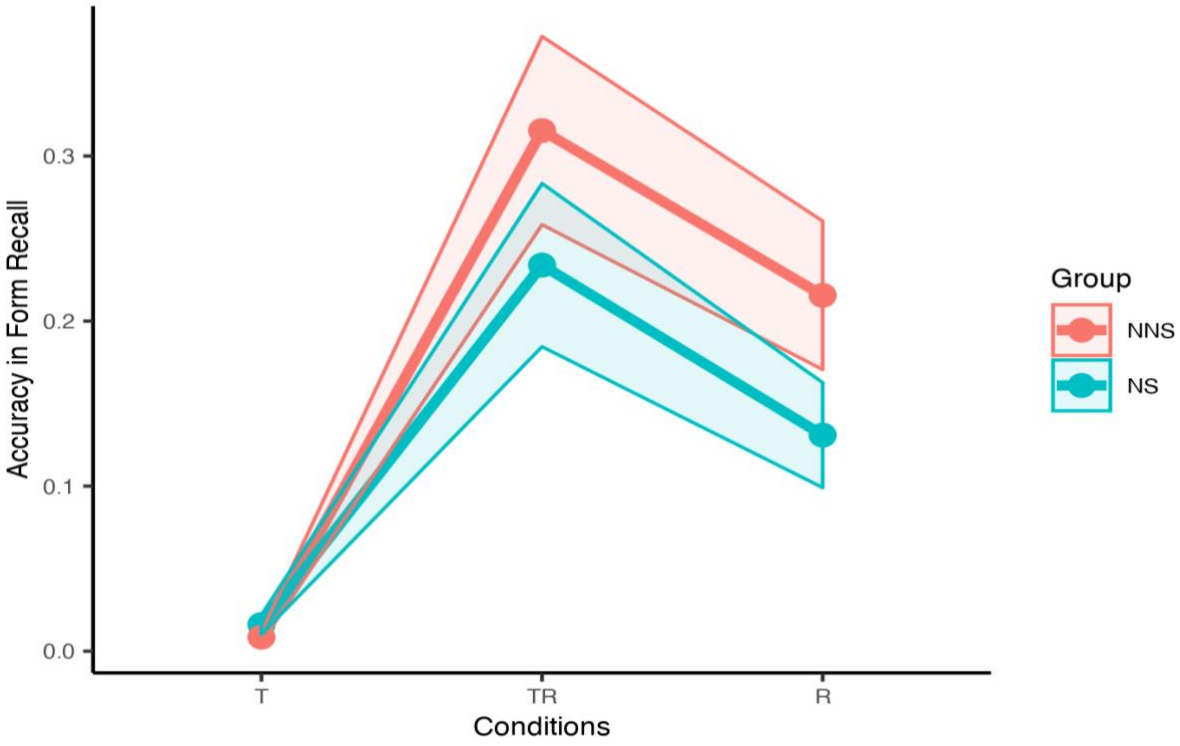
Table 5.6 Effects of Treatment Conditions on Collocational Learning

	Form Recall			**Form Recognition**		
Predictors	*Odds Ratios*	*std. Error*	*p*	*Odds Ratios*	*std. Error*	*p*
(Intercept)	0.02	0.37	**< .001**	0.32	0.23	**< .001**
Group [NNS]	0.52	0.48	.174	1.02	0.23	.947
condition [R-only]	9.20	0.25	**< .001**	13.36	0.10	**< .001**

condition [TR]	18.65	0.25	**< .001**	17.70	0.10	**< .001**
Familiarity	0.68	0.17	**.020**			
Group [NNS] * condition [R-only]	3.49	0.39	**.001**			
Group [NNS] * condition [TR]	2.89	0.38	**.006**			

Random effects	**Variance**	**SD**	**Variance**	***SD***
Subject (Intercept)	1.26	1.12	0.7459	0.86
Item (Intercept)	1.31	1.14	1.00	1.00

Figure 5.1 Interaction Between Group and Condition in Form Recall Scores (fitted)



RQ2: What is the effect of each learning condition (Training plus Reading, Reading-only) on online processing during reading?

Several models were fit to a range of eye-tracking measures to analyse the effect of treatment condition on reading behavior in the reading phase. Analyses considered both the whole phrase (phrase-level analysis) and the noun in the ADJ-N collocations (word-level). Table 5.7 summarises mean phrase- and word-level measures in the reading phase for both groups across different conditions.

Table 5.7 Summary of Reading Patterns of Phrases and Final Words for NS and NNS in the Reading Phase across Different Conditions and Repetitions (Rep)

		R-only				**TR**			
		Rep1	Rep2	Rep3	Rep4	Rep1	Rep2	Rep3	Rep4
	NS								
Phrase level	Total RT	787.05 (491.48)	663.79 (411.55)	591.03 (314.26)	552.07 (328.31)	731.57 (439.08)	600.12 (346.50)	583.02 (327.72)	524.27 (275.02)
	Fixation count	3.64 (2.23)	3.11 (1.75)	2.85 (1.38)	2.66 (1.45)	3.40 (1.91)	2.77 (1.38)	2.80 (1.46)	2.50 (1.17)
Word level	Skipping rate	0.11 (0.32)	0.17 (0.38)	0.18 (0.38)	0.20 (0.40)	0.13 (0.34)	0.21 (0.41)	0.17 (0.37)	0.22 (0.41)
	First Fixation RT	231.35 (85.66)	213.64 (67.33)	221.21 (80.44)	211.84 (69.77)	219.13 (81.93)	227.80 (74.84)	218.38 (82.66)	217.80 (77.04)
	First Pass RT	260.77 (118.40)	241.81 (103.98)	236.84 (95.31)	238.06 (113.40)	251.76 (115.00)	255.74 (111.32)	238.52 (105.36)	244.72 (117.57)
	Total RT	404.06 (308.55)	326.77 (180.50)	306.46 (182.67)	284.07 (169.11)	363.71 (211.99)	310.43 (161.17)	300.07 (177.52)	289.61 (160.70)
	NNS								
Phrase level	Total RT	1085.24 (664.04)	915.04 (604.05)	918.07 (567.41)	790.60 (475.25)	1031.90 (666.20)	906.625 (545.33)	877.43 (587.81)	760.47 (422.72)
	Fixation count	4.62 (2.52)	3.92 (2.32)	4.00 (2.27)	3.43 (1.93)	4.46 (2.67)	3.93 (2.27)	3.79 (2.29)	3.31 (1.70)
Word level	Skipping rate	0.08 (0.27)	0.10 (0.30)	0.11 (0.31)	0.13 (0.34)	0.09 (0.29)	0.13 (0.34)	0.13 (0.34)	0.12 (0.32)
	First Fixation RT	257.68 (89.40)	255.35 (98.94)	251.02 (96.22)	248.88 (90.78)	250.74 (92.52)	244.86 (86.42)	246.56 (87.47)	236.90 (83.24)
	First Pass RT	338.67 (203.25)	318.72 (162.50)	316.38 (243.89)	300.89 (148.85)	322.00 (188.67)	315.14 (174.19)	300.47 (138.80)	290.78 (142.20)
	Total RT	552.10 (415.51)	466.77 (345.87)	469.01 (362.13)	406.18 (261.68)	487.59 (328.63)	462.60 (301.08)	432.60 (285.25)	378.00 (232.78)

Note. Duration measures are reported in milliseconds. Skipping rate is reported as a probability. Values in parentheses are standard deviations.

Table 5.8 reports the analysis for total reading time and fixation count for the entire collocational phrase in the reading phase. Findings from both eye-tracking measures were

consistent. That is, the NNS group had longer total reading times and made more fixations on the entire phrase than the NS. The TR condition elicited less total reading time and fewer fixations than the R-only condition for both groups. Repetition had a faciliatory effect, in that every additional exposure to an item resulted in fewer fixations and less total readings time. However, the only individual comparison between repetitions that reached significance in fixation count was between the first and fourth repetitions. The only comparison that did **not** reach significance in total reading time was between the second and third repetitions. Neither Group nor Condition interacted with Repetition, suggesting that effect of repetition was equally robust in Group and Condition. Familiarity was a significant predictor, such that more familiar phrases had less total reading time. However, familiarity did not influence fixation count. Item Type also played a role in total reading time, such that items of lower type frequency (LT), required more processing effort for both groups. Trial Number was significant with phrases having less total reading times and fewer fixations as the experiment progressed.

Table 5.8 Effect of Condition and Repetition on Reading Time at the Phrase Level

	Total Reading Time				**Fixation Count**			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*	*Log-Mean*	*std. Error*	*Statistic*	*p*
(Intercept)	5.14	0.26	19.48	**<0.001**	-0.05	0.26	-0.18	0.854
Group [NNS]	0.37	0.07	5.40	**<0.001**	0.28	0.07	4.19	**<0.001**
Repetition [2]	-0.07	0.02	-3.06	**0.002**	-0.04	0.03	-1.73	0.084
Repetition [3]	-0.06	0.03	-2.00	**0.045**	-0.02	0.03	-0.51	0.610
Repetition [4]	-0.14	0.03	-4.25	**<0.001**	-0.11	0.04	-2.87	**0.004**
Condition [TR]	-0.05	0.01	-4.55	**<0.001**	-0.05	0.01	-3.96	**<0.001**
Familiarity [log]	-0.14	0.06	-2.62	**0.009**	-0.08	0.05	-1.46	0.143
Item Type [LT]	0.07	0.03	2.21	**0.027**				

Trial Number [log]	-0.08	0.02	-5.65	<**0.001**	-0.10	0.02	-5.99	<**0.001**
Length Phrase [log]	0.62	0.09	7.03	<**0.001**	0.59	0.08	7.20	<**0.001**
Random Effects								
σ^2	0.20				0.26			
τ_{00}	0.07 $_{subject}$				0.06 $_{subject}$			
	0.01 $_{Item}$				0.01 $_{Item}$			
ICC	0.27				0.21			

Table 5.9 presents the model outputs for the analysis of the noun from the ADJ-N collocations across measures. The table shows that the NS group was more likely to skip phrase-final words than the NNS group. While skipping rate in the TR condition was greater than in the R-only condition, this difference was only marginally significant. Repetition had a significant effect on the likelihood of skipping, with every additional exposure increasing the probability of skipping the final word of the collocation. Subsequent analysis of the effect of Repetition on skipping showed that 2 vs. 3 was the only comparison which did not reach significance. Analysis of the remaining measures also showed that that NNS group had longer total reading time, longer first pass reading time, and a longer first fixation duration than NS group. While TR led to shorter total reading time and shorter first pass reading time for both groups, this effect was only significant in the total reading time measure. There was no interaction between Condition and Repetition, suggesting that the TR yielded shorter total reading time than R-only across all repetitions, and that Repetition had the same effect for both conditions in decreasing reading time. Repetition significantly decreased total reading time (with the exception of 1 vs. 3 and 2 vs. 3 comparisons), first pass reading time (with the exception of 3 vs. 4 comparisons), and first fixation duration (with the exception of 3 vs. 4

comparisons) for both groups. Item Type only improved the fit of the first pass reading time model ($\chi^2 = 4.20$, $p < .05$) and the first fixation duration model ($\chi^2 = 9.91$, $p < .05$), suggesting that HT items required less processing effort than LT items in the early stages of processing. However, while Item Type reached a statistical significance in first fixation duration, it was only marginally significant in first pass reading time. The English proficiency level (LexTALE score) significantly improved the model's fit for first pass reading time ($\chi^2 = 6.19$, $p < .05$); higher LexTALE scores predicted shorter first pass reading times. However, inclusion of LexTALE did not improve the fit of models in other measures.

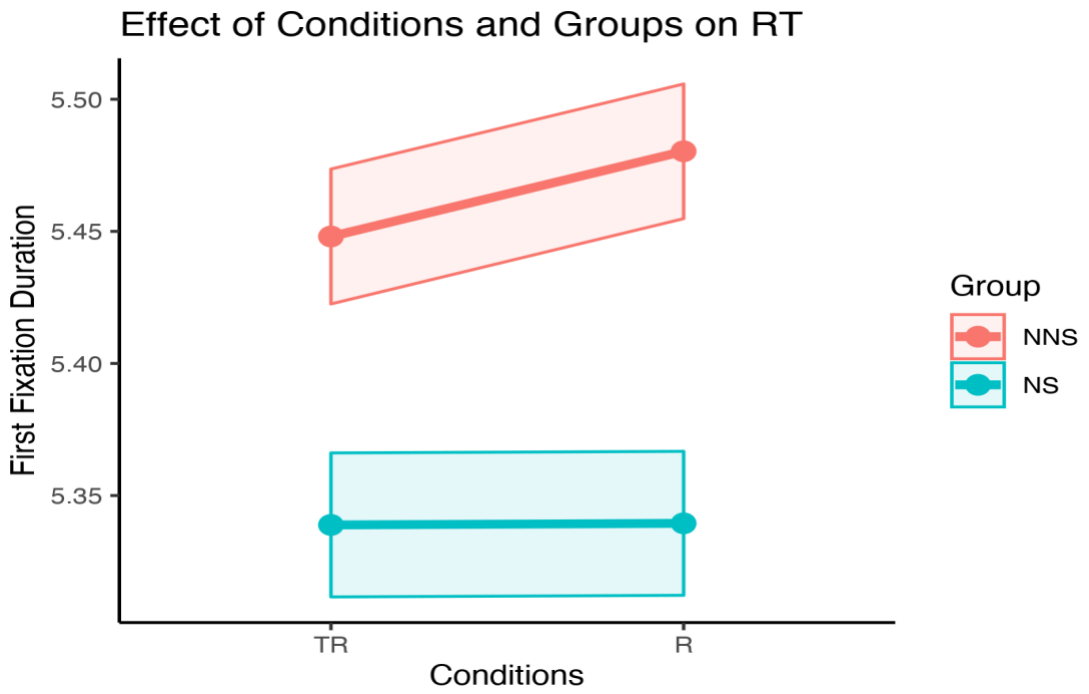
While total reading time and first pass reading time did not reveal significant interactions (Group x Condition, Group x Repetition, Condition x Repetition), first fixation duration demonstrated a significant interaction between Group and Condition. Analysis of this interaction, as plotted in Figure 5.2, indicated that the effect of Condition on first fixation durations was different across groups. While the NS group experienced no differences across learning conditions ($SE = 0.01$, $p > .05$), the NNS group experienced longer first fixation durations in the R-only condition than in the TR condition ($SE = 0.01$, $p < .05$).

Table 5.9 Omnibus Linear Mixed Effects Model Output for Eye-Tracking Measurements Used for Final Word

	Skipping Rate				**Total Reading Time**				**First Pass RT**				**First Fixation Duration**			
Predictors	Log-Odds	std. Error	Statistic	p	Estimates	std. Error	Statistic	p	Estimates	std. Error	Statistic	p	Estimates	std. Error	Statistic	p
(Intercept)	3.26	0.48	6.81	**<0.001**	5.26	0.13	40.84	**<0.001**	6.79	0.77	8.77	**<0.001**	5.11	0.06	81.22	**<0.001**
Group [NNS]	-0.69	0.22	-3.13	**0.002**	0.31	0.05	5.67	**<0.001**	0.13	0.06	2.30	**0.021**	0.14	0.04	3.90	**<0.001**
Repetition [2]	0.92	0.17	5.38	**<0.001**	-0.05	0.03	-1.97	**0.048**	-0.07	0.02	-3.31	**0.001**	-0.04	0.02	-2.24	**0.025**
Repetition [3]	1.03	0.22	4.74	**<0.001**	-0.06	0.03	-1.90	0.057	-0.12	0.03	-4.67	**<0.001**	-0.06	0.02	-2.64	**0.008**
Repetition [4]	1.33	0.25	5.37	**<0.001**	-0.12	0.04	-3.22	**0.001**	-0.14	0.03	-4.73	**<0.001**	-0.08	0.02	-3.17	**0.002**
Condition [TR]	0.13	0.07	1.80	0.072	-0.05	0.01	-4.06	**<0.001**	-0.02	0.01	-1.74	0.081	-0.00	0.01	-0.05	0.958
Length Word [log]	-2.44	0.18	-13.54	**<0.001**	0.38	0.06	6.50	**<0.001**	0.28	0.04	7.12	**<0.001**	0.09	0.02	3.81	**<0.001**
Trial Number [log]	-0.35	0.11	-3.24	**0.001**	-0.07	0.02	-3.97	**<0.001**	0.03	0.01	2.20	**0.028**	0.02	0.01	1.57	0.117
Item Type [LT]									0.04	0.02	1.94	0.052	0.05	0.01	3.35	**0.001**
LexTALE [log]									-0.43	0.17	-2.56	**0.011**				
Group [NNS] * Condition [TR]													-0.03	0.02	-2.03	**0.043**
Random Effects																
σ^2	3.29				0.23				0.14				0.10			

τ_{00}	0.62 $_{subject}$	0.04 $_{subject}$	0.02 $_{subject}$	0.02 $_{subject}$
	0.07 $_{Item\_number}$	0.01 $_{Item\_number}$	0.01 $_{Item\_number}$	0.00 $_{Item\_number}$
ICC	0.17	0.19	0.17	0.16

Figure 5.2 Interaction between Condition and Group on First Fixation Duration (fitted and log-transformed)



RQ3a: What is the impact of reading times (i.e., amount of study) in the training phase on reading in the reading phase?

This analysis more directly considers the impact of the training on subsequent reading behaviour. More precisely it looks at the amount of time spent studying (i.e., fixating) the novel collocations in the training phase and how this relates to the subsequent reading behaviour. First, I considered if there was variability in how long participants studied the novel collocations in the training phase. A summary of participants' performance in training is shown in Table 5.10. The table shows that NSs had a greater in-group variability in reading times when studying the items than NNSs. Table 5.11 reports the model outcomes for analysing the differences in performance between groups in the training phase. Total reading

time and fixation count were not significant, suggesting that both groups allocated similar amounts of time and showed similar looking patterns when studying the items. Although Item Type was significant in the reading phase, with HT collocations eliciting shorter reading times, it did not play a role in the training phase.

Table 5.10 Summary of Studying Patterns of Phrases in the Training Phase for NS and NNS

	T-only		**TR**	
	NS	NNS	NS	NNS
Total reading time (ms)	2057.03 (1870.15)	2152.41 (1227.62)	2042.05 (1912.60)	2142.95 (1245.53)
Range	184 -16023	213 - 6784	223 - 16480	207 - 8705
Fixation count	8.18 (7.33)	8.43 (4.88)	8.15 (7.22)	8.58 (5.14)
Range	1 - 57	1 - 30	1 - 63	1 - 29

Table 5.11 Model Outcome for Eye-tracking Measurements in the Training Phase

	Total Reading Time				**Fixation Count**			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*	*Log-Mean*	*std. Error*	*Statistic*	*p*
(Intercept)	6.20	0.31	20.31	**<0.001**	0.64	0.31	2.10	**0.036**
Group [NNS]	0.12	0.13	0.94	0.347	0.12	0.13	0.92	0.357
Familiarity [log]	-0.15	0.06	-2.32	**0.020**	-0.09	0.06	-1.50	0.133
Length Phrase [log]	0.49	0.10	4.88	**<0.001**	0.51	0.10	5.13	**<0.001**
Item Type [LT]	-0.03	0.04	-0.74	0.457	-0.03	0.03	-0.90	0.369
Random Effects								
σ^2	0.16				0.13			
τ_{00}	0.24 $_{subject}$				0.24 $_{subject}$			
	0.01 $_{Item\_}$				0.01 $_{Item\_}$			
ICC	0.61				0.66			

Second, I examined whether training performance related to subsequent reading behaviour. Analyses were only carried out on the TR condition because it is the only condition in which there are reading times in both the training and reading phases. Eye-tracking measures in the training phase (total reading time, first pass reading time, and fixation count) were considered as predictors of the reading behaviour in the reading phase. Analysis of the effect of training on reading patterns is highlighted in Table 5.12 for the phrase level and Table 5.13 for the word-level. Table 5.12 shows that the only eye-tracking measure from the training session that improved the model was total reading time; total reading time in the training phase had a subsequent effect on the reading phase with longer fixation durations in the training session associated with longer fixation durations and a higher number of fixations in the reading phase. The effect did not interact with Group, indicating that training benefited both groups equally in the reading session. As can be seen in Table 5.13 (word-level), only first pass reading in the training phase showed a relationship with subsequent reading ($\chi^2 = 5.76$, $p < .05$). Increased first pass reading time in the training phase resulted in an increased first pass reading time in the reading session (replicating the pattern observed at the phrase-level). This suggests that increased attention paid to an item during study increased its salience during reading, which might explain the positive correlation between processing in the two sessions. No eye tracking measures from the training predicted skipping during reading. LexTALE only improved model fit for first pass reading time at the word-level, with higher proficiency resulting in a shorter first pass reading times. Replicating the pattern from above (RQ2), effects of Repetition and Item Type remained robust in the reading phase (in both phrase and word-level) even after accounting for looking patterns in the training phrase.

Table 5.12 Effects of Training on Reading Behavior at the Phrase Level

	Total Reading Time			**Fixation Count**		
Predictors	*Estimates*	*std. Error*	*p*	*Incidence Rate Ratios*	*std. Error*	*p*
(Intercept)	5.01	0.30	**<0.001**	0.79	0.31	0.445
Group [NNS]	0.37	0.07	**<0.001**	1.32	0.06	**<0.001**
Total reading time from Training session [log]	0.04	0.02	**0.035**	1.07	0.02	**0.002**
Repetition [2]	-0.06	0.03	**0.048**	0.96	0.04	0.219
Repetition [3]	-0.06	0.04	0.149	0.98	0.05	0.689
Repetition [4]	-0.14	0.05	**0.003**	0.89	0.05	**0.028**
Familiarity [log]	-0.14	0.06	**0.012**	0.92	0.06	0.144
Length Phrase [log]	0.52	0.09	**<0.001**	1.56	0.09	**<0.001**
Item Type [LT]	0.08	0.03	**0.008**	1.07	0.03	**0.039**
Trial Number [log]	-0.08	0.02	**<0.001**	0.91	0.02	**<0.001**
Random Effects						
σ^2	0.20			0.27		
τ_{00}	0.06 subject			0.05 subject		
	0.01 Item			0.01 Item		
ICC	0.25			0.19		

Table 5.13 Effects of Training on Reading Behavior at The Word Level

	Total Reading Time				**First Pass Reading Time**				**First Fixation Duration**			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*	*Estimates*	*std. Error*	*Statistic*	*p*	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	5.10	0.12	42.03	**<0.001**	6.44	0.79	8.10	**<0.001**	5.04	0.08	61.12	**<0.001**
Group [NNS]	0.29	0.05	5.47	**<0.001**	0.12	0.06	2.13	**0.033**	0.11	0.04	2.98	**0.003**
Repetition [2]	-0.09	0.02	-3.91	**<0.001**	-0.01	0.02	-0.34	0.736	0.01	0.02	0.58	0.564

Repetition [3]	-0.15	0.02	-6.57	<**0.001**	-0.06	0.02	-3.11	**0.002**	-0.01	0.02	-0.75	0.451
Repetition [4]	-0.23	0.02	-9.72	<**0.001**	-0.07	0.02	-3.69	<**0.001**	-0.03	0.02	-2.15	**0.031**
Length Word [log]	0.33	0.06	5.56	<**0.001**	0.27	0.04	7.07	<**0.001**	0.09	0.03	3.52	<**0.001**
First Pass Reading Time in Training Session [log]					0.03	0.01	2.32	**0.021**				
Item Type [LT]					0.05	0.02	2.39	**0.017**	0.05	0.02	3.20	**0.001**
LexTALE [log]					-0.38	0.17	-2.16	**0.031**				

Random Effects

σ^2	0.22	0.14	0.10
τ_{00}	0.04 $_{subject}$	0.02 $_{subject}$	0.02 $_{subject}$
	0.01 $_{Item}$	0.00 $_{Item}$	0.00 $_{Item}$
ICC	0.19	0.16	0.16

RQ3b: What is the impact of reading times in the training/reading phase on performance in the testing phase?

The second part of RQ3 considers how looking patterns in the training/reading phase impacted performance on the form recognition and form recall. Specifically, I examined if there was an effect of studying behaviour (i.e., amount of study) on the response accuracies in the form recall and recognition tasks, that was above and beyond any effects in the reading phase. I analysed eye tracking measures in the training phase contributed by the T-only and TR conditions separately. Table 5.14 presents the model outcome for accuracy in recall and recognition tasks as a function of eye tracking measures in the training session. There was no improvement for any eye tracking measures in the model for the TR condition. This suggests

that reading patterns in the reading phase may have overshadowed any effects from the studying behaviour for TR items. The effect of training in the T-only condition on the testing phase was apparent even though testing did not occur immediately after training. In T-only, while groups performed similarly in the form recall, there was a significant interaction between the Group and first pass reading time during the training phase ($\chi^2 = 7.39$, $p < .05$), suggesting that groups were affected by first pass reading time differently. Analysis of this interaction indicates that first pass reading time in the training session significantly predicted form recall for NNS group;[11] longer first pass reading time was associated with higher form recall scores. However, the effect of first pass reading time was not as robust for the NS group: first pass reading time did not play a significant role in NS's recall. First pass reading time was not significant in form recognition. For form recognition, fixation count was a significant predicator, such that increased number of fixations improved accuracy of recognition for both groups.

As a final analysis, the relationship between online processing in the reading phase (R-only) and learning outcomes was explored further. The Summed Total Reading Time in the reading phase was calculated by adding total reading times (phrase-level) of all four repetitions for each item. Summed Total Reading Time made no improvement in either the model for form recall ($\chi^2 = 0.21$, $p > .05$) or the model for form recognition ($\chi^2 = 0.06$, $p > .05$).

[11] Each group was analysed separately.

Table 5.14 Effect of Time on the Training Phase on Form recall and Form Recognition

	T-only condition								TR condition							
	Form Recall				Form Recognition				Form Recall				Form Recognition			
Predictors	*Log-Odds*	*std. Error*	*Statistic*	*p*	*Log-Odds*	*std. Error*	*Statistic*	*p*	*Log-Odds*	*std. Error*	*Statistic*	*p*	*Log-Odds*	*std. Error*	*Statistic*	*p*
(Intercept)	-2.41	0.00	-2458.89	**<0.001**	-2.47	0.91	-2.71	**0.007**	0.63	1.02	0.62	0.536	1.07	1.13	0.95	0.344
Group [NNS]	-13.02	0.00	-13288.70	**<0.001**	-0.19	0.24	-0.80	0.427	0.45	0.29	1.53	0.126	-0.06	0.27	-0.21	0.830
First Pass RT in Training [log]	-0.32	0.00	-330.72	**<0.001**	0.14	0.11	1.21	0.225	-0.05	0.13	-0.38	0.701	0.10	0.14	0.71	0.478
Item Type [LT]	-0.98	0.00	-1000.41	**<0.001**					0.03	0.30	0.10	0.922				
Group [NNS] * First Pass RT in Training	1.83	0.00	1866.40	**<0.001**												
Fixation Count in Training [log]					0.70	0.16	4.28	**<0.001**					0.18	0.17	1.02	0.306
Familiarity [log]					-0.79	0.38	-2.05	**0.041**	-1.56	0.53	-2.94	**0.003**	-0.34	0.58	-0.59	0.556
Random Effects																
σ^2	3.29				3.29				3.29				3.29			
τ_{00}	2.76 $_{subject}$				0.49 $_{subject}$				0.87 $_{subject}$				0.60 $_{subject}$			
	2.02 $_{Item\_number}$				0.28 $_{Item\_number}$				0.74 $_{Item\_number}$				0.91 $_{Item\_number}$			
ICC	0.59				0.19				0.33				0.31			

5.6 Discussion

Previous research has investigated the intentional and/or incidental learning of collocations. However, very little research has utilized eye-tracking (as a measure of online processing) in combination with offline tests (as a measure of learning outcomes) to assess the effectiveness of different learning conditions in a single study. Further, the effect of an attention-drawing technique – pre-reading exposure – on the learning of collocations had yet to be explored.

The present study investigated three main questions. First, I assessed learning gains (immediate post-test of form recall and recognition) for novel collocations by NSs and NNSs in different learning conditions: training only (i.e., pre-reading exposure only, T-only), reading only (i.e., reading only with no pre-reading exposure, R-only), and training plus reading (i.e., pre-reading exposure plus reading, TR). Second, I investigated whether the learning conditions modulated processing of the novel collocations during reading comprehension. To achieve this, eye movement patterns were examined for items from different learning conditions and from different item types (high type frequency, HT vs. low type frequency, LT). Finally, I examined whether amount of study time (reading time in the training phase) affected processing of the same items when they were introduced in the reading phase (TR items) and whether it affected learning gains (TR and T-only items).

The main findings of the current study that will be discussed in relation to the wider literature can be summarised as follows. The TR condition was the most effective learning condition in terms of both recall and recognition, and the R-only was more effective than the T-only condition. The TR condition elicited shorter total reading times (at both word-level and phrase-level) and fewer fixations (at only phrase-level) than the R-only condition for both groups. HT items elicited shorter total reading times (phrase-level) and shorter first

fixation duration (word-level) than LT items for both groups. There was an interaction between Group and Condition in first fixation duration (word-level), such that NNSs read TR items faster (in the reading phase) than R-only, while the NSs read both TR and R-only items at similar speeds. Total reading time in the training phase predicted total reading time and fixation count in the reading phase (phrase-level). First pass reading time in the training phase predicted first pass reading time in the reading session (word-level). Eye tracking measures from the training phase only improved the model of learning gains for T-only, with an interaction between Group and first pass reading time, such that first pass reading time predicted form recall results for only the NNS group but not the NS group. For form recognition, fixation count in the training phase predicted form recognition scores in the T-only condition for both NSs and NNSs. Summed total reading times (additive total reading times across all repetitions in the reading phase) predicted neither form recall nor form recognition scores.

5.6.1 *The Effect of Condition on Learning Gains*

The findings demonstrated that while all learning conditions led to more gains relative to the control group (with no training or reading), the TR condition led to the greatest learning gains and the T-only condition led to the smallest learning gains. This is a pattern that was observed in both NS and NNS groups, as well as in both recall and recognition tasks. This highlights the benefit of pre-reading exposure for learning collocations. It demonstrates that intentionally-focused attention to decontextualized collocations, when combined with incidental exposure, results in greater learning gains than incidental or intentional exposure alone. The current finding aligns with previous findings showing the benefit of attention-drawing techniques leading to more gains than reading-only conditions (Choi, 2017; Sonbul & Schmitt, 2013; Toomer & Elgort, 2019).

This finding for collocations expands upon those of Pellicer-Sánchez et al. (2021) for single words, demonstrating that explicit pre-reading exposure plus incidental exposure is more beneficial than intentional learning alone (T-only) and incidental learning alone (R-only). The greater learning gains from pre-reading exposure in Pellicer-Sánchez et al. was evident in the current study, although the type of pre-reading exposure was different. In the current study, there was merely an explicit exposure to the items, while in Pellicer-Sánchez et al. participants also engaged in a pre-reading activity with the items. These findings support Schmitt's (2008) view that incidental learning is more effective when utilized to reinforce knowledge of previously encountered or partially known words.

The superior learning gains associated with the TR condition cannot easily be explained by Schmidt's (2001) *Noticing Hypothesis*. TR items had faster reading times compared to the R-only items that were encountered in the reading for the first time. Thus, pre-reading study did not increase attention to the items. Alternatively, it can be argued that the superiority of TR items in learning gains over other learning conditions could be attributed to the further extra exposures provided by the TR condition. In the TR condition, learners engaged with the target items initially in explicit exposures during pre-reading and then again in contextual exposures during reading. The repeated exposures in the reading phase seem to reinforce knowledge of collocations that was initially encoded in memory through explicit exposures, resulting in the greater learning gains of TR compared to R-only and T-only. At the beginning stages of vocabulary learning, intentional learning is most effective for acquiring initial form-meaning knowledge; it is essential for this knowledge, once established, to be reinforced with repeated contextualized exposures (Schmitt, 2008).

The superiority of the R-only condition over the T-only condition is likely due to repeated contextual exposures of the items, lending support to many studies confirming that repeated incidental exposures can still generate substantial learning gains, even in the absence

of any intentional learning of the target items (e.g., collocations: Durrant & Schmitt, 2010; Pellicer-Sánchez, 2017; Sonbul & Schmitt, 2013; Toomer & Elgort, 2019; Webb, Newton, and Chang, 2013; single words: Chen & Truscott, 2010; Horst et al., 1998; Rott, 1999; Pigada & Schmitt, 2006; Webb, 2007). Incidental exposures of items presented multiple times in context were more beneficial than intentional exposures of decontextualised items. This finding casts doubts on the claims that L2 learners fail to notice formulaic sequences in input (Wray, 2002), and that intentional learning necessarily promotes greater L2 vocabulary gains than incidental learning (i.e., reading) (Laufer, 2003). The advantage of R-only is likely due to the incidental learning of repeatedly encountered items, which is line with usage-based approaches (Tomasello 2003; Bybee 2013) supporting the role of frequency of occurrence in the representation and acquisition of lexical patterns.

Frequency of occurrence was also beneficial in the development of procedural knowledge. Repeated exposure was found to speed up reading times. This replicates recent eye-tracking findings that less processing effort is expended as a function of repeated contextual exposures during reading (Conklin & Carrol, 2020; Elgort et al., 2018; Godfroid et al., 2018; Joseph et al., 2014; Mohamed, 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021). Notably the repetition effect was comparable for the NSs and NNSs (in line with Godfroid et al., 2018) and in the TR and the R-only conditions (in line with Pellicer-Sánchez et al., 2021). This effect is likely due to increasing familiarity with the items with increasing exposures, which results in a faster decoding of the target items. More specifically, the first encounter with a multiword sequence forms an initial memory trace that is later strengthened with each additional encounter (Logan, 1988).

In contrast to the current findings, Pellicer-Sánchez et al. (2021) found that instruction-only (i.e., pre-reading exposure only) was, overall, more beneficial in terms of lexical gains than the reading-only condition. It may be that the pre-reading exposure

implemented in the Pellicer-Sánchez et al. (2021) study (i.e., learning and activities) involved tasks that were more closely matched to the post-test. The pre-reading activities may also have encouraged greater engagement with the items. Further, the participants in the current study did not spend much time studying the items in the training phase (i.e., average per item NS = 2.05s and NNS = 2.15s). Time spent on the items in the Pellicer-Sánchez et al. study was not measured, but because there was study and an explicit activity, it is likely that participants spent more time on the items in the pre-reading phase. Thus, the lack of explicit exercises in the training phase in the current study may have meant that there was less engagement with the target items, which could explain the superiority of R-only over T-only.

In the current study, gains in T-only were relatively small compared to those in Sonbul and Schmitt (2013), which could be attributed to the relatively short training time in the current study. Participants in Sonbul and Schmitt (2013) studied each item for 10 seconds, while in the current study time was up to the individual and was around two seconds. While the gains were less than in Sonbul and Schmitt, T-only was more beneficial for both NSs and NNSs than no training at all (i.e., versus the control group), suggesting that even short explicit exposure is beneficial. Interestingly, learning gains from the T-only condition were observed in spite of the time interval between learning and testing (i.e., after learners were exposed to T-only, they engaged in the reading phase of R-only and TR items which took 20-35 minutes before testing). The short exposure appears to be sufficient for learners to acquire receptive knowledge of collocations, supporting the claim that even one exposure leaves an imprint in memory (Goldberg, 2007; Logan, 1988; Rott, 2009). In the TR condition, it appears that the initial memory traces from the training are strengthened later, during reading. In the R-only condition, memory traces are strengthened with repeated exposures. This could account for the superiority of TR and R-only relative to T-only. Notably, the effect of learning condition on collocation gains was the same for both NSs and

NNSs (TR > R-only > T-only) which suggests that both NS and NNS are equally affected by the amount and type of input (Hoey, 2005).

Interestingly, the NNS group had greater gains in form recall in the R-only condition than the NS group (29% vs. 22%). This difference could be due to NSs processing textual input for meaning, whereas NNSs are more sensitive to language form (Sonbul and Schmitt, 2013). Similar to the current results, Sonbul and Schmitt (2013) found that NSs' knowledge of the form of collocations benefitted less from incidental exposure (reading-only) than from decontextualized input (T-only), while NNSs benefitted equally from both decontextualized input and incidental exposure. A form-focused versus meaning-focused approach to reading is supported by the eye-tracking measures. First fixation duration reflects the earliest point where a word form related effect can emerge (Liversedge, Paterson, & Pickering, 1998), tapping into initial and more automatic stages of processing. Notably, NSs did not exhibit differences in first fixation durations for items that were encountered in the reading for the first time (R-only) versus items that were studied in the pre-reading exposure (TR). However, NNS exhibited longer first fixation duration for items that that were encountered in the reading for the first time compared to items that were studied in the pre-reading exposure, suggesting that sensitivity to the form of previously encountered items was robust in the NNS group. That is to say, differences in form recall performance may reflect a different manner of processing during incidental exposures, which may have given the NNSs an advantage on the recall task. However, this conclusion is somewhat speculative and would need more research to support it.

Alternatively, another account as to why incidental exposure led to a superior recall for NNSs than NSs may be related to the unnaturalness of the items used. The current study used a set of adjective noun pairs that may be novel for both NSs and NNSs. However, due to NSs' greater formulaic language repertoire and greater experience with English, the

unnaturalness of these novel pairs may be more pronounced for NSs than NNSs. This might explain why recall difficulty (from incidental exposures) was more apparent for NSs than NNSs, suggesting that NSs may have needed more incidental exposures to overcome the unnaturalness of the novel pairs. While many studies on vocabulary learning indicated no difference in learning word forms incidentally between NSs and NNSs (e.g., Godfroid et al., 2018; Pellicer-Sánchez et al., 2021), this study indicated that learning new L2 collocations may put NNSs at an advantage due to their less overall collocational knowledge. In spite of NSs' form recall results, NSs maintained a faster processing of the target items than NNSs, confirming Godfroid et al.'s (2018) findings. NSs' robust processing advantage could simply reflect NS-NNS differences in proficiency, resulting in NSs' superior familiarity with individual word forms and their overall advanced reading fluency.

5.6.2 *The Effect of Condition on Processing*

In terms of the second research question, on the effect of learning condition on online processing during reading, we see a difference in performance in the TR and R-only conditions. Items that were studied in the training phase exhibited faster processing times in the subsequent reading phase than items which were not studied in the training phase for both groups (i.e., faster processing was evident in total reading time for both the phrase and word and fixation count for phrases). The absence of an effect for TR in early measures (at least for NSs) suggests that while TR items were more easily integrated into text than R items, they were not necessarily recognized more easily. Thus, the advantage of training may be attributed to post-lexical processes. The observation that the effect of training was more noticeable in the late measures suggests that intentionally-focused exposure (pre-reading) leads to easier lexical integration during incidental exposures. This processing advantage, attributed to pre-reading exposure, expands Pellicer-Sánchez et al.'s (2021) results to collocations. Pellicer-Sánchez et al.'s (2021) found a processing advantage for pre-taught

pseudowords, reflecting an easier lexical integration. The processing advantage of TR existed across all four encounters. This suggests that more encounters are needed before the items which were encountered in the reading for the first time to start to exhibit similar processing times to the items which were encountered in the pre-reading.

If both pre-reading and enhancement are considered to be attention drawing techniques, the current findings on pre-reading and Choi's (2017) finding on input-enhancement suggest that different types of attention drawing techniques may elicit different processing patterns. Choi found that typographically enhanced collocations led to increased attention, as indicated by more and longer fixations relative to unenhanced collocations. Similar to what Pellicer-Sánchez et al. (2021) concluded, this points to the fact that different attention-drawing techniques, while all beneficial, modulate attention differently. It appears that increased attention is observed when an attention-drawing technique is present during reading (e.g., underlining, bold-facing, colour). Thus, in a task simulating a meaning-focused reading activity (i.e., incidental learning), it may be that boldfacing a few segments of the text during reading promotes strategic processing (i.e., trying to understand why some items are enhanced). This could attract more attention, thus leading to longer processing and greater recall. Conversely, in pre-reading, the additional exposures provided by the training phase contributes to a processing advantage during incidental exposure. Thus, the advantage in learning gains is a function of increased familiarity due to further exposures in reading, indexed by a decrease in attention (i.e., less time is spent as items become more familiar). Further, by the end of the training session, participants should have formed initial memory traces linking the form and meaning of these novel collocations. It is likely that when these items are then encountered in reading, they are integrated into the text more easily, rendering the phrases more contextually congruous.

In addition to the difference in processing across conditions, there was a difference across item types. HT items (e.g., *deep pond*) were read faster than LT items (e.g., *tempting pastries*). Due to the high level of proficiency of NNSs in this study, NNSs followed a comparable reading pattern as NSs whereby HT items were read faster than LT. This effect was beyond the frequency of the collocation nodes – the adjectives (i.e., the nouns had similar frequencies in HT and LT while the adjectives in HT had higher frequencies than the adjectives in LT). This pattern was attested in many eye-tracking measures. Thus, it warrants further exploration.

The usage-based notion of schema can be used to account for the processing advantage of HT items compared to LT items.[12] The notion of schema refers to generic and abstracted templates that are instantiated from repeated use of occurring expressions (i.e., the adjective + noun schema generalizes/abstracts over various instances of adjective noun pairs in English). When encountering the words, *deep* (HT items) or *tempting* (LT items), both target nouns *pond* and *pastries* are unlikely to be activated for either NSs or NNSs in the initial exposures. However, both language users (NSs and NNSs) needed less time to tackle the unnaturalness of the HT non-collocate pairs items than the LT non-collocate pairs (i.e., HT was more facilitated at both word and phrase levels). Unlike LT items, HT items contain collocation nodes which are attested in the ADJ-N schema (i.e., the collocation node in HT items (*deep*) is of high type frequency since it collocates with many noun variants). Since the HT items comprise adjectives that are already licensed by the schema (i.e., more productive), generalizing the ADJ-N schema over novel instances may be more facilitative in HT items than LT items. Thus, the processing discrepancy in Item Type can be attributed to how

[12] In usage-based approaches, constructions exist in varying degrees of schematisation (i.e., the word *raisin* vs. Subject-Verb-Object-Object) (Wulff, 2019). The notion of *schema* refers to, "a cognitive representation comprising a generalization over perceived similarities among instances of use" (Kemmer & Barlow, 2000, p. xxiii)

natural a candidate the collocation node is in the schema. In other words, since the novel HT items are more likely candidates in the ADJ-N schema than novel LT items due to their high type frequencies,[13] it is likely that HT items would be more easily activated than LT items. Although entirely speculative, this could be one possible explanation why HT items elicited less processing time at the automatic word processing level (first fixation duration) and re-analysis phrase processing level (total reading time).

The fact that Item Type effects surfaced only in the online reading level, but not in the training level or offline testing level, suggests that these effects may be a function of automatic/implicit processes. They could therefore be more difficult to detect when using tasks or engaging in activities that elicit declarative knowledge and thus, can only be detected in online processing. This suggestion is also sustained in the word-level analysis, which indicated that Item Type effects emerged only at the immediate processing level (i.e., first fixation duration). However, this is merely a speculative account and further research is clearly needed to better understand what underpins this finding.

5.6.3 *The Effect of Amount of Study on Processing and Learning Gains*

The final research question considered whether the amount of time spent studying items during the training phase affects processing during reading and performance on the form recognition and recall tasks. Importantly, the amount of attention paid in training to items was indicative of processing time during reading. More specifically, longer total reading time in training predicted longer total reading time and more fixations for the phrase during reading, and longer first pass reading time in training predicted longer first pass reading time for the word level in reading. This suggests that the more attention that is paid to items in training, the greater their salience in incidental exposures, thereby attracting more

[13] Bybee (2010) notes that "the higher the type frequency the greater the productivity or likelihood that a construction will be extended to new items" (p. 67).

attention. Fixation patterns in the training phase were not indicative of performance on the testing phase for the TR condition, while they were for the T-only condition. In the T-only condition, first pass reading time in training predicted form recall for NNSs (although not for the NSs). Alternatively, the greater variability in fixation patters for the NSs during the study phase may be masking the effect in the NS group. Future research should attempt to explore the possible differential effects of pre-reading on memory performance by NSs and NNSs.

Notably, a higher fixation count in the training phase was associated with better form recognition for both NSs and NNSs, suggesting that more fixations improved recognition. Thus, it appears that number of fixations (indicating more exposures) was more beneficial than duration of fixations (indicating the amount of attention) in developing receptive knowledge of collocation form. However, the current study did not manipulate the number of exposures. Thus, such a tentative conclusion calls for more research to better isolate the effect of different eye tracking measures for different aspects of collocational learning. Although the current study did not manipulate the number of exposures, it points to similar directions to Godfroid et al.'s (2018) study which controlled for the number of exposures. Godfroid et al. found that repeated exposures but not summed total reading time was predictive of the learning of word form (form recognition).

Finally, I also considered the relationship between Summed Total Reading Time in R-only condition and learning outcomes. Summed total reading time was not indicative of performance in either form recall or form recognition, supporting the recent findings of Godfroid et al. (2018) and Pellicer-Sánchez et al. (2021). In these studies, summed/ cumulative total reading time across a number of occurrences was associated with the learning of word meaning but not the learning of word form. A possible interpretation is that different components of lexical knowledge are likely to be developed at different rates, suggesting that the learning of each component may pose different degrees of cognitive

effort. For example, developing knowledge of form may occur earlier than developing knowledge of the form-meaning connection (e.g., Mohamed, 2018). Thus, Godfroid et al. (2018) suggest that longer total reading times indicates a greater processing effort to map meaning to the word form (Godfroid et al., 2018). In light of Godfroid et al.'s (2018) conclusion, while the learning of word meaning benefits more from longer reading times, the learning of word form benefits more from repeated exposures. Thus, summed total reading time –when operationalized as a measure of attention – seems to better capture success in developing a form-meaning connection.

5.6.4 *Limitations*

A number of important limitations need to be noted. First the NNSs had a range of L1s. While care was taken to ensure the items used were not phrases in English (i.e., corpus, norming, and association database), it is possible that NNSs may have encountered these phrases in their L1s. However, given that the results from NSs and NNSs were extremely similar, this is unlikely to be the case or at least not a prevalent issue. Another limitation is related to the time interval between the T-only condition and the testing phase. After learners were exposed to T-only and TR items in the training, they engaged in the reading phase of R-only and TR items which took about 20-35 minutes, before they were tested. Testing occurred after the reading phase (of R-only and TR items), while there was a time interval between exposure and testing for the T-only items. Thus, any learning outcomes from T-only items may have been disadvantaged by such a time gap. Future replications should counteract this issue by counterbalancing the order in which treatments are given (e.g., in one list, T-only is introduced before reading and in another it is introduced after reading). The final limitation is related to operationalizing collocation type frequency (HT vs. LT); collocation type frequency was identified in this study by using the BNC (i.e., the main criteria was based on phrase frequency and MI values) and was treated dichotomously. However, a

dichotomous variable tends to reduce power (Baayen, 2010). Future research is thus suggested to identify type frequency as a continuous variable by using more sophisticated statistical tools.

5.7 Conclusion

Chapter 5 presented the second study in the thesis, which evaluated the effectiveness of different learning conditions for the learning and processing of collocations. The innovative findings of the study contribute to existing knowledge on the learning and processing of collocations. Results showed that the combination of intentional and incidental learning was more beneficial than the incidental or intentional exposures alone for learning collocations and that incidental learning was more beneficial than intentional learning. The results underscore the importance of pre-reading for developing knowledge of collocations. Pre-reading also led to processing advantage when re-encountering collocations during reading compared to when they were encountered in reading for the first time. More specifically, the processing advantage of pre-reading was attributed to additional exposures provided by pre-reading that appears to facilitate integration of collocations into the wider text. The results showed the amount of attention that learners paid to collocations in pre-reading was positively associated with the amount of attention paid to the same collocations in reading, suggesting that increased attention to the items in pre-reading may have increased their salience in later reading. In terms of the relationship between eye-movement patterns in pre-reading and the type of knowledge acquired, the results showed that longer fixation durations improved form recall (although only for NNSs) and more fixations improved form recognition (for both groups).

The next chapter presents the final study in the thesis. It aims to replicate the current findings with another type of formulaic language – opaque collocations.

Chapter 6. The Learning and Processing of Opaque Collocations

6.1 Introduction

Study 1 and Study 2 focused on the learning of completely transparent formulaic language – binomials and collocations. However, formulaic language exists on a spectrum of varying degrees of compositionality/semantic transparency. The line between transparent and opaque collocations is a gradient (e.g., ranging from very literal phrases such as *break his ankle,* to less literal phrases such as *break a habit* or *break a record,* to more idiomatic phrases such as *break a leg*). Laufer and Waldman (2011) describe the semantic transparency of collocations as relative. For example, the sense of *face* in the collocation *face a problem* is different to the typical sense of the verb, but is easier to understand than the idiom *face the music* (Laufer & Waldman, 2011).

Macis and Schmitt (2016) distinguished three categories of collocations on the basis of their semantic properties: 1) literal collocations (e.g., *powerful computer*: 'a computer with powerful performance'); 2) figurative collocations (e.g., *hot ticket*: 'a ticket that is in much demand'); and 3) duplex collocations (e.g., collocations that have both literal and figurative meanings and can also be considered shorter idiomatic phrases such as *top drawer* where the figurative meaning is 'something of the highest quality'). The individual components of figurative collocations contribute to the whole meaning of the phrase to varying degrees (e.g., *break a habit*). Despite their wide use in language, figurative collocations are rarely covered in English teaching materials (Macis & Schmitt, 2016) and have received little attention in the literature. More specifically, many studies failed to distinguish between figurative and more compositional collocations when studying collocations.

Idioms have figured prominently in research on the effect of semantic transparency on learning/processing. Research has often found that while NSs process the literal and

figurative meanings of idioms similarly, NNSs usually process the literal meaning faster than the figurative meaning. Conklin and Schmitt (2008) used self-paced reading to examine the reading patterns of idioms (e.g., *take the bull by the horns*) presented in a story context biasing either the literal meaning ('attack a problem') or the figurative meaning ('wrestle an animal'). Neither NSs nor NNSs showed any differences in reading times when an idiom was used figuratively or literally. The findings observed for L2 learners in the Conklin and Schmitt study were not consistent with Cieślicka's (2006) study. Cieślicka (2006) investigated L2 learners' processing of the literal and figurative meanings of idioms (e.g., *had cold feet*) using a lexical decision task on words associated with the idiom interpretations. Cieślicka found larger priming effects for word targets associated with the literal meanings (e.g., *toes*) than targets associated with the figurative meanings (e.g., *nervous*). The study by Siyanova-Chanturia, Conklin, and Schmitt (2011) was consistent with Cieślicka's findings. Siyanova-Chanturia, Conklin, and Schmitt (2011) used eye-tracking to investigate the processing of idioms used in contexts that either biased a figurative use (e.g., *at the end of the day*: 'eventually') or a literal use (*at the end of the day*: 'in the evening'). While NSs demonstrated no difference in the reading patterns for both literal and figurative uses, NNSs read the figurative use more slowly than the literal one.

While considerable research explored the figurativeness of idioms, very few studies have investigated the role of semantic transparency on collocations. Gyllstad and Wolter (2016) conducted a semantic judgement task (i.e., judging how meaningful and natural a phrase is) to investigate the effect of semantic transparency on the processing of collocations. They had three conditions: free combinations (i.e., fully transparent items, e.g., *write a letter*); collocations (i.e., the collocation node appeared in a restrictive and figurative sense while the noun was transparent, e.g., *draw a conclusion, run a risk*); and baseline items (e.g., *carry a car*). Collocations were processed slower compared to free combinations by

both NSs and NNSs. Since there were no differences in the critical conditions in terms of phrase frequency, the processing difficulty associated with collocations was attributed to the lower degrees of semantic transparency.

In sum, the picture is still unclear as to the role of semantic transparency for the learning of collocations. Both Webb et al. (2013), Gyllstad and Wolter (2016), and Pellicer-Sánchez (2017) called for more research to investigate the role of semantic transparency in the learning of collocations.

Less transparent collocations are more problematic for L2 learners than their more transparent counterparts (Macis & Schmitt, 2017). This highlights the role of transparency in the learning of collocations. However, previous research has dealt primarily with the learning of transparent collocations, with semantic transparency rarely being taken into account. As reported in Chapter 5, previous studies on collocations have either focused on fully transparent ones or failed to control for varying degrees of transparency in the collocations under investigation (e.g., Laufer & Girsai, 2008; Peters, 2009; Webb & Kagimoto, 2009; Zhang, 2017). Thus, it remains unclear whether patterns of results for the learning of transparent collocations, also extends to collocations with less transparency.

Schmitt and Carter (2004) argue that the process of learning transparent and opaque phrases is different. In comparison with transparent collocations (*e.g., broke his ankle*), the learning of opaque collocations (e.g., *break a habit*) does not depend on knowledge of component words (i.e., the verb in *break a habit* carries a figurative meaning related to 'bringing something to an end'). Learners could easily understand the meaning of transparent collocations if they know the meanings of the word parts. However, the meaning of the opaque phrase *break a habit/the rules* is not as easily decoded, since the word *break* carries a figurative meaning. Thus, it is possible that the learning of less transparent collocations may require more explicit attention from learners than the more transparent collocations.

6.2 The Present Study

The aim of Study 3 was to investigate the generalisability of Study 2 to the learning of less transparent collocations by controlling the semantic transparency of the items used. More specifically, the current study investigated whether the processing and learning of opaque collocations would benefit from the same treatments that were beneficial for transparent collocations as observed in Study 2. Study 3 had the same research design and methodology as Study 2; only a different type of collocation is investigated. More specifically, Study 2 used novel collocations that were fully transparent (e.g., *political caricaturists*), while Study 3 used novel collocations of varying degrees of transparency (e.g., *open appetite* meaning 'big appetite'). Study 3 aimed to address the following research questions:

1. Does Training plus Reading lead to more learning gains than Reading-only or Training-only in terms of form recognition, form recall, and meaning recognition?
2. What is the effect of each learning condition (Training plus Reading, Reading-only) on online processing during reading?
3. What is the impact of reading times (i.e., amount of study) in the training phase on a) reading in the reading phase and b) performance in the testing phase?

6.3 Methods

6.3.1 *Participants*

Ninety-four participants were recruited for this study (experimental group: NS = 35; NNS = 39; control group: NS = 20). The experimental group took part in all the three phases of the experiment – training, reading, and testing, while the control group took part in only the testing phase. Participants received course credit or were paid £7 for their participation. The participants were mostly undergraduate students (88.30 %), but few were postgraduate students (11.70 %). The NNS came from a variety of L1 backgrounds (Bulgarian =1, Chinese

= 19, Czech =1, Greek = 3, German =1, Gujarati = 2, Korean =1, Malay =1, Portuguese = 2, Romanian =1, Russian =1, Spanish = 4, Vietnamese = 2). Participants' proficiency in English was assessed using the Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012). Participants were asked to complete a language background questionnaire. The language background questionnaire asked participants to provide information about their age, the age they began studying English, their length of stay in an English-speaking country, estimates of their proficiency in reading, writing, listening, speaking, and understanding on a 7-point scale (1 = very low; 7 = native). Table 6.1 summarises the demographic and language proficiency data.

Table 6.1 Means and Standard Deviations of Demographic Data and Self-ratings of Proficiency on a 7-Point Scale (1 = very low; 7 = native)

	NS	**NNS**
Age	20.66 (4.31)	21.89 (2.98)
Age of English acquisition	0.11 (0.52)	5.84 (3.31)
Time lived in an English-speaking country (in years)	19.20 (5.94)	2.80 (6.12)
LexTALE score (%)	92.42 (7.36)	74.34 (12.72)
Self-rating (1 to 7)		
Overall English proficiency	7 (0.00)	5.32 (0.83)
Proficiency in speaking	7 (0.00)	5.25 (0.88)
Proficiency in understanding	7 (0.00)	5.56 (0.94)
Proficiency in reading	6.94 (0.23)	5.69 (0.97)
Proficiency in writing	7.00 (0.00)	5.12 (1.15)

6.3.2 *Materials*

Twenty-four novel semi-transparent adjective-noun collocations were created, having the following characteristics. First, the novel collocations had varying degrees of transparency. The meaning of a collocation was not easily deducible from its constituents

(e.g., *choking heat* meaning 'extreme heat'; *open appetite* meaning 'big appetite'; *scented legacy* meaning 'great legacy'; *stone mentality* meaning 'narrow-minded'). Second, the figurative sense of the novel collocations was carried by the adjectives. Third, while none of the novel collocations existed in English (as confirmed by the zero occurrences in the BNC), they existed in Arabic – a language that is unrelated to English. Drawing on collocations existing in another language should increase their authenticity. It is important to note that none of the NNSs spoke Arabic or another Semitic language, so the Arabic collocations should be unknown to them, similar to the NSs. Fourth, the items were controlled for their backward transitional probability (i.e., the probability that a noun would be preceded by a particular adjective), such that none of the items had a noun that was predicted by an adjective in an existing collocation. Finally, none of the word pairs in the collocations were semantic associates, as confirmed by the University of South Florida database for association norms (Nelson et al., 1998). A complete list of the items is presented in Appendix 6A.

To ensure that the collocations used in this study were indeed novel in English and to estimate the degree of transparency of the phrases, three rating-based norming studies were carried out. Participants who took part in the norming studies did not take part in main study. The ratings of the novel collocations were considered as covariates in the analyses to control for any minor differences between the items in terms of subjective familiarity and transparency.

The first norming study was carried out to confirm that the items were unknown in English. Fourteen NS participants rated the familiarity of the phrases on a scale from 1 (I have never heard/used this phrase) to 5 (I have very frequently heard/used this phrase). A set of English collocations (e.g., *short time*) of different frequencies were included as filler phrases in the norming study for comparison and to ensure use of the full rating scale. An

independent sample *t*-test confirmed that there was a significant difference between the English collocations ($M = 4.66$) and the novel ones ($M = 1.95$), $t = 19.47$, $p < .001$.

The second norming study was carried out to assess the transparency of the phrases in isolation. Twelve English native speakers were asked to rate how easy it is to guess the intended meaning of the phrase (e.g., *pasted accusation* means 'false accusation') on a scale from 1 to 5 (1= very difficult to guess, 5 = very easy to guess). The transparent novel collocations from Study 2 were used a filler phrases for comparison and to ensure use of the full rating scale. An independent sample *t*-test confirmed that the items used in the current study ($M = 3.15$) were significantly less transparent than the items used in Study 2 ($M = 4.76$), $t = 11.41$, $p < .001$.

The third norming study was carried out to assess the transparency of the phrases when used in sentences. Sixty-six English native speakers were asked to rate how easy it is to guess the intended meaning of the phrases based on the sentences in which they were presented on a scale from 1 to 5 (1= very difficult to guess, 5 = very easy to guess). In this study, the phrases were presented in the same sentences in which they appeared in the actual experiment (e.g., "*The white coup resulted in a rapid change of the government without any violence*". Based on this sentence, how easy is it to guess that *white coup* means 'peaceful coup'?). The purpose of norming items for transparency in the experimental sentences was to account for variation in the contextual clues potentially aiding their interpretation.

6.3.3 *Treatment Conditions*

The design of this study was the same as the one adopted for Study 2. It was a within-subject design such that each participant (from the experimental group) encountered all of the treatment conditions –Training only (T-only), Reading only (R-only), and Training plus Reading (TR). Items in T-only appeared along with their meanings in a decontextualised context and participants were explicitly instructed to familiarise themselves with the phrases

and their meanings. Items in R-only were presented four times in different sentences (i.e., each sentence appeared non-sequentially in a different block). Items in TR were taught and appeared in the reading as well. The presentation of items across different learning conditions was counterbalanced across different lists (i.e., the phrase *shy critique* appeared in R-only in List 1, in T-only in List 2, and in TR in List 3). There were eight items in each learning condition per list.

Similar to Study 2, Study 3 had three phases: training, reading, and testing. However, the main difference between the two studies was in the way collocations were presented in the training phase. In the training phase in Study 3, T-only and TR items appeared with their meanings on the screen and participants were asked to learn them. In Study 2, only the collocations were presented without their meanings because they were all transparent. Thus, it was expected that the training phase (i.e., pre-reading) in Study 3 would encourage greater engagement with the items than the reading phase. Figure 6.1 illustrates how the collocations were presented in the training phase in Study 3.

Figure 6.1 Display of Items in the Training Phase

```
hot applause: clapping loudly
dry reception: unwelcoming
shaky confidence: low self-esteem
cutting proof: conclusive proof
paralyzed tongue: inability to speak
pasted accusation: false accusation
shiny poet: outstanding poet
scented legacy: great legacy
```

In the reading phase, R-only and TR items occurred in four sentences with four exposures per an item across four blocks (see Appendix 6B). The testing phase occurred at the end of the experiment to assess participants' collocational knowledge.

6.3.4 *Measurement Instruments*

Online processing/attention during the training and reading phase was measured using eye-tracking. Individual words and the whole phrase were both taken into account in the analysis. The analysis considered the same eye-tracking measures as in Study 2: likelihood of skipping, first fixation duration, first pass reading time, total reading time, and fixation count.

Participants' collocational knowledge for the items was measured by three immediate tests that assessed form recall, form recognition, and meaning recognition. The tasks are presented in in Appendix 6C. In the form recall task, the adjective from each phrase was given as a prompt and participants were asked to provide a one-word completion (e.g., *open* _____). The form recognition task provided the adjective and a set of semantically similar distractor choices to select from (e.g., for the phrase *open appetite:* the choices were *desire, craving, appetite,* and *hunger*). The distractors were selected if 1) they were not collocations in English and 2) they were semantically associated with the correct answer. The meaning recognition task was a multiple-choice task that assessed participants' ability to identify the correct meaning of the phrase (e.g., the choices provided for *open appetite* were *sudden feeling of wanting something, strong willingness,* and *big appetite*). The option 'I don't know' was provided in the form recognition and meaning recognition tasks to minimize guessing.

6.3.5 *Procedure*

The study was carried out in accordance with the research ethics procedures at the University of Nottingham, which provided ethical approval. The study was conducted in a quiet psycholinguistics laboratory at the University of Nottingham. Upon arrival at the lab,

the study was explained in general terms and participants signed an informed consent form. Participants were randomly assigned to either the control group (i.e., did only the testing part of the experiment) or one of the three lists for the experimental group. Participants in the experimental group took part in all the three phases (training, reading, and testing). The three phases of the experiment were completed in one session lasting approximately 45 minutes. Eye movements were monitored in the training and reading phases using an EyeLink 1000 Plus desktop-mounted eye-tracker (SR Research, Canada), with data being sampled at 1000Hz. Participants were seated in front of a computer monitor where they placed their head against a chin- and forehead-rest to minimise their head movements while reading. Text was presented in black Courtier New font, size 14, double-spaced. Participants performed a nine-point calibration prior to the start of the experiment and between every reading block. Each screen was followed by a drift correction and recalibration was carried out as needed.

The eye-tracking experiment started with the training phase. In the training phase, participants were instructed to familiarize themselves with the phrases and their meanings, as they would see them later used in sentences. They were told that there was no time constraint and that they should press the space bar whenever they felt like they were done. Then, eight phrases were presented with their meanings. Once participants pressed the space bar, another set of eight phrases appeared on the screen.

The reading phase was preceded by four practice sentences followed by comprehension questions. The reading phase had four blocks, with 16 sentences in each block. There was a one exposure of an item per block. Each sentence was presented on a separate screen. Participants were instructed to read for comprehension and to press the spacebar once they finished reading. One third of the sentences were followed by a yes-no comprehension question to ensure that participants were reading for comprehension. The questions were unrelated to the novel collocations and asked about other details in the context

of the sentences. Participants' accuracy on the comprehension questions confirmed their understanding of the sentences (the mean accuracy was 94.00% for NSs and 87.03% for NNSs). Once the reading phase was completed, participants filled out the biographical questionnaire and did the LexTALE task.

The final part of the experiment was the testing phase. In this phase, all of the participants performed the form recall, form recognition, and meaning recognition tasks. The tasks were administered in this order to limit any order effects (i.e., the possibility of one task having an effect on performance on another task).

6.3.6 *Analysis*

Data from one NS was excluded from the analysis because the background questionnaire data showed that the participant spoke Arabic as a second language. Because the collocations in the experiment exist in Arabic, it may have impacted the participant's performance. Following Carrol and Conklin (2020), the data were cleaned using the Eyelink Data Viewer's four-stage procedure: 1.) very short fixations of < 40 ms were merged with neighbouring fixations within 0.5 degrees; 2.) those = 40 ms were merged with those within 1.25 degrees; 3.) any three consecutive fixations of < 140 ms were merged into one; and 4.) any fixations shorter than 80 ms or longer than 800 ms were deleted. Each trial for each participant was visually inspected to remove any cases of trackloss or cases in which the entire collocation was skipped. This left 6400 data points for analysis after removing 0.22% of the data.

The analyses considered both the phrases (the whole adjective-noun phrase) and the final words (the noun) for the reading phase, but only the whole phrase for the training phase. For the word-level analyses, the following measures were examined: likelihood of skipping, first fixation duration, first pass reading time, total reading time and fixation count. For the

phrase-level analyses, total reading and fixation count were examined. Word-level reading time analyses were carried out on non-skipped items.

The experimental data was analysed by fitting a series of linear mixed effects models using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in R, version, 3.6.1, (R Core Team, 2019), with the *p*-values estimated by the *lmerTest* package (Kuznetsova *et al.* 2015). For each model, the random slopes and intercepts for subjects and items were included as random effects (Barr, Levy, Scheepers, & Tily, 2013). For the offline data (form recall, form recognition, and meaning recognition task results), generalized linear mixed effects models were fitted with *binomial* distributions (Jaeger, 2008) (since the outcome variable was binary: 1 = correct, 0 = incorrect). Each eye-tracking measure was considered in its own model. For the likelihood of skipping, generalized logistic mixed effects models were fitted, and for fixation count generalized linear models with *poisson* distributions were fitted. Interactions and post-hoc comparisons were inspected using the *emmeans* package (Lenth, 2019) and plotted using the *effects* package (Fox, 2003). *P*-values for all post-hoc analyses were adjusted using Bonferroni correction. Multicollinearity of each model was assessed using the package *performance* (Lüdecke et al., 2020). The VIF values for the models indicated no issues with multicollinearity among the predictor variables.

Model predictors were added one by one in stepwise manner in different models. Likelihood ratio tests with the package *afex* (Singmann et al., 2019) was used to see whether inclusion of each predictor improved the fit. Predictors and their interactions were kept in the final model only if they significantly improved its fit. The following main predictors were considered in each model: Group (NS or NNS), Condition (treatment: T-only, R-only, or TR), Repetition (first, second, third or fourth).The following covariates were also considered: Word Transparency (transparency ratings for the phrases in isolation), Context Transparency (transparency ratings of the phrases based on the sentences in which they were presented),

Familiarity (i.e., subjective familiarity norming), LexTALE score,[14] Phrase and Word Length (in letters), Trial Number (to control for practice or fatigue effects), Frequency (frequency of the second word),[15] and Summed Frequency (the sum of frequencies of the first word + second word of the phrase). Word frequency was converted to Zipf scale values (van Heuven, Mandera, Keuleers and Brysbaert, 2014). All the other continuous predictors (for reading time variables) were log-transformed to reduce skewness and to ensure that variables had the same scale. Familiarity, Frequency, Summed Frequency, did not significantly improve the fit for any of the models.

6.4 Results

The results are reported in relation to each of the research questions.

RQ1: Does TR lead to more learning gains than R-only or T-only in terms of form recognition, form recall, and meaning recognition?

To assess the effect of the treatments, accuracy on the form recall, recognition, and meaning recognition tasks for the experimental groups (NS and NNS) were compared to the control group (NS). Table 6.2 summarises the response accuracy of the groups. Table 6.3 summarises the model outcome assessing the effect of training/exposure. Table 6.3 shows that response accuracy for form recognition, and meaning recognition were significantly greater in both the NS and NNS groups compared to the control group. Follow up post-hoc analysis showed no significant difference between NSs and NNSs in either task (*p's* > .05). However, response accuracies for form recall did not differ for the NS and NNS groups relative to the control group.

[14] Because there was a correlation between LexTALE scores and the self-rated proficiency (r= .40, p < .01), the LexTALE scores were used as they provide a more objective measure of proficiency.
[15] In the word-level analysis, only the frequency of the second word (target word) was considered. Phrasal frequency was not considered as the phrases were novel and did not occur in the BNC.

In subsequent analyses, each of the treatment conditions (T-only, R-only, and TR) were considered separately. For form recall, there was no difference between NSs, NNSs, and the control group in any of the learning conditions (T-only, R-only, and TR). In form recognition and meaning recognition, NSs and NNSs had significantly more correct responses than the control group in all of the learning conditions (T-only, R-only, and TR), with no differences between the NS and the NNS groups.

Table 6.2 Mean Collocational Gains and Standard Deviation for the Trained/Exposed (NS and NNS) and the Untrained/Unexposed Control Group

	Form Recall		***Form Recognition***		***Meaning Recognition***	
Predictors	Means	SD	Means	SD	Means	SD
Control Group	0.00	0.00	7.70	2.45	15.30	2.26
NNS Group	7.59	3.82	18.56	3.12	21.00	3.26
NS Group	7.65	3.96	18.50	3.34	21.50	3.22

Note. Maximum score is 24.

Table 6.3 Overall Effect of Treatment on Collocational Gains

	Form Recognition			**Form Recall**			**Meaning Recognition**		
Predictors	*Log-Odds*	*std. Error*	*p*	*Log-Odds*	*std. Error*	*p*	*Log-Odds*	*std. Error*	*p*
(Intercept)	-0.87	0.26	**0.001**	-17.89	13.35	0.180	0.71	0.38	0.058
Group [NS]	2.97	0.29	**<0.001**	17.40	13.35	0.192	2.83	0.42	**<0.001**
Group [NNS]	2.94	0.28	**<0.001**	17.38	13.35	0.193	2.32	0.40	**<0.001**

Next, analyses were carried out on only the experimental groups (NSs and NNSs) to examine the effectiveness of each of the learning conditions on participants' learning gains. As shown in Tables 6.4 and 6.5, there were no differences between NSs and NNSs in their performance on any of the tasks (form recall, form recognition, meaning recognition). There

was a main effect of Condition across all tasks such that TR and R-only conditions yielded significantly more learning gains than T-only, and TR yielded significantly more learning gains than R-only. Accuracy scores per each condition are reported in Table 6.6. The effect of learning conditions can be summarised as TR > R-only > T-only. There was no interaction between Group and Condition in any of the tasks, suggesting that the pattern TR > R-only > T-only was maintained in both groups. In all of the tasks (form recall, form recognition, and meaning recognition), the inclusion of Familiarity or Word Transparency did not improve the fit of the models nor their interactions.

Table 6.4 Summary of the Mean Collocation Gains per Condition with Standard Deviations in Parentheses

Types of knowledge	**NS**			**NNS**		
	T-only	R-only	TR	T-only	R-only	TR
Form recall	0.61 (0.77)	3.09 (1.96)	3.94 (1.85)	0.60 (0.84)	2.92 (1.83)	4.08 (2.20)
Form recognition	4.41 (2.15)	6.91 (1.01)	7.18 (1.02)	4.72 (1.65)	6.79 (1.25)	7.05 (1.32)
Meaning recognition	6.71 (1.74)	7.12 (1.30)	7.65 (0.72)	6.72 (1.74)	6.77 (1.19)	7.49 (1.04)

Note. Maximum score is 8.

Table 6.5 Effects of Condition on Collocational Knowledge

	Form Recall			**Form Recognition**			**Meaning Recognition**		
Predictors	*Odds Ratios*	*std. Error*	*p*	*Odds Ratios*	*std. Error*	*p*	*Odds Ratios*	*std. Error*	*p*
(Intercept)	0.03	0.36	**<0.001**	1.54	0.27	0.109	19.25	0.41	**<0.001**
Group [NNS]	0.98	0.30	0.955	0.96	0.29	0.884	0.58	0.43	0.202
Condition [R-only]	16.10	0.19	**<0.001**	6.84	0.12	**<0.001**	1.42	0.15	**0.020**
Condition [TR]	35.18	0.19	**<0.001**	9.54	0.12	**<0.001**	5.01	0.16	**<0.001**

Random Effects			
σ^2	3.29	3.29	3.29
τ_{00}	1.56 $_{subject}$	1.36 $_{subject}$	2.91 $_{subject}$
	1.21 $_{Item}$	0.50 $_{Item}$	1.15 $_{Item}$
ICC	0.46	0.36	0.55

Table 6.6 Summary of The Mean Accuracy Scores per Condition with Standard Deviations in Parentheses

Condition	**Form recall**	**Form recognition**	**Meaning Recognition**
T-only	0.60 (0.80)	4.58 (1.90)	6.71 (1.74)
R-only	3.00 (1.90)	6.85 (1.14)	6.93 (1.25)
TR	4.01 (2.05)	7.11 (1.19)	7.56 (0.90)

Note. Maximum score is 8.

RQ2: What is the effect of each learning condition (TR, R-only) on online processing during reading?

Eye-tracking measures were analysed to examine the effect of Condition (R-only vs. TR) on processing during the reading phase. Each eye-tracking measure was considered in its own model. Table 6.7 summarises mean phrase- and word-level measures for both groups across different conditions and in different repetitions. Model outputs for the whole phrase are reported in Table 6.8. Total reading time and fixation count had consistent patterns. NNSs made more and longer fixations than NSs. The TR condition elicited shorter total reading times and fewer fixations for both NSs and NNSs than the R-only condition. Each repetition resulted in shorter total reading times and fewer fixations. Post-hoc analysis showed that all of the individual comparisons between each repetition was significant for total reading time and fixation count ($p < .05$). There was an interaction between Repetition and Condition in

the total reading time model. The interaction suggests that effect of repetition was more robust in the R-only items compared to the TR items in the first two repetitions. The difference in total reading time between the first and second and between the first and third repetitions was larger for the untrained items (R-only) than the trained items (TR). There was an effect of Context Transparency, by which increased transparency led to shorter total reading times and fewer fixations.

Table 6.7 Summary of Reading Patterns of Phrases and Final Words for NS and NNS in the Reading Phase across Different Conditions and Repetitions (Rep)

		R-only				**TR**			
		Rep1	Rep2	Rep3	Rep4	Rep1	Rep2	Rep3	Rep4
	NS								
Phrase level	Total RT	999.86 (593.43)	777.93 (519.35)	616.43 (369.33)	565.94 (340)	921.49 (610.02)	707.36 (463.47)	591.67 (327.40)	532.09 (326.36)
	Fixation count	4.51 (2.44)	3.58 (2.07)	2.96 (1.64)	2.74 (1.48)	4.09 (2.41)	3.31 (1.92)	2.90 (1.49)	2.58 (1.28)
Word level	Skipping rate	0.10 (0.30)	0.18 (0.38)	0.17 (0.38)	0.18 (0.39)	0.15 (0.36)	0.18 (0.40)	0.13 (0.33)	0.17 (0.38)
	First Fixation RT	222.49 (83.23)	224.91 (89.94)	208.64 (74.37)	211.28 (78.47)	215.17 (80.37)	211.44 (75.08)	212.11 (79.58)	209.52 (75.91)
	First Pass RT	262.59 (124.84)	262.99 (124.95)	248.92 (115.40)	251.08 (122.32)	249.87 (122.01)	240.53 (103.19)	231.00 (93.78)	232.96 (101.87)
	Total RT	517.77 (319.30)	455.05 (326.14)	372.03 (258.62)	335.00 (193.11)	489.84 (338.72)	401.52 (286.75)	324.53 (187.94)	307.77 (191.00)
	NNS								
Phrase level	Total RT	1757.70 (1164.13)	1272.19 (916.32)	1016.66 (631.03)	935.16 (565.72)	1501.31 (1109.61)	1136.35 (749.68)	992.63 (668.33)	808.36 (490.01)
	Fixation count	7.07 (4.50)	5.32 (3.59)	4.36 (2.75)	3.91 (2.18)	6.05 (4.13)	4.73 (2.90)	4.23 (2.76)	3.50 (2.13)
Word level	Skipping rate	0.10 (0.31)	0.09 (0.30)	0.86 (0.28)	0.90 (0.28)	0.09 (0.30)	0.14 (0.34)	0.10 (0.31)	0.16 (0.36)
	First Fixation RT	270.47 (112.50)	264.30 (101.10)	256.07 (92.08)	256.08 (93.14)	268.63 (114.51)	264.00 (102.21)	260.38 (112.07)	248.50 (101.95)
	First Pass RT	358.96 (188.57)	342.90 (179.94)	320.06 (151.51)	311.33 (140.68)	375.38 (230.36)	328.02 (186.29)	321.31 (166.05)	307.12 (159.30)
	Total RT	912.48 (732.15)	695.99 (524.96)	544.79 (380.41)	489.19 (319.38)	842.77 (630.38)	606.82 (439.41)	544.31 (385.51)	459.31 (311.71)

Note. Duration measures are reported in milliseconds. Skipping rate is reported as a probability. Values in brackets are standard deviations.

Table 6.8 Effect of Condition and Repetition on Reading Time and Fixation Count at the Phrase Level

	log(Total Reading Time)				**Fixation Count**			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*	*Log-Mean*	*std. Error*	*Statistic*	*p*
(Intercept)	6.79	0.38	17.65	<**0.001**	1.48	0.38	3.87	<**0.001**
Group [NNS]	0.47	0.07	6.34	<**0.001**	0.36	0.07	4.95	<**0.001**
Rep [2]	-0.18	0.04	-4.41	<**0.001**	-0.12	0.03	-3.86	<**0.001**
Rep [3]	-0.31	0.05	-6.00	<**0.001**	-0.21	0.04	-4.67	<**0.001**
Rep [4]	-0.35	0.06	-5.76	<**0.001**	-0.30	0.05	-5.50	<**0.001**
Condition [TR]	-0.14	0.03	-5.16	<**0.001**	-0.09	0.01	-6.36	<**0.001**
Trial Number [log]	-0.14	0.04	-4.05	<**0.001**	-0.15	0.03	-4.69	<**0.001**
Length (Phrase)[log]	0.51	0.14	3.78	<**0.001**	0.49	0.14	3.61	<**0.001**
Context Transparency [log]	-0.79	0.14	-5.56	<**0.001**	-0.71	0.15	-4.73	<**0.001**
Rep [2] * Condition [TR]	0.03	0.04	0.88	0.381				
Rep [3] * Condition [TR]	0.12	0.04	2.94	**0.003**				
Rep [4] * Condition [TR]	0.03	0.04	0.78	0.437				
Random Effects								
σ^2	0.23				0.23			
τ_{00}	0.10_{subject}				0.09_{subject}			
	0.01_{Item}				0.01_{Item}			
ICC	0.32				0.30			

Model outputs for the analysis of the noun from the adjective-noun collocations are reported in Table 6.9. Analysis of skipping rate showed that NSs were more likely to skip phrase-final words than NNSs. Analysis of skipping rate demonstrated a significant interaction between Group and Condition. The interaction, plotted in Figure 6.2, indicates

that the effect of Condition on skipping was different across groups. A post-hoc analysis exploring this interaction showed that while NSs were more likely to skip final words in R-only condition than NNSs ($p < .05$), there was no difference in the skipping rate between the groups in the TR condition ($p > .05$). The analyses also showed that NNSs had more skipping in TR condition than in R-only condition ($p < .05$), while NSs had similar skipping rates across the conditions ($p > .05$). There was an effect of Repetition on skipping rate, with the likelihood of skipping phrase-final words increasing with every additional exposure. The difference in skipping rate between the second and the third repetitions was the only comparison that did *not* reach significance.

Analyses of durational measures (total reading time, first pass reading time and first fixation duration) showed that NSs had shorter fixations than NNSs. While the TR condition resulted in shorter total reading time and shorter first pass reading time, the effect of Condition on first fixation duration was not significant. The absence of an interaction between Condition and Repetition in all measures indicated that a faciliatory effect of Condition was maintained across all repetitions. There was an effect of Repetition, whereby reading times for all measures significantly decreased as the number of exposures increased. Post-hoc analyses showed that the individual comparisons between repetitions was significant ($p < .05$) for all measures (except 3 vs. 4 comparison in first fixation duration). The analysis of total reading time showed that Repetition interacted with Group. The interaction indicated that while total reading time decreased with each repetition for both groups, the difference in total reading time between first repetition and the second repetition was lager in the NNS group compared to the NS group. Context Transparency had an effect on total reading time and first pass reading time, with more transparent items being read faster. LexTALE had a significant effect, with higher proficiency resulting in shorter total reading times and shorter first pass reading times.

Figure 6.2 Interaction between Condition and Group on Likelihood of Skipping (skipped = 1, non-skipped = 0)

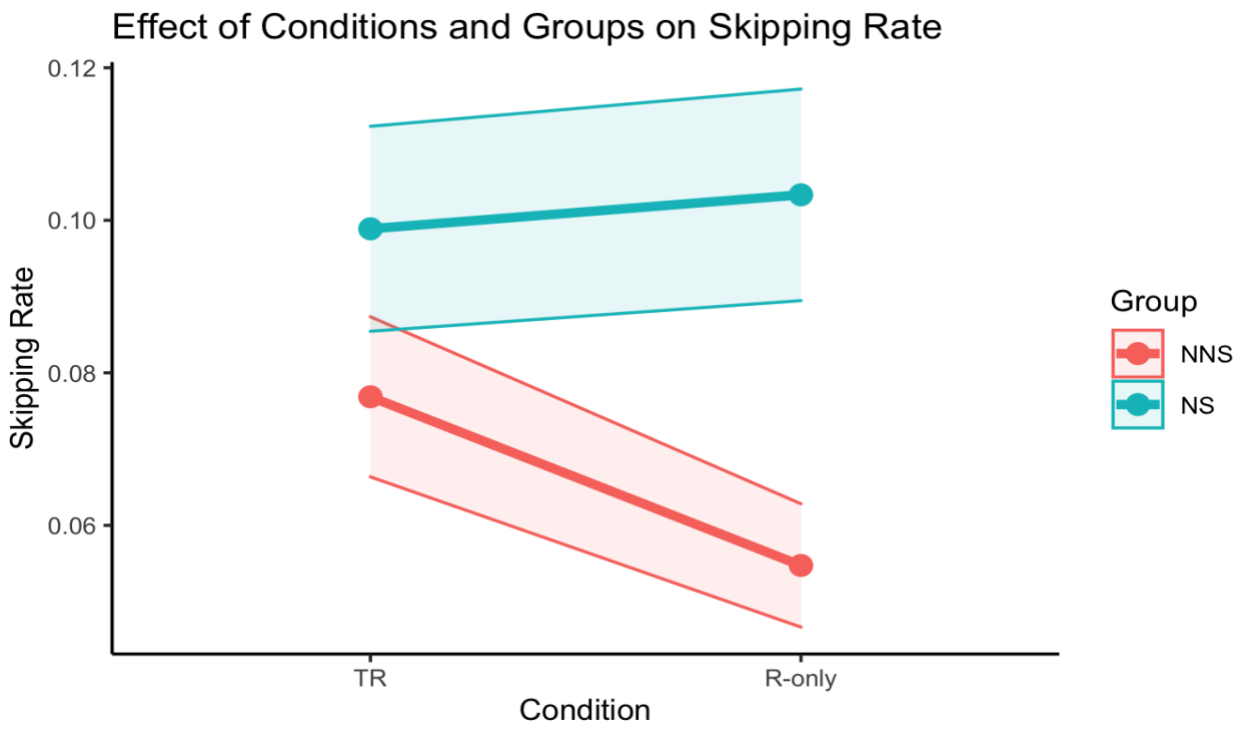


Table 6.9 Omnibus Linear Mixed Effects Model Output for Eye-tracking Measurements Used for Final Word

	Skipping Rate				**Total Reading Time**				**First Pass RT**				**First Fixation Duration**			
Predictors	*Log-Odds*	*std. Error*	*Statistic*	*p*	*Estimates*	*std. Error*	*Statistic*	*p*	*Estimates*	*std. Error*	*Statistic*	*p*	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	4.25	0.72	5.88	**<0.001**	10.08	0.99	10.16	**<0.001**	6.70	0.58	11.56	**<0.001**	5.06	0.08	66.06	**<0.001**
Group [NNS]	-0.69	0.19	-3.64	**<0.001**	0.35	0.08	4.25	**<0.001**	0.20	0.04	4.39	**<0.001**	0.19	0.03	6.30	**<0.001**
Condition [TR]	-0.05	0.12	-0.40	0.692	-0.08	0.02	-5.11	**<0.001**	-0.03	0.01	-2.66	**0.008**	-0.02	0.01	-1.64	0.101
Rep [2]	1.17	0.24	4.99	**<0.001**	-0.10	0.05	-2.00	**0.046**	-0.14	0.03	-4.29	**<0.001**	-0.09	0.03	-3.27	**0.001**
Rep [3]	1.31	0.33	4.03	**<0.001**	-0.20	0.06	-3.31	**0.001**	-0.22	0.04	-4.93	**<0.001**	-0.16	0.04	-4.37	**<0.001**
Rep [4]	1.93	0.39	4.94	**<0.001**	-0.27	0.07	-3.76	**<0.001**	-0.28	0.05	-5.23	**<0.001**	-0.20	0.04	-4.54	**<0.001**
Trial Number [log]	-1.02	0.23	-4.37	**<0.001**	-0.09	0.04	-2.11	**0.035**	0.12	0.03	3.48	**0.001**	0.10	0.03	3.74	**<0.001**
Length (Word) [log]	-1.96	0.20	-9.95	**<0.001**	0.36	0.06	6.36	**<0.001**	0.21	0.04	5.35	**<0.001**				
Group [NNS] * Condition [TR]	0.41	0.18	2.26	**0.024**												
Context Transparency [log]					-0.82	0.15	-5.39	**<0.001**	-0.30	0.11	-2.66	**0.008**				
LexTALE [log]					-0.77	0.21	-3.60	**<0.001**	-0.34	0.12	-2.80	**0.005**				
Group [NNS] * Rep [2]					-0.14	0.05	-2.88	**0.004**								

Group [NNS] * Rep [3]		-0.13	0.05	-2.87	**0.004**			
Group [NNS] * Rep [4]		-0.17	0.05	-3.62	**<0.001**			
Random Effects								
σ^2	3.29	0.28				0.17	0.12	
τ_{00}	0.32 $_{subject}$	0.06 $_{subject}$				0.02 $_{subject}$	0.01 $_{subject}$	
	0.06 $_{Item}$	0.01 $_{Item}$				0.00 $_{Item}$	0.00 $_{Item}$	
ICC	0.10	0.20				0.11	0.12	

RQ3a: What is the impact of reading times (i.e., amount of study) in the training phase on reading in the reading phase?

This analysis examines the relationship between the amount of time spent studying the collocations in the training phase and processing in the reading phase. Table 6.10 presents the descriptive statistics for participants' performance in training and Table 6.11 presents the model output for the analyses. Analyses of total reading time and fixation count showed that NSs and NNSs had similar looking patterns for the phrases in training and Familiarity had no effect on 'studying' time. Word Transparency was significant, suggesting that more transparent items elicited shorter total reading times and fewer fixations for both NSs and NNSs in the 'study' phase.

Table 6.10 Summary of Studying Patterns of Phrases in the Training Phase for NS and NNS

	T-only		**TR**	
	NS	NNS	NS	NNS
Total reading time (ms)	3788.64 (2376.60)	4999.38 (3660.91)	3888.28 (2501.19)	5189.05 (3700.09)
Range	852 - 14596	450 - 23336	336 - 16300	526 - 25807
Fixation count	16.14 (9.46)	19.82 (14.11)	16.31 (10.17)	20.68 (14.13)
Range	4 - 73	2 - 90	2 -76	3 - 100

Table 6.11 Model Outcome for Eye-tracking Measurements in the Training Phase

	Total RT at Training (log)				**Fixation Count at Training**			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*	*Log-Mean*	*std. Error*	*Statistic*	*p*
(Intercept)	7.91	0.26	30.76	**<0.001**	2.58	0.23	11.18	**<0.001**
Group [NNS]	0.24	0.13	1.92	0.055	0.19	0.12	1.63	0.103
Familiarity	0.09	0.07	1.28	0.201	0.11	0.06	1.70	0.088

Word Transparency	-0.15	0.06	-2.46	**0.014**	-0.11	0.05	-2.18	**0.029**
Length (Phrase)	0.03	0.02	1.87	0.062	0.02	0.02	1.16	0.247
Random Effects								
σ^2	0.14				0.06			
τ_{00}	0.27 $_{subject}$				0.24 $_{subject}$			
	0.02 $_{Item}$				0.02 $_{Item}$			
ICC	0.68				0.81			

The analyses were then carried out on the TR items to examine the effect of studying times (in the training phase) on reading behaviour (in the reading phase). The amount of time spent studying the phrases in the training (total reading time, first pass reading time, and fixation count) were treated as predictors of the reading patterns in the reading phase. Table 6.12 presents the model outcome for analysing the effect of training on the reading patterns for all phrase-level measures. Analyses of total reading time and fixation count showed that NSs had shorter and fewer fixations than NNSs. The effect of total reading time in the training phase on total reading time in the reading phase interacted with Group. As plotted in Figure 6.3, longer total reading time in the training phase was associated with longer total reading time in the reading phase but only for the NNS group. There was also an interaction between the amount of time spent studying the items in training and Repetition in the reading phase. This interaction emerged in both total reading time and fixation count. As demonstrated in Figures 6.4 and 6.5, the interactions suggest that the effect of studying times in training (total reading time and fixation count from training) started to decline with more exposures. Longer total reading times and more fixations in trainings were associated with longer total reading times and more fixations in the reading phase but only in the initial exposures (the first and second). The effect of training on the duration and number of fixations in the reading phase disappeared by the third repetition.

Analyses of the effect of training on reading patterns at the word-level is presented in Table 6.13. Eye-tracking measures from the training session that improved the model fit were first pass reading time and total reading time. First pass reading time during training improved the model for skipping rates in reading: longer first pass reading times in the training phase led to less skipping in subsequent reading. Total reading time during training improved the model for total reading time in reading, with greater total reading time in the training phase leading to marginally longer total reading time in the reading phase.

Replicating the pattern from above (RQ2), effects of Repetition remained robust in the reading phase (for both the whole phrase and the final word) even after accounting for looking patterns in the training phrase. Context Transparency and LexTALE were significant, with increased proficiency and greater transparency resulting in shorter total reading time (for the phrase and word), fewer fixations (for the phrase), and shorter first pass reading times (for the word).

Table 6.12 Model Output for the Effects of Training on Reading Behavior at the Phrase Level

	Total Reading Time (log)			**Fixation Count**		
Predictors	*Estimates*	*std. Error*	*p*	*Incidence Rate Ratios*	*std. Error*	*p*
(Intercept)	9.89	1.22	**<0.001**	133.96	1.15	**<0.001**
Group [NNS]	-0.68	0.41	0.098	1.17	0.09	0.067
Total RT at Training [log]	0.09	0.05	**0.048**			
Rep [2]	0.27	0.34	0.426	0.93	0.06	0.246
Rep [3]	0.96	0.34	**0.005**	0.94	0.08	0.447
Rep [4]	0.84	0.34	**0.015**	0.88	0.09	0.159
Length (Phrase)[log]	0.50	0.17	**0.003**	1.58	0.15	**0.002**
Context Transparency [log]	-0.77	0.19	**<0.001**	0.41	0.18	**<0.001**

LexTALE [log]	-0.96	0.24	**<0.001**	0.49	0.24	**0.002**
Group [NNS] * Total RT at training [log]	0.11	0.05	**0.026**			
Total RT at training [log] * Rep [2]	-0.06	0.04	0.125			
Total RT at training [log] * Rep [3]	-0.16	0.04	**<0.001**			
Total RT at training [log] * Rep [4]	-0.17	0.04	**<0.001**			
Fixation Count at training				1.00	0.00	**0.003**
Trial Number [log]				0.85	0.05	**<0.001**
Rep [2] * Fixation Count at training				1.00	0.00	0.578
Rep [3] * Fixation Count at training				1.00	0.00	0.101
Rep [4] * Fixation Count at training				0.99	0.00	**0.009**

Random Effects

σ^2	0.23	0.24
τ_{00}	0.07_{subject}	0.07_{subject}
	0.01_{Item}	0.01_{Item}
ICC	0.28	0.26

Figure 6.3 Interaction between Total Reading Time in Training and Group on Total Reading Time in Reading

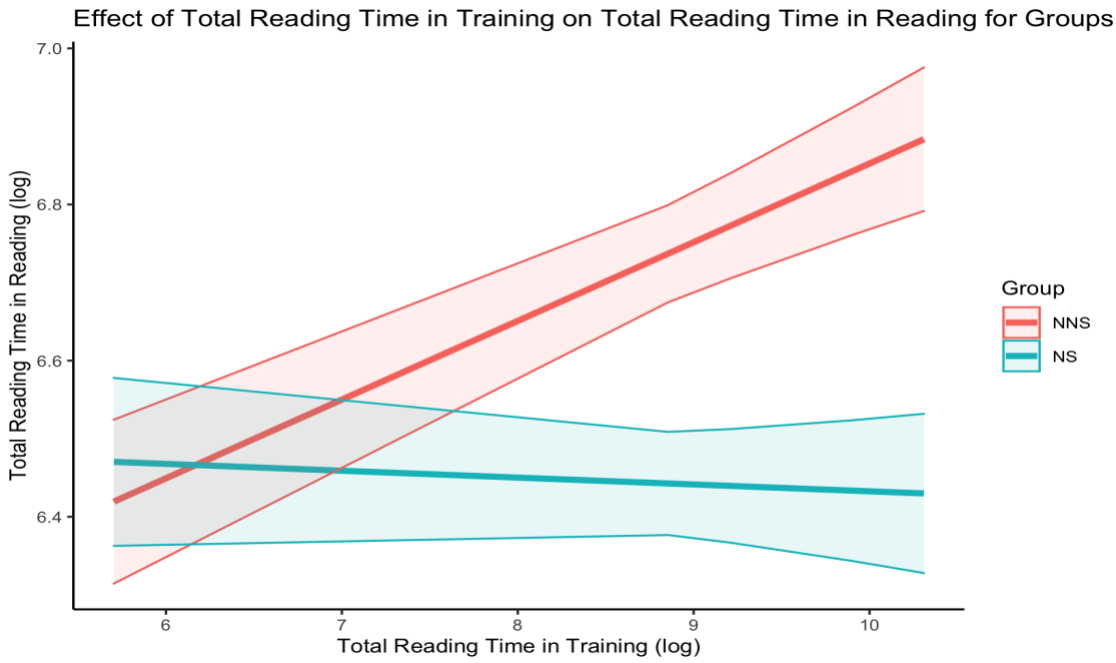


Figure 6.4 Interaction between Total Reading Time in Training and Repetition on Total Reading Time in Reading

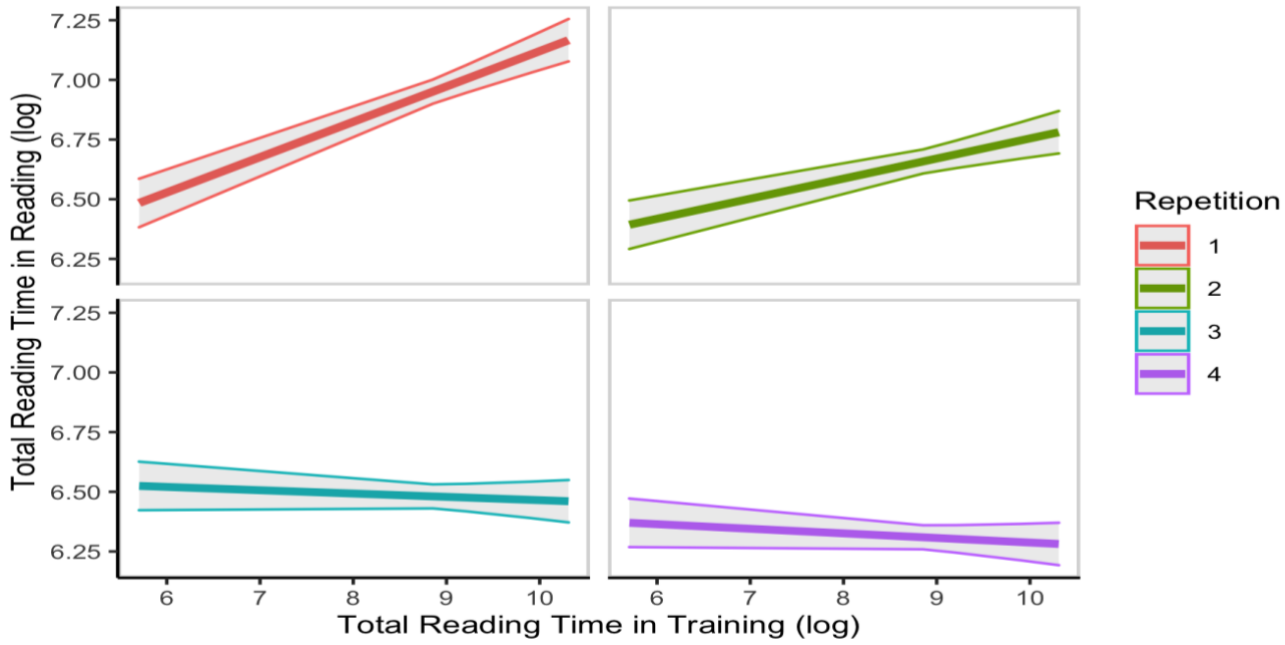


Figure 6.5 Interaction between Fixation Count in Training and Repetition on Fixation Count in Reading

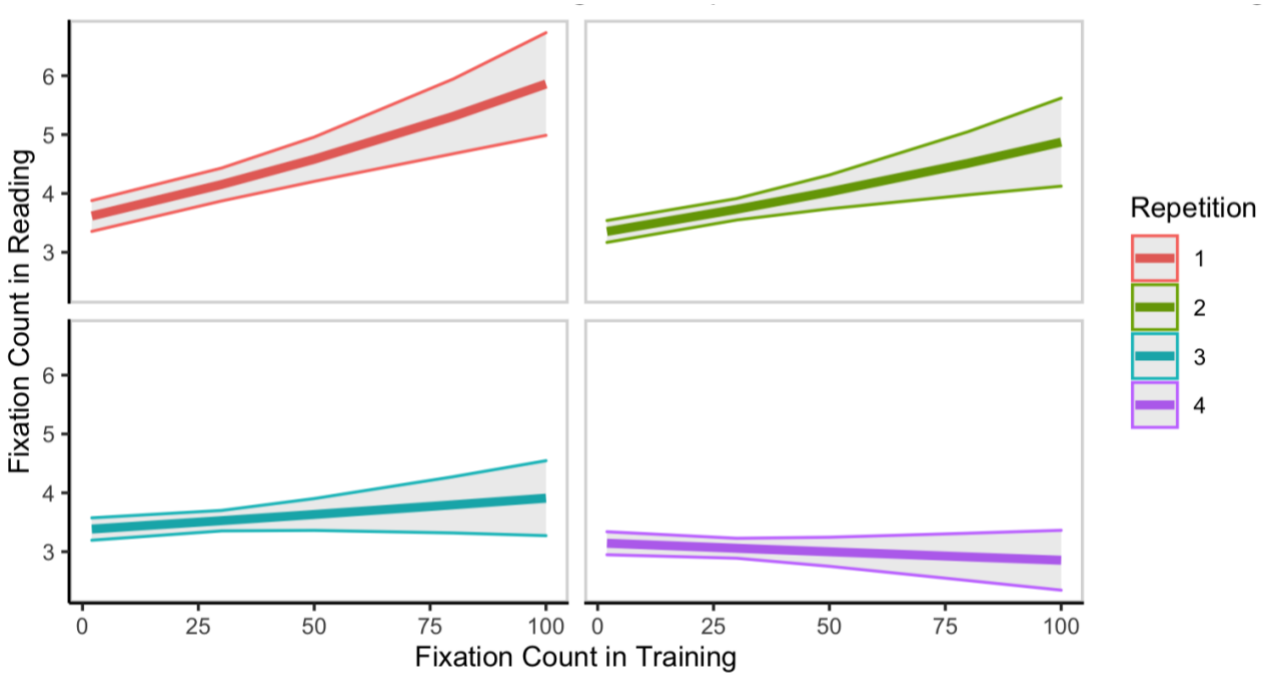


Table 6.13 Model Output for the Effects of Training on Reading Behavior at the Word Level

	Skipping Rate				**Total Reading Time**				**First Pass RT**				**First Fixation Duration**			
Predictors	*Log-Odds*	*std. Error*	*Statistic*	*p*	*Estimates*	*std. Error*	*Statistic*	*p*	*Estimates*	*std. Error*	*Statistic*	*p*	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	2.79	0.64	4.33	**<0.001**	9.82	1.08	9.09	**<0.001**	6.55	0.67	9.84	**<0.001**	4.85	0.18	27.68	**<0.001**
Group [NNS]	-0.22	0.17	-1.29	0.196	0.22	0.08	2.78	**0.005**	0.22	0.05	4.61	**<0.001**	0.19	0.03	5.56	**<0.001**
Rep [2]	0.34	0.18	1.96	0.050	-0.25	0.03	-7.58	**<0.001**	-0.16	0.05	-3.65	**<0.001**	-0.08	0.04	-2.25	**0.025**
Rep [3]	-0.10	0.19	-0.52	0.600	-0.36	0.03	-10.91	**<0.001**	-0.24	0.06	-3.81	**<0.001**	-0.14	0.05	-2.75	**0.006**
Rep [4]	0.39	0.17	2.22	**0.027**	-0.49	0.03	-14.86	**<0.001**	-0.31	0.08	-4.11	**<0.001**	-0.20	0.06	-3.14	**0.002**
First Pass RT at training [log]	-0.21	0.07	-3.06	**0.002**												
Length (Word) [log]	-1.74	0.22	-7.87	**<0.001**	0.33	0.07	4.69	**<0.001**	0.20	0.04	4.58	**<0.001**				
Total RT at Training [log]					0.05	0.03	1.84	0.065	0.02	0.02	0.84	0.399	0.02	0.02	1.39	0.163
Context Transparency [log]					-0.88	0.20	-4.37	**<0.001**	-0.37	0.14	-2.72	**0.006**				
LexTALE [log]					-0.84	0.22	-3.79	**<0.001**	-0.33	0.13	-2.55	**0.011**				
Trial Number [log]									0.13	0.05	2.66	**0.008**	0.10	0.04	2.57	**0.010**
Random Effects																
σ^2	3.29				0.27				0.17				0.12			
τ_{00}	0.24 $_{subject}$				0.06 $_{subject}$				0.02 $_{subject}$				0.02 $_{subject}$			
	0.05 $_{Item}$				0.01 $_{Item}$				0.00 $_{Item}$				0.00 $_{Item}$			
ICC	0.08				0.21				0.11				0.13			

RQ3b: What is the impact of reading times (in the training/reading phase) on performance in the testing phase?

This analysis more directly considers the impact of training on learning gains. It looks at how the amount of time spent studying (i.e., fixating) the phrases is related to response accuracy in the form recall, form recognition, and meaning recognition tasks. Eye-tracking measures in the training phrase (for T-only conditions) were considered as predictors of performance in the testing phase. Eye-tracking measures from the TR items were not included in this analysis because the TR items were involved in both the training and reading phases, making it difficult to disentangle any effects of the training phase from effects of the reading phase.

Table 6.14 presents the model outcome for accuracy in form recall, form recognition, and meaning recognition, as predicted by eye-tracking measures in the training phase. Total reading time in the training phase predicted form recall and meaning recognition but had no effect on form recognition. More precisely, longer total reading times in the training phase was associated with better form recall and meaning recognition. Transparency had an effect on form recognition, indicating that more transparent items were recognized more easily.

The final analysis considered the relationship between online processing in the reading phase (R-only) and the learning outcomes. Summed Total Reading Time (i.e., reading time across the four occurrences) was considered as a predicator of performance in form recall, form recognition, and meaning recognition. Summed Total Reading Time was calculated by adding the total reading times of all four repetitions for each phrase and for each participant. Table 6.15 presents the model outcome for accuracy in form recall, form recognition, and meaning recognition, as predicted by Summed Total Reading Time for R-only items. Summed Total Reading Time was significant in the form recognition and meaning recognition models but was not significant in the form recall model. The effect of

reading behaviour on form recognition indicated that longer total reading times during reading improved accuracy of form recognition for both groups. The effect of reading behaviour on meaning recognition showed that faster processing of the phrases was indicative of greater accuracy in meaning recognition.

Table 6.14 Effect of the Training Phase (T-only) on Form recall, Form Recognition, and Meaning Recognition

	Form Recall				**Form Recognition**				**Meaning Recognition**			
Predictors	*Log-Odds*	*std. Error*	*Statistic*	*p*	*Log-Odds*	*std. Error*	*Statistic*	*p*	*Log-Odds*	*std. Error*	*Statistic*	*p*
(Intercept)	-8.84	2.39	-3.69	**<0.001**	-1.14	0.48	-2.35	**0.019**	-3.67	2.44	-1.50	0.132
Group [NNS]	-0.25	0.36	-0.69	0.488	0.14	0.28	0.49	0.627	-0.12	0.46	-0.26	0.793
Total RT at training [log]	0.74	0.28	2.60	**0.009**					0.75	0.30	2.50	**0.013**
Word Transparency [log]					1.41	0.53	2.69	**0.007**				
Familiarity [log]					-0.12	0.40	-0.30	0.763				
Random Effects												
σ^2	3.29				3.29				3.29			
τ_{00}	0.22_{subject}				0.77_{subject}				2.13_{subject}			
	0.56_{Item}				0.00_{Item}				0.57_{Item}			
ICC	0.19				0.19				0.45			

Table 6.15 Effect of Reading Patterns (R-only) on Form recall, Form Recognition, and Meaning Recognition

	Form Recall				**Form Recognition**				**Meaning Recognition**			
Predictors	*Log-Odds*	*std. Error*	*Statistic*	*p*	*Log-Odds*	*std. Error*	*Statistic*	*p*	*Log-Odds*	*std. Error*	*Statistic*	*p*
(Intercept)	-2.79	2.58	-1.08	0.279	-3.40	2.62	-1.30	0.195	15.63	3.65	4.28	**<0.001**
Group [NNS]	-0.30	0.37	-0.82	0.412	-0.44	0.33	-1.35	0.178	0.18	0.48	0.39	0.698

Summed Total RT [log]	0.27	0.32	0.82	0.410	0.70	0.33	2.09	**0.036**	-1.58	0.44	-3.56	**<0.001**
Random Effects												
σ^2	3.29				3.29				3.29			
τ_{00}	1.19 $_{subject}$				0.35 $_{subject}$				1.45 $_{subject}$			
	1.70 $_{Item}$				0.35 $_{Item}$				1.28 $_{Item}$			
ICC	0.47				0.17				0.45			

6.5 Discussion

An overall aim of this study was to see if the patterns of results observed for transparent collocations in Study 2 would also be observed for collocations that were opaque. Thus, the research questions and methods in Studies 2 and 3 were largely the same; the studies only differed in the items they investigated and what was presented in the training phase. Study 3 presented both the collocations and their meanings in the training phase while Study 2 presented only the collocations. Study 3 extended the type of knowledge assessed in Study 2 to include meaning recognition. The findings of Study 3 are discussed in relation to the findings of Study 2.

6.5.1 *The Effect of Condition on Learning Gains*

The first research question addressed the effectiveness of different learning conditions on the learning of opaque collocations. As expected, the offline tasks showed that TR, R-only, and T-only led to greater response accuracy for both the NSs and NNSs than the control group in form recognition and meaning recognition, demonstrating the effectiveness of learning conditions. However, the learning conditions (i.e., response accuracy for the NSs and NNSs) did not yield better performance than the control group in form recall. Difficulty recalling the form of the novel collocations, which was not noted in Study 2, could be due to the decrease in transparency of the collocations in Study 3. This supports the notion that less literal multiword utterances are more difficult to learn than their more literal counterparts (Irujo, 1986; Macis & Schmitt 2017, Siyanova & Schmitt, 2007).

Replicating the same pattern from Study 2, the TR condition led to the greatest learning gains; the T-only condition led to the smallest learning gains; and the R-only condition leading to better performance than the T-only condition but worse than the TR condition. There were no differences in performance by NSs and NNSs in any of the tasks.

The superior learning gains associated with the TR condition in form recall and form recognition in Study 2 was also true of meaning recognition in Study 3. This suggests that acquiring knowledge of the form and meaning of collocations is helped by both initial explicit exposures as well as by repeated exposures in reading. The combination of intentional and incidental learning was more beneficial than the incidental or intentional exposures alone for both transparent and opaque collocations.

The superiority of TR could be explained by explicit exposure creating initial memory traces of the form and meaning, which were later strengthened during reading. In other words, the explicit exposures of decontextualized collocations in combination with contextual exposures to the recently learned collocations may have promoted greater engagement with the target items. According to Schmitt (2008), higher levels of sustained engagement with the target words maximises lexical gains. Using different sets of collocations, both Studies 2 and 3 demonstrated an advantage of a pre-reading 'study' activity on the development of collocational knowledge, supporting and expanding the findings of Pellicer-Sánchez et al. (2021) on the role of pre-reading instruction for the development of vocabulary knowledge.

Both Studies 2 and 3 showed that contextual encounters (R-only) led to greater gains than explicit exposures alone (T-only). Study 3 expanded the advantage of R-only over T-only to meaning recognition. Study 3 indicated that learners gain knowledge of the non-literal meaning of collocations after repeated exposures. For example, encountering *dry reception* repeatedly in context allowed learners to acquire the intended non-literal meaning ('unwelcoming'), which was different from the prototypical sense of the word *dry* ('not wet'). However, it was expected that meaning recognition would benefit more from the T-only than R-only because the training phase focused on linking collocational forms to their meanings. It is thought that "acquisition of word meanings requires explicit (conscious) learning processes" (Ellis, 1995, p. 2). Although collocations' meanings were provided in the training phase, contextual exposures generated greater gains in meaning

recognition than explicit exposures. This is in contrast to Pellicer-Sánchez et al. (2021) who found that instruction-only (i.e., pre-reading exposure only) benefited meaning recognition more than reading-only condition. A possible explanation for the different findings could be differences in the pre-reading activities in the two studies. Pre-reading in the Pellicer-Sánchez et al. study involved two activities, the second of which ensured that participants had successfully mapped the form and meaning of the items before beginning the reading phase (i.e., participants needed to correctly match the items with their meanings before engaging in the reading). Pre-reading in the current study only involved exposures to the items and their meanings; the amount of time spent studying each item was up to the individual and took, on average, 3.78 seconds by NSs and 4.99 seconds by NNSs. Thus, a lower level of engagement with the target items in pre-reading could explain the smaller amount of learning provided by T-only compared to R-only. Also, the Pellicer-Sánchez et al. study investigated individual word learning as opposed to collocations in the current study.

6.5.2 *The Effect of Condition on Processing*

The second research question addressed the effect of learning condition on processing during the reading phase. The findings indicated that there was an effect of learning condition on reading behaviour. In support of the results of Study 2, collocations which had been taught ahead of reading (TR) were processed faster compared to collocations which were encountered in reading for the first time. Notably, the faster reading times associated with pre-reading indicated that the learning advantage of pre-reading was not caused by increased attention to the phrases during reading. Instead, the learning advantage is likely due to additional exposures (provided by pre-reading exposure + contextual exposures). The effect of TR was noted in late measures (i.e., total reading time for both the phrase and word, fixation count for phrases), indicating that the advantage of pre-reading is attributed to easier integration of the collocation's meaning into the wider text, confirming the pattern observed in Study 2. While there was an absence of an effect of TR in early measures in Study 2, Study 3 demonstrated an advantage of TR already in early measures (i.e., first pass reading

time and skipping rates for words). This might be due to the fact that pre-reading in the current study presented both the phrases and their figurative meanings (i.e., *heavy tea:* 'strong tea'). Thus, while there was more focus on form in Study 2, both forms and meanings were presented in Study 3 to encourage learners to establish a form-meaning link for the opaque collocations ahead of reading. Thus, collocations whose form-meaning link had been potentially established ahead of reading in Study 3 (TR items) provided an advantage in the initial stages of processing (reflective of lexical access), as well as at the later stages of processing (reflective of integration and/or re-analysis).

Eye-movements in both Studies 2 and 3 revealed that NSs had faster reading than NNSs. This is line with Cop, Drieghe, and Duyck (2015) who showed that reading in an L2 elicited longer reading times and less skipping than reading in an L1. Longer reading times in an L2 may reflect greater processing effort when reading in an L2 compared to L1 (Conklin, Alotaibi, Pellicer-Sánchez, & Vilkaitė-Lozdienė, 2020). However, the current study showed that pre-reading helped NNSs exhibit a more native-like reading. Specifically, the effect of pre-reading was noted in the likelihood of skipping for NNSs. NSs had faster reading times (in all measures) and higher skipping rates than NNSs. While NNSs had lower skipping rates than NSs in R-only, they had similar skipping rates to NSs in TR. Thus, pre-reading made skipping the final words of the collocations more likely, suggesting that studying the novel combinations before reading increased their predictability in reading for NNSs. This signals the contribution of pre-reading on the representation of the phrases in memory. In other words, studying the collocations beforehand contributed to the degree to which they were entrenched in learners' memory.

The effect of repetition was noted in all measures for both phrases and individual words, similar to the pattern of results observed in Study 2: reading times decreased with every additional encounter. This suggests that each encounter contributes to forming stronger connections between the form and the meaning of collocations, as is reflected in faster processing (Conklin, 2019). The repetition effect was involved in an interaction with Condition (for phrases) and Group (for words),

indicating that that the repetition effect was larger for R-only items and for the NNSs group. With respect to the interaction with Group, the pattern of results supports the assumption that NNSs benefits more from additional exposures than NSs (Conklin, 2019). With respect to the interaction with Condition, collocations which were encountered for the first time during reading (R-only) needed longer times to process in the first encounter compared to collocations in which form and meaning had likely been linked ahead of reading (TR). It might be that the longer reading times for items that were encountered for the first time (R-only) could be indicative of laying down memory traces for the new forms and/or creating form-meaning connections of the newly encountered opaque items. This could explain the large difference between the first encounter and the second encounter noted in R-only items (i.e., the reading patterns of R-only items in the second encounter were similar to those of TR items in the first encounter).

The degree of transparency influenced reading patterns, such that more transparent items were processed faster. The measure of transparency that improved the models was Context Transparency (i.e., how guessable the meaning of a phrase was based on the sentence in which it was used). Increased transparency led to shorter total reading times (for phrases and words), fewer fixations (for phrases), and shorter first pass reading times (for words). Ratings of transparency in isolation did not improve the fit of any models, highlighting the role of context for processing. One possible interpretation is that context plays a significant role in resolving the lexical ambiguity of the adjectives used. For example, in a collocation such as *break a promise*, where the verb is used figuratively, Gyllstad and Wolter (2016) assume competition between literal and figurative meanings that could result in processing difficulty. Context plays an influential role in overcoming the processing difficulty that could arise from the two competing meanings, by supporting the activation of one meaning over the other (Gyllstad and Wolter, 2016) (i.e., supporting the activation of the figurative over the literal meaning in the case of the current study). This suggests that the processing of the opaque collocations in the current study was facilitated by the degree to which context

supports a more figurative reading of the adjectives (e.g., *dry* meaning 'unwelcoming'). This means that fast processing, as a function of transparency, is reflective of the easier resolution of multiple sense interpretations. The role of context in processing opaque collocations supports studies on idioms, showing that strongly idiom-biasing contexts facilitated the processing of idiomatic meanings (e.g., Cacciari & Corradini, 2015; Fanari, Cacciari, & Tabossi, 2010).

6.5.3 *The Effect of Amount of Study on Processing and Learning Gains*

The third research question examined whether the amount of time spent studying items during the training phase had an effect on reading patterns during the reading phase. The findings indicated that the amount of attention paid to items in the training phase predicted the amount of attention given to the same items in reading, consistent with the pattern of results that emerged in Study 2. More specifically, longer total reading times and more fixations for a phrase in training were associated with longer total reading times and more fixations for the same phrase during reading. In addition, longer first pass reading times in training were associated with less skipping for the final word of a phrase, and longer total reading times in training were associated with (marginally) longer total reading times for the final word of a phrase during reading.

It is important to note that effects of the training phase on later phrase reading interacted with Repetition and Group. Total reading time and fixation count in training had an influence on the reading patterns of TR phrases that was limited to the initial exposures. This suggests that increased attention to items during training elicits increased processing effort during the initial exposures in reading. By the third encounter, all TR phrases had similar processing patterns regardless of their processing during training (i.e., number of fixations and fixation durations were the same, regardless of the amount of attention in training). Further, the amount of time spent studying TR items during the training phase influenced the reading patterns of the NNSs but not the NSs. Fixation duration during training did not influence NSs' performance in the reading phase but did influence NNSs' reading.

The interactions involving Group and Repetition suggest that influence of training that was seen in Study 2 was not as pervasive in Study 3. Study 2 showed that increased attention during the training phase was associated with increased attention in the reading phase for both groups and across all encounters, suggesting that increased attention while 'studying' the items may increase their salience in later reading. This may be attributed to differences in the items used in the two studies. The novel collocations in Study 3 were opaque collocations, while the collocations in Study 2 were transparent. The unnatural form/meaning pairing of the items in Study 3 (e.g., *stone mentality* meaning 'narrow-minded') may have increased their salience, thereby drawing more attention to the items during reading, independent of their processing in the training. The overall longer reading times for the phrases in Study 3 compared to the phrases in Study 2 may support this suggestion. The mean cumulative total reading times of the opaque phrases was 2949 ms for the NSs and 4974 ms for the NNSs while for transparent phrases it was 2588 ms for the NSs and 3706 ms for the NNSs.[16] Thus, it may be that the salience of opaque collocations overrides the influence of training on subsequent reading patterns. In other words, because opaque collocations are salient, the amount of attention in training does not make them more salient and therefore training had limited influence in increasing their salience during reading (i.e., only in the first encounter).

It could be that novel opaque collocations are more unnatural (and thus more salient) for the NSs than NNSs during reading, due to NSs' greater L1 proficiency. This (assumed) increased salience for the NSs may help explain the difference between the two groups. Although entirely speculative, it is one possible explanation for why the amount of attention in training influenced the reading patterns of the NSs for transparent (Study 2) but not opaque collocations (Study 3). More research is needed to better understand the role of a deliberate study of both transparent and opaque collocations on their later reading patterns.

[16] It is important to note that this is not a perfect comparison because the phrases were not matched across the two studies.

The final research question examined whether the amount of time spent studying an item (during the training phase) or reading it (during the reading phase) had an effect on performance in form recall, form recognition, and meaning recognition. To disentangle effects of training from reading in context, this question considered phrases in the T-only and R-only conditions, and not those in the TR condition.

As for the relationship between attention (total reading time) in the training phase and collocational gains, the pattern of results showed that the fixation patterns for T-only items were indicative of performance in form recall and meaning recognition tasks. Specifically, increased attention to the phrases in training (longer total reading time for T-only items) improved performance on form recall and meaning recognition. Similar to Study 2, fixation durations in training did not predict performance on form recognition. It is important to point out that participants had overall better performance on form recognition than form recall in both Studies 2 and 3, consistent with previous research (e.g., Laufer & Goldstein, 2004; Webb et al. 2013; Sonbul & Schmitt, 2013). The results suggest that form recognition may be easier than other components of collocational knowledge (e.g., form recall). Thus, one possible explanation for why total reading time did not predict performance for form recognition is that developing a sense of form recognition may not necessarily require increased attention (i.e., greater effort reflective of longer total reading times) to the collocational forms. It might be that any attention is sufficient to develop form recognition; increased attention has no additional benefit for form recognition. Greater attention (i.e., longer fixation durations in the deliberate 'study' phase) could further entrench the collocation in memory and/or be indicative of greater depth of processing, which benefits form recall and identification of the figurative meaning. In other words, increased attention during intentional learning may lead to better performance on form recall and meaning recognition but have little benefit to form recognition.

The final analysis considered the relationship between Summed Total Reading Time for the items that were presented only in the R-only condition and learning gains. While fixation durations for items that were presented in the T-only condition had no influence on form recognition, Summed Total Reading Time for R-only items had an effect on form recognition, with longer total reading times improving accuracy of form recognition for both NSs and NNSs. One possible explanation for an effect of Summed Total Reading Time for R-only items, but not for total reading time for T-only items, on form recognition could be related to the type of knowledge being assessed – form recognition. As discussed earlier, form recognition might be considered "a more shallow type of word knowledge" (Godfroid et al., 2018, p. 575) and may be acquired at the earliest stages of the learning process (Mohamed, 2017). As was discussed above, increased attention in training may not impact form recognition (i.e., minimal attention is needed for such a task, increasing attention does not improve performance). Thus, form recognition may not be associated with increased attention during explicit exposures (training). Increased attention during explicit exposures may instead capture success in mastering more difficult components of word/collocation knowledge. In contrast, in an incidental learning task (R-only), it is likely that a 'shallow type of knowledge' is acquired, and therefore we see a relationship between processing and developing a sense of form recognition. However, this finding must be interpreted with caution since it was not evident in Study 2. Increased attention during reading better captured success in developing form recognition for opaque collocations (Study 3), but not transparent ones (Study 2). Further research is needed to explore the type of knowledge (shallow vs. deep) that is acquired in an explicit training task and implicitly while reading for transparent and opaque collocations.

Overall, the pattern of results suggests that longer total reading times in incidental exposures benefit the development of collocation form, while longer total reading times during the training phase benefit the development of more explicit-declarative knowledge of opaque collocations (form

recall and meaning recognition). However, further research is needed to understand what underpins these findings.

Shorter Summed Total Reading Time for R-only items was indicative of greater accuracy of meaning recognition. This suggests that the faster the items were read, the more easily their meanings were identified. Thus, when participants had to infer the meaning of items from context, shorter reading times were related to better meaning recognition. A possible explanation is that shorter total reading times reflected ease in linking a collocation's form with its meaning, which benefited later meaning recognition. Longer total reading times may reflect a greater effort in making form-meaning connections, which may not necessarily have been successful during reading, thereby making later meaning recognition more difficult.

In contrast to the current pattern of results, studies on vocabulary have found that total reading time was positively associated with recognition of word meaning (e.g., Godfroid et al., 2018; Pellicer-Sánchez et al., 2021; Mohamed, 2018). This discrepancy may reflect differences in the items used. Godfroid et al. (2018), Pellicer-Sánchez et al. (2020), and Mohamed (2018) used nonwords in which learners had no prior knowledge of their form or meanings (e.g., *blef*), whereas the current study used collocations consisting of real, familiar words with one word having a figurative meaning (e.g., *dry reception*). Furthermore, the meanings of the novel collocations in the current study were very easy to guess, as those in the control group were able to identify the meaning of the collocations with a 63% success rate. This suggests that Summed Total Reading Time in the current study may reflect an ease of meaning identification of the phrase, rather than an effort in learning the meaning of the phrase.

6.5.4 *Limitations*

Three important limitations need to be noted. First, we cannot rule out the possibility that task specific effects may have influenced the results. Although efforts were made to ensure that the tasks were equally difficult and that the 'right' answer was not obvious in any task, it is possible that there

were differences in the task that made one easier than the other. For example, the form recognition task (i.e., selecting the correct option from similar distractor ones) may have been more difficult than the meaning recognition task because the options may have been more difficult to distinguish from each other.[17] Contextual clues in the sentences may have aided performance in the meaning recognition task but not the form recognition task. The second limitation is related to the fact that NNSs did not speak a common L1. Although care was taken that no participant spoke Arabic (i.e., the source language of the items), the possibility that participants may have encountered some of the items in their L1s could not be ruled out. However, the fact that NSs and NNSs had similar findings makes this possibility unlikely. Third, the conclusion provided about the relationship between the eye-tracking patterns and different components of collocational knowledge was highly speculative. Further research is needed to explore the connection between different patterns of eye-movements and the type of collocational knowledge acquired.

6.5.5 *Conclusion*

The present study adds to our understanding of the learning and processing of a rarely explored type of formulaic language – opaque collocations. The findings underscore the importance of pre-reading for learning. A learning condition that combined deliberate study before reading and reading exposures was the most effective for the learning of opaque collocations. This supports the findings observed for transparent collocations (Study 2), indicating that both transparent and opaque collocations benefit from the same type of exposure. Pre-reading exposure also led to a processing advantage in later reading relative to items encountered in reading for the first time. This advantage was attributed to additional exposures. Study 3 suggest that the influence of attention during training in later reading patterns was less pervasive than in Study 2. This was likely due to the difference in

[17] An example of options provided in the form recognition task for *open appetite*: open _____ (a. desire, b. craving, c. appetite, d. hunger, e. don't know). An example of options provided in the meaning recognition task for *open appetite*: open appetite (a. sudden feeling of wanting something, b. strong willingness, c. big appetite, e. don't know)

items used in the two studies (transparent vs. opaque). Finally, the findings suggest that increased attention during reading was associated with success in form recognition while increased attention during pre-reading was associated with success in form recall and meaning recognition.

Chapter 7. Conclusion

Each of the three studies in this thesis presents findings that enrich our understanding of the learning and processing of formulaic language. This chapter highlights the main findings and contributions of the studies, as well identifying some of their limitations where further research is needed.

7.1 Scope of the Thesis

The main goal of the present thesis was to examine some of the key factors that are likely to affect the learning of formulaic language in an L2. Most studies in the field (reviewed in Chapters 3 and 5) that have investigated the factors implicated in formulaic language learning are limited in a number of ways.

First, most studies have only used offline measures. This was generally done using offline post-tests to assess the effect of a particular learning intervention on formulaic language learning. Offline measures fail to capture the "partial and incremental growth in knowledge occurring as a function of an individual's experience with each word" (Joseph et al., 2014, p. 245). An important contribution of the present thesis (in particular in Studies 2 and 3) is that it links offline declarative knowledge with real-time processing to gain greater insights into how knowledge is incrementally developed and how attention to formulaic language in acquisition is related to processing and aspects of formulaic knowledge. The use of both online (e.g., self-paced reading and eye-tracking) and offline measures offers a more nuanced understanding of formulaic language learning.

Second, to control for prior knowledge of the target items, studies used a pre-test (e.g., Peters, 2009) or pseudowords (e.g., Pellicer-Sánchez, 2017). The use of pre-test might have alerted learners to the target items. The salience of the pseudowords might have drawn more attention to them. Studies 2 and 3 addressed this by using novel word combinations to control for learners' prior

knowledge. The BNC as well as norming studies and the control groups' results confirmed that the word combinations used were indeed novel (i.e., none of the collocations were phrases in English).

Third, the majority of studies assessing the effects of intentional and incidental approaches to formulaic language learning assigned a different learning treatment to separate groups. Between-group variability (i.e., different participants' profiles) might interact with the effect of the treatment under investigation. More specifically, between-group designs make it difficult to disentangle any effect of the group itself from the treatment. Studies presented in the thesis had participants as their own controls encountering all levels of the variables under investigation (i.e., within subject design), potentially increasing the validity of comparisons.

Fourth, while research has examined the effect of attention-drawing techniques by making input salient through the use of typographical enhancement techniques (e.g., bolding), research has not really explored the effectiveness of another technique commonly used in the classroom – explicit exposure to the target items before reading. Further, in many studies there has been a lack of consideration and/or control of transparency, an important feature of formulaic language. This has left the effect of transparency on formulaic language learning and processing unclear.

The research presented in this thesis addresses these limitations and, therefore, contributes to our existing knowledge of formulaic language learning and processing. In addition to addressing some of the limitations in previous research, the thesis makes a number of important contributions to our understanding of formulaic language acquisition and processing.

The present thesis aimed to explore the role of congruency and evaluate the most effective learning condition for developing binomial and collocational knowledge. Study 1 investigated congruency in binomial learning and processing. This study adds to our understanding of the learning and processing of a rarely explored type of formulaic language – binomial phrases. This study, presented in Chapter 4, is the first to provide suggestive evidence that congruency facilitates identifying the conventional word order of binomials. Influence of the L1 was limited to binomials

that were the same in the L1 and L2 (i.e., congruent) and not the ones that had been translated from the L1 but did not exist in the L2 (i.e., incongruent). Furthermore, both NSs and NNSs seemed to learn the conventional word order of binomials from very few incidental exposures (two exposures). The number of exposures did not play an influential role in the learning outcome. However, this study is limited in that it assessed the learning of formulaic language under one learning treatment (meaning-focused reading treatment).

Study 2, presented in Chapter 5, explored the most effective learning condition to develop knowledge of transparent collocations. Although use of pre-reading activities is common in the classroom, its effect on the learning of formulaic language has not been extensively studied. The study showed, for the first time, that a short pre-reading exposure led to substantial learning gains of formulaic sequences. Pre-reading, when followed-up with reading exposures, led to superior learning gains than reading alone or pre-reading alone. Reading exposures of collocations appearing in different contexts (i.e., reading only) was more beneficial than pre-reading study alone. The benefit of pre-reading was also demonstrated in processing on subsequent reading. Items previously encountered in pre-reading had a processing advantage when encountered in reading, relative to items encountered in reading for the first time. Pre-reading appears to facilitate lexical integration during incidental exposures, and this was attributed to the further exposures provided by pre-reading. Triangulating eye fixation patterns with post-test data provided innovative perspectives into the influence of pre-reading on later reading or learning outcomes. In particular, the amount of time spent studying items during pre-reading was positively associated with the amount of time spent reading the items, suggesting that attention in pre-reading may have increased the salience of the items. Increased learner attention (longer fixation durations) in pre-reading was associated with better recall, and more fixations were associated with better recognition of the collocations. This is the first attempt in the literature to provide empirical evidence for how attention in pre-reading

predicts processing in subsequent reading or later learning outcomes, thus contributing to existing knowledge on the learning and processing of formulaic language.

Study 3 aimed to replicate the findings from Study 2 with another type of formulaic language – opaque collocations. The role of semantic transparency has rarely been taken into account when exploring collocations. Thus, this study extends our knowledge by exploring a type of formulaic language that has received little attention – opaque collocations. Study 3 confirms previous findings in Study 2 and contributes additional evidence suggesting that a combined pre-reading plus reading learning condition was most effective for developing knowledge of opaque collocations and that reading exposures alone led to superior learning gains than pre-reading study only. Similar to Study 2, pre-reading led to a processing advantage in later reading. The influence of pre-reading on later reading patterns was more apparent on opaque items (i.e., both early and late measures) than transparent items (i.e., only late measures), which was likely due to presenting the opaque sequences with their meanings in the pre-reading task. Further, transparency and context modulated processing for opaque collocations. Increased transparency, based on the context in which the item was used, led to faster reading times. Study 3 showed that the influence of attention during training in later reading patterns was less pervasive than in Study 2. This was likely due to the difference in the items used in the two studies (transparent vs. opaque). In terms of the connection between eye-tracking patterns and collocational knowledge, the study showed that increased attention during pre-reading was indicative of better performance in form recall and meaning recognition, while increased attention during reading was indicative of better performance in form recognition. By making an original contribution to the body of knowledge on the relationship between attention in intentional or incidental exposures and the development of different aspects of formulaic knowledge, findings of Studies 2 and 3 will serve as a basis for future studies.

7.2 General Discussion and Pedagogical Implications

The results presented in this thesis have important pedagogical implications. The overall pattern of results emerging in Study 1 suggest that the L1 has a positive influence on the learning of L2 formulaic language. Response time and accuracy data in Study 1 suggest that the recognition of binomial phrases is facilitated when the conventionalised word order of the phrases in the L1 and L2 is the same, which points to the key role of language transfer in L2 learning (e.g., Ionin & Montrul, 2010; Ellis, 2006).

As was discussed in Chapter 2, learning formulaic sequences in an L2 is complicated by the fact that learners already have a set of L1 formulaic sequences entrenched in memory. L1-L2 interference could hinder the learning or understanding of formulaic sequences. Nesselhauf (2003, 2005) pointed out that between 48% and 56% of the collocation errors made by L2 learners were due to L1 knowledge transfer. Thus, L2 teachers needs to take into account when learners can benefit from knowledge transfer (i.e., when formulaic language is the same across the languages) and when knowledge transfer will be problematic for learners. If teachers identify a set of congruent formulaic language, learners may quickly be able to incorporate these into their developing knowledge of the L2 and improve their fluency. When formulaic language is incongruent in the L1 and L2 or does not exist in the L1, teachers should consider that more effort may be required to learn such sequences. Because of the difficulty associated with learning incongruent formulaic sequences, Laufer and Girsai (2008) suggested that incongruent word combinations should receive more attention in L2 teaching than congruent ones. Crucially, while this may be possible in EFL classrooms in which all learners often share the same L1, it is likely to be more challenging in ESL classrooms where learners often do not share their L1. Notwithstanding this challenge, learners should be made aware that L1 word combinations are not always acceptable in the L2 and that the L1 and L2 will contrast in many respects, including which word combinations are acceptable.

Interestingly, Studies 2 and 3 show that NSs and NNSs had similar performance when learning formulaic sequences that do not exist in either L1 or L2. The results from NSs and NNSs in Studies 2 and 3, were extremely similar in both studies, suggesting that both NSs and NNSs are influenced similarly by the amount and type of exposure (Hoey, 2005). These studies refute some of the claims reported in the literature (e.g., Wray, 2002). Wray (2002) suggest fundamental differences in which NSs and NNSs process, learn and use formulaic sequences. Wray claims that NSs notice formulaic sequences by following a holistic approach while NNSs fail to notice these sequences as they follow a more analytical approach by analysing the internal parts of formulaic sequences. However, Studies 2 and 3 demonstrate that the effect of learning condition on collocation gains was the same for both NSs and NNSs, suggesting that both NSs and NNSs are equally affected by the amount and type of input and that they both notice formulaic sequences in natural input. Further, Study 1 show that frequency had the same effect for both NSs and NNSs. Thus, the studies presented in the thesis suggest that if there are any differences in performance between NSs and NNSs, it might be due to insufficient L2 exposure rather than fundamentally different processing of formulaic language. This supports usage-based theory (discussed in Chapter 2), which says that acquisition of formulaic language by NSs and NNSs are greatly influenced by the amount of exposure received by both populations.

Studies 2 and 3 suggest that amount of exposure (i.e., repetition) demonstrate a processing advantage during natural reading. However, it is important to note that a facilitatory effect of repetition for processing collocations was more evident when comparing differences in processing between the first and the fourth repetitions than when comparing differences between the first and second, second and third repetitions, or the third and fourth repetitions (Study 2). Further, frequency of exposure (i.e., two vs. five) had little effect on the learning process of binomials (Study 1). These findings suggest that a greater range of occurrences is necessary in order for significant repetition effects to emerge during natural reading. It is also important to acknowledge that other factors may

determine the effect of repetition for enhancing phraseological knowledge (e.g., learner individual variability, items' distinctive characteristics, type of exposure). This shows that the role of repetition is complicated by many other factors that determine its impact on formulaic language learning.

Studies 2 and 3 suggest that a combination of intentional learning (i.e., pre-reading activity) and contextual incidental learning (i.e., reading sentences for comprehension) is more effective in promoting collocational knowledge than either one of these alone. This means that adopting an approach, whereby both intentional and incidental learning are used, is the most effective. This has two important implications. First, both intentional and incidental approaches to learning are essential and complement each other (Webb, 2019). Thus, explicit instruction combined with incidental exposure is the best way to enhance language learning in the classroom (Hulstijn, 2001). Second, rather than treating incidental and intentional approaches as being in competition, future research should explore ways of integrating them to maximize learning. This can be done by evaluating the effectiveness of a variety of methods where intentional and incidental approaches might be usefully integrated.

It is likely that a complementary, intentional-incidental approach, would promote greater depth of knowledge[18] that enhances language learning in general, as well as formulaic language in particular. Schmitt (2014) explained that acquiring an additional aspect of word knowledge (e.g., form, meaning, and use) contributes to depth of knowledge. It might be that an intentional approach (i.e., explicit attention) establishes a form-meaning link and when this is followed up by an incidental approach (i.e., repeated exposures) knowledge about use is enhanced. Together the intentional and incidental learning could contribute to deeper knowledge of the formulaic sequences in the input. Such a hypothesis aligns well with proposals in the wider literature. According to Schmitt (2008), intentional learning is the most effective approach for developing initial knowledge of form-meaning

[18] Depth of knowledge refers to the quality of the knowledge acquired (i.e., how well all word knowledge aspects are developed) (Schmitt, 2014).

links. Crucially, it is important to enhance this knowledge through repeated contextual exposures (Schmitt, 2008). Schmitt (2008) believes that incidental exposure is most effective to strengthen knowledge of previously encountered or partially known lexical items. Studying vocabulary in isolation without contextual exposures limits developing knowledge to that of the form and meaning (Schmitt, 2014). Thus, it is important to reinforce this knowledge through repeated exposures that further entrench the word or the collocational pattern in memory. The use of pre-reading exposure may help form initial memory traces of the form-meaning link of the collocational pattern. The repeated contextual exposures may reinforce and further entrench those initial memory traces, resulting in acquiring an additional aspect of knowledge (i.e., how the collocations are used in context). The overall results of Study 2 and Study 3, linking attention in pre-reading to later learning gains, demonstrate the benefit of pre-reading in developing collocational knowledge.

Further, the results presented in the thesis suggest that a complementary approach of both incidental and intentional learning leads to a processing advantage for formulaic language. This suggests that pre-reading not only improves declarative knowledge of collocations (evidenced by performance on the post-tests), but it also improves procedural knowledge (evidenced by reading patterns). Specifically, pre-reading exposure facilitates subsequent integration of formulaic sequences into the surrounding sentence context during reading. This is in line with recent eye-tracking research demonstrating the benefit of pre-reading for facilitating word-to-text integration for individual words (e.g., Pellicer-Sánchez et al., 2021).

It is important to note that the advantage of pre-reading in processing was more apparent for opaque sequences (Study 3). This suggests that the use of pre-reading in the classroom may be more beneficial for opaque formulaic language. The non-compositional nature of opaque formulaic sequences (e.g., *kick the bucket*) may cause comprehension difficulties while reading for L2 learners. These difficulties may interrupt the flow of reading (i.e., having to stop reading to look up the meaning in the dictionary). Thus, introducing opaque sequences alongside their meanings in a pre-

reading activity may facilitate their later lexical access in subsequent reading and their integration into the overall context. It is important to point out that the transparent collocations were only presented as a list and not with their meanings because their meanings were apparent from the words that made them up. However, it may be that presenting transparent sequences with their meanings might also benefit subsequent processing by strengthening the link between the form and meaning. This is an interesting question for future research.

From a pedagogical perspective, it appears that implementing an attention-drawing device, such as pre-reading, before a reading activity is more beneficial than teaching a list of formulaic language or reading alone. It is also thought that implementing an attention-drawing device (e.g., pre-reading) may address some practical issues associated with incidental or intentional techniques when they are used alone in the classroom. In terms of relying only on incidental learning, the rate of learning uptake stimulated from incidental learning is low (Schmitt, 2008) because formulaic sequences may not occur frequently enough in authentic materials for substantial learning to occur (Pellicer-Sánchez & Boers, 2019). Furthermore, it is impractical for educators to manipulate texts such that a target formulaic sequence recurs a sufficient number of times in a short period. The low frequency of formulaic sequences in authentic reading materials means that it is unlikely that learners would notice repeated sequences of words (i.e., it is difficult to recognise whether a word combination is a recurring formula in an authentic text).

Crucially, "the first process encouraging learning is noticing, that is giving attention to an item" (Nation, 2013, p. 98) and it is thought to be the main driver for L2 learning (Schmidt, 1990, 2001). Thus, it is important to increase learners' awareness of what words co-occur. The use of pre-reading in the classroom might be one way to address this issue, by attracting learners' attention to the target word combinations before reading. An attention-drawing device, such as pre-reading, is also easier to implement, in teaching materials, compared to input flooding. Introducing formulaic sequences prior to reading should increase their salience during reading, encouraging learners to

recognize the phraseological dimension of formulaic sequences in the text. Results presented in the thesis (i.e., Study 2) pointed to a positive association between processing in pre-reading and reading, indicating that that longer total reading times and more fixations for a phrase in pre-reading were associated with longer total reading times and more fixations for the same phrase during reading. This suggests that increased attention to a formulaic sequence during pre-reading may increase its salience during reading.

While the thesis shows that intentional learning alone is not better than intentional with incidental learning, it is important to note that adopting a purely intentional learning approach may be equally as effective as a combined approach, if learners engage in multiple intentional tasks. Intentional learning is thought to be more effective than incidental learning (Laufer, 2003). However, in practical terms, it is difficult to explicitly teach the myriad of formulaic sequences due to limited classroom time (Boer & Lindstromberg, 2012). Pre-reading can, on the other hand, depend on less effortful resources than the typical intentional learning tasks reviewed in Chapter 5 (e.g., tasks that involve L1-L2 translation or gap-filling exercises). Thus, pre-reading does not need to consume much classroom time. In the current studies, time NNSs spent studying each collocation in the pre-reading phase did not exceed 2.14 seconds for transparent collocations and 5.09 seconds for opaque ones (and their definitions).

Studies 2 and 3 demonstrated that the learning gains came from the combination of pre-reading and the four contextual exposures. This means the learning gained through pre-reading needs to be supplemented with repeated contextual exposures. It also means that the additional exposures in different contexts were necessary to consolidate the initial knowledge encoded in pre-reading. Notably, it is expected that pre-reading exposure can reduce the number of incidental exposures necessary for successful acquisition. Another attention-drawing technique – input enhancement – has been shown to reduce the number of exposures to gain collocational knowledge. For example, Szudarski and Carter (2016) showed that the use of input-enhancement (in the form of

underlining) reduced the number of exposures needed to develop knowledge about collocations. Future studies should explore whether pre-reading exposure can reduce the number of subsequent exposures needed in context to promote knowledge of formulaic sequences. It is important to note that the number of exposures in the pre-reading task may not be equated with the number of exposures in the reading (i.e., incidental) task. It is expected that the intentional nature of pre-reading would promote greater level of learners' engagement with the target formulaic sequences.

Lastly, it is important to acknowledge that although the use of pre-reading as an instructional method has practical advantage, the process of determining which items to be presented in the pre-reading task may be challenging. From a pedagogical perspective, both learners' proficiency level and the frequency of formulaic sequences need to be taken into account when deciding which phrases to introduce in the pre-reading task.

7.3 Limitations and Future Directions

As with any body of research, there are some limitations to bear in mind when considering the findings and implications of the current studies. Limitations specific to each study were discussed in the relevant empirical chapters. This section identifies more general limitations, as well as considering possible directions for future research.

First, the studies presented in the present thesis focused on only two types of formulaic sequences – binomials and collocations. Thus, any conclusions that can be drawn about the learning and processing of formulaic language are limited to adjective-noun collocations and binomials. As discussed in Chapter 3, formulaic language consists of different types characterised by different properties (e.g., frequency, predictability, fixedness, and semantic transparency). Because of the diverse nature of formulaic language, findings about one type will not necessarily be true of other types. Crucially, L2 learners will encounter the different types. Thus, to better inform pedagogical practice, we would need to explore the effects of frequency, congruency, and learning context (i.e.,

incidental and intentional) for the different types of formulaic language. It might be the case that different types of formulaic sequences benefit more or less from different learning interventions.

Second, the conclusion that a combination of both intentional and incidental learning is the best approach for learning formulaic sequences is drawn from immediate post-test results. However, Studies 2 and 3 did not allow us to see how durable the knowledge gained through pre-reading plus reading would be. If the aim is to inform pedagogical practice, then it is worth investigating how durable the learning gained from any of the learning conditions would be.

The present thesis explored only one approach to combining intentional and incidental learning. Future research should also explore the effectiveness of alternative ways of combining the two. For example, an alternative method would be the use of post-reading activities. Hill and Laufer (2003) demonstrated the benefits of post-reading activities for developing knowledge of vocabulary. Szudarski (2012) showed that the addition of a post-reading activity (i.e., form-focused instruction), after the reading treatment, led to more collocational gains than the reading treatment alone. However, it is unclear whether a post-reading activity is more beneficial than a pre-reading activity. It would also be interesting to compare pre-reading with other types of attention-drawing techniques (e.g., typographical enhancement). Choi (2017) found an advantage for typographical enhancement in the learning of collocations. More research is needed to better understand the effectiveness of typographical enhancement, pre-reading, and post-reading for the learning of formulaic language. This would provide important insights into how to best promote knowledge of formulaic language. It would also be a big step toward discovering the best way to introduce formulaic language in the classroom.

The present thesis did not provide firm conclusions about the role of repetition. As discussed in the previous section, attention-drawing techniques may reduce the number of exposures required for incidental learning to occur. In Studies 2 and 3, each item was repeated the same number of times (four exposures). In order to better understand how the addition of pre-reading, or other attention-

drawing techniques, reduces the number of encounters needed for collocational gains, future research should manipulate the number of times items appear in the text. Furthermore, many other questions relevant to repetition remain unanswered in the current thesis. For example, how far should the repetitions be spaced, or would massed repetition be more beneficial?

The thesis suggests that the recall of opaque formulaic sequences is more difficult than transparent ones. However, the non-compositional nature of opaque formulaic sequences (e.g., *show someone the ropes*) is thought to increase their salience, and thus, make them more likely to be noticed than transparent formulaic sequences (e.g., *tell a lie*) (Boers & Lindstromberg, 2009). It is important to note that the comparison of the learning of transparent (Study 2) and opaque collocations (Study 3) should be approached with caution because semantic transparency was assessed across two studies and not within one. Although efforts were taken to ensure that the participants were similar across the two studies, it is impossible to rule out the possibility that group differences underpin any differences between the two studies. It would be worthwhile investigating how the hypothesized increased salience of opaque sequences, relative to transparent sequences, influences attention and learning in a single study.

The studies presented in the thesis examined the learning of formulaic language from silent reading. Future research should investigate if other input modalities (e.g., listening, reading while listening, watching videos) are equally beneficial. Feng and Webb (2020) showed that different input modes (i.e., written, audio, and audio-visual input modes) were equally effective for developing knowledge of vocabulary. The effect of input modality on the learning of formulaic language has just recently started to attract attention. Webb and Chang (2020) showed that reading while listening led to greater collocational gains than reading only and listening only. One explanation is that prosodic patterns signal a formulaic sequence as one intonation unit in spoken input (Lin, 2019), which could facilitate identifying the target sequence in the text. Nevertheless, research on the effect of the modality of input on the learning of formulaic language is in its infancy. Comparing the learning of

formulaic language from a single input mode (i.e., reading or listening), a bi-modal input mode (i.e., reading while listening), or a multi-input mode (i.e., captioned videos) would be a useful avenue for future research. This comparison would help educators and learners understand the advantages of different input modes for learning formulaic language and could inform language teaching practice.

Finally, Studies 2 and 3 did not look into the role of L1 knowledge in the learning of collocations. NNSs from these studies did not speak a common L1, which made it difficult to investigate how congruency may influence the learning of collocations under different learning conditions. Peters' (2016) study showed that non-congruency may impede the learning of collocations even when they were studied intentionally. It remains an open question whether the learning of congruent or incongruent formulaic sequences may be influenced differently by different learning conditions.

In sum, the studies presented in the present thesis explored factors that affect the learning and processing of formulaic sequences. However, the studies leave open important questions for future research. Thus, the research included in this thesis represents a step in our understanding of how to best teach and learn formulaic language in an L2. More research is needed to better understand what influences the learning of formulaic language in an L2 and how to best introduce formulaic language in the classroom.

References

Abel, B. (2003). English idioms in the first language and second language lexicon: A dual representation approach. *Second language research, 19*(4), 329-358.

Alemi, M., & Ebadi, S. (2010). The Effects of Pre-reading Activities on ESP Reading Comprehension. *Journal of Language Teaching & Research, 1*(5).

Allen, D. B., & Conklin, K. (2013). Cross-linguistic similarity and task demands in Japanese-English bilingual processing. *PloS one, 8*(8), e72631.

Al-Thubaity, A. (2015). A 700M + Arabic corpus: KACST Arabic corpus design and construction. *Language Resources and Evaluation, 49*(3), 721–751.

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1-L2 differences, Special issue on multiword units in language. *Topics in Cognitive Science, 9*(3), 621–636.

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

Baayen, R. H. (2010). A real experiment is a factorial experiment? *The Mental Lexicon, 5*, 149 –157. doi:10.1075/ml.5.1.06baa

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations?. *System, 21*(1), 101-114.

Bannard, C., & Lieven, E. (2012). Formulaic language in L1 acquisition. *Annual Review of Applied Linguistics*, *32*, 3-16.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19,* 241–248. doi:10.1111/j.1467- 9280.2008.02075.x

Bardovi-Harlig, K. (2019). Formulaic language in second language pragmatics research. In A. Siyanova-Chanturia and A. Pellicer-Sánchez (Eds). *Understanding formulaic language: A second language acquisition perspective* (pp. 97-114). London, UK: Routledge.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models.* arXiv:1506.04967 [Stat.ME]. Retrieved from http://arxiv.org/abs/1506.04967

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version, 1*(7), 1–23.

Beck, S. D., & Weber, A. (2016). Bilingual and monolingual idiom processing is cut from the same cloth: the role of the L1 in literal and figurative meaning activation. *Frontiers in psychology*, *7*, 1350.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language Is a Complex Adaptive System: Position Paper. *Language Learning, 59* (Supp. 1), 1-26.

Benor, S., & Levy, R. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language*, *82*(2), 233-278.

Bialystok, E. (2011). Coordination of executive functions in monolingual and bilingual children. *Journal of experimental child psychology*, *110*(3), 461-468.

Bod, R. (2000). *The storage vs. computation of three-word sentences*. Paper presented at AMLaP2000, University of Leiden, Leiden, the Netherlands.

Bod, R. (2001). *Sentence memory: Storage vs. computation of frequent sentences*. Paper presented at CUNY 2001, University of Pennsylvania, Philadelphia, PA.

Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *The linguistic Review*, *23*(3), 291-320.

Bod, R., Hay, J., & Jannedy, S. (Eds.). (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, *27*(2), 448–469.

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), 245-261.

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

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics, 32*(1).

Bybee, J. (1995). Regular morphology and the lexicon. *Language and cognitive processes*, *10*(5), 425-455.

Bybee, J. (1998). The emergent lexicon. *Chicago Linguistic Society*, *34*, 421-435.

Bybee, J. (2001). Frequency effects on French liaison. In J. Bybee & P. Hopper (eds), *Frequency and the emergence of linguistic structure*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711-733.

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford University

Press.

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–236). London: Routledge.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Cacciari, C., & Corradini, P. (2015). Literal analysis and idiom retrieval in ambiguous idioms processing: A reading-time study. *Journal of Cognitive Psychology, 27*(7), 797-811.

Carrol, G., & Conklin, K. (2014a). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition, 17*(4), 784-797.

Carrol, G., & Conklin, K. (2014b). Eye-tracking multi-word units: some methodological questions. *Journal of Eye Movement Research, 7*(5).

Carrol, G., & Conklin, K. (2017). Cross language lexical priming extends to formulaic units: Evidence from eye-tracking suggests that this idea 'has legs'. *Bilingualism: Language and Cognition*, *20*(2), 299-317.

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech, 63(*1), 95-122.

Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in translation: The influence of L1 on the processing of idioms in L2. *Studies in Second Language Acquisition, 38*(3), 403–443.

Carter, R. (1988). Vocabulary, cloze and discourse: An applied linguistic view. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 161- 180). Harlow: Longman.

Chan, T. P., & Liou, H. C. (2005).Effects of web-based concordancing instruction on EFL students? learning of verb-noun collocations. *Computer Assisted Language Learning, 18,* 231–251.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693-713.

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research, 21*,403–426.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39*.

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

Cieślicka, A. B., & Heredia, R. R. (2017). How to" Save Your Skin" When Processing L2 Idioms: An Eye Movement Analysis of Idiom Transparency and Cross-Language Similarity among Bilinguals. *Iranian Journal of Language Teaching Research, 5*(3), 81-107.

Cobb, T. (n. d.). Compleat Lexical Tutor (Version 2) [Online application software]. Retrieved from https://www.lextutor.ca

Conklin, K. (2019). Processing single- and multi-word items. In S. Webb (Ed). Routledge *Handbook of Vocabulary Studies* (pp. 174-188). Routledge.

Conklin, K., Alotaibi, S., Pellicer-Sánchez, A., & Vilkaitė-Lozdienė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research, 36*(3), 257-276.

Conklin, K. & Carrol, G. (2019). First language influence on the processing of formulaic language in a second language. In A. Siyanova-Chanturia and A. Pellicer-Sánchez (Eds). *Understanding formulaic language: A second language acquisition perspective* (pp. 62–77). Routledge.

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research, 32*(3), 453-467.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and non-native speakers? *Applied Linguistics, 29*, 72–89.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review* of *Applied Linguistics*, 32, 45-61.

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS ONE, 10*, e0134008. doi:10.1371/journal.pone.0134008

Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, *2*(3), 223-235.

Cristoffanini, P., Kirsner, K., & Milech, D. (1986). Bilingual lexical representation: The status of Spanish-English cognates. *The Quarterly Journal of Experimental Psychology Section A*, *38*(3), 367-393.

Day, R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, 7, 541–551.

De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp. 67–79). London: Longman.

De Groot, A. M. (2011). *Language and cognition in bilinguals and multilinguals: An introduction*. Psychology Press.

Dijkstra, A., & Van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. Jacobs (Eds.), *Localist Connectionist Approaches to Human Cognition* (pp. 189–225). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, *5*(3), 175-197.

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory, 6*, 125-155.

Dupuy, B., & Krashen, S. (1993). Incidental vocabulary acquisition in French as a foreign language. *Applied Language Learning, 4*(1-2), 55-63.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163-188.

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye movement study. *Studies in Second Language Acquisition, 40*, 341–366.

Ellis, R. (1999). Factors in the incidental acquisition of second language vocabulary from oral input. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 35- 62). Amsterdam: John Benjamins Publishing Company.

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1103–1123.

Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning, 8*(2-3), 103-128.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in second language acquisition*, 91-126.

Ellis, N. C. (1997). Vocabulary acquisition: Word structure, collocation, grammar, and meaning. In M. McCarthy & N. Schmidt (Eds.) (1997), *Vocabulary: description, acquisition and pedagogy* (pp. 122-139). Cambridge: Cambridge University Press.

Ellis, N.C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second*

language instruction. Cambridge, UK: Cambridge University Press.

Ellis, N. C. (2002a). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*(2), 143–188.

Ellis, N. C. (2002b). Reflections on frequency effects in language acquisition: A response to commentaries. *Studies in Second Language Acquisition*, *24*, 297–339.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied linguistics*, *27*(2), 164-194.

Ellis, N. C. (2008). Usage-based and form-focused language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372– 405). London: Routledge.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*, 17-44.

Ellis, N. C., & Larsen-Freeman, D. (2009). *Language as a complex adaptive system* (Vol. 3). John Wiley & Sons.

Ellis, N. C., & Ogden, D. C. (2017). Thinking About Multiword Constructions: Usage-Based Approaches to Acquisition and Processing. *Topics in cognitive science*, *9*(3), 604-620.

Ellis, N. C. & Robinson, P. (2008). An introduction to Cognitive Linguistics and Second Language Acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, (pp. 3–24). London: Routledge.

Ellis, R. (2015). *Understanding second language acquisition*. Oxford, UK: Oxford University Press.

Ellis, R. S. (2007). *Entropy, large deviations, and statistical mechanics*. Springer.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse*, *20*(1), 29-62.

Fanari, R., Cacciari, C., & Tabossi, P. (2010). The role of idiom length and context in spoken idiom comprehension. *European Journal of Cognitive Psychology*, *22*(3), 321-334.

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective?. *Studies in Second Language Acquisition*, *42*(3), 499-523.

Fernández, B. G., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *ITL-international journal of applied linguistics*, *166*(1), 94-126.

File, K. A., & Adams, R. (2010). Should vocabulary instruction be integrated or isolated? *TESOL Quarterly, 24*,222–249. https://doi.org/10.5054/tq.2010.219943

Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development*, *67*(6), 3192-3218.

Floyd, P., & Carrell, P. L. (1987). Effects on ESL reading of teaching cultural content schemata. *Language learning*, *37*(1), 89-108.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behavior*, *12*(6), 627-635.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*, Martin Bygate, Peter Skehan and Merrill Swain (eds.), 75–94. London: Longman.

Fox, J. (2009). Effect displays in R for generalised linear models. *Journal of Statistical Software, 8*(15), 1–27. https://doi.org/10.18637/jss.v032.i01.

Gabrys-Biskup, D. (1992). L1 influence on learners' renderings of English collocations. A Polish/German empirical study. In P. J. L. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 85–93). London: Macmillan.

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & cognition*, *8*(2), 149-156.

Gibbs, R.W., Bogdanovich, J.M., Sykes, J.R., & Barr, D.J. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language, 37*, 141-15

Gibbs Jr, R. W., & Gonzales, G. P. (1985). Syntactic frozenness in processing and remembering idioms. *Cognition*, *20*(3), 243-259.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press on Demand.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., …Yoon, H. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition, 21*,563–584.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive linguistics*, *15*(3), 289-316.

Grabe, W., & Stoller, F. L. (2019). *Teaching and researching reading*. Routledge.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145– 160). Oxford: Oxford University Press.

Granger, S., & Meunier, F. (Eds.). (2008). *Phraseology: An interdisciplinary perspective*. John Benjamins Publishing.

Greenbaum, S. (1988). Some verb-intensifier collocations in American and British English. In S. Greenbaum (Ed.), *Good English and the grammarian* (pp. 113- 124). London: Longman.

Grosjean, F. (2010). *Bilingual*. Harvard University Press.

Gurevich, O., Johnson, M. A., & Goldberg, A. E. (2010). Incidental verbatim memory for language. *Language and Cognition*, *2*(1), 45-78.

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

Hernández, M., Costa, A., & Arnon, I. (2016). More than words: multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience*, *31*(6), 785-800.

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics in Language Teaching, 41*(2), 87-106.

Hilpert, M. (2008). New evidence against the modularity of grammar: Constructions, collocations, and speech perception. *Cognitive Linguistics*, 19(3), 491-511.

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: Routledge.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, *61*(3), 355- 382.

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.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 161–186). Oxford: Oxford University Press.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In Anaud, P. J., & Béjoint, H. (Eds.) *Vocabulary and applied linguistics* (pp. 113–125). London: Macmillan.

Hulstijn, J.H. (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 University Press, Cambridge.

Hulstijn, J. H. (2003). Incidental and intentional learning. K. J. Doughty & M. H. Long (Eds), *The handbook of second language acquisition*. (pp. 349–381). Malden, MA: Blackwell. doi: 10.1002/9780470756492.ch12

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Ionin, T., & Montrul, S. (2010). The role of L1 transfer in the interpretation of articles with definite plurals in L2 English. *Language learning, 60*(4), 877-925.

Irujo, S. (1986). Don't put your leg in your mouth: Transfer in the acquisition of idioms in a second language. *TESOL Quarterly, 20*(2), 287-304.

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*, 434–446.

Jahangard, A., Moinzadeh, A., & Karimi, A. (2012). The effect of grammar vs. vocabulary pre-teaching on EFL learners' reading comprehension: A schema-theoretic view of reading. *Journal of English language teaching and learning, 3(*8), 91-113.

James, W. (1890). *The principles of psychology*. New York: Holt.

Jenkins, J. R., M. L. Stein, & K. Wysocki. (1984). Learning vocabulary through reading. *American Educational Research Journal, 21*(4), 767-787.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21, 47-77.

Joseph, H. S. S. L., Wonnacott, E., Forber, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition, 133*,238–248.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive science*, *20*(2), 137-194.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. *Probabilistic linguistics*, 21.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: Benjamins.

Kemmer , S. , & Barlow , M . (2000). Introduction . In M. Barlow & S. Kemmer (Eds.), *Usage- based models of language* (pp. vii–xxviii). Stanford, CA : CSLI Publications .

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second language research*, *18*(2), 137-171.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, *33*(2), 149-174.

Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In Schmitt, N. (Ed.), *Formulaic sequences* (pp. 37–54). Amsterdam, the Netherlands: John Benjamins.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: Tests in linear mixed effects models. R package version 2.0-29. Retrieved from https://CRAN.Rproject.org/package5lmerTest.

Kweon, S. O., & Kim, H. R. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a foreign language*, *20*(2), 191-215.

Kyriacou, M., Conklin, K., & Thompson, D. (2020). Passivizability of idioms: Has the wrong tree been barked up?. *Language and speech*, *63*(2), 404-435.

Lambon Ralph, M. A., & Ehsan, S. (2006). Age of acquisition effects depend on the mapping between representations and the frequency of occurrence: Empirical and computational evidence. *Visual Cognition, 13*, 928–948.

Lambrecht, K. (1984). Formulaicity, frame semantics, and pragmatics in German binomial expressions. *Language*, 753-796.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: descriptive application* (Vol. 2). Stanford university press.

Langacker, R. W. (2009). *Investigations in cognitive grammar* (Vol. 42). Walter de Gruyter.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian modern language review, 59*(4), 567-587.

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.

Laufer, B. & Girsai, N. (2008). Form-focused instruction in L2 vocabulary learning. *Applied Linguistics, 29*(4), 694–716.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language learning, 54*(3), 399-436.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647-672.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44*, 325-343.

Lenth, R. (2019). Emmeans: Estimated marginal means, aka least-squares means. Retrieved from https://CRAN.R-project.org/package=emmeans

Lewis, M. (1997). *Implementing the lexical approach*. Hove: LTP.

Lin, P. (2012). Sound evidence: The missing piece of the jigsaw in formulaic language research. *Applied Linguistics*, *33*(3), 342–347.

Lin, P. (2019). Formulaic language and speech prosody. In A. Siyanova-Chanturia and A. Pellicer-Sánchez (Eds). *Understanding formulaic language: A second language acquisition perspective* (pp. 78-94). Routledge.

Liversedge, S. E, Paterson, K. B., & Pickering, M. J. (1998). Eye move- merits and measures of reading time. In G. Underwood (Ed.), *Eye movements in reading and scene perception* (pp. 55-76). Oxford, England: Elsevier.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological review*, *95*(4), 492.

Lüdecke, D., Makowski, D., Waggoner, P., & Patil, I. (2020). Package 'performance'.

Macis, M., & Schmitt, N. (2016). The figurative and polysemous nature of collocations and their place in ELT. *ELTJ*.

Macis, M., & Schmitt, N. (2017). Not just 'small potatoes': Knowledge of the idiomatic meanings of collocations. *Language Teaching Research, 21*(3), 321-340.

MacWhinney, B. (1997). Second language acquisition and the competition model. *Tutorials in bilingualism: Psycholinguistic perspectives*, 113-142.

MacWhinney, B. 2017. Entrenchment in second-language learning. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*, 343–366. Boston, MA & Berlin: APA & Walter de Gruyter.

Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Matlock, T., & Heredia, R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In: Heredia R and Altarriba J (eds) *Bilingual sentence processing*. Amsterdam: Elsevier, 251–74.

Matthews, D., & Bannard, C. (2010). Children's production of unfamiliar word sequences

is predicted by positional variability and latent classes in a large sample of child-directed speech. *Cognitive science*, *34*(3), 465-488.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305-315.

McDonald, S. A., & Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological science*, 14(6), 648-652.

McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, 43(16), 1735-1751.

McDonough, K., & Trofimovich, P. (2013). Learning a novel pattern through balanced and skewed input. *Bilingualism: Language and Cognition*, *16*(3), 654-662.

Meara, P., & Jones, G. (1990). Eurocentres vocabulary size test 10KA. *Zurich: Eurocentres*.

Meunier, F., & Granger, S. (Eds.). (2008). *Phraseology in foreign language learning and teaching*. John Benjamins Publishing.

Millar, N. (2011). The processing of malformed formulaic language. *Applied linguistics*, *32*(2), 129-148.

Millisecond Software. (2015). Inquisit 5. Retrieved from https://www.millisecond.com.

Mohamed, A. A. (2018). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, *40*(2), 269-293.

Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling—Sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 183.

Montero Perez, M., Peters, E., & Desmet, P. (2015). Is less more? Effectiveness and per-

ceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL*, *26*(01), 21–43. https://doi.org/10.1017/S0958344013000256

Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach.* Oxford: Clarendon Press.

Morgan, E., & Levy, R. (2016). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition*, 157, 384-402.

Nagy, W., R. Anderson, & P. Herman. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24.2, 237-270.

Nagy, W., Herman, P., & Anderson, R. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233-253.

Nation, I. S. P. (2001). *Learning Vocabulary in Another language*. Cambridge: Cambridge University Press.

Nation, I. S. (2013). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.

Nattinger, J. R. and DeCarrico, J. S. 1992. *Lexical Phrases and Language Teaching*. Oxford: OUP.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*(2), 223-242.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: Benjamins.

Northbrook, J., & Conklin, K. (2019). Is what you put in what you get out? Textbook-derived lexical bundle processing in beginner English learners. *Applied Linguistics*, *40*(5), 816-833

Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent?. *Studies in second language acquisition*, *18*(1), 1-25.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication, 191*, 225.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*(1), 97-130.

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research, 21*(3), 381-402.

Pellicer-Sánchez, A., & Boers, F. (2019). Pedagogical perspectives to the teaching and learning of formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 153-173). New York, NY: Routledge.

Pellicer-Sánchez, A., Conklin, K., & Vilkaitė-Lozdienė, L. (2021). The effect of pre-reading instruction on vocabulary learning: An investigation of L1 and L2 readers' eye movements. *Language Learning, 71*(1), 162-203.

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, A. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.

Peters, E. (2009). Learning collocations through attention-drawing techniques: A qualita- tive and quantitative analysis. In A. Barfield & H. Gyllstad (Eds.*), Researching collocations in another language: Multiple interpretations* (pp. 194–207). Basingstoke: Palgrave Macmillan.

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research, 18*(1), 75-94.

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research, 20*(1), 113-138.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case

study. *Reading in a foreign language, 18*(1), 1-28.

Pine, J. M., Lieven, E. V., & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics, 36*(4), 807-830.

Pinker, S. (1991). Rules of language. *Science, 253*(5019), 530-535.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6*(11), 456-463.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language, 5*(2), 271-275.

Pritchett, L. K., Vaid, J., & Tosun, S. (2016). Of black sheep and white crows: Extending the bilingual dual coding theory to memory for idioms. *Cogent Psychology*, *3*(1), 1135512.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*,372-422.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension pro- cesses [Special issue]. *Language and Cognition Processes, 4*, 21- 49.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: multilevel statistical learning by 12-month-old infants. *Infancy, 4*(2), 273-284.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human perception and performance*, *3*(1), 1.

Schmid, H. J. (2007). Entrenchment, salience, and basic levels. *The Oxford handbook of cognitive linguistics*, *117138*.

Schmid, H. J. (2015). A blueprint of the entrenchment-and-conventionalization model. *Yearbook of the German Cognitive Linguistics Association*, *3*(1), 3

Schmid, H. J. (2017). A framework for understanding linguistic entrenchment and its psychological foundations. In H.-J. Schmid (Ed). *Language and the human lifespan series. Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 9–35). Berlin: De Gruyter Mouton.

Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, 11, 129-158.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language teaching research*, *12*(3), 329-363.

Schmitt, N. (2010). *Researching vocabulary*. London: Palgrave Macmillan.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language learning*, *64*(4), 913-951.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1–22). Amsterdam: John Benjamins.

Shu, H., Anderson, R. C., & Zhang, Z. (1995). Incidental learning of word meanings while reading: a Chinese and American cross-cultural study. *Reading Research Quarterly, 30*(1), 76-95.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). afex: Analysis of Factorial Experiments. R

package version 0.16-1. https://CRAN.R- project.org/package=afex

Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., & Van Heuven, W. J. (2017). Representation and processing of multi-word expressions in the brain. *Brain and language, 175*, 111-122.

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*, 251-272.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Language, Memory and Cognition, 37*, 776–784.

Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, *36*(5), 549-569.

Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics in Language Teaching, 45*, 119−139.

Siyanova, A., & Schmitt, N. (2008). L2 Learner Production and Processing of Collocation: A Multi-study Perspective. *The Canadian Modern Language Review / La revue canadienne des langues vivantes 64*(3), 429 458.

Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of native speaker and second language learner intuition of collocation frequency. *Language Learning*, *65*(3), 533-562.

Siyanova-Chanturia, A., & Van Lancker Sidtis, D. (2019). What on-line processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 38–61). London, UK: Routledge.

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on

off-line/on-line collocational processing. Bilingualism: *Language and Cognition*, *18*(3), 419-437.

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*(1), 121-159.

Sorhus, H. B. (1977). To hear ourselves—Implications for teaching English as a second language. *ELT Journal*, *31*(3), 211-221.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative methods. *Functions of language*, *2*(1), 1–33.

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, *18*(5), 523-534.

Szudarski, P. (2012). Effects of meaning-and form-focused instruction on the acquisition of verb-noun collocations in L2 English. *Journal of Second Language Teaching & Research*, *1*(2), 3-37.

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), 245-265.

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast?. *Memory & cognition*, *37*(4), 529-540.

Taglieber, L. K., Johnson, L. L., & Yarbrough, D. B. (1988). Effects of preceding activities on EFL reading by Brazilian college students. *TESOL quarterly*, *22*(3), 455-472.

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting bilingual and monolingual idiom processing. *Bilingual figurative language processing*, 171-207.

Titone, D., Holzman, P. S., & Levy, D. L. (2002). Idiom processing in schizophrenia: Literal implausibility saves the day for idiom priming. *Journal of Abnormal*

Psychology, *111*(2), 313.

Tomasello, M. (2003). *Constructing a language: A Usage-Based theory of language acquisition*. Cambridge: Harvard University Press.

Toomer, M., & Elgort, I. (2019). The Development of Implicit and Explicit Knowledge of Collocations: A Conceptual Replication and Extension of Sonbul and Schmitt (2013). *Language Learning, 69*(2), 405-439.

Tremblay, A., & Baayen, H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 151-173). London/New York: Continuum

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.

Tudor, I. (1990). Pre-reading format and learner proficiency level in L2 reading comprehension. *Journal of research in reading, 13*(2), 93-106.

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning, 69*(3), 559-599.

Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language, 93*(2), 185–238.

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. 153–172). Amsterdam: John Benjamins.

Van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in

bilingual word recognition. *Journal of memory and language*, *39*(3), 458-483.

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*, 1176–1190. doi: 10.1080/17470218.2013.850521

Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609-624.

Vilkaitė, L. (2017). Incidental acquisition of collocations in L2: Effects of adjacency and prior vocabulary knowledge. *ITL-International Journal of Applied Linguistics*, *168*(2), 248-277.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied linguistics*, *28*(1), 46-65.

Webb, S. (2014). Repetition in incidental vocabulary learning. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–6). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781405198431.wbeal1425

Webb, S. (2019). Incidental vocabulary learning. In Webb, S. (Ed.), *The routledge handbook of vocabulary studies*. Routledge.

Webb, S., & Chang, A. (2020). How does mode of input affect the Incidental learning of collocations? *Studies in Second Language Acquisition*, 1-22. doi:10.1017/S0272263120000297

Webb, S. & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly, 4*(1), 55-76.

Webb, S. & Kagimoto, S. (2011). Learning Collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*, *32*(3), 259-276.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language*

Learning, *63*(1), 91-120.

Whaley, C. P. (1978). Word—nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 143-154.

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.

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

Wolter, B., & Yamashita, J. (2017). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for performance? *Studies in Second Language Acquisition*. Advance online publication. doi:10.1017/ S0272263117000237.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, UK: CambridgeUniversity.

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford 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.

Wray, A & Perkins, M (2000). The functions of formulaic language and integrated model. *Language & Communication, 20*(1), 1-28.

Wulff, S. (2019). Acquisition of formulaic language from a usage-based perspective. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 19–37).

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.

Yeganehjoo, M., & Thai, Y. N. (2012). Lexical access in production of idioms by proficient L2 learners. *The Southeast Asian Journal of English Language Studies, 18,* 87–104

Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency. *Bilingualism across the lifespan*, 55-72.

Zhang, X. (2017). Effects of receptive-productive integration tasks and prior knowledge of component words on L2 collocation development. *System, 66,* 156-167.

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*. Cambridge: MA: MIT Press.

Appendices

Appendix 4A: The Multiple-Choice Vocabulary Test (Study 1)

Please choose the correct meaning for each word.

1. **POOR**: We are poor.
   a. have no money
   b. feel happy
   c. are very interested
   d. do not like to work hard
   e. don't know

2. **SHOE**: Where is your shoe?
   a. the person who looks after you
   b. a thing you keep your money in
   c. the thing you use for writing
   d. a thing you wear on your foot
   e. don't know

3. **PATIENCE**: He has no patience.
   a. will not wait happily
   b. has no free time
   c. has no faith
   d. does not know what is fair
   e. don't know

4. **CIRCLE**: Make a circle.
   a. rough picture
   b. space with nothing in it
   c. round shape
   d. large hole
   e. don't know

5. **DINOSAUR**: The children were pretending to be dinosaurs.
   a. robbers who work at sea
   b. very small creatures with human form but with wings
   c. large creatures with wings that breathe fire
   d. animals that lived long ago
   e. don't know

6. **DASH**: They dashed over it.
   a. moved quickly
   b. moved slowly
   c. fought
   d. looked quickly
   e. don't know

7. **QUIZ**: We made a quiz.
   a. thing to hold arrows
   b. serious mistake
   c. set of questions
   d. box for birds to nest in
   e. don't know

8. **VOCABULARY**: You will need more vocabulary.
   a. words
   b. skill
   c. money
   d. guns
   e. don't know

9. **COMPOST**: We need some compost.
   a. strong support
   b. help to feel better
   c. hard stuff made of stones and sand stuck together
   d. rotted plant material
   e. don't know

10. **PEEL**: Shall I peel it?
    a. let it sit in water for a long time
    b. take the skin off it
    c. make it white
    d. cut it into thin pieces
    e. don't know

11. **THRESHOLD**: They raised the threshold.
    a. flag
    b. point or line where something changes
    c. roof inside a building
    d. cost of borrowing money
    e. don't know

12. **CAVALIER**: He treated her in a cavalier manner.
    a. without care
    b. politely
    c. awkwardly
    d. as a brother would
    e. don't know

13. **BRISTLE**: The bristles are too hard.
    a. questions
    b. short stiff hairs
    c. folding beds
    d. bottoms of the shoes
    e. don't know

14. **GIMMICK**: That's a good gimmick.
    a. thing for standing on to work high above the ground
    b. small thing with pockets to hold money
    c. attention getting action or thing
    d. clever plan or trick
    e. don't know

15. **ECLIPSE**: There was an eclipse.
    a. a strong wind
    b. a loud noise of something hitting the water

c. the killing of a large number of people
d. the sun hidden by a planet
e. don’t know

16. **AUTHENTIC**: It is authentic.
a. real
b. very noisy
c. old
d. like a desert
e. don’t know

17. **WHIM**: He had lots of whims.
a. old gold coins
b. female horses
c. strange ideas with no motive
d. sore red lumps
e. don’t know

18. **OCTOPUS**: They saw an octopus.
a. a large bird that hunts at night
b. a ship that can go under water
c. a machine that flies by means of turning blades
d. a sea creature with eight legs
e. don’t know

19. **UPBEAT**: I'm feeling really upbeat about it.
a. upset
b. good
c. hurt
d. confused
e. don’t know

20. **CROWBAR**: He used a crowbar.
a. heavy iron pole with a curved end
b. false name
c. sharp tool for making holes in leather
d. light metal walking stick
e. don’t know

Participant Code……

Appendix 4B: Language Background Questionnaire (Studies 1, 2, and 3)

English language background information

Age ___ Gender _____ Country of Birth _____

Native language (language learned from birth) _____, proficiency 1 (very low) 2 3 4 5 6 7 (native-like)

Other language _____, proficiency 1 2 3 4 5 6 7

Other language _____, proficiency 1 2 3 4 5 6 7

Other language _____, proficiency 1 2 3 4 5 6 7

Highest level of education secondary school/high school ☐

Undergraduate 1st year ☐, 2nd year ☐, 3rd year ☐, 4th year ☐

Postgraduate MA/MSc ☐, PhD ☐

Age of first contact with English ___, and where home ☐, family ☐, school ☐, other _____

At primary school the main language of instruction was English ☐, your native language ☐, other _____

At secondary/high school the main language of instruction was English ☐, your native language ☐, other _____

Amount of time formally studied English ___ years & months ___

Total amount of time lived in UK ___ years & months ___

Total amount of time lived in *other* English speaking countries: ___ years & months ___

Please rate your *current* ability in English on the following:

Speaking (excluding accent) 1 (very low) 2 3 4 5 6 7 (native-like)

Understanding spoken language 1 2 3 4 5 6 7

Reading 1 2 3 4 5 6 7

Writing 1 2 3 4 5 6 7

How often do you *currently* do the following things in English vs. another language(s)?

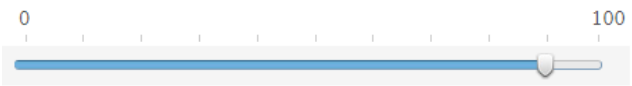
Other English

Speak with:

…. your family

…. your friends

…. people at work / university



Write to:

…. your family

…. your friends
…. people at work / university

Read things for:
…. academic / work purposes
…. pleasure (e.g. books, newspapers, web, etc.)

Watch / hear / listen to:
…. TV / films / videos (e.g. YouTube) / etc.
…. podcasts / audio books / radio / music / etc.
…. lectures / classes / seminars / meetings / etc.

Appendix 4C: Binomial Phrases (Study 1)

Items (in their preferred word order)	Type
sunrise and sunset	Arabic
coffee and dates	Arabic
land and sea	Arabic
rice and meat	Arabic
prophets and messengers	Arabic
sun and moon	Arabic
heaven and hell	Arabic
hunger and thirst	Arabic
fig and olives	Arabic
poetry and prose	Arabic
seller and buyer	Arabic
camels and sheep	Arabic
individual and society	Arabic
Jews and Christians	Arabic
boys and girls	Congruent
brothers and sisters	Congruent
Adam and Eve	Congruent
success and failure	Congruent
food and drink	Congruent
king and queen	Congruent
profit and loss	Congruent
men and women	Congruent
north and south	Congruent
past and present	Congruent
rich and poor	Congruent
gold and silver	Congruent
strengths and weaknesses	Congruent
reading and writing	Congruent
bride and groom	English
cause and effect	English
fish and chips	English
church and state	English
costs and benefits	English
art and design	English
accident and emergency	English
facts and figures	English
knife and fork	English
law and order	English
rules and regulations	English
salt and pepper	English
pain and suffering	English

trial and error	English
mind and body	Control
town and country	Control
hands and knees	Control
head and shoulders	Control
income and expenditure	Control
indoor and outdoor	Control
time and money	Control
science and education	Control
age and sex	Control
sweet and sour	Control
iron and steel	Control
stocks and shares	Control
urban and regional	Control
words and phrases	Control

Appendix 4D: The Binomial Stimuli in the Exposure Phase (Study 1)

List 1

The idea of keeping church and state separate can be traced back to Thomas Jefferson. Jefferson pressed for a total separation of church and state in which the government remains neutral toward religion while still maintaining people's religious rights.

Winston Churchill once described fish and chips as 'the good companions'. England's popular dish of fish and chips first appeared in UK shops in 1860 and was sold by a Jewish immigrant. By the 19th century, the presence of fish and chips shops had increased significantly to meet the demands of a growing industrial population. It started as a popular take-away meal among the British working class, and then later fish and chips became popular among all social classes. Nowadays, the popularity of fish and chips has reached many other English-speaking countries around the world.

When children discover the relationship between cause and effect early on, they are better able to understand how their actions can lead to certain consequences. One way to do this is by demonstrating cause and effect while playing with them.

The National Health Service in the UK report that 93.4 percent of patients in the UK are discharged from accident and emergency units within 4 hours of arrival. On the other hand, 95 percent of patients going to accident and emergency in the US are seen within 3 hours. Do the statistics reported for accident and emergency in the US mean that the American healthcare system is better or faster? The answer is no, because the average time reported for UK accident and emergency cases refers to the total amount of time from arrival to the end of treatment. By contrast, the 3 hours wait time reported for US accident and emergency cases refers to the average waiting time before being seen by a doctor.

I used to season my steak with a pinch of salt and pepper in addition to a mixture of other spices. However, I recently followed the advice of several chefs in adding only salt and pepper when seasoning steak.

The United States is full of rules and regulations which many people might find ridiculous. For example, in Oklahoma, some rules and regulations state that it is illegal to have a sleeping donkey in the bathtub at night. Another example is in Nevada, where the state's rules and regulations mean it is against the law to ride a camel on the highway. Apparently, there are even some rules and regulations in Florida which make it illegal to tie an elephant to a parking lot without paying the parking fee. Regardless of whether these rules and regulations are still needed, I think that the stories behind them are very interesting.

British people highly value table manners, which for them involves eating with a knife and fork at most meals. There are some exceptions when a knife and fork are not used

-when eating fruit, crisps or sandwiches for example.

Political proponents of strict law and order argue that a war on crime is the only way to prevent it. Conversely, many left-wing politicians believe that this kind of approach to law and order doesn't address the real causes of crime. Many American presidential candidates have talked about law and order to promote their presidential Campaigns. For example, President Nixon promised to restore law and order by cracking down on illegal drugs. More recently, calling himself 'the law and order candidate', President Trump promised to tackle crime and violence and vowed to make America safe again.

Skinner argued that trial and error can explain how children learn language. Subsequently, the famous linguist Chomsky rejected the notion that language learning is based on trial and error and argued that instead it is driven by an innate biological ability.

In Hinduism, the concept of Karma states that pain and suffering occur because of negative behaviour in either the current or a past life. Hindus consider Karma to be an integral part of life, which means that pain and suffering are also unavoidable. They also believe that while spiritual practices can never entirely prevent pain and suffering they do offer ways to learn from these experiences. According to Hinduism, overcoming human desires can break the cycle of Karma and allow one to move beyond pain and suffering to a state of inner peace and purity. Hindus can also choose to willingly endure pain and suffering in the hope of a better incarnation in another life.

One common form of an arranged marriage is a consensual arranged marriage where both the bride and groom approve of being brought together by their family members. Another less common form of an arranged marriage is a forced marriage, where both bride and groom are given no choice in the matter.

If you choose your career path wisely, a degree in art and design can offer many work opportunities. These are some of the highest paid jobs that an art and design graduate can get according to PayScale.com. With 10 years' experience in art and design you can be a creative director, with a median annual salary of 103,000 US dollars. Those who majored in art and design can also have a profitable career in videography and get film editing jobs. In addition, acquiring programming skills alongside your experience in art and design can make your resume very attractive for industry management positions.

The Spanish Constitution of 1978 restored the monarchy and the King and Queen sit on the throne. After they were reinstated, the press met with the King and Queen to talk about the power of the monarchy.

Different versions of the story of Adam and Eve appear in different religious texts. According to the Bible, God created Adam and Eve and then placed them in Eden. In this narrative, God forbade Adam and Eve from eating fruit from a forbidden tree. However, the serpent tempted Adam and Eve to eat from that tree. Because they ate the forbidden fruit, God punished Adam and Eve by expelling them from the garden of Eden.

Starting in 1861 the Civil War raged across the north and south of the United States for about four years. However, an important outcome of the war between the north and south is that it brought slavery to an end.

Miller's Death of a Salesman provides a powerful portrayal of success and failure in pursuit of the 'American dream'. Miller's depiction of the various characters demonstrates how success and failure is achieved. In particular, he examines how success and failure is determined by an individual's choices. He shows how success and failure can come about by dramatizing the reactions of the characters in various situations. The author denies that success and failure can be inherited or in-born.

Their musical band is made up of brothers and sisters who live in San Francisco, which has a vibrant music scene. The band was started by the brothers and sisters in late 2002 when they were all in their twenties.

The first time in history currency was exchanged was in the form of gold and silver in 1500 BC. In 1066, the Norman conquerors issued gold and silver coins to use in international trade. In 1285, Venice and England recognized gold and silver as official currency. In 1792, the United States started adapting a gold and silver monetary system that was linked to the value of the dollar. However, converting the dollar into assets such as gold and silver was suspended by President Nixon in 1971.

The difference in pay between men and women has attracted a lot of media attention lately. The BBC reported that the gender pay gap between men and women has reached an extreme level in the UK, with three quarters of firms paying higher wages to their male staff.

In this era of social media, texting, and blogs, how students engage with reading and writing is changing at a rapid pace. In the digital age, students do more reading and writing outside of the classroom than they do in it. They are also exposed to other visual forms of communication that fall outside of conventional reading and writing norms. For example, the use of emoticons and emojis is on the rise and may need to be addressed when teaching reading and writing in today's classroom. Teachers need to keep up with these changes so that they can teach their students reading and writing skills that are relevant in today's society.

During Ramadhan, Muslims all over the world do not have any food and drink from dawn until dusk. Depending on the region and season, the amount of time that people have to refrain from partaking in food and drink differs greatly.

Worldwide the gap in wealth between the rich and poor has widened considerably. A recent economic report shows that inequality between the rich and poor is the most extreme in the US and Russia. What makes the gap amongst the rich and poor worse in these countries is the fact that only a very small percentage of the population has all of the wealth. Such extreme differences between the rich and poor have led

to a shrinking middle class. serious political reformations must be made in order to lessen the gap between the rich and poor worldwide.

Psychologists from the City University of London compared boys and girls in terms of their toy preferences. The study showed that in all age groups boys and girls tended to select gender specific toys.

As a financial advisor, I often receive requests from clients to prepare a profit and loss statement that projects areas of growth as well as any potential risks. They use my profit and loss calculations to help them make key business decisions. In addition to reviewing profit and loss accounts, I also prepare annual budget reports. The 7 years of experience that I have covers all areas of profit and loss accounting and has provided me with the skills required to handle the job. The current economic turbulence makes my job in preparing profit and loss projections very demanding.

The survey asked 35,071 adult Americans about their beliefs in heaven and hell in telephone interviews. Roughly 72 percent and 58 percent of respondents reported believing in heaven and hell respectively.

Photographers have captured images of the sunrise and sunset from all over the world. A world-famous photographer described what he considered to be the top locations for sunrise and sunset photos. Topping his list as the world's most impressive sunrise and sunset viewing spot, is Mount Bromo, in Indonesia. In order to see the breathtaking sunrise and sunset here, travellers need to hike for 2 hours over the mountain. His second choice for capturing the sunrise and sunset is Namib-Naukluft in Namibia.

The use of elements of both poetry and prose was evident in most of Shakespeare's early comedies. Later, Shakespeare began to exploit the interplay between poetry and prose in some of his tragedies, such as Hamlet.

Muslims believe that God has sent prophets and messengers to the world at different times in history. Muslims believe that prophets and messengers are sent to the people to preach similar religious beliefs. The Quran mentions some of these prophets and messengers by name, while others are not mentioned at all. Muslims believe that these prophets and messengers are the greatest human beings of all time. All Muslims agree that Muhammed is the final of Allah's prophets and messengers who have been sent to bring God's message to the world.

One non-violent act of political protest is to reject eating and drinking and not give in to hunger and thirst until a set of demands are met. Sometimes, protestors risk dying of hunger and thirst to bring public attention to injustice or to provoke sympathy from an offending party.

The observation of eclipses has led to interesting stories about the relationship between the sun and moon in different cultures. In some aboriginal Australian cultures,

the eclipse was seen as an event where the sun and moon procreate. In contrast, some ancient cultures in Africa saw the eclipse as a fight between the sun and moon to resolve old grudges. For the Inuits, the sun and moon were siblings, and the eclipse was a chance for them to meet. Finally, in ancient Chinese cultures, the eclipse was interpreted as the sun and moon being eaten by a dragon.

Tourists who travel by land and sea enjoy not only the destinations, but also the beauty between them. Travellers who go by land and sea can spontaneously stop and explore interesting places and experience new cultures.

Historically, Bedouin tribes in Arabia mainly depended on camels and sheep as a source of food. Not many animals are suited to such a harsh climate as the desert, but camels and sheep were able to survive. Bedouin tribes mainly used their camels and sheep to produce milk and cheese. They also used the wool from the camels and sheep to make clothes. Each member of the Bedouin tribe helped to take care of the camels and sheep because they were an invaluable resource.

Kabsah is a traditional rice and meat dish in Saudi Arabia. While the rice and meat Kabsah is often associated with Saudi Arabia, it is also served in other Arabic countries.

The view of man as a social animal has made studying the relationship between the individual and society an important area of research in sociology. Sociologists have made different classifications based on the role played by the individual and society in establishing collective shared values. Researching how the individual and society dictate social behaviour has also been a topic of interest to social psychologists. Because exploring the relationship between the individual and society generally involves the question of values, the topic has also been of interest to philosophers. The relationship between the individual and society is essential in shaping self-identity, as well as group-identity.

There are many shared religious beliefs amongst Jews and Christians that make the two similar in many ways. For example, Jews and Christians share important religious figures, such as Abraham and Jesus, and Moses.

Guides for travellers to Saudi Arabia say that coffee and dates are more than just a snack. In fact, serving coffee and dates is an important aspect of hospitality in the Saudi Arabian culture. Offering a guest snacks other than coffee and dates is frowned upon. All segments of society offer coffee and dates at a variety of occasions, from family gatherings to tribal meetings. The custom of offering coffee and dates seems to span the generations, being done by the old and young alike.

Students should be informed about the costs and benefits of student loans in order to make the best decisions about how to finance their education. Universities should also be involved in evaluating the costs and benefits of student loans and helping students make good choices.

The Chicago Alzheimer's Association recently released a YouTube video detailing various facts and figures about Alzheimer's. Their set of facts and figures showed that in America, 5.7 million people have been diagnosed with Alzheimer's. Their projections for the facts and figures indicate that by 2050, 14 million people will be living with Alzheimer's. The facts and figures suggest that 1 in 3 seniors will die due to Alzheimer's or related diseases. The organization made a closing plea in their facts and figures video asking people to donate money to support the fight against Alzheimer's.

Asking job candidates about their strengths and weaknesses is very common in interviews. When candidates are asked about their strengths and weaknesses in interviews, they are always advised to relate the answer to the job description.

In Egypt, the division between past and present is apparent almost immediately. In fact, we can visibly see past and present divided by a few kilometres of land as the pyramids sit alongside Cairo's skyscrapers. Some of the world's most ancient remains are in walking distance from Cairo, so Egyptians are surrounded by the past and present in their daily lives. Living with such a mix of the past and present shows our need to modernize, while at the same time preserving our traditions. However, being surrounded by the past and present can provide inspiration to people.

The relationship between a seller and buyer should be based on trust and mutual understanding. However, legal documents should be signed by both the seller and buyer in order to prevent any future conflict.

A recent medical study showed that eating figs and olives has many health benefits. The research indicated that figs and olives promote metabolism and increase energy. It also found that figs and olives help reduce cholesterol levels. Other research showed that figs and olives are a good remedy for constipation and stomach aches. Doctors now recommend that figs and olives be incorporated into our regular diet.

LIST 2

'I believe in an America where the separation of church and state is absolute', said J.F. Kennedy in 1960. He made it clear that he wanted to ensure the US was a country where church and state had no influence over each other. The idea of keeping church and state separate can be traced back to Thomas Jefferson. Jefferson pressed for a total separation of church and state in which the government remains neutral toward religion while still maintaining people's religious rights. Jefferson's attitude toward the separation of church and state was based on his interpretations of the first amendment.

Winston Churchill once described fish and chips as 'the good companions'. England's popular dish of fish and chips first appeared in UK shops in 1860 and was sold by a Jewish immigrant.

Young children find the concept of cause and effect very difficult to understand. Parents should be encouraged to introduce the concept of cause and effect to their children at an early age. When children discover the relationship between cause and effect early on, they are better able to understand how their actions can lead to certain consequences. One way to do this is by demonstrating cause and effect while playing with them. For example, if a dog toy has a button that makes it squeak, parents can point out the cause and effect relationship between pressing the button and the squeaking sound.

The National Health Service in the UK report that 93.4 percent of patients in the UK are discharged from accident and emergency units within 4 hours of arrival. On the other hand, 95 percent of patients going to accident and emergency in the US are seen within 3 hours.

I used to season my steak with a pinch of salt and pepper in addition to a mixture of other spices. However, I recently followed the advice of several chefs in adding only salt and pepper when seasoning steak. This is because they say that adding spices other than salt and pepper may actually hide the flavour of the beef. Instead, most chefs and cookbooks recommend adding a generous amount of salt and pepper before grilling steak. The recommendations are to add the salt and pepper either just before grilling or 24 hours in advance.

The United States is full of rules and regulations which many people might find ridiculous. For example, in Oklahoma, some rules and regulations state that it is illegal to have a sleeping donkey in the bathtub at night.

British people highly value table manners, which for them involves eating with a knife and fork at most meals. There are some exceptions when a knife and fork are not used -when eating fruit, crisps or sandwiches for example. Chicken and pizza may also be eaten without a knife and fork in more informal situations. In restaurants, when taking a break from eating, people usually leave their knife and fork crossed in the centre of the plate. When they have finished eating, people often place their knife and fork parallel, with the handles facing the 4 o'clock position.

Political proponents of strict law and order argue that a war on crime is the only way to prevent it. Conversely, many left-wing politicians believe that this kind of approach to law and order doesn't address the real causes of crime. ~n

Many early psychological studies found that animals apply trial and error as a problem-solving strategy. These studies concluded that learning by trial and error only takes place if a certain action is repeated under the same conditions several times. Skinner was among the first psychologists to claim that trial and error can also influence human behavior, especially in language learning. Skinner argued that trial and error can explain how children learn language. Subsequently, the famous linguist Chomsky rejected the notion

that language learning is based on trial and error and argued that instead it is driven by an innate biological ability.

In Hinduism, the concept of Karma states that pain and suffering occur because of negative behaviour in either the current or a past life. Hindus consider Karma to be an integral part of life, which means that pain and suffering are also unavoidable.

In an arranged marriage, the bride and groom are brought together by people other than the couple themselves. The criteria for matching a bride and groom are usually left to family members or sometimes a dedicated matchmaker. One common form of an arranged marriage is a consensual arranged marriage where both the bride and groom approve of being brought together by their family members. Another less common form of an arranged marriage is a forced marriage, where both bride and groom are given no choice in the matter. However, since this form of an arranged marriage deprives both the bride and groom of their liberty and autonomy, the United Nation considers it a serious human rights violation.

If you choose your career path wisely, a degree in art and design can offer many work opportunities.
For example, with 10 years' experience in art and design you can be a creative director, with a median annual salary of 103,000 US dollars.

The Spanish Constitution of 1978 restored the monarchy and the King and Queen sit on the throne. After they were reinstated, the press met with the King and Queen to talk about the power of the monarchy. The journalists arrived at the venue before the King and Queen and waited patiently for them to arrive. Once the press conference started, the King and Queen answered all of their questions in a very transparent way. Because of their complete openness and transparency, the King and Queen are well liked in Spain.
Different versions of the story of Adam and Eve appear in different religious texts. According to the Bible, God created Adam and Eve and then placed them in Eden.

Before the American Civil War, the relationship between the north and south was not very stable due to a dispute over taxes. It was also apparent that the north and south had different views on slavery. Starting in 1861 the Civil War raged across the north and south of the United Statesfor about four years. Over five hundred thousand people in the north and south died as a result of the war. However, an important outcome of the war between the north and south is that it brought slavery to an end.

Miller's Death of a Salesman provides a powerful portrayal of success and failure in pursuit of the 'American dream'. Miller's depiction of the various characters demonstrates how success and failure is achieved.

Their musical band is made up of brothers and sisters who live in San Francisco, which has a vibrant music scene. The band was started by the brothers and sisters in late 2002 when they were all in their twenties. Every Saturday night the brothers and sisters performed their alternative rock music in locations all over the city. However, recently the band has become less popular

and the brothers and sisters have not received many requests to perform. It's probably because the style of music the brothers and sisters play has become less popular.

In 1792, the United States started adapting a gold and silver monetary system that was linked to the value of the dollar. However, converting the dollar into assets such as gold and silver was suspended by President Nixon in 1971.

The difference in pay between men and women has attracted a lot of media attention lately. The BBC reported that the gender pay gap between men and women has reached an extreme level in the UK, with three quarters of firms paying higher wages to their male staff. Airline companies and banks account for the greatest pay gap between men and women in the UK. While in many cases these figures reflect differences in pay between men and women who perform different work,it still raises questions about discrimination. For example, there could be a gender bias in the recruitment process, which places men and women in positions that differ in rank.

In this era of social media, texting, and blogs, how students engage with reading and writing is changing at a rapid pace. In the digital age, students do more reading and writing outside of the classroom than they do in it.

During Ramadhan, Muslims all over the world do not have any food and drink from dawn until dusk. Depending on the region and season, the amount of time that people have to refrain from partaking in food and drink differs greatly. For example, this year people in Argentina had to refrain from food and drink for 10 hours. In contrast, in Iceland they had to avoid food and drink for over 20 hours. Hopefully, the fact that Iceland is usually cold makes refraining from food and drink more endurable, but it is undoubtedly extremely challenging.

Worldwide the gap in wealth between the rich and poor has widened considerably. A recent economic report shows that inequality between the rich and poor is the most extreme in the US and Russia.

There is still considerable debate about whether behavioural differences between boys and girls are determined by genetic or environmental factors, or both. Psychologists from the City University of London compared boys and girls in terms of their toy preferences. The study showed that in all age groups boys and girls tended to select gender specific toys. Crucially, as the age of the children increased, the boys and girls selected gendered toys at greater rates. The study doesn't conclusively show whether the choices made by boys and girls are biologically determined or socially ingrained from an early age.

As a financial advisor, I often receive requests from clients to prepare a profit and loss statement that projects areas of growth as well as any potential risks. They use my profit and loss calculations to help them make key business decisions.

The Pew Research Center revealed interesting statistics about beliefs in heaven and hell in their 2016 survey. The survey asked 35,071 adult Americans about

their beliefs in heaven and hell in telephone interviews. Roughly 72 percent and 58 percent of respondents reported believing in heaven and hell respectively. The belief that there is a heaven and hell was widely accepted by both Christians and Muslims in the survey. People who did not report believing in heaven and hell were mainly those who reported not following any religion.

A world-famous photographer described what he considered to be the top locations for sunrise and sunset photos. Topping his list as the world's most impressive sunrise and sunset viewing spot, is Mount Bromo, in Indonesia.

In English literature, at the end of the 16th century combining poetry and prose became a popular form of writing. Shakespeare's plays combined poetry and prose to create works that have stood the test of time. The use of elements of both poetry and prose was evident in most of Shakespeare's early comedies. Later, Shakespeare began to exploit the interplay between poetry and prose in some of his tragedies, such as Hamlet. For example, Shakespeare employed literary devices from both poetry and prose to convey Hamlet's distress.

Muslims believe that God has sent prophets and messengers to the world at different times in history. Muslims believe that prophets and messengers are sent to the people to preach similar religious beliefs.

One non-violent act of political protest is to reject eating and drinking and not give in to hunger and thirst until a set of demands are met. Sometimes, protestors risk dying of hunger and thirst to bring public attention to injustice or to provoke sympathy from an offending party. If the participants enduring hunger and thirst are prisoners, they could be subjected to force-feeding by prison authorities. However, in 1975 the World Medical Association declared that if the prisoner is capable of making a rational judgment about the consequences of his hunger and thirst strike, doctors must not engage in force-feeding. In other words, one's personal choice of enduring hunger and thirst should be respected, and any forms of force-feeding should be rejected.

The observation of eclipses has led to interesting stories about the relationship between the sun and moon in different cultures. In some aboriginal Australian cultures, the eclipse was seen as an event where the sun and moon procreate.

A famous travel blogger recently wrote that travelling by land and sea is the only way to see the world. Tourists who travel by land and sea enjoy not only the destinations, but also the beauty between them. Travellers who go by land and sea can spontaneously stop and explore interesting places and experience new cultures. In addition, travelling by land and sea usually involves less strict security procedures than at airports. A major drawback of travelling by land and sea is that it requires more time off work, which isn't always possible.

Historically, Bedouin tribes in Arabia mainly depended on camels and sheep as a source of food. Not many animals are suited to such a harsh climate as the desert, but camels and sheep were able to survive.

Kabsah is a traditional rice and meat dish in Saudi Arabia. While the rice and meat Kabsah is often associated with Saudi Arabia, it is also served in other Arabic countries.It is made by cooking rice and meat with a special mixture of spices. A popular way of cooking Kabsah is by grilling rice and meat in a deep covered hole in the ground. Another way to prepare Kabsah is to cook seasoned rice and meat on a flat burning stone.

The view of man as a social animal has made studying the relationship between the individual and society an important area of research in sociology. Sociologists have made different classifications based on the role played by the individual and society in establishing collective shared values.

There are many shared religious beliefs amongst Jews and Christians that make the two similar in many ways. Both Jews and Christians share important religious figures, such as Abraham and Jesus, and Moses. In addition, both Jews and Christians consider Jerusalem to be a holy place. Beliefs about the afterlife and 'judgment day' are similar for Jews and Christians with both believing in an eternal life after death. It is also a common belief for both Jews and Christians that mankind and the universe were created by God.

Guides for travellers to Saudi Arabia say that coffee and dates are more than just a snack. In fact, serving coffee and dates is an important aspect of hospitality in the Saudi Arabian culture.

Students should be informed about the costs and benefits of student loans in order to make the best decisions about how to finance their education. Universities should also be involved in evaluating the costs and benefits of student loans and helping students make good choices. In many cases, students leave university 27,000 pounds in debt, which means that re-examining the costs and benefits of student loans is an urgent issue. One way to look at the costs and benefits and potential solutions to the growing debt would be to consider how university education is financed in different countries. For example, looking at the situation in Australia might show how costs and benefits could be addressed differently in England.

The Chicago Alzheimer's Association recently released a YouTube video detailing various facts and figures about Alzheimer's. Their set of facts and figures showed that in America, 5.7 million people have been diagnosed with Alzheimer's.

Asking job candidates about their strengths and weaknesses is very common in interviews. When candidates are asked about their strengths and weaknesses in interviews, they are always advised to relate the answer to the job description. Another recommendation is to frame both their strengths and weaknesses in ways that could be considered positive. Turning negatives into positives when answering questions about strengths and weaknesses is always a good thing to keep in mind. When faced with a question about strengths and weaknesses, it is also important to emphasize your skills and to show how you are working on addressing

some of your limitations.

In Egypt, the division between past and present is apparent almost immediately. In fact, we can visibly see past and present divided by a few kilometres of land as the pyramids sit alongside Cairo's skyscrapers.

The relationship between a seller and buyer should be based on trust and mutual understanding. However, legal documents should be signed by both the seller and buyer in order to prevent any future conflict. A sale contract is an official document in which the seller and buyer reach mutual agreement. This contract is signed by both the seller and buyer in order to save both their legal rights. Signing the contract ensures that transfer of ownership between the seller and buyer goes as smoothly as possible.

A recent medical study showed that eating figs and olives has many health benefits. The research indicated that figs and olives promote metabolism and increase energy.

Appendix 4E: Model Outcome of Accuracy Scores for the Control Items

	Accuracy for control items			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	-16.18	7.22	-2.24	**0.025**
Group [NS]	1.06	0.14	7.55	**<0.001**
Item Type [Congruent]				
Item Type [English]				
Session [Post-test]	-0.11	0.11	-1.00	0.319
Session [Delayed test]	-0.14	0.11	-1.27	0.204
Familiarity ratings - NNS	0.82	1.13	0.73	0.467
Familiarity ratings - NS	3.01	1.47	2.05	**0.041**
Group [NS] * Item Type [Congruent]				
Group [NS] * Item Type [English]				
Item Type [Congruent] * Session [Post-test]				
Item Type [English] * Session [Post-test]				
Item Type [Congruent] * Session [Delayed test]				
Item Type [English] * Session [Delayed test]				

Random Effects

σ^2	3.29
τ_{00}	0.09 $_{subject}$
	0.41 $_{Item}$

ICC	0.13

Appendix 4F: Analysis of Accuracy as a Function of Repetition

Means and Standard deviations of Accuracy Scores by Repetition

Group	Repetition	Item Type					
		Arabic		Congruent		English	
		M	*SD*	*M*	*SD*	*M*	*SD*
NNS	2	9.81	1.94	11.1	2.27	9.55	2.95
NNS	5	10.10	2.28	11.00	2.25	9.87	2.07
NS	2	9.83	1.98	13.20	1.20	13.60	0.68
NS	5	10.30	2.09	13.10	1.12	13.50	0.76

Note. The maximum score was 14.

Model Outcome of Accuracy Scores as a Function of Repetition

	Accuracy			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	-3.59	2.24	-1.60	0.110
Group [NS]	0.04	0.17	0.22	0.822
Item Type [Cong]	-0.09	0.25	-0.35	0.726
Item Type [Eng]	-0.39	0.27	-1.44	0.150
Time of Testing [Delayed test]	-0.38	0.11	-3.43	**0.001**
Reps [5]	0.08	0.07	1.14	0.256
Familiarity ratings NNS	0.81	0.56	1.45	0.147
Familiarity ratings NS	0.37	0.13	2.91	**0.004**
Group [NS] * Item Type [Cong]	1.35	0.23	5.90	**<0.001**
Group [NS] * Item Type [Eng]	2.61	0.28	9.21	**<0.001**
Item Type [Cong] * Time of Testing [Delayed test]	0.40	0.17	2.32	**0.020**

Item_Type [Eng] * Time_of_Testing [Delayedtest]	0.29	0.16	1.78	0.076

Random Effects

σ^2	3.29
$\tau_{00\ subject}$	0.18
$\tau_{00\ Item}$	0.26
ICC	0.12
$N_{subject}$	65
N_{Item}	42
Observations	5459
Marginal R^2 / Conditional R^2	0.180 / 0.278

Appendix 4G: Analysis of RT as a Function of Test Session for Experimental and Control Items

	Log RT			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	7.23	0.15	48.74	**<0.001**
Group [NS]	-0.74	0.06	-12.86	**<0.001**
Item Type [Ar]	0.04	0.04	1.01	0.310
Item Type [Cong]	-0.18	0.03	-5.12	**<0.001**
Item Type [Eng]	-0.09	0.04	-2.70	**0.007**
Session [Post-test]	-0.20	0.02	-10.33	**<0.001**
Session [Delayed test]	-0.35	0.02	-18.13	**<0.001**
Length [log]	0.35	0.05	6.49	**<0.001**
Item Type [Ar] * Session [Post-test]	-0.13	0.03	-4.64	**<0.001**
Item Type [Cong] * Session [Post-test]	-0.07	0.03	-2.76	**0.006**
Item Type [Eng] * Session [Post-test]	-0.06	0.03	-2.27	**0.023**
Item Type [Ar] * Session [Delayed test]	-0.01	0.03	-0.40	0.689
Item Type [Cong] * Session [Delayed test]	0.01	0.03	0.34	0.733
Item Type [Eng] * Session [Delayed test]	0.03	0.03	0.97	0.334
Random Effects				
σ^2	0.11			
$\tau_{00\ subject}$	0.04			
$\tau_{00\ Item}$	0.01			
ICC	0.30			

Appendix 4H: Analysis of RT as a Function of Repetition

	Log RT			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*
(Intercept)	6.89	0.17	40.68	**<0.001**
Group [NS]	-0.56	0.07	-8.47	**<0.001**
Time of Testing [Delayed test]	-0.05	0.01	-4.50	**<0.001**
Item Type [Cong]	-0.16	0.03	-4.92	**<0.001**
Item Type [Eng]	0.01	0.03	0.17	0.866
Length [log]	0.36	0.06	5.78	**<0.001**
Reps [5]	-0.01	0.01	-1.35	0.179
Group [NS] * Time of Testing [Delayed test]	-0.00	0.02	-0.04	0.964
Group [NS] * Item Type [Cong]	-0.07	0.03	-2.65	**0.008**
Group [NS] * Item Type [Eng]	-0.28	0.03	-10.24	**<0.001**
Random Effects				
σ^2	0.10			
$\tau_{00\ subject}$	0.05			
$\tau_{00\ Item}$	0.01			
ICC	0.36			

Appendix 5A: List of the Novel Collocations Used in Study 2

Items	Item Type
scared tone	LC
social surveillance	HC
deceptive graph	LC
private cabin	HC
amazing sightseeing	LC
exhausted survivor	LC
important chat	HC
thick pillars	LC
frightened baby	LC
excited spectators	LC
political caricaturists	HC
armed kidnapper	HC
rural jungles	HC
brilliant salesman	LC
foreign vendors	HC
personal airplane	HC
illustrative visuals	LC
pregnant minors	HC
respected poet	LC
minimum fine	HC
financial tragedy	HC
large theatre	HC
persuasive critique	LC
brief diary	HC
wide sidewalk	HC
grateful bride	LC
detailed brochure	HC
ambiguous lyrics	LC
supportive teammates	LC
physical clashes	HC
evident talent	LC
relaxed rapport	LC
nuclear battery	HC
deep pond	HC
front patio	HC
old laptop	HC
tempting pastries	LC
awful backache	LC
perplexing plot	LC
noticeable flaws	LC
disabled tourists	HC

local souvenirs	HC
young bloggers	HC
severe tornado	HC
cynical viewers	LC
superficial bruises	LC
constant horror	LC
disputable legality	LC

Appendix 5B: The Novel Collocations Presented in Sentences (Study 2)

Sentences	Block	List
Officials promised that pregnant minors would receive more comprehensive health care.	1	1
Buying a personal airplane was the pilot's dream.	1	1
Many apps that are meant to connect people allow for social surveillance that may not be easily prevented.	1	1
Rival groups protested outside of Parliament but no physical clashes were reported to the police.	1	1
The government is planning to impose more tax regulations on foreign vendors next year.	1	1
They designed a detailed brochure to highlight their start-up company's mission.	1	1
Our city has a large theatre that often features Broadway plays.	1	1
Isabel kept a brief diary in her chest of drawers where she documented major life events.	1	1
Jim owns a private cabin with an outdoor jacuzzi.	1	1
The new WhatsApp update enables users to pin an important chat to the top of the window.	1	1
Noah paid only the minimum fine for speeding after his appeal in court.	1	1
At the recent press conference, the government promised to prevent another financial tragedy by adjusting its interest rates.	1	1
Some of the techniques political caricaturists use to convey their messages are symbolism and analogy.	1	1
Leading up to the Eifel tower is a wide sidewalk that is always full of people.	1	1
The newspaper reported that an armed kidnapper took three hostages near the US-Mexico border.	1	1
The CNN showed that the fungus attacking bananas swept across the rural jungles of Central and South America.	1	1
After being trapped in a collapsed mine for days the exhausted survivor was anxious to tell his story.	1	1
The pianist said that her evident talent was inherited from her mother.	1	1
Robert replied with a scared tone when he heard the news.	1	1
The album generated a lot of controversy because of the ambiguous lyrics that could be interpreted as hate speech.	1	1
His grandfather was a respected poet in the nineteen fifties.	1	1
Before he was promoted to manager, Bill was a brilliant salesman at the company.	1	1
The teacher and students developed a relaxed rapport that positively impacted the learning environment.	1	1
The professor offered a persuasive critique of the opposition's view.	1	1
The newlyweds explored the amazing sightseeing opportunities in India on their honeymoon.	1	1
Stella tried to comfort her frightened baby who cried during the whole flight.	1	1
When they were kids, they played handball against the thick pillars in the town's square.	1	1
Morris knows that he wouldn't have gotten to this level without having supportive teammates around him over the years.	1	1
Ashley promised Samuel to remain his grateful bride for the rest of her life.	1	1
The book comes with illustrative visuals highlighting the main ideas.	1	1

There were more excited spectators attending the game than security was prepared for.	1	1
In their statistics class, their teacher demonstrated how a deceptive graph can be used to convey a misleading message.	1	1
The article demonstrated some cases of pregnant minors who experienced bullying at school.	2	1
The footballer's personal airplane can fit up to six people.	2	1
The Guardian showed how the internet is used for social surveillance by big tech companies.	2	1
After the security forces attacked one of the demonstrators, physical clashes broke out all over the city.	2	1
Getting the components for their product depends on foreign vendors from many different countries.	2	1
After working hard to make a detailed brochure for the event, the event manager decided to go with a poster instead.	2	1
Our hotel was just next to a large theatre that made it difficult to find a parking spot.	2	1
When he was learning English, he carried a brief diary where he wrote down the new words he came across.	2	1
Liam and Lucy were trying to find a private cabin to rent in the forest.	2	1
Debra had an important chat about fire safety with her son.	2	1
In some cases, the minimum fine for driving without insurance is £500.	2	1
There have been many debates in the media about who to blame for the financial tragedy that broke out recently.	2	1
He is one of the great political caricaturists in the country and is considered to be a true artist.	2	1
Tracey always goes jogging on the wide sidewalk that runs along the park.	2	1
Everyone hoped that the armed kidnapper would be found and arrested.	2	1
The criminal was hiding in the rural jungles of Brazil and Peru for months before he got caught.	2	1
The detectives interviewed the exhausted survivor for 3 hours.	2	1
Alison has an evident talent for playing a variety of musical instruments.	2	1
Joshua spoke with a scared tone when he was questioned about the theft.	2	1
Any violent or even ambiguous lyrics were banned from the radio station.	2	1
Isaac quoted the respected poet John Donne at the beginning of his speech.	2	1
They knew Ed would be a brilliant salesman from the moment they interviewed him.	2	1
Albert and Elina have had a relaxed rapport for the whole time they have worked together.	2	1
Katelyn published a persuasive critique of some prominent theories in language acquisition.	2	1
The information about Brazil showed the amazing sightseeing that was available all over the country.	2	1
After seeing a big dog the frightened baby cried for the rest of the afternoon.	2	1
Ben and his friends loved climbing up the thick pillars near the main library.	2	1
Mohammad felt so lucky to be surrounded by such supportive teammates in his department.	2	1
After the celebration, the grateful bride bought a nice gift for her new mother-in-law.	2	1
The teacher utilizes many illustrative visuals that contribute a lot to the interactive learning environment.	2	1
The stadium was full of excited spectators who cheered and chanted for the whole match.	2	1
Due to a deceptive graph in the article, Jamie drew an incorrect conclusion.	2	1
Schools offer some programs to provide support for pregnant minors who are not living with their parents.	3	1
Charlie bought a personal airplane for about a million pounds.	3	1
George Orwell discussed social surveillance in his novel "1984".	3	1

There were lots of physical clashes amongst the players in the championship game.	3	1
The new law permits foreign vendors to obtain multi-entry visas.	3	1
Our manger asked us to prepare a detailed brochure highlighting our advertising strategy for the board of directors.	3	1
The Arena of Verona is a large theatre in Italy.	3	1
Alexander used the nickname Big Mac when writing a brief diary about himself.	3	1
Lewis plans to move into his private cabin after retirement.	3	1
Antonio had an important chat with his flatmate about smoking inside the flat.	3	1
Each person responsible for the vandalism paid the minimum fine of £10,000.	3	1
Many people believe that the financial tragedy of 2008 was the worst since the great depression.	3	1
There is a high demand for political caricaturists to work at the new magazine.	3	1
Glen and Lauren met at the coffee shop located on the wide sidewalk in front of the memorial statue.	3	1
The woman was able to successfully escape from the armed kidnapper before the police had time to help her.	3	1
The biologist went to collect some samples from the rural jungles in the Amazon Rainforest.	3	1
All of the reporters surrounded the exhausted survivor wanting to be the first to get his story.	3	1
Julie has been unemployed for more than a year despite possessing evident talent as a cook.	3	1
Gabriella had a scared tone when reporting on the search for the missing girl.	3	1
Many people believe that the ambiguous lyrics in the popular song mask a more serious message.	3	1
Daniel still couldn't believe the fact that he had become a respected poet and his work was being translated into many languages.	3	1
The brilliant salesman achieved his monthly target in just a week.	3	1
She had a relaxed rapport with her colleague that eventually turned into a great friendship.	3	1
The student presented a persuasive critique of the novel.	3	1
The Hawaiian Islands have many amazing sightseeing destinations for nature lovers.	3	1
After hearing the loud crashing sound the frightened baby refused to be put down for a nap.	3	1
During their trip to Athens, they got a room with a view of the thick pillars of ancient Greece.	3	1
The new employee felt that her colleagues were not being supportive teammates but hoped that they would be after they got to know her.	3	1
I couldn't be successful without the love of my grateful bride Anna, said Alex.	3	1
The many illustrative visuals that the presenter used captured the attention of the audience.	3	1
A mass of excited spectators crowded onto the pitch at the end of the match.	3	1
Sometimes people accidentally produce a deceptive graph by changing the scale on the axis.	3	1
A State Health official stated that pregnant minors should be fully informed about their options.	4	1
For the rich and famous a personal airplane is a necessary status symbol.	4	1
The conference we attended on Wednesday focused on social surveillance in the digital world.	4	1
The Head Teacher said that bullying is the main cause of many of the physical clashes in the school.	4	1
Many people transfer funds to foreign vendors using PayPal services.	4	1
According to the detailed brochure we picked up at the resort's reception, we are not allowed to stay on the beach after midnight.	4	1
Ramon was really excited about the large theatre being built in the city.	4	1

Ted enjoyed looking at the brief diary he used to keep when he was a child.	4	1
They just bought a private cabin that is 250 km away from the city.	4	1
Peter couldn't find an important chat he had on the app a few weeks ago.	4	1
The officer said that minimum fine doubles from £70 to £140 if it is not paid within 14 days.	4	1
Ethan was lucky to find a good job after graduation despite the financial tragedy that hit the country.	4	1
There are a lot of political caricaturists that use various online platforms to showcase their work.	4	1
Nelson was upset when people parked their cars on the wide sidewalk because it created a hazard for his wheelchair.	4	1
It was reported that the armed kidnapper forced the victim to drive him to the airport.	4	1
Environmentalists are worried about species extinction in the rural jungles of Southeast Asia.	4	1
The doctor said that the exhausted survivor should recover in a couple of days.	4	1
The performer demonstrated an evident talent as a ballet dancer.	4	1
Florence was so amazed that her newborn recognized her scared tone and immediately knew to stop crying.	4	1
The musician said that the reason the band members write ambiguous lyrics is because they want listeners to reflect their meaning.	4	1
This year, the Nobel Prize for literature went to a widely respected poet who is also a peace activist.	4	1
He quickly established himself as a brilliant salesman for the giant pharmaceutical company.	4	1
Philip likes to maintain a relaxed rapport with his clients.	4	1
Will was learning how to write a persuasive critique in his composition class.	4	1
They spent the day doing some amazing sightseeing in Paris and were ready for a break in a cafe.	4	1
It took Alexia a month to realize that her frightened baby was scared of the dark.	4	1
His photo of waves hitting the thick pillars in Taiwan got more than 5 million likes on Instagram.	4	1
Mark thinks that the rugby squad's competitive nature prevents them from being supportive teammates for each other.	4	1
When the expensive venue flooded the day before the celebration, the grateful bride was so happy that the church offered her their hall.	4	1
Adam has been working on making an app that enhances text with illustrative visuals that transforms it into an infographic picture.	4	1
About thirty-thousand excited spectators packed into the stadium at the yesterday's match.	4	1
The TV channel apologized for running a deceptive graph to promote the presidential candidate's campaign.	4	1
It has been hypothesised that a nuclear battery could be used to run cars in the future.	1	2
There is a deep pond in the middle of our village's park where many children go swimming.	1	2
Leo likes to drink his morning coffee on his front patio every day.	1	2
Aaron couldn't believe that his old laptop still worked really well.	1	2
Brian was happy that he found a guide for disabled tourists on the Trip Advisor website.	1	2
One of the best places to buy local souvenirs in London is the Camden Lock Market.	1	2
George attended an event that showcased a couple of the young bloggers at the University.	1	2
The Weather Service put out a warning for a severe tornado over the weekend in our area.	1	2
Jim owns a private cabin with an outdoor jacuzzi.	1	2
The new Whatsapp update enables users to pin an important chat to the top of the window.	1	2
Noah paid only the minimum fine for speeding after his appeal in court.	1	2

At the recent press conference, the government promised to prevent another financial tragedy by adjusting its interest rates.	1	2
Some of the techniques political caricaturists use to convey their messages are symbolism and analogy.	1	2
Leading up to the Eifel tower is a wide sidewalk that is always full of people.	1	2
The newspaper reported that an armed kidnapper took three hostages near the US-Mexico border.	1	2
The CNN showed that the fungus attacking bananas swept across the rural jungles of Central and South America.	1	2
Claude couldn't resist the tempting pastries even though he was on a diet.	1	2
Robin has had an awful backache for more than two weeks.	1	2
Despite the perplexing plot in the novel, the students really enjoyed reading it.	1	2
The review committee spotted many noticeable flaws in the in the paper Clark wrote.	1	2
The news programme targets mainly cynical viewers aged 20 to 35.	1	2
Luckily, Scott just had some superficial bruises after falling off of his bike.	1	2
He applied for a refugee visa to flee the constant horror of war in his country.	1	2
Marshal refused the company's job offer because of the disputable legality of some of its projects.	1	2
The newlyweds explored the amazing sightseeing opportunities in India on their honeymoon.	1	2
Stella tried to comfort her frightened baby who cried during the whole flight.	1	2
When they were kids, they played handball against the thick pillars in the town's square.	1	2
Morris knows that he wouldn't have gotten to this level without having supportive teammates around him over the years.	1	2
Ashley promised Samuel to remain his grateful bride for the rest of her life.	1	2
The book comes with illustrative visuals highlighting the main ideas.	1	2
There were more excited spectators attending the game than security was prepared for.	1	2
In their statistics class, their teacher demonstrated how a deceptive graph can be used to convey a misleading message.	1	2
Researchers have invented a nuclear battery that could last for years.	2	2
Bill took off his ring and threw it into the deep pond after he had got divorced.	2	2
They were out on their front patio when their house was robbed.	2	2
Alexis had to throw out her old laptop because she wasn't able to sell it.	2	2
Many people volunteered to help disabled tourists visiting the historic village.	2	2
Dan had a hard time finding local souvenirs when he was in Leeds.	2	2
The majority of young bloggers are shifting away from online platforms such as Facebook and Twitter.	2	2
They were watching a documentary about a severe tornado that rampaged through 38 miles of eastern Asia.	2	2
Liam and Lucy were trying to find a private cabin to rent in the forest.	2	2
Debra had an important chat about fire safety with her son.	2	2
In some cases, the minimum fine for driving without insurance is £500.	2	2
There have been many debates in the media about who to blame for the financial tragedy that broke out recently.	2	2
He is one of the great political caricaturists in the country and is considered to be a true artist.	2	2
Tracey always goes jogging on the wide sidewalk that runs along the park.	2	2
Everyone hoped that the armed kidnapper would be found and arrested.	2	2

The criminal was hiding in the rural jungles of Brazil and Peru for months before he got caught.	2	2
Susan went to the French restaurant mainly for the tempting pastries they made.	2	2
Craig has been suffering from an awful backache for a long time.	2	2
Because of the perplexing plot of the movie, Eli had to watch it again.	2	2
The movie had many noticeable flaws and the reviewers spotted them.	2	2
They were disappointed about the comments the cynical viewers posted on their YouTube video.	2	2
Rick only suffered superficial bruises to his arm and left wrist during the rugby match.	2	2
The BBC documentary showed the constant horror of global climate change.	2	2
Many people are worried about the disputable legality of the company's use of customer's data.	2	2
The information about Brazil showed the amazing sightseeing that was available all over the country.	2	2
After seeing a big dog the frightened baby cried for the rest of the afternoon.	2	2
Ben and his friends loved climbing up the thick pillars near the main library.	2	2
Mohammad felt so lucky to be surrounded by such supportive teammates in his department.	2	2
After the celebration, the grateful bride bought a nice gift for her new mother-in-law.	2	2
The teacher utilizes many illustrative visuals that contribute a lot to the interactive learning environment.	2	2
The stadium was full of excited spectators who cheered and chanted for the whole match.	2	2
Due to a deceptive graph in the article, Jamie drew an incorrect conclusion.	2	2
I just read an article that highlights how a nuclear battery can be a source of environmentally friendly energy.	3	2
Amelia was surprised to learn the deep pond dried up after the hot summer they had.	3	2
Having a front patio was a deal breaker for Arthur when he was trying to buy a house.	3	2
Joey decided to take his old laptop to an electronics recycler.	3	2
The World Health Organisation says that hotels and resorts should provide disabled tourists with easy access to all facilities.	3	2
The Howard family have got a great selection of local souvenirs and antiques in their living room.	3	2
During the Arab Spring, many young bloggers expressed viewpoints encouraging economic reform.	3	2
Last month a severe tornado caused significant damage to cars and buildings.	3	2
Lewis plans to move into his private cabin after retirement.	3	2
Antonio had an important chat with his flatmate about smoking inside the flat.	3	2
Each person responsible for the vandalism paid the minimum fine of £10,000.	3	2
Many people believe that the financial tragedy of 2008 was the worst since the great depression.	3	2
There is a high demand for political caricaturists to work at the new magazine.	3	2
Glen and Lauren met at the coffee shop located on the wide sidewalk in front of the memorial statue.	3	2
The woman was able to successfully escape from the armed kidnapper before the police had time to help her.	3	2
The biologist went to collect some samples from the rural jungles in the Amazon Rainforest.	3	2
Annabelle politely refused to reveal her recipe for the tempting pastries she bakes.	3	2
Ron woke up with an awful backache that did not go away until the next day.	3	2
Marlin thinks that the perplexing plot is confusing because of the numerous protagonists.	3	2
The new smart phone had a few noticeable flaws that meant sales were lower than projected.	3	2

Despite what cynical viewers are saying, the show was great.	3	2
Tommy was relieved that the superficial bruises healed very quickly.	3	2
Everyone in the town is in constant horror because of the on-going gang war.	3	2
Because of the disputable legality of the project, the CEO had to cancel it	3	2
The Hawaiian Islands have many amazing sightseeing destinations for nature lovers.	3	2
After hearing the loud crashing sound the frightened baby refused to be put down for a nap.	3	2
During their trip to Athens, they got a room with a view of the thick pillars of ancient Greece.	3	2
The new employee felt that her colleagues were not being supportive teammates but hoped that they would be after they got to know her.	3	2
I couldn't be successful without the love of my grateful bride Anna, said Alex.	3	2
The many illustrative visuals that the presenter used captured the attention of the audience.	3	2
A mass of excited spectators crowded onto the pitch at the end of the match.	3	2
Sometimes people accidentally produce a deceptive graph by changing the scale on the axis.	3	2
Russian scientists designed a nuclear battery that could power a spacecraft.	4	2
Freddie loved the deep pond in the Japanese garden in town.	4	2
They used their front patio mainly for keeping their rubbish bins.	4	2
Johnny was so happy that he was able to trade his old laptop for some gift cards.	4	2
In a recent survey, almost all disabled tourists were disappointed about the lack of accessible public toilets.	4	2
They started a family business making local souvenirs and selling them in the town square.	4	2
It's common for young bloggers to make money by placing ads on their sites.	4	2
Multiple injuries were recorded yesterday due to a severe tornado that hit different parts of the city.	4	2
They just bought a private cabin that is 250 km away from the city.	4	2
Peter couldn't find an important chat he had on the app a few weeks ago.	4	2
The officer said that minimum fine doubles from £70 to £140 if it is not paid within 14 days.	4	2
Ethan was lucky to find a good job after graduation despite the financial tragedy that hit the country.	4	2
There are a lot of political caricaturists that use various online platforms to showcase their work.	4	2
Nelson was upset when people parked their cars on the wide sidewalk because it created a hazard for his wheelchair.	4	2
It was reported that the armed kidnapper forced the victim to drive him to the airport.	4	2
Environmentalists are worried about species extinction in the rural jungles of Southeast Asia.	4	2
Katrina starts her day in a coffee shop having a cup of coffee and some tempting pastries while reading the news.	4	2
Harold was gardening until he got an awful backache which forced him to stop and rest.	4	2
The movie's perplexing plot features many twists and turns.	4	2
Despite a few noticeable flaws with his camera, the photographer still thinks Canon is the best company to buy a camera from.	4	2
The magician's trick on Britain's Got Talent was exposed by some cynical viewers on Facebook.	4	2
Neil had a couple of superficial bruises due to his recent car accident.	4	2
The island inhabitants lived in a constant horror because of the nearby volcano on their island.	4	2
The Head of the Legal Department warned the company about the disputable legality of some of the clauses in its contracts.	4	2

They spent the day doing some amazing sightseeing in Paris and were ready for a break in a cafe.	4	2
It took Alexia a month to realize that her frightened baby was scared of the dark.	4	2
His photo of waves hitting the thick pillars in Taiwan got more than 5 million likes on Instagram.	4	2
Mark thinks that the rugby squad's competitive nature prevents them from being supportive teammates for each other.	4	2
When the expensive venue flooded the day before the celebration, the grateful bride was so happy that the church offered her their hall.	4	2
Adam has been working on making an app that enhances text with illustrative visuals that transforms it into an infographic picture.	4	2
About thirty-thousand excited spectators packed into the stadium at the yesterday's match.	4	2
The TV channel apologized for running a deceptive graph to promote the presidential candidate's campaign.	4	2
It has been hypothesised that a nuclear battery could be used to run cars in the future.	1	3
There is a deep pond in the middle of our village's park where many children go swimming.	1	3
Leo likes to drink his morning coffee on his front patio every day.	1	3
Aaron couldn't believe that his old laptop still worked really well.	1	3
Brian was happy that he found a guide for disabled tourists on the Trip Advisor website.	1	3
One of the best places to buy local souvenirs in London is the Camden Lock Market.	1	3
George attended an event that showcased a couple of the young bloggers at the University.	1	3
The Weather Service put out a warning for a severe tornado over the weekend in our area.	1	3
Officials promised that pregnant minors would receive more comprehensive health care.	1	3
Buying a personal airplane was the pilot's dream.	1	3
Many apps that are meant to connect people allow for social surveillance that may not be easily prevented.	1	3
Rival groups protested outside of Parliament but no physical clashes were reported to the police.	1	3
The government is planning to impose more tax regulations on foreign vendors next year.	1	3
They designed a detailed brochure to highlight their start-up company's mission.	1	3
Our city has a large theatre that often features Broadway plays.	1	3
Isabel kept a brief diary in her chest of drawers where she documented major life events.	1	3
Claude couldn't resist the tempting pastries even though he was on a diet.	1	3
Robin has had an awful backache for more than two weeks.	1	3
Despite the perplexing plot in the novel, the students really enjoyed reading it.	1	3
The review committee spotted many noticeable flaws in the in the paper Clark wrote.	1	3
The news programme targets mainly cynical viewers aged 20 to 35.	1	3
Luckily, Scott just had some superficial bruises after falling off of his bike.	1	3
He applied for a refugee visa to flee the constant horror of war in his country.	1	3
Marshal refused the company's job offer because of the disputable legality of some of its projects.	1	3
After being trapped in a collapsed mine for days the exhausted survivor was anxious to tell his story.	1	3
The pianist said that her evident talent was inherited from her mother.	1	3
Robert replied with a scared tone when he heard the news.	1	3
The album generated a lot of controversy because of the ambiguous lyrics that could be interpreted as hate speech.	1	3

His grandfather was a respected poet in the nineteen fifties.	1	3
Before he was promoted to manager, Bill was a brilliant salesman at the company.	1	3
The teacher and students developed a relaxed rapport that positively impacted the learning environment.	1	3
The professor offered a persuasive critique of the opposition's view.	1	3
Researchers have invented a nuclear battery that could last for years.	2	3
Bill took off his ring and threw it into the deep pond after he had got divorced.	2	3
They were out on their front patio when their house was robbed.	2	3
Alexis had to throw out her old laptop because she wasn't able to sell it.	2	3
Many people volunteered to help disabled tourists visiting the historic village.	2	3
Dan had a hard time finding local souvenirs when he was in Leeds.	2	3
The majority of young bloggers are shifting away from online platforms such as Facebook and Twitter.	2	3
They were watching a documentary about a severe tornado that rampaged through 38 miles of eastern Asia.	2	3
The article demonstrated some cases of pregnant minors who experienced bullying at school.	2	3
The footballer's personal airplane can fit up to six people.	2	3
The Guardian showed how the internet is used for social surveillance by big tech companies.	2	3
After the security forces attacked one of the demonstrators physical clashes broke out all over the city.	2	3
Getting the components for their product depends on foreign vendors from many different countries.	2	3
After working hard to make a detailed brochure for the event, the event manager decided to go with a poster instead.	2	3
Our hotel was just next to a large theatre that made it difficult to find a parking spot.	2	3
When he was learning English, he carried a brief diary where he wrote down the new words he came across.	2	3
Susan went to the French restaurant mainly for the tempting pastries they made.	2	3
Craig has been suffering from an awful backache for a long time.	2	3
Because of the perplexing plot of the movie, Eli had to watch it again.	2	3
The movie had many noticeable flaws and the reviewers spotted them.	2	3
They were disappointed about the comments the cynical viewers posted on their YouTube video.	2	3
Rick only suffered superficial bruises to his arm and left wrist during the rugby match.	2	3
The BBC documentary showed the constant horror of global climate change.	2	3
Many people are worried about the disputable legality of the company's use of customer's data.	2	3
The detectives interviewed the exhausted survivor for 3 hours.	2	3
Alison has an evident talent for playing a variety of musical instruments.	2	3
Joshua spoke with a scared tone when he was questioned about the theft.	2	3
Any violent or even ambiguous lyrics were banned from the radio station.	2	3
Isaac quoted the respected poet John Donne at the beginning of his speech.	2	3
They knew Ed would be a brilliant salesman from the moment they interviewed him.	2	3
Albert and Elina have had a relaxed rapport for the whole time they have worked together.	2	3
Katelyn published a persuasive critique of some prominent theories in language acquisition.	2	3
I just read an article that highlights how a nuclear battery can be a source of environmentally friendly energy.	3	3

Amelia was surprised to learn the deep pond dried up after the hot summer they had.	3	3
Having a front patio was a deal breaker for Arthur when he was trying to buy a house.	3	3
Joey decided to take his old laptop to an electronics recycler.	3	3
The World Health Organisation says that hotels and resorts should provide disabled tourists with easy access to all facilities.	3	3
The Howard family have got a great selection of local souvenirs and antiques in their living room.	3	3
During the Arab Spring, many young bloggers expressed viewpoints encouraging economic reform.	3	3
Last month a severe tornado caused significant damage to cars and buildings.	3	3
Schools offer some programs to provide support for pregnant minors who are not living with their parents.	3	3
Charlie bought a personal airplane for about a million pounds.	3	3
George Orwell discussed social surveillance in his novel “1984”.	3	3
There were lots of physical clashes amongst the players in the championship game.	3	3
The new law permits foreign vendors to obtain multi-entry visas.	3	3
Our manger asked us to prepare a detailed brochure highlighting our advertising strategy for the board of directors.	3	3
The Arena of Verona is a large theatre in Italy.	3	3
Alexander used the nickname Big Mac when writing a brief diary about himself.	3	3
Annabelle politely refused to reveal her recipe for the tempting pastries she bakes.	3	3
Ron woke up with an awful backache that did not go away until the next day.	3	3
Marlin thinks that the perplexing plot is confusing because of the numerous protagonists.	3	3
The new smart phone had a few noticeable flaws that meant sales were lower than projected.	3	3
Despite what cynical viewers are saying, the show was great.	3	3
Tommy was relieved that the superficial bruises healed very quickly.	3	3
Everyone in the town is in constant horror because of the on-going gang war.	3	3
Because of the disputable legality of the project, the CEO had to cancel it	3	3
All of the reporters surrounded the exhausted survivor wanting to be the first to get his story.	3	3
Julie has been unemployed for more than a year despite possessing evident talent as a cook.	3	3
Gabriella had a scared tone when reporting on the search for the missing girl.	3	3
Many people believe that the ambiguous lyrics in the popular song mask a more serious message.	3	3
Daniel still couldn’t believe the fact that he had become a respected poet and his work was being translated into many languages.	3	3
The brilliant salesman achieved his monthly target in just a week.	3	3
She had a relaxed rapport with her colleague that eventually turned into a great friendship.	3	3
The student presented a persuasive critique of the novel.	3	3
Russian scientists designed a nuclear battery that could power a spacecraft.	4	3
Freddie loved the deep pond in the Japanese garden in town.	4	3
They used their front patio mainly for keeping their rubbish bins.	4	3
Johnny was so happy that he was able to trade his old laptop for some gift cards.	4	3
In a recent survey, almost all disabled tourists were disappointed about the lack of accessible public toilets.	4	3
They started a family business making local souvenirs and selling them in the town square.	4	3

It’s common for young bloggers to make money by placing ads on their sites.	4	3
Multiple injuries were recorded yesterday due to a severe tornado that hit different parts of the city.	4	3
A State Health official stated that pregnant minors should be fully informed about their options.	4	3
For the rich and famous a personal airplane is a necessary status symbol.	4	3
The conference we attended on Wednesday focused on social surveillance in the digital world.	4	3
The Head Teacher said that bullying is the main cause of many of the physical clashes in the school.	4	3
Many people transfer funds to foreign vendors using PayPal services.	4	3
According to the detailed brochure we picked up at the resort’s reception, we are not allowed to stay on the beach after midnight.	4	3
Ramon was really excited about the large theatre being built in the city.	4	3
Ted enjoyed looking at the brief diary he used to keep when he was a child.	4	3
Katrina starts her day in a coffee shop having a cup of coffee and some tempting pastries while reading the news.	4	3
Harold was gardening until he got an awful backache which forced him to stop and rest.	4	3
The movie’s perplexing plot features many twists and turns.	4	3
Despite a few noticeable flaws with his camera, the photographer still thinks Canon is the best company to buy a camera from.	4	3
The magician’s trick on Britain's Got Talent was exposed by some cynical viewers on Facebook.	4	3
Neil had a couple of superficial bruises due to his recent car accident.	4	3
The island inhabitants lived in a constant horror because of the nearby volcano on their island.	4	3
The Head of the Legal Department warned the company about the disputable legality of some of the clauses in its contracts.	4	3
The doctor said that the exhausted survivor should recover in a couple of days.	4	3
The performer demonstrated an evident talent as a ballet dancer.	4	3
Florence was so amazed that her newborn recognized her scared tone and immediately knew to stop crying.	4	3
The musician said that the reason the band members write ambiguous lyrics is because they want listeners to reflect their meaning.	4	3
This year, the Nobel Prize for literature went to a widely respected poet who is also a peace activist.	4	3
He quickly established himself as a brilliant salesman for the giant pharmaceutical company.	4	3
Philip likes to maintain a relaxed rapport with his clients.	4	3
Will was learning how to write a persuasive critique in his composition class.	4	3

Appendix 5C: The Form Recall and Form Recognition Tasks (Study 2)

I. Form Recall Task:

You just encountered a number of new phrases. We would like to see if you remember them. We have given you the first word of each phrase. Please provide its one-word completion. For example, for a phrase like '*abject poverty*' you would see 'abject' and should provide the word 'poverty'.

1. nuclear _____
2. deep _____
3. front _____
4. old _____
5. disabled _____
6. local _____
7. young _____
8. severe _____
9. tempting _____
10. awful _____
11. perplexing _____
12. noticeable _____
13. cynical _____
14. superficial _____
15. constant _____
16. disputable _____
17. private _____
18. important _____
19. minimum _____
20. financial _____
21. political _____
22. wide _____
23. armed _____
24. rural _____
25. amazing _____
26. frightened _____
27. thick _____
28. supportive _____
29. grateful _____
30. illustrative _____
31. excited _____
32. deceptive _____
33. pregnant _____
34. personal _____

35. social _____

36. physical _____

37. foreign _____

38. detailed _____

39. large_____

40. brief _____

41. exhausted_____

42. evident _____

43. scared _____

44. ambiguous _____

45. respected _____

46. brilliant _____

47. relaxed _____

48. persuasive _____

II. Form Recognition Task:

Now we would like to see if you recognise the new phrases that you encountered. We have given you the first word of the phrase and you need to select the word that completes it. If you don't know the answer, please don't guess: choose 'I don't know' instead.

1. nuclear _____
    a. hardware
    b. vessel
    c. battery
    d. shell
    e. I do not know

2. deep _____
    a. harbour
    b. canyon
    c. bay
    d. pond
    e. I do not know

3. front _____
    a. terrace
    b. porch
    c. yard
    d. patio
    e. I do not know

4. old _____
    a. mobile
    b. gadget
    c. laptop
    d. notebook
    e. I do not know

5. disabled _____
    a. travellers
    b. tourists
    c. drivers
    d. fans
    e. I do not know

6. local _____
    a. collections
    b. antiques
    c. memorials
    d. souvenirs
    e. I do not know

7. young _____
    a. bloggers
    b. columnists
    c. reporters
    d. journalists
    e. I do not know

8. severe _____
    a. volcano
    b. storm
    c. hurricane
    d. tornado
    e. I do not know

9. tempting _____
    a. cakes
    b. pastries
    c. pies
    d. cookies
    e. I do not know

10. awful _____
    a. stomach-ache
    b. headache
    c. toothache
    d. backache
    e. I do not know

11. perplexing _____
    a. plot
    b. story
    c. topic
    d. scenario
    e. I do not know

12. noticeable _____
    a. errors
    b. flaws
    c. mistakes
    d. shortcomings

e. I do not know

13. cynical _____
    a. observers
    b. viewers
    c. audience
    d. readers
    e. I do not know

14. superficial _____
    a. wounds
    b. injuries
    c. sores
    d. bruises
    e. I do not know

15. constant _____
    a. horror
    b. fright
    c. anxiety
    d. agitation
    e. I do not know

16. disputable _____
    a. legitimacy
    b. legality
    c. validity
    d. lawfulness
    e. I do not know

17. private _____
    a. hut
    b. shack
    c. lodge
    d. cabin
    e. I do not know

18. important _____
    a. chat
    b. dialogue
    c. gossip
    d. discussion
    e. I do not know

19. minimum _____
    a. punishment
    b. penalty
    c. fine
    d. sentence
    e. I do not know

20. financial _____
    a. catastrophe
    b. tragedy
    c. misfortune
    d. failure
    e. I do not know

21. political _____
    a. satirists
    b. editors
    c. critics
    d. caricaturists
    e. I do not know

22. wide _____
    a. sidewalk
    b. path
    c. lane
    d. alley
    e. I do not know

23. armed _____
    a. hijacker
    b. kidnapper
    c. attacker
    d. abductor
    e. I do not know

24. rural _____
    a. forests
    b. parks
    c. woodlands
    d. jungles
    e. I do not know

25. amazing _____
    a. attraction
    b. tour
    c. sightseeing
    d. journey
    e. I do not know

26. frightened _____
    a. newborn
    b. baby
    c. infant
    d. kid

e. I do not know

27. thick _____
    a. columns
    b. piers
    c. pillars
    d. shafts
    e. I do not know

28. supportive _____
    a. colleagues
    b. classmates
    c. partners
    d. teammates
    e. I do not know

29. grateful _____
    a. newlywed
    b. groom
    c. bride
    d. fiancée
    e. I do not know

30. illustrative _____
    a. visuals
    b. diagrams
    c. figures
    d. charts
    e. I do not know

31. excited _____
    a. audience
    b. spectators
    c. crowd
    d. watchers
    e. I do not know

32. deceptive _____
    a. infographic
    b. graph
    c. table
    d. data
    e. I do not know

33. pregnant _____
    a. teenagers
    b. minors
    c. adolescents
    d. ladies
    e. I do not know

34. personal _____
    a. jet
    b. helicopter
    c. airplane
    d. aircraft
    e. I do not know

35. social _____
    a. regulation
    b. inspection
    c. monitoring
    d. surveillance
    e. I do not know

36. physical _____
    a. struggles
    b. conflicts
    c. clashes
    d. confrontations
    e. I do not know

37. foreign _____
    a. vendors
    b. sellers
    c. dealers
    d. brokers
    e. I do not know

38. detailed _____
    a. flyer
    b. brochure
    c. pamphlet
    d. leaflet
    e. I do not know

39. large _____
    a. stage
    b. auditorium
    c. arena
    d. theatre
    e. I do not know

40. brief _____
    a. journal
    b. record
    c. memoir
    d. diary
    e. I do not know

41. exhausted _____
    a. warrior
    b. fighter
    c. survivor
    d. champion
    e. I do not know

42. evident _____
    a. skill
    b. talent
    c. genius
    d. capability
    e. I do not know

43. scared _____
    a. pitch
    b. sound
    c. tone
    d. voice
    e. I do not know

44. ambiguous _____
    a. lyrics
    b. lines
    c. verses
    d. poems
    e. I do not know

45. respected _____
    a. novelist
    b. author
    c. poet
    d. playwright
    e. I do not know

46. brilliant _____
    a. clerk
    b. agent
    c. retailer
    d. salesman
    e. I do not know

47. relaxed _____
    a. affinity
    b. rapport
    c. connection
    d. understanding
    e. I do not know

48. persuasive _____
    a. analysis
    b. review
    c. commentary
    d. critique
    e. I do not know

Appendix 6A: List of the Novel Opaque Collocations Used in Study 3 and their Meanings

hot applause: clapping loudly

dry reception: unwelcoming

shaky confidence: low self-esteem

cutting proof: conclusive proof

paralyzed tongue: inability to speak

pasted accusation: false accusation

shiny poet: outstanding poet

scented legacy: great legacy

heavy tea: strong tea

stone mentality: narrow-minded

wide imagination: unlimited imagination

cloudy vision: unpredictable future

iron retention: excellent memory

flaming dilemma: serious dilemma

choking heat: extreme heat

white coup: peaceful coup

shy critique: gentle criticism

poetic scenery: breathtaking scenery

cruel defeat: crushing loss

sharp headache: severe headache

musical harmony: beautiful harmony

pink outlook: positive outlook

symbolic fare: low fare

open appetite: big appetite

Appendix 6B: The Novel Collocations Presented in Sentences (Study 3)

Sentences	Block	List
Ashley likes to drink a cup of heavy tea every morning.	1	1
Because of his stone mentality, Phillip did not fit well with his culturally diverse co-workers.	1	1
It takes someone with a very wide imagination to be a successful fantasy novelist.	1	1
There is still cloudy vision about who will win the election.	1	1
Because of his iron retention, Sam remembered the phone number after hearing it only once.	1	1
Many people believe that the world is facing a flaming dilemma about how to best address global climate change.	1	1
Many places experienced choking heat for extended periods last summer.	1	1
Many people believe that the white coup saved the country from an inevitable civil war.	1	1
The professor only offered a shy critique of the opposing side's view.	1	1
John always chooses places with poetic scenery when deciding where to go on holiday.	1	1
The political party suffered a cruel defeat in the last election.	1	1
Craig has been suffering from a sharp headache all day.	1	1
The couple lived in musical harmony for many years.	1	1
The research team was studying whether having a pink outlook on life improved blood pressure.	1	1
Arthur was amazed about the symbolic fare he paid for a round trip which was only £1.50.	1	1
Because of her open appetite, Sophia ordered a three course meal.	1	1
Albert couldn't get to sleep after drinking a very heavy tea after dinner.	2	1
People with a stone mentality tend to judge others who are different from them very harshly.	2	1
The creative writing teacher was impressed by the wide imagination of one of her students.	2	1
Financial forecasters are having cloudy vision when it comes to predicting the dollar's value.	2	1
The student believes that having iron retention helps him a lot in medical school.	2	1
The article discusses the flaming dilemma doctors face about making data freely available for wider study and protecting patients' identities.	2	1
Because of the persistent choking heat, we bought a portable air conditioner.	2	1
The white coup resulted in a rapid change of the government without any violence.	2	1
After summarising the main theories of language acquisition, Katelyn only gave a shy critique of them.	2	1
Sri Lanka is known for its poetic scenery, which makes it a popular a destination for nature lovers.	2	1
Despite their cruel defeat in yesterday's match, the football team is still determined to win the league.	2	1
Ron woke up with a sharp headache that did not go away until the next day.	2	1
The vegetarian couple wanted to live in musical harmony with animals.	2	1
Freddie maintains a pink outlook no matter how big a problem he faces.	2	1
They only had to pay a symbolic fare because they worked for the airline.	2	1
After fasting yesterday, Lucas woke up with open appetite and ate a full English breakfast.	2	1
On their trip to Turkey, the waiter offered them a heavy tea after finishing their meal.	3	1

Will didn't like conflict, so he never discussed sensitive topics with people who had a stone mentality unless he already knew their views agreed with his.	3	1
Stella needed to use her wide imagination when directing the sci-fi film.	3	1
The unstable political situation resulted in cloudy vision about the future of the country's economy.	3	1
Florence uses her iron retention to recite the lyrics of hundreds of songs.	3	1
There isn't a good substitute for air travel, so plane emissions pose a flaming dilemma for environmentalists.	3	1
One of the main causes of the recent choking heat is global warming.	3	1
The protestors were trying to promote a white coup to ensure a non-violent transition of power.	3	1
The student wrote a shy critique of the novel.	3	1
The poetic scenery could be enjoyed from the holiday cottage's small balcony.	3	1
The party's popularity decreased after their cruel defeat in the 2016 election.	3	1
Harold had been studying until he got a sharp headache that forced him to stop and rest.	3	1
The sales manager said that the department's success was due to his staff working in musical harmony over the last six months.	3	1
Teddy said that developing a pink outlook on life was the only thing that cured his depression.	3	1
The government has implemented a symbolic fare to encourage commuters to use public transport.	3	1
The meal Aiden ordered came with many sides and required an open appetite to finish it.	3	1
Serving freshly brewed heavy tea in different flavours is popular in many Starbucks in India.	4	1
When protesting about climate change, Thomas met people with a stone mentality who didn't believe scientists' claims.	4	1
It did not require a wide imagination to predict the end of the movie.	4	1
Because he didn't have much time to study for the exam, Ben has cloudy vision about his chances of getting a passing grade.	4	1
Walter thinks having iron retention is the reason why he can speak many languages.	4	1
The new government has to face the flaming dilemma about how to deal with an increasing unemployment rate.	4	1
As choking heat becomes more frequent in the city, people are installing central air conditioning in their houses.	4	1
Since the failure of the white coup, the rebel group is planning to launch more violent action to take down the government.	4	1
The political party was disappointed that their opposition only received a shy critique in the media.	4	1
The nice weather and the poetic scenery made their holiday to the mountains very memorable.	4	1
The president's new health plan saved him from a cruel defeat in last year's election.	4	1
Arnold keeps getting a sudden sharp headache for no apparent reason.	4	1
Ryder really liked the musical harmony of the mix of colours in the painting.	4	1
Edward still retained the enthusiasm and pink outlook he had when he was in college.	4	1
By using her student ID, Isabella only pays a symbolic fare when taking the bus.	4	1
Weightlifting always leaves Jayden with an open appetite afterwards.	4	1
Brian's performance was met with hot applause from the audience.	1	2
Despite the dry reception he got from his flatmate when he moved in, they are now very close friends.	1	2
Liam blamed his shaky confidence for being bad at public speaking.	1	2
The professor's paper didn't contain cutting proof that the theory is generalizable to the whole population.	1	2
When she walked on stage, Dorothy got a paralyzed tongue and she needed a moment to calm down.	1	2
Even though he was proven innocent later, the employee was fired because of the pasted accusation of theft.	1	2

His grandfather was a shiny poet in the nineteen fifties.	1	2
The esteemed scientist left behind a scented legacy after 20 years of working at the University.	1	2
Ashley likes to drink a cup of heavy tea every morning.	1	2
Because of his stone mentality, Phillip did not fit well with his culturally diverse co-workers.	1	2
It takes someone with a very wide imagination to be a successful fantasy novelist.	1	2
There is still cloudy vision about who will win the election.	1	2
Because of his iron retention, Sam remembered the phone number after hearing it only once.	1	2
Many people believe that the world is facing a flaming dilemma about how to best address global climate change.	1	2
Many places experienced choking heat for extended periods last summer.	1	2
Many people believe that the white coup saved the country from an inevitable civil war.	1	2
When the TED talk presenter finished, hot applause filled the lecture theatre.	2	2
When the politician made a speech on campus, he got a dry reception because most students do not agree with his political views.	2	2
Aaron found that his new job helped him a lot with his shaky confidence and made him feel more positive.	2	2
The environmentalist presented cutting proof for the root cause of the global warming.	2	2
Marshal is very private, so he got a paralyzed tongue when his flatmates started asking a lot of personal questions.	2	2
The police investigation revealed that it was just a pasted accusation and declared the man's innocence immediately.	2	2
Isaac quoted the shiny poet John Donne at the beginning of his speech.	2	2
The actor was very talented and left a scented legacy of great films.	2	2
Albert couldn't get to sleep after drinking a very heavy tea after dinner.	2	2
People with a stone mentality tend to judge others who are different from them very harshly.	2	2
The creative writing teacher was impressed by the wide imagination of one of her students.	2	2
Financial forecasters are having cloudy vision when it comes to predicting the dollar's value.	2	2
The student believes that having iron retention helps him a lot in medical school.	2	2
The article discusses the flaming dilemma doctors face about making data freely available for wider study and protecting patients' identities.	2	2
Because of the persistent choking heat, we bought a portable air conditioner.	2	2
The white coup resulted in a rapid change of the government without any violence.	2	2
The stadium burst into hot applause as the players came onto the pitch.	3	2
The new manager got a very dry reception when he joined the company but is now well liked.	3	2
Cognitive behavioural therapy has proven effective in treating Harry's shaky confidence by improving his social skills.	3	2
The police had no cutting proof that the suspects were responsible for the burglary.	3	2
In the interview, Robert got a paralyzed tongue when they asked him about his reasons for leaving his last job.	3	2
The movie told the story of an innocent man serving life in prison because of a pasted accusation of murder.	3	2
Daniel still couldn't believe that he had become a shiny poet and his work was being translated into many languages.	3	2
George Washington had a scented legacy as a great leader.	3	2
On their trip to Turkey, the waiter offered them a heavy tea after finishing their meal.	3	2
Will didn't like conflict, so he never discussed sensitive topics with people who had a stone mentality unless he already knew their views agreed with his.	3	2
Stella needed to use her wide imagination when directing the sci-fi film.	3	2

The unstable political situation resulted in cloudy vision about the future of the country's economy.	3	2
Florence uses her iron retention to recite the lyrics of hundreds of songs.	3	2
There isn't a good substitute for air travel, so plane emissions pose a flaming dilemma for environmentalists.	3	2
One of the main causes of the recent choking heat is global warming.	3	2
The protestors were trying to promote a white coup to ensure a non-violent transition of power.	3	2
The musician was interrupted by hot applause when he started playing his most famous song.	4	2
Jim and Nicola felt disappointed about the dry reception they got from their new neighbour when they knocked on his door to welcome him to the building.	4	2
The techniques used by the psychiatrist helped with Scott's shaky confidence by building up his self-image.	4	2
After the researcher provided cutting proof of the old scroll's archaeological significance, she was granted the funding to study it.	4	2
He always gets a paralyzed tongue whenever someone asks him about his childhood, which was a traumatic time for him.	4	2
Morris felt completely devastated by the pasted accusation of fraud, even though his family and friends stood by him.	4	2
This year the Nobel Prize for literature went to a very shiny poet who is also an activist for world peace.	4	2
Nelson Mandela established a scented legacy and is a symbol of the struggle for justice.	4	2
Serving freshly brewed heavy tea in different flavours is popular in many Starbucks in India.	4	2
When protesting about climate change, Thomas met people with a stone mentality who didn't believe scientists' claims.	4	2
It did not require a wide imagination to predict the end of the movie.	4	2
Because he didn't have much time to study for the exam, Ben has cloudy vision about his chances of getting a passing grade.	4	2
Walter thinks having iron retention is the reason why he can speak many languages.	4	2
The new government has to face the flaming dilemma about how to deal with an increasing unemployment rate.	4	2
As choking heat becomes more frequent in the city, people are installing central air conditioning in their houses.	4	2
Since the failure of the white coup, the rebel group is planning to launch more violent action to take down the government.	4	2
Brian's performance was met with hot applause from the audience.	1	3
Despite the dry reception he got from his flatmate when he moved in, they are now very close friends.	1	3
Liam blamed his shaky confidence for being bad at public speaking.	1	3
The professor's paper didn't contain cutting proof that the theory is generalizable to the whole population.	1	3
When she walked on stage, Dorothy got a paralyzed tongue and she needed a moment to calm down.	1	3
Even though he was proven innocent later, the employee was fired because of the pasted accusation of theft.	1	3
His grandfather was a shiny poet in the nineteen fifties.	1	3
The esteemed scientist left behind a scented legacy after 20 years of working at the University.	1	3
The professor only offered a shy critique of the opposing side's view.	1	3
John always chooses places with poetic scenery when deciding where to go on holiday.	1	3
The political party suffered a cruel defeat in the last election.	1	3
Craig has been suffering from a sharp headache all day.	1	3
The couple lived in musical harmony for many years.	1	3
The research team was studying whether having a pink outlook on life improved blood pressure.	1	3

Arthur was amazed about the symbolic fare he paid for a round trip which was only £1.50.	1	3
Because of her open appetite, Sophia ordered a three course meal.	1	3
When the TED talk presenter finished, hot applause filled the lecture theatre.	2	3
When the politician made a speech on campus, he got a dry reception because most students do not agree with his political views.	2	3
Aaron found that his new job helped him a lot with his shaky confidence and made him feel more positive.	2	3
The environmentalist presented cutting proof for the root cause of the global warming.	2	3
Marshal is very private, so he got a paralyzed tongue when his flatmates started asking a lot of personal questions.	2	3
The police investigation revealed that it was just a pasted accusation and declared the man's innocence immediately.	2	3
Isaac quoted the shiny poet John Donne at the beginning of his speech.	2	3
The actor was very talented and left a scented legacy of great films.	2	3
After summarising the main theories of language acquisition, Katelyn only gave a shy critique of them.	2	3
Sri Lanka is known for its poetic scenery, which makes it a popular a destination for nature lovers.	2	3
Despite their cruel defeat in yesterday's match, the football team is still determined to win the league.	2	3
Ron woke up with a sharp headache that did not go away until the next day.	2	3
The vegetarian couple wanted to live in musical harmony with animals.	2	3
Freddie maintains a pink outlook no matter how big a problem he faces.	2	3
They only had to pay a symbolic fare because they worked for the airline.	2	3
After fasting yesterday, Lucas woke up with open appetite and ate a full English breakfast.	2	3
The stadium burst into hot applause as the players came onto the pitch.	3	3
The new manager got a very dry reception when he joined the company but is now well liked.	3	3
Cognitive behavioural therapy has proven effective in treating Harry's shaky confidence by improving his social skills.	3	3
The police had no cutting proof that the suspects were responsible for the burglary.	3	3
In the interview, Robert got a paralyzed tongue when they asked him about his reasons for leaving his last job.	3	3
The movie told the story of an innocent man serving life in prison because of a pasted accusation of murder.	3	3
Daniel still couldn't believe that he had become a shiny poet and his work was being translated into many languages.	3	3
George Washington had a scented legacy as a great leader.	3	3
The student wrote a shy critique of the novel.	3	3
The poetic scenery could be enjoyed from the holiday cottage's small balcony.	3	3
The party's popularity decreased after their cruel defeat in the 2016 election.	3	3
Harold had been studying until he got a sharp headache that forced him to stop and rest.	3	3
The sales manager said that the department's success was due to his staff working in musical harmony over the last six months.	3	3
Teddy said that developing a pink outlook on life was the only thing that cured his depression.	3	3
The government has implemented a symbolic fare to encourage commuters to use public transport.	3	3
The meal Aiden ordered came with many sides and required an open appetite to finish it.	3	3
The musician was interrupted by hot applause when he started playing his most famous song.	4	3
Jim and Nicola felt disappointed about the dry reception they got from their new neighbour when they knocked on his door to welcome him to the building.	4	3
The techniques used by the psychiatrist helped with Scott's shaky confidence by building up his self-image.	4	3

After the researcher provided cutting proof of the old scroll's archaeological significance, she was granted the funding to study it.	4	3
He always gets a paralyzed tongue whenever someone asks him about his childhood, which was a traumatic time for him.	4	3
Morris felt completely devastated by the pasted accusation of fraud, even though his family and friends stood by him.	4	3
This year the Nobel Prize for literature went to a very shiny poet who is also an activist for world peace.	4	3
Nelson Mandela established a scented legacy and is a symbol of the struggle for justice.	4	3
The political party was disappointed that their opposition only received a shy critique in the media.	4	3
The nice weather and the poetic scenery made their holiday to the mountains very memorable.	4	3
The president's new health plan saved him from a cruel defeat in last year's election.	4	3
Arnold keeps getting a sudden sharp headache for no apparent reason.	4	3
Ryder really liked the musical harmony of the mix of colours in the painting.	4	3
Edward still retained the enthusiasm and pink outlook he had when he was in college.	4	3
By using her student ID, Isabella only pays a symbolic fare when taking the bus.	4	3
Weightlifting always leaves Jayden with an open appetite afterwards.	4	3

Appendix 6C: The Form Recall, Form Recognition, and Meaning Recognition Tasks Used in Study 3

I. The Form Recall Task

You just encountered a number of new phrases on the reading experiment. We would like to see if you remember them. We have given you the first word of each phrase. Please provide its one-word completion. For example, for a phrase like ‘abject poverty’ you would see ‘abject’ and should provide the word ‘poverty’. If you can’t remember how to complete the phrase, just put X in the space.

1. hot ……
2. dry ……
3. shaky ……
4. cutting ……
5. paralyzed ……
6. pasted ……
7. shiny ……
8. scented ……
9. heavy ……
10. stone ……
11. wide ……
12. cloudy ……
13. iron ……
14. flaming ……
15. choking ……
16. white ……

17. shy ……

18. poetic ……

19. cruel ……

20. sharp ……

21. musical ……

22. pink ……

23. symbolic ……

24. open ……

II. The Form Recognition Task

Now we would like to see if you recognise the new phrases that you encountered. We have given you the first word of the phrase and you need to select the word that completes it. If you don't know the answer, please don't guess: choose 'I don't know' instead.

1. hot _____
   a. applause
   b. cheers
   c. praise
   d. welcome
   e. I do not know

2. dry _____
   a. greeting
   b. reaction
   c. hospitality
   d. reception
   e. I do not know

3. shaky _____
   a. trust
   b. confidence
   c. belief
   d. reliance
   e. I do not know

4. cutting _____
   a. confirmation

b. evidence
c. proof
d. sign
e. I do not know

5. paralyzed _____
   a. language
   b. tongue
   c. speech
   d. speaking
   e. I do not know

6. pasted _____
   a. accusation
   b. charge
   c. allegation
   d. claim
   e. I do not know

7. shiny
   a. novelist
   b. author
   c. playwright
   d. poet
   e. I do not know

8. scented
   a. legacy
   b. heritage
   c. reputation
   d. fame
   e. I do not know

9. heavy
   a. tea
   b. drink
   c. juice
   d. coffee
   e. I do not know

10. stone
    a. character
    b. mentality
    c. attitude
    d. personality
    e. I do not know

11. wide
    a. fantasy
    b. imagination
    c. creativity
    d. curiosity
    e. I do not know

12. cloudy
    a. vision
    b. intuition
    c. perception
    d. understanding
    e. I do not know

13. iron
    a. memory
    b. retention
    c. remembering
    d. recalling
    e. I do not know

14. flaming
    a. crisis
    b. disaster
    c. dilemma
    d. problem
    e. I do not know

15. choking
    a. warmth
    b. weather
    c. temperature
    d. heat
    e. I do not know

16. white
    a. revolution
    b. rebellion
    c. uprising
    d. coup
    e. I do not know

17. shy

a. critique
b. analysis
c. evaluation
d. assessment
e. I do not know

18. poetic
    a. countryside
    b. landscape
    c. scenery
    d. view
    e. I do not know

19. cruel
    a. downfall
    b. defeat
    c. loss
    d. failure
    f. I do not know

20. sharp
    a. headache
    b. backache
    c. stomach-ache
    d. toothache
    e. I do not know

21. musical
    a. coordination
    b. agreement
    c. rapport
    d. harmony
    e. I do not know

22. pink
    a. viewpoint
    b. perspective
    c. outlook
    d. position
    e. I do not know

23. symbolic
    a. price
    b. fare
    c. cost

d. payment
e. I do not know

24. open
    a. desire
    b. craving
    c. appetite
    d. hunger
    e. I do not know

III. The Meaning Recognition Task:

Now we would like to see if you recognize the meaning of the new phrases that you encountered. You need to choose the meaning that matches each phrase. If you don't know the answer, please don't guess: choose 'I don't know' instead.

1. hot applause
    a. clapping that causes heat through friction
    b. clapping loudly
    c. phony clapping
    d. don't know

2. dry reception
    a. a reception area with dry heat
    b. a dusty reception area
    c. unwelcoming
    d. don't know

3. shaky confidence
    a. someone who is unreliable
    b. lack confidence in someone's ability
    c. low self-esteem
    d. don't know

4. cutting proof
    a. conclusive proof
    b. proof that results in hurting someone
    c. disappointing proof
    d. don't know

5. paralyzed tongue
    a. numb tongue muscles
    b. inability to speak
    c. comments that stop others from performing well

d. don’t know

6. pasted accusation
   a. an accusation that posted publicly
   b. implicit accusation
   c. false accusation
   d. don’t know

7. shiny poet
   a. outstanding poet
   b. well-dressed poet
   c. attractive poet
   d. don’t know

8. scented legacy
   a. biography of someone’s achievements
   b. great legacy
   c. inheritance from criminal activity
   d. don’t know

9. heavy tea
   a. exceptionally large tea leaves
   b. dark coloured tea
   c. strong tea
   d. don’t know

10. stone mentality
   a. narrow-minded
   b. unemotional
   c. defensive
   d. don’t know

11. wide imagination
   a. delusional thinking
   b. unlimited imagination
   c. inaccurate imagination
   d. don’t know

12. cloudy vision
   a. unpredictable future
   b. vision obscured by the clouds
   c. poor eyesight
   d. don’t know

13. iron retention
   a. keep something by locking it very securely
   b. type of disease that affects memory
   c. excellent memory
   d. don’t know

14. flaming dilemma
    a. serious dilemma
    b. sudden crisis
    c. dilemma about preventing wildfires
    d. don’t know

15. choking heat
    a. extreme anger
    b. fever resulting from choking
    c. extreme heat
    d. don’t know

16. white coup
    a. worthless accomplishment
    b. peaceful coup
    c. brutal coup
    d. don’t know

17. shy critique
    a. gentle criticism
    b. apprehension about criticizing someone
    c. sufficient criticism
    d. don’t know

18. poetic scenery
    a. poem praising the landscape
    b. breathtaking scenery
    c. scenery that many poets write about
    d. don’t know

19. cruel defeat
    a. crushing loss
    b. defeat that inflicts physical pain
    c. complete disappointment
    d. don’t know

20. sharp headache
    a. severe headache
    b. serious problem
    c. severe inconvenience
    d. don’t know

21. musical harmony
    a. combination of sounded musical notes
    b. beautiful harmony
    c. when people singing in a harmonious way
    d. don’t know

22. pink outlook
    a. negative outlook
    b. a pleasant view from a window
    c. positive outlook
    d. don’t know

23. symbolic fare
    a. typical food
    b. replica of food
    c. low price
    d. don’t know

24. open appetite
    a. big appetite
    b. sudden feeling of wanting something
    b. strong willingness
    c. don’t know