Formulaic language:

Distribution, processing, and acquisition

Laura Vilkaitė

MA, BA

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

May 2016

Abstract

Formulaic sequences are very frequently used in language as a preferred way to convey certain meanings. This thesis looks at distribution, processing, and incidental acquisition of formulaic sequences, by presenting four separate studies on different aspects of formulaicity.

Study 1 investigated the distribution of four different categories of formulaic sequences (collocations, idiomatic phrases, lexical bundles, and phrasal verbs) and showed that those four categories vary considerably in terms of frequency. Also, register seems to affect the number of formulaic sequences used, as well as the categories of formulaic sequences preferred. Importantly, this study raised an issue of form variation of formulaic sequences (especially collocations) which seemed to be an under-researched area. Therefore, the following studies investigated the effect of form variation (focusing on non-adjacency) on collocation processing and their incidental acquisition.

Studies 2 and 3 used an eye-tracking technique to investigate how native and non-native speakers of English process adjacent and non-adjacent verb-noun collocations. The results suggest that native speakers process both adjacent and non-adjacent collocations faster than matched control phrases, albeit the collocation effect seems to be larger for adjacent collocations. As for non-native speakers, there is a clear collocation effect for adjacent collocations and it is moderated by prior vocabulary knowledge. However, there seems to be almost no effect for non-adjacent collocations. This finding suggests that even advanced non-native speakers process non-adjacent collocations differently than native speakers.

Finally, Study 4 tried to take the findings from the previous studies to a classroom. It investigated whether there is any difference between the chances of incidentally acquiring adjacent and non-adjacent collocations from reading. The results suggest low but durable gains for both adjacent

and non-adjacent collocations, with no significant differences between these two groups of items.

Overall, the results presented in the thesis support the idea that formulaic language is ubiquitous, but suggest that some of the criteria that have been widely used for identifying formulaic sequences might need to be reconsidered. It seems that collocations, at least for native speakers, retain their formulaic status even when presented non-adjacently. They seem to be successfully learned as non-adjacent dependencies as well. While this finding cannot be easily generalized to other types of formulaic sequences, it seems to suggest that a lot more research on form variation of formulaic sequences is needed in order to better understand the scope of the phenomenon.

Acknowledgements

The three years of my PhD have been a truly enriching experience. I had a chance to learn lot, both about formulaic language and about myself. The work presented in this thesis would have been impossible without the help of many teachers, colleagues, and friends. I owe my gratitude to them all.

My deepest gratitude is to my supervisor, Prof Norbert Schmitt. He was the one who encouraged me to embark on this journey at the first place. He also gave me the freedom to focus on the research I was interested in, at the same time following me closely and offering me his help, whenever I needed it. I hope I will manage to stay as excited about research as Norbert is.

I would also like to thank Dr Ana Pellicer-Sánchez, my second supervisor. Starting from my MA, she has always been there to listen and give advice on both technical questions and personal decisions.

My thanks also go all the past and current members of the Vocabulary Research Group: Hana Almutairi, Samuel Barclay, Mélodie Garnier, Beatriz González Fernández, Benjamin Kremmel, Duyen Le Thi, Marijana Macis, Niloofar Rahimi, Dr Michael Patrick Hindley Rodgers, Dr Kholood Saigh, and Dr Hilde van Zeeland. Our group discussions have taught me a lot about vocabulary and research in general. They all have definitely made me a more critical researcher.

Thanks also go to the Bilingualism Research Group and especially to Dr Kathy Conklin and Dr Walter van Heuven. They welcomed me to the group and gave me valuable feedback on my psycholinguistic studies. I have learned a lot about psycholinguistics from them.

I would also like to thank Dr Gareth Carrol for his patience and enthusiasm during our numerous discussions about statistics and data analysis.

My sincere thanks also go to Dr Jurgis Pakerys, Dr Rita Juknevičienė, Dr Jonė Grigaliūnienė and Ieva Gelžinytė from Vilnius University for their help and support with my last study. Without their goodwill my work would have been impossible.

My special thanks go to my two dearest friends in Nottingham, Ben and Bea, for keeping me sane and believing in me when I did not. Also, I cannot thank them enough for making me bring some great memories back home with me.

And finally, my special thanks go to my family: my parents Loreta and Arūnas, and my sister Rūtenė. Their encouragement and love made me see the positive side even during the most challenging times.

Ačiū!

Declaration

I declare that the work presented in this thesis is my own and was conducted during my time as a PhD student at the University of Nottingham. Several parts of this thesis have been published or submitted to peer-reviewed journals:

- Study 1, presented in Chapter 3, was published in a Lithuanian journal *Taikomoji kalbotyra* (Vilkaitė, 2016);
- Study 2, presented in Chapter 4, was published in the *Journal of Experimental Psychology: Learning, Memory and Cognition* (Vilkaitė, in press);
- Study 3, presented in Chapter 5, has been submitted as a paper co-authored with Norbert Schmitt and is currently under review.

All chapters included in this thesis are more detailed versions of the published papers. The published versions should be used for any citations or page references.

Contents

List of Figures

List of Tables

Abbreviations

BNC	British National Corpus
COCA	Corpus of Contemporary American English
EFL	English as a foreign language
ESL	English as a second language
FL	formulaic language
FS	formulaic sequence
L1	first language
L2	second language
MI	mutual information
MCT	multiple choice test
NS	native speaker
NNS	non-native speaker
SLA	second language acquisition
VLT	Vocabulary Levels Test
% DIFF	percentage difference

People and language create each other, grow from each other, and change and act under the influence of the other.

N. C. Ellis, 2015, p. 16

Chapter 1

Introduction

Traditionally, language was seen as a system consisting of words, which are arbitrary signs, and rules, which determine how these words can be combined together to transmit messages. This kind of system implies two different modes of processing: one for memorized words and one for combinatory rules (see, for example, Pinker, 1998). The rules-based processing was considered to be the dominant one, used whenever a rule is available. However, the evidence from various empirical studies has shown that this idea is rather problematic. It seems that we memorize a large number of sequences of more than one word, and we are able to access them without constructing them from scratch each time (Van Lancker Sidtis, 2012; Wray, 2002). Therefore, the rules-and-words approach has been replaced with the ideas of the importance of chunks in language learning (A. M. Peters, 1983; Tomasello, 2005) and potential multiple representations of the same item (Langacker, 2000, Wray, 2002).

One of the main issues with the traditional approach was its inability to deal with the idea of formulaic language (FL) – recurring word combinations, that are not constructed each time of use but rather have a stereotypical form, conventionalized meaning, and are highly familiar to a speech community (Van Lancker Sidtis, 2012). Various studies have shown that a high percentage of the language we use is formulaic. Widely cited figures are those calculated by Erman and Warren (2000), who claimed that pre-constructed phrases account for as much as 59% of spoken language and 52% of written texts. Obviously, these numbers have to be interpreted with caution and various researchers using different methodologies arrived at different figures, ranging from 17–30% (Sorhus, 1977) to 80% (Altenberg, 1998). However, even when making more cautious generalizations, it was suggested that formulaic sequences (FS) make up from one third to a half of our language (Conklin & Schmitt, 2012; Schmitt, 2010).

It seems that FL is used so widely because it serves various purposes. Nattinger and DeCarrico (1992) claimed that conventionalized lexical expressions organize discourse both at the micro and at the macro level and make an important part of speakers' pragmatic competence. Furthermore, FSs seem to aid processing, especially when there is a time pressure, and allow achieving the best chances of being understood (Wray, 2002). It seems that psychologically, the use of FL helps to overcome the limits of working memory (Conklin & Schmitt, 2012), increasing language users' fluency. Moreover, phraseology[1] is closely related to culture (Colson, 2008). Idioms and collocations metaphorically encode certain concepts representing collective mentality of the community of language users (Teliya, Bragina, Oparina, & Sandomirskaya, 1998). Therefore, the use of FSs might also help to signal one's identity and belonging to a group (Wray, 2002).

The importance of FL is now widely acknowledged and FL has received a lot of attention in research, with a number of books published on the topic (e.g., Barfield & Gyllstad, 2009; Corrigan, Moravcsik, Ouali, & Wheatley, 2009a, 2009b; Schmitt, 2004; Wray, 2002, 2008) and a special journal issue dedicated to the field (ARAL 2012 issue). However, there are still a number of controversies and questions to be answered. One of the problems with researching FL is the lack of agreement on definitions. Therefore, a plethora of terms exist, sometimes they overlap, and sometimes the same term is used to refer to different phenomena. Wray (2012) pointed out that even if formulaicity is a complex phenomenon, we might have made it even more complex by the way of conceptualizing it. Gries (2008) also noted that authors interpret the notion of phraseology differently, but they sometimes fail to provide a clear definition of which type of FSs they are focusing on.

The issues with definitions may lead to problematic overgeneralizations, when studies carried out on one facet of FL being generalized to the FL as a whole (Wray, 2012). Even studies that look at the same category of FSs can arrive at different results if they adopt different definitions or use different methods (O'Donnell, Romer, & Ellis, 2013). On the

[1] In the present thesis the terms 'formulaic language' and 'phraseology' are used synonymously.

one hand, there seems to be no reason to believe that there are no common mechanisms underlying the use of different categories of FSs: they are all a part of our language knowledge and their use is driven by the same cognitive system. On the other hand, we cannot discard the possibility that different types of FSs (while all being different from novel language) are formulaic for different reasons and are processed and acquired slightly differently. Therefore, at the current stage of the research, it seems wiser to avoid broad generalizations and try to focus on individual categories of FL instead. This approach might be less ambitious, but it might prove more fruitful.

The present thesis sets out to look at formulaicity from three different perspectives and using three different methodologies: corpus analysis, eye-tracking technique, and an acquisition study. While the title of the thesis suggests that it focuses on FL overall, the actual studies narrowed the research focus down considerably. For the Study 1, four different categories of FSs were analysed, leading to a selection of only one type of collocations (verb-noun collocations) for the subsequent studies.

1.1 Distribution of formulaic sequences

Considering the problems of variation in terminology and definitions, the thesis starts out from an attempt to divide FSs into categories, provide clear definitions, and investigate their relative importance (as approximated by frequency) in language. From all the categories of FSs identified and discussed in previous research, four well-researched categories were chosen (i.e., collocations, idioms, lexical bundles, and phrasal verbs). The aim of this choice was to adopt clear-cut definitions and to consider the relative frequency of each of the four categories. While the results of the study provided some interesting insights, they also raised some issues about the extraction of FSs. While a lot of definitions of FSs imply fixedness as a defining feature, there seemed to be a considerable amount of form variation, difficult to capture when automatically extracting FSs from the corpus. Therefore, the remaining part of the thesis addressed the question of the form variation of FSs.

1.2 Form variation of formulaic sequences

When starting to look at the variation (or flexibility) of formulaic sequences, it became clear that it has been widely acknowledged by various researchers. Moon (1998a), for example, claimed that fixed expressions are often unstable in their forms. In corpus linguistics, Gray and Biber (2013) looked at discontinuous lexical frames with open slots, while Cheng and colleagues introduced a concept of a concgram that allows capturing variation of form (Cheng, Greaves, Sinclair, & Warren, 2009). Van Lancker Sidtis (2012) also discussed the idea that a speaker knows a stereotypical form of an expression along with all information about its use, but this core form is modified during the actual use. Overall, there have been attempts to address the questions of flexibility of FSs from the theoretical and corpus perspectives. However, there is still not a lot of research on processing and acquisition of FSs that are used in their non-core form.

This led me to choose the topic of variation in FSs for the subsequent studies. Variation is a very broad term, and it can include positional variation (*draw a conclusion* versus *a conclusion drawn from…)*, morphological variation (*draw a conclusion* versus *drawing conclusions*), lexical variation (*reach / draw / come to a conclusion*), and adjacency variation (*draw a conclusion* versus *draw an important conclusion*). As all the types of variation were impossible to address in one thesis, I decided to look at one type only, namely adjacency. I investigated the effect of adjacency on verb-noun collocations during processing and incidental acquisition.

1.3 Structure of the thesis

The present thesis reports on four independent studies, focusing on three different aspects of FL: distribution and frequency, processing, and incidental acquisition. It starts from a broad literature review in Chapter 2, which aims to provide a general background for all four studies. After discussing various definitions of FSs, I propose the one that will be used for the thesis. I also summarize different identification criteria for FSs and define four categories that will be the focus of Study 1 more in detail. Following that, the research on processing FL is discussed. Finally, the chapter provides an overview of the difficulties of learning FSs in a second language (L2) and the effectiveness of teaching them directly or acquiring them incidentally.

However, Chapter 2, even if providing the background for the thesis, does not focus on methodology. The following chapters have their own brief background sections, each presenting the relevant methodologies: corpus analysis in Chapter 3, eye-tracking in Chapter 4, and studies on incidental acquisition in Chapter 6.

Each study is written up as a self-standing report which presents a detailed description of preparation, analysis, results, and interpretations of the findings. Chapter 3 presents a corpus study which looked at the distribution of collocations, idioms, lexical bundles, and phrasal verbs in language as well as in four different registers. Chapter 4 reports an eye-tracking experiment which looked at how native speakers of English process non-adjacent collocations. Chapter 5 presents an eye-tracking study extending the findings of Study 2 to non-native speakers. While the two eye-tracking studies were carried out using the same design and the same materials, they are presented as two different studies, instead of two parts of one study. This is the case for the following reasons. Firstly, they were conceptualized as separate studies building up on each other from the beginning, initially focusing on the NSs only. This was the case because of the lack of research on psychological reality of non-adjacent collocations. Therefore, it seemed that the question should initially be examined in first language (L1), before even thinking about adding a non-native speaker (NNS) group. Also, some additional factors were analysed for the NNSs data (e.g., vocabulary knowledge). Therefore, Study 3 extends Study 2 to a different population and is presented separately. Chapter 6 presents a classroom study on incidental acquisition of adjacent and non-adjacent collocations from reading.

Finally, Chapter 7 brings all the results together and places them in the broader context of research on FL. This chapter also discusses some limitations of the research presented in the thesis and offers some potential directions for future research.

Chapter 2

Formulaic language: What do we know so far?

This chapter provides a broad overview of the research on FL. As the thesis looks at FL from three different perspectives (distribution, processing, and acquisition), this chapter is divided into three main sections. The first one presents definitions of FL, focusing on the criteria applied to identify different categories of FL. The second part will look at the psychological reality of FSs. It will review both empirical studies and theories that can account for formulaicity in language. Finally, the last part focuses on the acquisition of FL in a second language (L2).

2.1 Defining formulaic language

Almost two decades ago, Sinclair claimed that with the corpus evidence we have, we should move away from conceptualising a word as a principal unit of language and consider multiword units of meaning to be the main linguistics elements (Sinclair, 1996, 1998, 2007). Sinclair noted that individual words very rarely carry a meaning and that meaning rather emerges from a co-selection of words (Krishnamurthy, Sinclair, Daley, & Jones, 2004; Sinclair, 1998, 2000, 2004a, 2004b). Following this approach, FL becomes a central aspect of language use.

As the co-selection of words was central for Sinclair, he made a distinction between an ‘open-choice’ and an ‘idiom’ principle in producing language. The open choice principle implies selecting individual words to construct messages, while the idiom principle relies on phraseological tendency of using words. Following the idea of the importance of the idiom principle, Sinclair and his colleagues proposed a concept of a ‘meaning shift unit’: a co-selection of words that conveys a meaning (Cheng et al., 2009). The meaning shift unit has a core (words co-selected together), semantic prosody, collocations, and colligations. There are now tools to investigate meaning shift units, by extracting concgrams (co-selections of words, allowing for positional variation and insertions between them). These concgrams have received some attention in research (Cheng, 2008; Cheng et al., 2009; Cheng, Greaves, & Warren, 2006, Greaves, 2009) and it might be the case that investigating concgrams (rather than n-grams or traditional collocations) would help to better uncover the phraseological tendency of language use (Cheng et al., 2009). However, at least for now, it is not clear how to best approach the extracted items. The research so far has mostly focused on case studies of one or a couple of examples, and did not really manage to look at the overall use of meaning shift units. Also, the procedure of extracting concgrams allows co-selected words to be separated by up to 12 words in between the core elements, which makes the psychological reality of the extracted items questionable. Furthermore, it is not entirely clear how to integrate this approach with various already studied categories of FSs. For instance, meaning shift units seem to subsume some of the well-researched categories (e.g., idioms or phrasal verbs), but not others (e.g., lexical bundles). Thus, even if the idea of the importance of the co-selection of words is convincing, the term seems to be tricky to operationalize and therefore is not chosen as the definition of FL for this thesis.

A different widely cited definition of FL was proposed by Wray (2002), who defined a formulaic sequence as follows:

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

This definition does not focus on the meaning of the sequence, but rather implies that FSs are memorized, processed, and produced as units, rather than constructed from single words. Therefore, this definition is very inclusive and it comprises various different categories of word sequences (including even memorized poems, lists of objects, songs, or songs in foreign languages that a speaker might not even understand). While the memorization of these listed items is not questionable, they are considerably different from FSs that are used within sentence limits. Also, these FSs seem to be not as commonly used in everyday language. Therefore, these sequences are beyond the scope of this thesis, which looks at the formulaicity within sentence (or utterance) limits.

While the aforementioned definition of FSs seems to be too inclusive, Wray also suggested an idea of a 'morpheme equivalent unit' (2002). Morpheme equivalent unit is "a word or word string, whether incomplete or including gaps for inserted variable items, that is processed like a morpheme, that is, without recourse to any form-meaning matching of any sub-parts it may have" (Wray, 2008, p. 12). While this definition is a bit more restrictive and was taken up by some researchers working on FL (e.g., Martinez, 2011), the idea of the morpheme equivalence seems to be problematic as well. First of all, a word can also be a morpheme equivalent unit, so this concept includes both individual words and formulaic sequences. Also, this definition suggests that morpheme equivalent units are processed "without any recourse to any form-meaning matching of any sub-parts" (Wray, 2008, p. 12). This idea might also be debatable, at least for some types of FSs, as their individual parts seem to be activated during processing (Siyanova-Chanturia & Martinez, 2015). Furthermore, the sequence itself can be variable and its component parts can have morphological markers. These morphological markers have to be processed in order to interpret the actual meaning of the sequence in a specific context. Taken all the above reasons into account, defining FSs as morpheme equivalent units seems to be difficult.

Overall, there is no universally adopted definition of FL. This results in various researchers using different terminology to refer to FSs and these terms pointing to slightly different aspects of formulaicity (e.g., its frequency, storage in the mental lexicon, and conventionalization). Therefore, the terms used include modifiers such as 'multi-word', 'unanalysed', 'prefabricated', or 'recurrent' (see Wray, 2002, p. 9 for a comprehensive list), depending on which aspect of FL researchers are focusing on. Due to the variety of different types of strings that are defined as formulaic, it is extremely complicated to provide a comprehensive definition that would be specific and yet inclusive.

For the purpose of this thesis, FL will be understood as sequences of two or more words recurring in a near proximity to each other within sentence boundaries, familiar to language users, and showing a processing advantage (though see Section 2.2 for a further discussion on the psychological reality of

FSs). In the present thesis, the criteria of semantic completeness or opacity are not adopted as necessary for FSs. Following Schmitt (2010), the term ‘formulaic language’ will be used to define the phenomenon of recurrent word sequences in language, while the term ‘formulaic sequence’ will be used to identify its specific instances.

2.1.1 Identification criteria

There are two main approaches to identification of FL (Barfield & Gyllstad, 2009; Granger & Paquot, 2008):

- **Phraseological approach.** This approach to FL is mostly meaning based. Phraseology is seen as a continuum between the most opaque and fixed items to free combinations. The representatives of this approach usually focus on idioms and do not have clear criteria to distinguish free combinations from transparent multi-word units.
- **Distributional approach**. Frequency plays a key role in this approach, which is strongly influenced by corpus linguistics. Phraseology is understood as a central element of language which is more important than single words.

These two different approaches to phraseology, not surprisingly, lead to different methods of analysing FL. In the present thesis, the distributional approach is adopted. Therefore, the frequency is taken to be central for identification, as well as processing of FL (see Section 2.2.3.1 on usage-based theories and Section 2.2.3.3. on statistical learning).

As FL is a heterogeneous phenomenon, listing specific criteria for identifying all the sequences that could be classified as formulaic is not a straight forward task. Moreover, FL is very often understood as a continuum between the most holistic units (e.g., idioms, memorized routines, proverbs) to the items, that are more fluid and transparent, such as collocations (Wray, 2002). However, Wray (2012) criticised the idea of seeing FL as a continuum, by claiming, that it is too simplistic, and there could be a number of continua: based on frequency, compositionality, salience, and fixedness of the phrases.

Though, even if we see FL as a certain continuum (or a set of different continua), it is still not easy to place various types FSs along this continuum.

Table 2.1 summarizes the criteria identified by different researchers. It is worth noting, though, that if a researcher does not mention a certain criterion, this does not mean they assume this criterion is not valid for identifying FSs. Rather it suggests that the researcher does not put emphasis on that particular feature.

Table 2.1 Criteria for identifying formulaic sequences

Researcher	Schmitt (2010)	Wray (2002)	Erman & Warren (2000)	Moon (1998a)	Pawley & Syder (1983)
Criteria	Multi-word items	FSs	Prefabs	Fixed expressions and idioms	Lexical items
Formal: - two or more words - frequency of use	✓		✓	✓	
Semantic: - non-compositional - restricted exchangeability	✓			✓	✓
Syntactic: - no (less) syntactic variation - behave as one syntactic unit	✓	✓	✓	✓	✓
Psychological: - stored holistically - produced as a unit	✓	✓	✓		
Phonetic: - single intonation curve - no pauses or hesitations - unique stress	✓	✓		✓	
Social: - conventionalization - institutionalization		✓		✓	✓

Table 2.1 is certainly non-definitive and it serves more as an illustration of various criteria that can be applied rather than a comprehensive list of all the criteria. It can be argued that these criteria identified by different researchers are not comparable, as they were used to describe different phrasal units (as indicated in the table itself) for different purposes. However, it shows how many different features of FSs are mentioned. All in all, there is no single criterion or an accepted list of criteria to apply when studying FL.

Even these listed criteria are not easy to apply. Svensson (2008), for example, looked closer at non-compositionality and found that this term is used referring to motivation of the meaning, transparency, analysability or literal versus non-literal meaning of phrases. Thus, there seems to be some variation not only in the criteria used, but also in how they are interpreted.

Various researchers have also criticized the criterion of fixedness, claiming that FSs show considerable flexibility (Moon, 1998a; Schmitt, 2005; Sinclair, 1991; Wray, 2002). Sinclair raised a point that many FSs allow internal lexical, syntactic variation or some variation in word order (Sinclair, 1991). Also, some FSs have open slots. Van Lancker Sidtis (2012) tried to provide an explanation of how FSs have a fixed form and at the same time can be flexible. She suggested the concept of 'formuleme' "which is the canonical form of an expression known to the speaker and can be manipulated using grammatical rules" (p. 347). So a formuleme is a unit stored in the mental lexicon along with all the information about its use and appropriateness and it is modified to fit the context each time of use. However, this concept of formuleme seems not to be taken on by many researchers. In any case, fixedness of FSs does not seem to necessarily be a defining criterion.

One other important issue with the identification of FSs is the fact that individual speakers can have different sets of FSs available to them. A. M. Peters (1983, p. 2) made a distinction between 'cultural formulas' that are formulaic for a speech community and 'idiosyncratic formulas', which have a formulaic status for a particular speaker. Wray (2008) also distinguished between speaker-internal and speaker-external FSs. This is something to keep in mind when discussing psycholinguistic validity of any type of formulaic sequences: FSs are somewhat idiosyncratic. Moreover, as Wray (1999) notes, even for the same speaker, the set of FSs in their lexicon may change over time. However, there is certainly at least some overlap between phrases that are formulaic to different speakers and the frequency of the phrase might be one indication of how likely that phrase might be a cultural formula, rather than the idiosyncratic one. Even though individual differences are indisputably important, they are beyond the scope of this thesis. The studies presented in this thesis will focus on FSs that are frequent and therefore presumably formulaic for a majority of speakers.

2.1.2 Categories of formulaic language

One of the reasons why so many different criteria for identifying FSs are proposed is that various researchers focus on different categories of FSs, for example: binomials, clichés, collocations, collocational frameworks, formulas, idioms, lexical bundles, phrasal verbs, proverbs, and spaced compounds. Some of them seem to overlap in their definitions to certain extent (e.g., some collocations can have an idiomatic meaning) and some of the terms are used to define more than one phenomenon (e.g., a term 'collocation' can refer to specific phrases or to a tendency of co-occurrence).

These different categories of FL are not all studied equally, both in terms of processing and in terms of acquisition and teaching. Some of them seem to be listed as formulaic but did not receive a lot of attention in second language acquisition (SLA) research. For example, proverbs are addressed from the lexicographical perspective, with dictionaries compiled (e.g., Simpson & Speake, 2008), but they do not seem to receive a lot of attention as a type of FL.

Some other categories have benefitted from a bit more research. For example, Renouf and Sinclair (1991) suggested the idea of collocational frameworks: patternings of grammatical words with open slots, such as *a ... of* or *too ... to*. These collocational frameworks were shown to be relevant in other languages (e.g., Butler (1998) studied collocational frameworks in Spanish) and in specific registers (e.g., Marco (2000) analysed collocational frameworks in medical research papers). However, with the exception of these few studies, there has not been much research on this category. Also, collocational frameworks could be conceptualized as non-adjacent collocations, making it questionable whether there is a need to establish them as a separate category.

Some categories, though, have received considerable attention in corpus analysis, psycholinguistic research, and pedagogical studies. Four of them (collocations, idioms, lexical bundles, and phrasal verbs) were selected as the categories to focus on in the present thesis. Each one of them is defined in the following sections. The discussion on collocations will be the most detailed, as they will be the focus of three out of the four studies presented in the thesis.

2.1.2.1 Collocations

The term 'collocation' has been defined in various ways. Some researchers use it to refer to any occurrence of words within a certain span (e.g., Sinclair, 1991), while others require a statistical expectancy between the words beyond their co-occurrence (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Also, some researchers conceptualize collocational knowledge as one of the aspects of the depth of word knowledge (Nation, 2001), while others see collocations as lexical items with the dimensions of knowledge on their own right (Brown, 2014).

Moon summarized that "collocation typically denotes frequently repeated or statistically significant co-occurrences, whether or not there are any special semantic bonds between collocating items" (1998a, p. 26). Depending on the approach a researcher is taking (phraseological or distributional), either the frequent co-occurrence of the words or certain restrictions of their use are emphasized. These two criteria are not necessarily mutually exclusive: at least a part of collocations extracted from corpora using statistical approaches would be somewhat restricted. In the phraseological perspective, the fact that collocations are expected to have some meaning restrictiveness, places them on a cline with idioms (figurative phrases). There have been various attempts to draw the line between free combinations, collocations, and idioms. Howarth (1998, see also Cowie, 1981), for example, proposed a classification of verb-noun phrases into:

1. free combinations (*blow a trumpet*): clear literal sense;
2. restricted collocations (*blow a fuse*): one of the components is used in a specialized sense;
3. figurative idioms (*blow your own trumpet*): both literal and figurative meaning available;
4. pure idioms (*blow the gaff*): only figurative reading.

So Howarth required a restricted use of one of the elements for a phrase to be classified as a collocation. Nesselhauf (2003) proposed a very similar classification for verb-noun collocations (free combinations – collocations – idioms), also suggesting that for a phrase to be classified as a collocation, a verb has to be used in a restrictive sense. These classifications imply that collocations show some restrictiveness of use, but their meaning is more transparent than the meaning of idioms. However, as Nesselhauf (2005) noted herself, any scheme of classification is problematic. First of all, they tend to be restricted to one type of collocations only (in this case, verb-noun collocations). Also, even the definitions in the schemes do not allow clearly separating collocations from other types of phrases (idioms or non-restricted phrases).

A part of the challenge of analysing collocations is that they are very fluid in use: they reflect tendencies and preferences rather than obligations of use (Brown, 2014; Wray, 2002). As Wray (2002) put it, collocations are "common, but often far from exclusive, pairings of words" (p. 63). Another issue with defining collocations is the fact that the items in this category vary considerably in terms of their structure. Some researchers make a distinction between grammatical collocations (collocations containing function words) and lexical collocations (collocations formed of at least two content words) (e.g., Durrant, 2009). Lexical collocations can be further classified based on the part of speech of their elements: e.g., verb-noun, noun-noun, adjective-noun collocations. Overall, as Gries (2013) summarized, despite the fact that the idea of collocation has been around from at least the 1950s (works of Firth (1957)), there is still no consensus of how to best identify and extract collocations or measure their associations.

For the purpose of the present thesis, a frequency-based approach to collocations is adopted. Collocations are understood as co-occurrences of two words that are frequent and mutually expected (e.g., *take care, provide information, environmental issues).* No additional semantic or restrictiveness criteria are applied, but only lexical collocations are the focus of the studies reported.

2.1.2.2 Idioms

Idioms (e.g., *spill the beans* or *kick the bucket*) are usually defined as multi-word units that have a figurative meaning, that is, the meaning of the whole sequence cannot be inferred from the meanings of its parts (Grant & Nation, 2006, Čermák, 2001). So idioms are different from collocations from the meaning perspective. Despite the fact that idioms received a lot of attention in research and lexicography, there are a number of issues in defining idioms. After reviewing numerous studies on idioms, Čermák (2001) concluded that two main problems with idiom research are the lack of communication between research groups working with different languages (and writing in different languages) as well as the lack of strict methodology leading to a lot of reliance on intuition. Moon (1998a) also noted inconsistencies in the use of the term: 'idiom' is used both for metaphorical or opaque phrases and for any conventionalized formulas in general.

While idioms are usually defined as fixed, opaque, non-compositional, and non-productive, various researchers have now pointed out that idioms can be modified (Omazic, 2008, Simpson & Mendis, 2003) and used creatively omitting or changing certain words (Philip, 2008). They allow quite a lot of lexical variation as well (Moon, 1998a). However, even if words in idioms can be modified, idioms have to retain certain content words (or at least underlying conceptual meanings) to retain the figurative meaning (Biber et al., 1999). Because of this figurative meaning, idioms seem to be quite salient for the language users and are usually taken as a prototypical example of FSs (Van Lancker Sidtis, 2012; Wray, 2002).

2.1.2.3 Lexical bundles

The idea of lexical bundles was suggested by Biber and his colleagues (Biber et al., 1999). Lexical bundles (e.g., *I don't know*, *on the other hand, I think I*) are defined as "extended collocations: bundles of words that show statistical tendency to co-occur" (Biber et al., 1999, p. 989). The research on the use of lexical bundles (Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2004; Nesi & Basturkmen, 2006) showed that they are usually employed for organizing discourse or expressing one's stance. Therefore, they are used very frequently.

Lexical bundles are identified automatically following clear criteria of frequency and dispersion, making their identification easy and uncontroversial. If FL is conceptualized as a continuum, lexical bundles in many ways are in the opposite side of idioms, as they are very frequent, not perceptually salient, literal in meaning, and do not necessarily represent a complete structural or semantic unit (Biber, 2009). Therefore, they would be considered a part of FL only from the distributional, but not from the phraseological perspective.

2.1.2.4 Phrasal verbs

Even if it seems that phrasal verbs (e.g., *get up, move on*) can be easily defined based on their structure (a verb plus an adverbial particle), there are at least two different approaches to phrasal verbs. While some researchers choose a structural criterion of a verb and an adverbial particle co-occurring together either contiguously or with insertions (Gardner & Davies, 2007, Liu, 2011), others apply an additional semantic criterion claiming that phrasal verbs need to “have meanings beyond the separate meanings of the two parts” (Biber et al., 1999, p. 404).

One of the distinctive features of phrasal verbs is their polysemy (Biber et al., 1999): they can retain their literal meaning, and also they tend to have more than one figurative meaning. Gardner and Davies (2007), for example, found that 100 most frequent phrasal verbs have 559 meaning senses. Hence at least those most frequent phrasal verbs seem to be idiomatic and very polysemous.

2.1.3 Summary of the issues with definition and extraction

Summarizing the main ideas of defining FL, it can be emphasized that FL is very heterogeneous. Therefore, it is difficult, if not impossible, to provide a definition that encompasses all of the different aspects of formulaicity and to list all the potential criteria for identification of FSs.

The four types of FSs addressed in this thesis (collocations, idioms, lexical bundles, and phrasal verbs) admittedly cover only a part of FL. However, as focusing on all the FSs was impossible, some decisions had to be made. These four categories seemed to be the best researched categories and thus a good starting point. However, I am aware that the findings based on these four categories only cannot be easily generalized to all other types of FSs.

2.2 Processing of formulaic language

One of the functions of FL is aiding the processing both for a hearer, or a reader, and for a speaker, or a writer (Wray, 2002). This section starts from summarizing the main evidence from various experimental studies that looked at processing of different types of FSs. Then it briefly discusses the question of holistic storage of FSs, which is sometimes taken as an explanation of the special status of FSs in the mental lexicon. Finally, it discusses the usage-based theory of language acquisition, as a theoretical framework that can account for the use of FSs in language, as well as two more specific approaches within the usage-based idea: lexical priming theory and statistical learning.

It is important to acknowledge that quite a lot of research on the psychological reality of FSs was based on language of impaired patients with aphasia (e.g., Mondini, Jarema, Luzzatti, Burani, & Semenza, 2002; Mondini, Luzzatti, Saletta, Allamano, & Semenza, 2005) or Alzheimer's disease (see Van Lancker Sidtis, 2012; Wray, 2010). However, some researchers even make a distinction between 'automatic language' (used by brain-damaged patients, fixed, limited in repertoire for each individual speaker, and not manipulated creatively) and 'formulaic language' (used by healthy language users) (see, for example, A. M. Peters, 1983; Van Lancker Sidtis, 2010). As the studies carried out for the present thesis looked at healthy language users only (i.e., formulaic but not automatic language), the research on patients' with brain damage is purposefully not presented here.

2.2.1 Evidence from psycholinguistic studies

Ellis claimed that "much of language learning is the gradual strengthening of associations between co-occurring elements of the language and fluent language performance is the exploitation of this probabilistic knowledge" (2002, p. 173). This idea seems to underlie the processing of FSs: they are encountered in language flow and links between the elements of the sequence are established and strengthened with each new encounter, making these sequences readily available when needed for language processing.

In recent years there have been a vast number of studies on processing FSs. There have also been attempts to look at processing of different FSs in one study. For example, Columbus (2011) set out to answer an ambitious question whether different types of FSs are processed differently. She focused on collocations, lexical bundles, and idioms and compared them to compositional sentences and semantically abnormal sentences. Columbus interpreted her results to show that FSs are read faster than control phrases, with differences between different types of FSs remaining to be determined. While the aim of the study is definitely worth attention, the actual study is rather problematic. As the study aimed for very natural sentences, all the target sentences were taken from a corpus to retain the authenticity. This made the control of potentially confounding variables (length, predictability, semantic associations, and transparency) very loose. Also, the control sentences were not matched with the target sentences, so the claims about faster reading of sentences containing FSs are also questionable. Therefore, while ideally we would like to compare different kinds of FSs, looking at different categories separately and using more controlled experimental designs seems to be a more reliable approach.

The following sections summarize experimental research on different types of FSs: collocations, idioms, lexical bundles, and phrasal verbs. These sections cannot, and is not aiming to, review all of the published studies. However, some clear trends emerge from the studies discussed.

2.2.1.1 Collocations

After corpus linguistics established the idea of the prevailing nature of collocations in language, various researchers have tried to investigate, whether collocations are only textual phenomenon or whether they are also psychologically real. These studies focused on different types of collocations, mostly lexical collocations, though, apart from a few examples, such as Sosa and MacFarlane (2002). In spite of using different experimental techniques, the studies seem point in the same direction: collocations are processed faster than matched novel phrases.

One of the first studies that looked at processing of collocations was Siyanova and Schmitt's (2008) study. They used a lexical decision task to analyse how collocations are processed by native speakers (NS) and non-native speakers (NNS) of English. While NSs were faster overall, both groups judged frequent collocations faster than infrequent ones. The researchers interpreted the results to show that collocations are processed faster than novel language because of the frequency of words co-occurring together.

Durrant and Doherty (2010) followed up on the previous study asking whether collocations are processed faster only because of the semantic association between the words, or whether there is something beyond the association that drives the processing advantage. Their lexical decision experiments showed that very frequent collocations are indeed processed faster even if the collocates are not associated. However, they also suggested that the mutual information (MI) score cut-off point of 3 (usually adopted in corpus linguistics) might be too low to extract psychologically real collocations.

Wolter and Yamashita (2014) tried analysing the effect of the L1 on processing collocations in an L2. They had higher and lower proficiency Japanese speakers, as well as NSs of English participate in a lexical decision experiment. They presented the participants with English-only collocations (non-existent in Japanese therefore incongruent for Japanese speakers learning English) and Japanese-only collocations translated into English, as well as randomly generated phrases. The elements of the target collocations were presented one above the other and the participants had to decide whether both letter strings were real words in English. The study found no facilitative effect for English-only or Japanese-only colocations for NNSs, suggesting no clear transfer from the L1 and no facilitation for incongruent items. However, curiously, English NSs showed a processing advantage not only when judging English-only collocations but also when responding to Japanese-only items. This was true, though, only for the verb-noun but not the adjective-noun collocations. The researchers took this result to indicate inherently different processing of adjective-noun and verb-noun phrases drawing on theoretical suggestions of generative grammar. However, these results could be due to some problems with the experiment itself. It would be interesting to see whether there were any specific target items potentially driving this effect and whether the same effect could be replicated with a different set of items (as it is somewhat surprising that NSs of English show a facilitative effect for Japanese-only collocations). Also, letter strings were presented one above the other. This might have encouraged word by word rather than phrase level processing. Because of these reasons, the results of this study should be evaluated with caution. However, there is a possibility that there are certain processing differences between adjective-noun and verb-noun collocations.

While the previously described studies used lexical decision tasks, Molinaro and Carreiras (2010) used the ERP technique to look at processing of literal and figurative collocations (that latter could also be classified as idioms). All of their target phrases were highly predictable. The results showed that both literal and figurative collocations are processed differently than novel phrases: collocations are pre-activated when their first element is encountered, which aids their processing. Figurative collocations also showed some additional later semantic processing. So this study suggests that there are some sort of associative links between collocates, allowing one of them to pre-activate the other, no matter if the meaning of the phrase is literal or figurative.

However, in the studies discussed above, collocations were mostly presented word by word (either as lexical decision task stimuli or in ERP studies) or without any context. Arguably, presenting collocations word by word could disrupt their processing, encouraging a word by word rather than a phrasal processing. Therefore, it is interesting to look at the studies that investigated collocation processing using more naturalistic reading tasks.

Siyanova-Chanturia, Conklin, and van Heuven (2011), for example, carried out an eye-tracking experiment investigating the effect of phrase frequency in language processing. They looked at how NSs and NNSs of English process binomials[2] (e.g., *salt and pepper*). Binomials are ideal stimuli for investigating phrasal frequency effects, because they can be reversed changing the frequency of the sequence dramatically, while the frequencies of the individual words remain exactly the same. The study showed that frequent

[2] While binomials might be classified as a separate category of FSs, they could also be interpreted as a specific type of collocations: they are also characterised by frequent co-occurrence, even if there are additional structural criteria to identify binomials.

phrases were processed faster by NSs and highly proficient NNSs. However, the frequency effect did not hold for less proficient NNSs. This result was interpreted to suggest that phrasal frequency leads to an entrenchment of the phrase in the memory. Therefore, phrasal frequency, not only individual word frequency, seems to play a role in language processing. However, the results of this study might be driven not only by the phrase frequency, but also by the perceived violation of binomial's structure, when its constituent parts are reversed (i.e., *salt and pepper* sounds more natural than *pepper and salt*). Other types of phrases that allow positional variation and do not have such a clear preferred order might be processed differently.

Some studies in eye-movement research looked at how transitional probabilities affect processing (Frisson, Rayner, & Pickering, 2005; McDonald & Shillcock, 2003a, 2003b). 'Transitional probability' is a measure calculated based on corpus data. It estimates the probability of words co-occurring together. If collocations are defined from the distributional perspective, word sequences with high transitional probabilities essentially are collocations. McDonald and Shillcock (2003a, 2003b) carried out two eye-tracking experiments with NSs of English and suggested that transitional probabilities are good predictors of reading ease, showing an effect on early eye-movement measures. However, Frisson et al. (2005) replicated and extended this research, and claimed that transitional probabilities do not have an effect on processing if the contextual predictability is controlled for. On the other hand, the authors acknowledged that contextual predictability (as measured by cloze tests) capture some information about transitional probabilities because learners probably draw on their distributional knowledge of language when completing the cloze tests. This study, thus, casts some doubts on the faster processing of collocations, showing that it is very difficult to disentangle the effects of a statistical association between words from their contextual predictability.

More recently, Sonbul (2015) carried out a study using both on-line (eye-tracking) and off-line (rating) measures with NSs and NNSs of English. She presented her participants with high-frequency, lower-frequency, low-frequency and non-attested adjective-noun pairs and looked at how phrase frequency affects processing of word sequences. She found that both NSs and NNSs were sensitive to collocation frequency in early measures of eye-movements, but not in late measures. This led her to conclude that readers initially spend more time dealing with unexpected words, but later they incorporate them in a more general adjective-noun schema and are able to process them easily.

One limitation of all the studies discussed so far is that they only looked at adjacent collocations and overlooked the fact that collocations are frequently used in a non-adjacent form (with words intervening between the collocates). This issue is further discussed in Chapter 4. Overall, the studies on collocation processing seem to suggest that words frequently co-occurring together are processed faster than novel phrases, even if those word sequences do not have a figurative meaning or are not semantically associated. Hence it seems that the phrase frequency and the probability of the words co-occurring are important driving factors for faster processing of collocations.

2.2.1.2 Idioms

Idioms have received a lot of interest of researchers for many years, mostly because of their figurative meaning which makes them very salient. The main questions about idiom processing are the following: are idioms processed holistically or compositionally and are both literal and figurative meanings of an idiom activated during its processing. As presenting an extensive review of idiom processing is beyond the scope of this thesis, this section only discusses a few studies than adopted different methodologies to look at idiom processing.

In a self-paced reading study Conklin and Schmitt (2008) looked at processing of idioms as compared to novel phrases. They created contexts with both literal and figurative meanings of the idioms. Their results suggest that both literal and figurative idioms are read faster than control phrases, both by NSs and by advanced NNSs.

In their lexical decision task, Tabossi, Fanari, and Wolf (2009) had NSs of Italian read both compositional and non-compositional idioms, clichés and novel phrases. They concluded that familiar phrases (both idiomatic and not) are processed faster than novel ones. However, the similarities between processing idioms and clichés have to be interpreted with caution, because the clichés in this study were chosen only based on native speaker intuition and presented in sentences, in contrast to idioms which were presented without any context.

Vespignani, Canal, Molinaro, Fonda, and Cacciari (2009) found idiom processing advantages for Italian NSs both in a self-paced reading task and an ERP study. They also showed that when processing idioms, we are predicting the upcoming words, which might be the reason why idioms show processing advantage. Rommers, Dijkstra, and Bastiaansen (2013) also carried out an ERP study with idioms. Their findings confirmed the hypothesis that idioms are processed differently than novel phrases. They also suggested a literal meaning of the idiom is not activated during the processing.

When trying to answer the question whether both meanings of an idiom are active during the processing, Siyanova-Chanturia, Conklin, and Schmitt (2011) analysed the processing of idioms used in their figurative and literal sense and novel phrases. They carried out an eye-tracking study and found that NSs read idioms faster, no matter if they were presented in their literal or figurative meaning, while NNSs showed no processing advantages. On the contrary, idioms used in figurative meanings impeded their processing. These results suggest that both literal and figurative meanings of idioms are activated simultaneously. One potential reason for the difference between Rommers et al.'s (2013) and Siyanova-Chanturia et al.'s (2011) studies is the different methodologies used: during eye-tracking experiments the sentences are presented in a more naturalistic reading situation.

Overall, it seems that idioms are processed faster than novel phrases due to their familiarity and predictability. Various models have been built to explain idiom comprehension, such as idiom list hypothesis (Bobrow & Bell, 1973), lexical representation hypothesis (Swinney & Cutler, 1979), configuration hypothesis (Cacciari & Tabossi, 1988), idiom decomposition hypothesis (Gibbs, Nayak, & Cutting, 1989), and Superlemma theory (Sprenger, Levelt, & Kempen, 2006, Kuiper, Van Egmond, Kempen, & Sprenger, 2007). The more recent theories moved away from initial ideas of idioms being simply listed in the mental lexicon as 'large words' (Bobrow & Bell, 1973), to a more complex view of idiom representation as both unitary and compositional at the same time (Sprenger et al., 2006).

2.2.1.3 Lexical bundles

While idioms have received a lot of attention, lexical bundles have been studied far less in psycholinguistics. Several studies that looked at their processing also showed a facilitative effect for familiar lexical bundles. It is worth noting, though, that this research did not necessarily focus on the lexical bundles as a category defined by Biber et al. (1999), but rather more generally on recurrent fixed sequences of words. Therefore, different studies referred to them differently, using terms such as clusters, formulaic sequences or n-grams.

Schmitt, Grandage, and Adolphs (2004) were probably the first ones to look at the processing of word clusters. They used a dictation task and suggested that some of the clusters were more likely to be stored holistically than the others, but neither the frequency nor the length of the cluster seemed to explain the differences between different items well. However, the FSs in the study varied in terms of frequency, length, and transparency and the analysis did not take these differences into account systematically.

Following on the previous study, Nekrasova (2009) carried out two studies looking at discourse organizing and referential lexical bundles. In the first study she gave a lexical completion task to a group of NS and NNS participants while in the second study the participants performed a dictation task. The results of the first study were in line with Schmitt et al.’s (2004) study, suggesting that some of the FSs seem to be more easily accessible for NSs and more proficient NNSs than others. However, the results of the second study were rather surprising with advanced NNSs recalling more lexical bundles than NSs. Nekrasova suggested that this was probably due to NNSs being used to focusing on listening and memorizing foreign language. While this might be the case, the results could also be interpreted to indicate that some lexical bundles are stored differently in NS and NNS mental lexicons. However, an important limitation to both Nekrasova’s studies is that the fact that the experiments did not include any control phrases. Hence no clear conclusions about faster processing of lexical bundles as compared to novel language can be drawn.

In a much better controlled study, Arnon and Snider (2010) looked at the phrase frequency effect on processing lexical bundles. They carried out two lexical decision experiments on phrases that differed in phrase frequencies, but were matched in word frequencies (e.g., *don't' have to worry* versus *don't have to wait*). They demonstrated that more frequent phrases are processed faster and this effect depends on the phrase frequency rather than on individual word frequencies. They also suggested that frequency is best treated as a continuum: there seems to be no threshold of when frequency becomes important or stops having an effect. These findings are in line with the usage-based models, suggesting that our experience with language affects the way language is represented in our brain.

There have also been a number of self-paced reading studies focusing on the processing of lexical bundles. Schmitt and Underwood (2004) found mixed results and suggested that self-paced reading task may interfere the processing of longer sequences and therefore is probably not the best methodology for investigating FSs. However, Tremblay, Derwing, Libben, and Westbury (2011) replicated the aforementioned study and concluded that lexical bundles show a processing advantage, which is better observable if the text is presented in larger chunks rather than word by word. Reali and Christiansen (2007) also showed the frequency effects on processing lexical bundles, suggesting that more frequent chunks are read faster. However, the sentences they used as stimuli seem to be rather artificial in trying to introduce ambiguity. Also, they used *Google* search engine to estimate the phrase frequency which is a debatable choice (see Kilgarriff, 2007 for a discussion).

Taken together, the studies on lexical bundle processing seem to suggest a faster processing of recurrent fixed phrases than of matched novel phrases. The reason for faster processing of lexical bundles seems to be their phrasal frequency, as they are not salient and some of them even semantically incomplete.

2.2.1.4 Phrasal verbs

The questions about processing phrasal verbs are essentially the same as the questions about processing idioms because phrasal verbs also have a figurative meaning. A few research studies looked a processing of phrasal verbs.

Matlock and Herediag (2002), for example, carried out two experiments looking at how NSs and NNSs process phrasal verbs. The first experiment was a sentence completion task, where the participants could complete the sentence following the ambiguous phrasal verb by interpreting it either as a phrasal verb or as a verb with a preposition. Their results suggested that NSs were more likely to interpret the word sequences as phrasal verbs than NNSs. In the second experiment they looked at the reading times of sentences containing prepositional verbs (*Paul went over the bridge with his bicycle*) and sentences containing phrasal verbs (*Paul went over the exam with his students*). Both NSs and early bilinguals read phrasal verbs faster than verb + preposition phrases. The authors took this finding to suggest that the figurative meaning of the phrasal verb is activated faster than the compositional meaning. However, the results are somewhat problematic, as the sentences were different for the different conditions, but only the whole sentence reading times were compared.

Gonnerman and Hayes (2005) in a priming task looked at how a verb in a phrasal verb primes the whole phrasal verb. They found that the priming effect depends on how semantically similar the meaning of the verb is to the meaning of the phrasal verb: the more the verb retained its meaning in the phrasal verb, the larger priming occurred. Blais and Gonnerman (2013) found the same result with NNSs. Thus taken together these two studies show that the more the verb loses its individual meaning in the sequence, the more lexicalized the phrasal verb becomes.

Kim and Kim (2012) in their self-paced reading study, did not compare literal and figurative readings of the phrasal verbs, but rather looked at the importance of phrasal frequency for the processing of phrasal verbs. They carried out a self-paced reading study on verbs used with the adverbial particle *out*. They found that their participants (both NSs and NNSs of English) were sensitive to phrasal frequency of phrasal verbs and processed the more frequent phrasal verbs faster. However, their study was limited to very small number of participants (11 NSs and 14 NNSs), as well as only one adverbial particle analysed. Even so, it still corroborates the idea that phrasal frequency plays a role in processing FSs.

There are a number of differences in the studies on processing phrasal verbs and the studies on processing other types of FSs. The most important one seems to be the fact that while other sequences were compared to novel control phrases, it is very difficult to create a semantically plausible control phrase for a phrasal verb. Therefore, the studies discussed above simply indicate that frequency and semantic transparency of a phrasal verb play a role in its processing, without making direct comparisons to novel language.

2.2.1.5 Summary of findings on different types of FSs

While the studies reviewed above (together with many others) are based on different FSs and used various methodologies, they all seem to reach very similar conclusions: FSs are processed faster and/or more accurately than novel phrases. This advantage might be driven by different characteristics of a FS: frequency and mutual expectancy of the words for some of them, and salience together with familiar figurative meaning for the others. However, what they all have in common is a facilitative effect when compared to novel language.

2.2.2 The question of holistic storage

One of the questions driving a lot of research on FL is whether FSs are stored and processed holistically (without any recurrence to their individual components) as opposed to novel language, which is analysed into its component parts. The idea of the holistic storage of FSs is very prominent: it is included in Wray's (2002) definition of a FS (see Section 2.1); implied in Sinclair's idea of the idiom principle (Sinclair, 2004b); and a number of researchers have interpreted their results as evidence for the holistic storage of FSs (e.g., Altenberg, 1998; Dahlmann & Adolphs, 2007; Jiang & Nekrasova, 2007; Underwood, Schmitt, & Galpin, 2004). If FSs were indeed processed holistically, this would make a distinction between what is formulaic and what is not relatively clear. However, the picture seems to be rather complicated.

Wray (2002) suggested that there are two different modes of processing language: analytic and holistic. When engaged in the analytic processing, language users rely on their knowledge of grammatical rules and their analytical capacity, while the holistic processing is based on processing pre-fabricated chunks that are accessed directly from memory without any segmentation. Both processes seem to be happening simultaneously. While for novel language the direct access is not available, for recurrent phrases the direct retrieval is faster. This idea of dual processing (both direct access and computation) is especially relevant for idiom research (e.g., Abel, 2003). There seems to be neurological evidence for dual processing mode as well (Van Lancker Sidtis, 2009, 2012).

Dual processing implies that the same phrase can be stored in the mental lexicon a number of times: both as an entire phrase and as its constituent parts. Because of that, the phrase can be analysed when there is a need for such an analysis, but it can also be accessed as a whole when the analysis is not necessary (and the holistic retrieval is faster). This idea led to the conclusion that "units of different sizes and internal complexity can be stored in the lexicon as morpheme equivalents" (Wray 2002, p. 278).

In her later work, though, Wray claimed that it is still questionable "whether a processing advantage in terms of speed indicates holistic storage or simply the faster mapping of components – and indeed whether these are really two different things or just different ways of conceptualizing the same thing" (Wray, 2012, p. 233-234). While she did not abandon the idea of the holistic storage and presented it as an alternative, some researchers seriously question this idea. For example, Siyanova-Chanturia (2015) noted that the metaphor of holistic storage is very frequently used in corpus linguistics and applied linguistics, but the evidence from psycholinguistic processing studies point in the opposite direction. She argues that some researchers who interpreted their findings as an indication of holistic storage of FSs (e.g., Altenberg, 1998; Dahlmann & Adolphs, 2007; Jiang & Nekrasova, 2007; Underwood et al., 2004), carried out studies that did not address the question of holistic storage. Therefore, their results can only be taken to show that frequent and familiar phrases are processed faster than control phrases. Siyanova-Chanturia and Martinez (2015) drew the same conclusion arguing that the evidence we have does not answer the question whether FSs are retrieved as wholes, but rather shows that phrase frequency, familiarity, and predictability play a role in language processing. However, Siyanova-Chanturia (2015) makes it very clear that the fact that FSs are not necessarily stored holistically does not undermine their importance. The finding that they are processed quantitatively faster than novel phrases can be taken as supporting evidence for usage-based approaches to language acquisition.

Drawing on Siyanova's and Martinez's (2015) observations, it seems that with the empirical evidence available, we cannot assume holistic storage of FSs. However, this does not necessarily mean that no sequences are stored holistically. Furthermore, Wray (2002) makes it clear that if one member of a category (e.g., an idiom) is stored holistically for a particular individual, this does not imply that all other items of that category (i.e., all idioms) will be stored holistically as well. In the same way, if this item is stored holistically for a particular speaker, it does not mean that it is processed holistically by everyone else speaking that language. This approach, while complicating the picture, might be more useful than seeing holistic or analytical processing as a simple dichotomy. It places familiarity for a particular speaker in the centre of language processing.

In the thesis, I am assuming a psychological reality but not the holistic storage of FSs. I am taking the position that at the current stage of research, we cannot make any strong claims about the holistic storage of all the FSs. Nonetheless, we cannot completely discard the idea that individual sequences or even some categories of FSs can be accessed holistically.

2.2.3 Theoretical approaches that account for FL

Because of the importance of FL and the fact that it is processed differently than novel language, FL should be integrated in theories of language acquisition and processing. This section looks at theories that can account for the importance of FSs in language use. It starts from briefly discussing usage-based approaches to language acquisition and then it looks at two more specific approaches, namely, lexical priming (Hoey, 2004, 2005, 2012) and statistical learning.

2.2.3.1 Usage-based approach

Usage-based models of language acquisition are informed by experimental research and corpus-analysis and they differ considerably from the traditional formal approaches to language and grammar (Tomasello, 2005). As Tomasello summarized:

> [a]s opposed to conceiving linguistic rules as algebraic procedures for combining words and morphemes that do not themselves contribute to meaning, this approach conceives linguistic constructions as themselves meaningful linguistic symbols – since they are nothing other than patterns in which meaningful linguistic symbols are used in communication (2005, p. 5).

Hence, a construction is taken to be a building block of language. We learn constructions from communicating with others: linguistic competence emerges from a large number of utterances that we remember and regularities that we extract from them (Ellis, 2015). For usage-based models, language processing and acquisition are driven by the same cognitive skills (e.g., analogy finding, distributional analysis, and automatization) as any other types of learning (Tomasello, 2005). Langacker (2000) claimed that there is no such specific system as grammar, rather when encountering utterances we compare them to the language experience we already have in order to derive their meanings. Each such comparison shifts our overall linguistic experience so the grammar is fluid. Thus, "[a] usage-based view takes grammar to be the cognitive organization of one's experience with language" (Bybee, 2006, p. 711). Because of that, the representation of language in the brain of two individuals will never be identical (Kemmer & Barlow, 2000).

Importantly, when we learn a language as infants, we are not exposed to individual words in isolation, but rather to sequences of words (Tomasello, 2005). Therefore, we start learning a language from utterances. Even if later on in our lives we generalize them into certain schemas, we still have multiple representations of certain units, including single words, abstract constructions, and very concrete specific utterances. Therefore, it seems that adults, as well as children, simultaneously have access to linguistic units that differ in terms of their complexity (Langacker, 2000; Tomasello, 2005; also, in line with Wray's (2002) heteromorphic distributed lexicon. Therefore, redundancy is seen as a natural feature of language stored in the mental lexicon (Tomasello, 2005). As a consequence, usage-based models do not conceptualise FSs as some sort of exceptions to the linguistic system but rather as units of language representation. According to Tomasello, “[i]t turns out that, upon inspection, a major part of human linguistic competence – much more than previously believed – involves the mastery of all kinds of routine formulas, fixed and semi-fixed expressions, idioms, and frozen collocations” (2005, p. 101).

As language knowledge is largely experience driven, frequency becomes central for usage-based theories. Frequency leads to an entrenchment: each occurrence of an event leaves some trace in our brain which facilitates further occurrences of the same event (Langacker, 2000). While frequency is very important for usage-based models, it is not the only explanatory factor for language acquisition and processing. Tomasello (2005) emphasised that pattern finding (as driven by distributional analysis) is as important as intention-reading. Infants only learn a language from exposure because they are capable of reading social intentions of their interlocutor. Hence the social aspect of language use is also crucial for usage-based approach. As Ellis and Larsen-Freeman (2006) claimed, language is constructed and guided by our social needs both in processing and in acquisition. Ellis (2002, 2012a) also claimed that there are various factors beyond frequency, playing a role in language acquisition, such as salience, significance for comprehension, prototypicality, generality, redundancy, communicative intent, and semantic basicness. So while frequency is an important factor, it is by no means the only driving force in language acquisition.

‘Usage-based approach’ is a very broad term that can be used when referring to different theories or grammatical approaches, as long as they take speakers’ experience into account. I will briefly discuss two lines of research that are fully in line with usage-based approach and that seem to be important when looking at the processing of FSs. One of them is corpus-informed (the lexical priming theory), while the other one is based on research in psychology and pyscholinguistics (statistical learning).

2.2.3.2 Lexical priming

Lexical priming is a linguistic theory developed by Hoey, when trying to account for collocational use in language (Hoey, 2004, 2005, 2012). The theory is driven by the findings of corpus linguistics showing the pervasive use of collocations. Hoey tried to draw on research in psycholinguistics (especially on semantic and repetition priming) to develop an explanation of this ubiquitous nature of collocations.

According to Hoey's theory, every time we encounter a word, information about its grammatical features, collocations, colligations, textual position, semantic associations, and genre is accumulated. Consequently, when we want to use that word, it is primed for certain usage patterns (Hoey, 2009). The lexical priming goes beyond the sentence level. It extends to the discourse level as well: a word is primed to occur (or not to occur) in certain genres or specific textual positions (O'Donnell, Scott, Mahlberg, & Hoey, 2012). Also, lexical priming accounts for the sociolinguistic context: speakers from different linguistic communities might be primed to use different words (Hoey, 2009). Lexical priming theory, therefore, accounts for both collocational use and individual differences, as each language user, depending on their experience with language, would be primed for slightly different uses (Hoey, 2004).

When it comes to processing FSs (at least the ones that are defined by their frequent co-occurrence), the lexical priming theory assumes that the elements of the sequence pre-activate each other and in such a way aid each other's processing. Hence, the lexical priming idea is very strongly related to predictability, which was shown to play a major role in language processing (see Kuperberg & Jaeger, 2016 for an overview).

While lexical priming theory draws on psychological research, its psychological reality has not been investigated so far (Hoey, 2012). There have been attempts to summarize the experimental evidence from research on artificial intelligence and psychology that is in line with the lexical priming (see Pace-Sigge, 2013), but the theory itself and its specific predictions have not been tested. So its psychological validity still remains to be established even if the existing body of research in psycholinguistics seems to support it.

2.2.3.3 Statistical learning

Statistical learning, as opposed to the lexical priming theory, is mostly studied in psychology. For the representatives of the statistical learning paradigm, learning is essentially based on finding patterns and extracting generalizations in order to be able to use them for making predictions in the future (Onnis, 2012). So statistical learning is not language specific, on the contrary, it is "a robust, domain-general, age-independent and not specifically human ability" (Shukla, Gervain, Mehler, & Nespor, 2012, p. 174). When it comes to language acquisition, though, statistical learning was defined as "our ability to make use of statistical information in the environment to bootstrap language acquisition" (Rebuschat & Williams, 2012, p. 1).

The information we can statistically extract from a language flow ranges from simple frequencies of words to complex dependencies and co-occurrences. There is now plenty of evidence that humans are able to track transitional probabilities and use this information for language processing. For example, infants seem to segment language flow into words using transitional probabilities between syllables as at least one of the potential cues (Johnson, 2012). Computational models that tried to account for child language segmentation and production using transitional probabilities have also been relatively successful (Bannard, Lieven, & Tomasello, 2009; McCauley & Christiansen, 2011). Studies show that regularities in unknown languages can be learnt from exposure only, even when all the other clues, such as prosody, are removed (Citron, Oberecker, Friederici, & Mueller, 2011).

Even if statistical learning studies usually focus on transitional probabilities between syllables or sounds, there is no reason to believe that the same principles do not apply at the lexical level as well. In case of FL, transitional probabilities can be tracked between words, making certain words more predictable than others. While, research on processing FSs has mostly been limited fixed adjacent forms of FSs, studies on learning dependencies between elements have also focused on non-adjacent items.

Linguistic patterns can be both adjacent (e.g., certain consonants following each other) or non-adjacent (e.g., grammatical categories encoded non-adjacently, such as person marking in *he sings*). Therefore, in order to learn a language, infants "possess a robust mechanism for tracking non-adjacent dependencies. This ability is essential for early grammatical development” (van Heugten & Shi, 2010, p. 223). Overall, studies on learning non-adjacent dependencies show the following results:

1. adjacent dependencies seem to be easier to learn than non-adjacent dependencies (Gómez, 2002; Newport & Aslin, 2004);
2. non-adjacent dependencies seem to be learned better if the two dependent elements are perceptually similar and the intervening element is different (Gebhart, Newport, & Aslin, 2009);
3. non-adjacent dependencies seem to be learned better if there is a lot of variation of the intervening elements (Gómez & Maye, 2005);
4. distance between the non-adjacent elements seems to play a role in learning (Santelmann & Jusczyk, 1998).

A lot of studies analysed infants’ ability to learn non-adjacent dependencies. It has been suggested that infants by default start extracting statistical co-occurrence information about the adjacent dependencies. Only if adjacent elements show very low transitional probabilities, infants look for regularities further away and possibly learn non-adjacent dependencies if they are very regular (Gómez, 2002). Research also shows a crucial role of experience in learning: infants as young as 12-months old were demonstrated to be able to track non-adjacent dependencies if they were first exposed to the same elements presented adjacently (Lany & Gómez, 2008). More recent studies, though, questioned the idea of the difficulty of learning non-adjacent dependencies. For example, Vuong, Meyer, and Christiansen (2011) looked at adults learning adjacent and non-adjacent dependencies simultaneously. They showed that both adjacent and non-adjacent dependencies were learned at the same time and that their learning was strongly correlated.

All these studies on statistical learning mostly looked at dependencies at the morpheme, syllable or even phoneme level, mostly mirroring learning of grammatical constructions. However, dependent elements with a semantic meaning (as in case of FL) might be even easier to notice and learn than meaningless elements in artificial languages. For instance, Vries, Petersson, Geukes, Zwitserlood, and Christiansen (2012) argued that “[b]y adding semantics to a structure, establishing non-adjacent dependencies is easier, as additional information about the underlying structure is revealed” (p. 2074).

While this claim may seem to be obvious (we learn and use language, because it carries certain meanings), in the studies on statistical learning semantics are usually ignored. Therefore, even if the studies on learning non-adjacent dependencies seem to show that we are able to track and store information about non-adjacent co-occurrences, there is yet a lot to be explored. One of the aspects to be investigated is how we learn dependencies at the lexical level.

2.2.4 Summary of the issues with processing FL

Summarizing the main points on processing FL, it can be concluded that there is a general consensus that FL has a processing advantage when compared to novel language. Even if FL, as already noted, is a very heterogeneous phenomenon, studies on processing various types of FSs seem to point in the same direction. This finding suggests that FSs are not just theoretical constructs of corpus linguistics, but rather psychologically real units. This facilitative effect for FSs seems to be (at least partially) due to their frequency, as well as their predictability and familiarity for the language users.

As frequency, predictability, and familiarity are inextricably connected to the language use and exposure to language, usage-based theories are able to explain the pervasive use of FSs in language. The two more specific theories discussed (lexical priming and statistical learning), while coming from the very different perspectives, make rather similar claims. Our exposure to language seems to create certain expectancies for the words to co-occur. These expectancies seem to be driving our language production and facilitating comprehension of familiar language. However, while it seems to be clearly established that co-occurrences of items play a role in processing, it is still much less clear of what happens when the elements co-occur non-adjacently. The research on statistical learning suggests that non-adjacent dependencies are learned. However, there is a lack of research on learning and processing non-adjacent dependencies at the lexical level.

2.3 Formulaic language and SLA

Learning an L2 is usually a very different process than learning the L1. Wray (2002) notes that learning situation, learners' age, and their language learning agenda usually differ considerably in the L1 and in an L2. L2 learners are usually adults, they receive much less exposure to the target language, and the language they are exposed to is not necessarily tailored to their abilities. Also, most of L2 learners are already literate in their L1. Moreover, L2 learners usually have much higher cognitive abilities and tend to approach language learning from a more analytic perspective. Furthermore, while children seem to start learning a language from chunks that they can map meanings on (Bannard & Lieven, 2012; A. M. Peters, 1983; Tomasello, 2000), adults learning an L2 seem to focus on words as essential building blocks. Therefore, classroom taught learners tend to overgeneralize the creativity of grammatical constructions (Wray, 1999). As Foster (2001) noted, if learners focus on individual words and grammar for constructing grammatically accurate novel language, they sound slow and non-fluent.

The way teachers approach language also contributes to the difficulty of acquiring FSs. Traditionally, teaching single words received most of the attention in a language classroom, because words are generally believed to be central elements of meaning and because it is simply easier to teach individual words (Schmitt, 2010). Collocational features were understood only as an aspect of word knowledge. For example, when conceptualizing the components of vocabulary knowledge, Nation (2001) listed these aspects:

- spoken form;
- written form;
- word parts;
- form and meaning link;
- concepts and referents;
- associations;
- grammatical functions;
- collocations;
- constraints of use.

So for Nation, collocational knowledge (or in a broader sense, the knowledge of the FSs that a word can be used in) was only one of the components of word knowledge. However, following Sinclairian ideas that a meaning is not attached to a single word but rather arises from a co-selection of words, a FS can be equally conceptualized as a lexical unit (see, e.g., Brown, 2014; Revier, 2009). It can be suggested that, for example, idioms, phrasal verbs or collocations are lexical units on their own rights that can be known to a learner to different degrees: a learner might only know a spoken or a written form of a FS, they might need to learn its grammatical patterns of use, a FS can have its own collocations, frequency profile, appropriateness for certain contexts or genres, one or more meanings, and associations. Hence this list of dimensions of the vocabulary knowledge can easily be extended to FSs as well.

Recently the attention in language teaching has shifted from individual words to a lexicon that includes both words and FSs. Focussing on FL seems to be important at all stages of learning an L2. It can be a useful starting point for beginner learners. For example, Schmidt (1992) suggested that FSs offer reliable building blocks which give learners time to construct further chunks of language. Nattinger and DeCarrico (1992) also claimed that learning FSs helps learners to improve fluency and increase motivation as it enables them to use phrases that are too complicated for them to construct from scratch. Furthermore, FL seems to be important beyond the beginner level. It was suggested that phrasal knowledge is a better predictor of reading success than the knowledge of individual words or syntactic knowledge (even if phrasal knowledge itself is strongly related to vocabulary knowledge and somewhat related to syntactic knowledge) (Kremmel, Brunfaut, & Alderson, 2015). Also, it seems that the use of FL has a massive influence on how the learners' lexical proficiency is judged: Crossley, Salsbury, and Mcnamara (2015) showed that collocational knowledge explained almost 90% of the variance in the judgements on candidate's lexical proficiency (considerably more than lexical diversity or frequency of the words used).

Despite the agreement that FSs are important in acquiring an L2, it is generally believed that the knowledge of FL is not simply naturally acquired with growing vocabulary but rather lags behind the general vocabulary knowledge (e.g., Bahns & Eldaw, 1993). The next section looks at the reasons why FL causes difficulties for language learners.

2.3.1 Main difficulties in acquiring formulaic sequences

Ellis (2008b) claimed that "[l]exical phrases are as basic to SLA as they are to the L1" (p. 97). According to him, in order to achieve fluency in both L1 and L2 we need to acquire memorized sequences. While for NSs, who have a lot of exposure to language, usage-based theories can explain the acquisition of FSs well, for the NNSs, the picture might be more complex. A lot of learners around the world have very little exposure outside the classroom, and this amount of input can simply not be enough for any kind of usage-based or statistical learning to occur. There have even been suggestions that corpus frequency is not as good a predictor for collocational knowledge, as it is for individual words (González Fernández & Schmitt, 2015). This claim, though, has to be interpreted with caution. Potentially, the frequency has a very similar effect on both individual words and FSs, but FSs are simply not as frequent as individual words and therefore the frequency effect is not easily observable.

Apart of the problem of not enough exposure to FL, there are several other challenges in learning FSs in an L2. The main ones seem to be a large number of items to learn, low salience of these items, and L1 influence. These three issues will be briefly discussed below.

2.3.1.1 Large number

More than three decades ago, Pawley and Syder (1983) claimed that in order to achieve a native-like fluency, L2 speakers have to master hundreds of thousands of memorized chunks. Bishop (2004) even claimed that most of L2 learners never achieve high proficiency the second language because of their failure to fully master the "memory based lexical system" (p. 15). This reason for learners having difficulties seems to be rather obvious: the numbers of FSs are simply overwhelming. For example, Gardner and Davies (2007) after studying the British National Corpus (BNC), found more than 12,000 different phrasal verbs. Liu (2003) compiled an initial list of idioms from various dictionaries for his study and ended up with a list of 9,683 items. These are already very high numbers, but they might be even higher for collocations or lexical bundles. So there is no way to address all these FSs in a classroom along with all the individual words.

Moreover, the individual sequences are not necessarily very frequent. For example, Grant and Nation (2006) showed that the ‘core idioms’ in English are so infrequent that they are probably not worth spending teaching time on. However, if they are infrequent, they will not be encountered often enough to be learned from exposure either. On the one hand, it is only natural that as there are frequent and less frequent words, there are also frequent and less frequent FSs (Nation & Webb, 2010). On the other hand, this leads to a question if those infrequent items are worth acquiring at all (Moon, 1998c).

As there are so many different FSs that it is virtually impossible to teach them all, teachers are usually encouraged to raise awareness about those sequences and to leave them for independent learning (Boers & Lindstromberg, 2012). While raising awareness is definitely a useful technique, it is not unproblematic. It has been pointed out in a number of occasions that raising awareness tends to lead to better awareness but not necessarily to more accurate use of FSs (Boers & Lindstromberg, 2012; Cortes, 2006)

2.3.1.2 Low salience

Another issue with learning FSs is low salience of many of the sequences. Even when encountering a FS, learners do not necessarily notice it as such (e.g., Bishop, 2004). This seems to be the reason why learners sometimes fail to understand non-transparent FSs without even realizing it. For example, Laufer (1989) noted that idioms can be deceptively transparent for learners. Martinez and Murphy (2011) carried out a reading study to investigate the question further. They gave their participants two short texts composed of the same words, but one text was much more idiomatic than the other. They found that the comprehension of the idiomatic text was significantly lower, but the participants overestimated their understanding. Spöttl and McCarthy (2004) also showed that learners find it difficult to evaluate their knowledge of FSs and tend to overestimate their knowledge. Overall, it seems that learners do not only fail to notice FSs but also do not realize they are failing to understand some important information.

This finding is problematic for incidental learning of FL. It is established that some knowledge of FSs can be (and is) acquired implicitly, when learners are exposed to language (see Section 2.3.2.2). However, in order for that learning to occur, a learner has to notice the form to be learned. In other words, learning can be unintentional, but it cannot be subliminal (Schmidt, 1990; 1993). Schmidt (1990) suggested that factors determining noticeability include salience and frequency. Different types of FL vary greatly in terms of their salience and frequency (e.g., lexical bundles are frequent and not salient, while idioms are salient but rather infrequent). Therefore, they might have different chances of being noticed and thus being learned.

2.3.1.3 L1 interference

Also, the L1 plays a role in acquiring FL. The L1 seems to have a somewhat different effect on different types of FSs. For example, Laufer (2000) showed that the similarity of idioms between the L1 and an L2 affects the use of idioms in the L2: idioms which are partially similar in both languages tend to be avoided. Bahns (1993), on the other hand, claimed that collocations are usually transferred from the L1 more willingly than idioms. Laufer and Waldman (2011) also found the L1 could account for one third of the persistent errors in collocations used by their advanced group of participants. So it seems that learners are aware of a special status idioms (which have a figurative meaning and hence are more salient) and avoid simply transferring them from the L1 to the L2. Collocations, on the other hand, seem to be willingly transferred causing numerous errors even for advanced speakers.

Transfer is only one type of influence that the L1 can have. A different challenge is learning FSs that do not exist in one's L1. For example, if learners do not have a concept of phrasal verbs, teaching phrasal verbs can be more difficult (Dagut & Laufer's, 1985; Paquot & Granger, 2012).

Overall, depending on the tendencies of use in the L1, FSs can be either avoided or transferred too willingly, and both of these strategies can lead to non-native-like use of language.

2.3.2 Learning formulaic sequences

Even if the importance of teaching FSs is established and a lot of researchers have encouraged focusing on FL in classroom (see, e.g., Lewis's (1997, 2000) ideas on the lexical approach), the question remains whether the same methods can be adopted for teaching FSs as for teaching individual words. Different studies have looked at teaching FSs directly, introducing them inductively, or learning them from exposure. There were studies that looked at teaching different types of FSs: idioms (Alali & Schmitt, 2012; Cooper, 1998; Irujo, 1986), lexical bundles (Cortes, 2004, 2006; Jones & Haywood, 2004), and phrasal verbs (Dagut & Laufer, 1985). However, it is beyond the scope of this thesis to review all of them. As this thesis will mostly focus on collocations, this section will summarize studies that analysed collocations. It will look at two approaches of dealing with collocations in a language classroom: focusing on them directly (Section 2.3.2.1), or leaving them to be picked up incidentally (Section 2.3.2.2).

2.3.2.1 Direct teaching

Even advanced students struggle to use appropriate collocations (Nesselhauf, 2003). Therefore, Nesselhauf (2003) argued that teaching collocations should not end with raising awareness: they should be included in teaching syllabi for direct teaching. Erman (2007) also pointed out that even if learners might have no problems decoding collocations, they might struggle to encode them. So it is worth focusing on collocations, even if they are transparent.

Numerous studies have addressed questions related to teaching collocations directly. For example, Webb and Kagimoto (2009) carried out a classroom study where learners were either presented with a receptive task (reading a translation and three corpus examples) or had to perform an easy productive task (inserting collocations in sentences). The study showed that a large number of collocations can be taught relatively quickly using the same methods as for teaching individual words. However, there was no difference between the effectiveness of the receptive and the productive tasks. This might have been the case because the productive task was rather easy and did not require a lot of attention, though.

In their later study, Webb and Kagimoto (2011) presented their learners with translations of collocations and explicitly asked them to learn the lexical items. The researchers wanted to investigate how the position of a node word, the number of collocates, and the synonymy of collocates affected learning. They found that the position of node does not make any difference for learning and that it is worth teaching multiple collocations of the same node. Teaching synonymous collocations, though, seemed to hinder learning. While the findings of both Webb and Kagimoto's studies seem to encourage teaching collocations explicitly, they both have the same important limitation: they did not include a delayed post-test. Therefore, it is not clear how durable the learning outcomes were.

Boers, Demecheleer, Coxhead, and Webb (2014) looked at the effectiveness of different vocabulary learning exercises. They compared various collocation matching exercises to the exercises where the collocations were presented intact. The rationale for their study was the idea that guessing an incorrect collocation in a matching exercise during the learning stage can leave a memory trace which would hinder further learning of that collocation. The authors did not find any significant differences between the groups doing different exercises, but the groups were rather small. However, more importantly, the overall gains were very low. As this was not an incidental learning study and the aim of learning the collocations was communicated to the learners, these low gains raise a question of the effectiveness of the direct teaching of collocations. If similar gains can be achieved incidentally, it remains questionable if it is worth using classroom time for teaching. However, one of the reasons for the low gains might have been the use of collocations with delexicalized verbs only, which may be more difficult to acquire.

Another question investigated in direct teaching research, was the effect of repetition. E. Peters (2014) looked at a number of repetitions needed to learn collocations and single words during vocabulary-focused tasks. Twelve words and 12 collocations were given to participants as a list of items to learn, followed by some vocabulary exercises. The results of the study showed an effect of frequency. Also, the participants learned more individual words than collocations. The findings suggest that the frequency of exposure plays a role in deliberate teaching as well as it does in incidental acquisition, and also that collocations are somewhat more difficult to learn than individual words.

Serrano, Stengers, and Housen (2015) extended the question of repetition by looking at the effect of spacing for the repetition. They compared differences of acquisition of FSs between an intense and a regular language teaching programs to investigate if spacing helps or hinders learning. They looked at beginner, intermediate, and advanced learners. Their results showed that beginners and intermediate learners benefited from the intensive program more than advanced learners. While overall the effect of spaced repetition is very interesting to address and the results seem to have clear pedagogical implications, the study has some limitations. First of all, the study does not report any comparisons between the groups in regular and advanced courses; the description of the task they had to perform is rather short and therefore it is not clear how many words the learners produced and how large the whole dataset was. Because of these reasons it is difficult to evaluate how reliable the findings of the study are.

E. Peters (2016) looked at the effect of congruency, type (adjective-noun, verb-noun, and phrasal verb-noun), and length of collocations. The participants were asked to learn collocations and complete some exercises with them. The results showed that the length of collocations played a clear role in learning. Congruency was important only at the form recall level, but not in the form recognition task (though the author did not specify if direct translational equivalents were available as distractors). The study also showed that adjective-noun collocations were the easiest ones to learn. So the study shows that apart from the repetition, there are a number of variables that also affect the chances of success of learning collocations.

Overall, it seems that teaching collocations directly is a relatively effective approach, with a rather large number of items learned quickly. But even for direct teaching, a number of occurrences, spacing of these occurrences, the item type and its congruency with the learners' L1 play a role in acquiring FSs. So very similar issues have to be considered as when looking at the incidental acquisition of collocations.

2.3.2.2 Incidental acquisition

Considering a large number of collocations to learn, all of them can never be addressed in a classroom. Therefore, it is important to look at how likely the learners are to acquire FSs incidentally.

Meunier (2012) claimed that "[e]xposure to a large amount of natural native-like input (be it oral or written) remains one of the prerequisites to promoting the acquisition of formulaic sequences" (p. 113). It has also been argued that FL is so closely related to culture, that successfully acquiring FSs is a culturally loaded process, so they are best learned from the contact with the L1 community (Dörnyei, Durrow, & Zahran, 2004). There have been attempts to investigate the effects of immersion in an L2 speaking community for acquiring FSs. For example, Adolphs and Durrow (2004) carried out a small-scale longitudinal study with two Chinese learners of English studying at a British university and concluded that there seems to be a relationship with learner's social integration and their use of native-like FSs. While this one case study is probably not enough to make general claims, the results seem to be intuitively convincing. Li and Schmitt (2009) also carried out a longitudinal study by collecting all the writing assignments produced by one Chinese student during an academic year at a British university. They looked at different FSs in her writing and concluded that in that year she improved both in terms of accuracy of using FSs and in terms of confidence of use. However, in this case, the learning was not purely incidental. The student had a lot of exposure to English language, but she also received corrective feedback from the researchers, as well as some direct teaching. However, even if both explicit and incidental learning seemed to have led to acquiring new sequences, direct personal feedback seemed to be the least effective source for learning.

While extensive exposure in an L2 speaking environment appears to be helpful for acquiring FSs, not many language learners have a chance to immerse themselves in an L2 speaking community. Luckily, there seem to be other ways of receiving adequate exposure. For example, Lin (2014) compared the most frequent FSs in spoken language of British television and in the BNC and found that television can give an adequate exposure of FSs. So presumably, watching television could provide chances for learning FSs (especially considering that learners were shown to learn lexical items from watching videos: see, for example, Rodgers (2013) for individual words and Montero Perez, Peters, and Desmet (2015) for both words and FSs). Reading can be another source of incidental learning. There are now a number of classroom studies that looked at incidental acquisition of collocations from reading (e.g., Pellicer-Sánchez, 2015; Sonbul & Schmitt, 2013; Szudarski, 2012; Webb, Newton, & Chang, 2013; these studies are discussed in detail in Chapter 6). Most of these studies show that collocations can be acquired incidentally from reading. The learning rates are rather low but they are comparable to incidental learning rates for individual words (Pellicer-Sánchez, 2015). So it seems that successfully acquiring FSs from exposure is possible without an immersion in the L2 community as well.

2.3.3 Summary of the findings on FL in SLA

Overall, there seems to be a consensus about the importance of FSs in order to become a fluent user of an L2. However, acquiring native-like command of FL is tricky because of a large number of items to be learned, difficulty to notice them in language flow, and also the influence of the L1, which has its own phraseological profile. Because of these difficulties, teachers should probably take an active role in introducing the idea of FL to their learners. The best approach to addressing FL seems to be a combination of direct teaching and creating conditions for incidental acquisition. Direct teaching, though, implies choices of which items to prioritise for teaching (which will be further discussed in Section 3.4.4). Even when those most frequent items are taught directly, there are still a lot of FSs that will be left for incidental acquisition. Therefore, a better understanding the processes of incidental acquisition of FSs would have important implications for language pedagogy. This question will be discussed in Chapter 6.

Chapter 3

Distribution of formulaic sequences

As discussed in the previous chapter, FL is heterogeneous in nature: it is made up of various categories of FSs with their own characteristics and behaviour. In order to better understand the formulaicity, we should understand the characteristics of each of these categories on their own. As a first step in that direction, it seems worth trying to break down the overall percentage of FL in language into percentage parts for each separate category of FSs. This chapter presents a study that looked at the frequency distribution of four categories of FSs: collocations, idiomatic phrases, lexical bundles, and phrasal verbs. While this analysis does not provide a complete picture, it should give a reasonable approximation of how widely used these categories are and how frequent they are relatively to each other.

3.1 Background of the study

3.1.1 Corpus analysis

Before the introduction of computerized corpora in 1960's (Hunston, 2012), the majority of linguistic studies were based on a small number of examples, quite commonly invented by a researcher. The development of computers and the ability to collect, store, and analyse millions of word occurrences had a large influence on linguistics and especially on the research on FL. When summarizing corpus research, Barlow (2011) listed three main ideas that became widely accepted after establishing corpus linguistics:

1. importance of FL – corpus analysis has shown the pervasiveness of collocations;
2. importance of frequency – linguists' attention has shifted from what is possible in language towards what is typical and used frequently;
3. variability in language – while invented data are neat and usually fit a theory well, corpus data are usually more variable.

Hence, even if the importance of FL was raised before the corpus era (e.g., Firth, 1957), there is no doubt that it was established by corpus linguistics.

Researchers using corpora in their studies approach them in two different ways: some of them start from existing theories and use corpora to find illustrative examples or to check their theories (corpus-based approach), while others start directly from corpora and look at what the data suggest (corpus-driven approach) (Tognini-Bonelli, 2001). For the corpus-based research, corpus data are only a complementary part of the research that has no determining force. Hence, corpus-based approach usually implies a certain dichotomy between the actual use of language and some kind of an underlying knowledge of the structure of language. Corpus-driven research, on the other hand, starts from corpus data. It seeks to assemble as much relevant information as possible to draw comprehensive conclusions inductively based on the evidence. Biber (2009) emphasizes that neither of these approaches is better than the other, but rather they conceptualize different standpoints that might lead to radically different findings.

Actual corpus studies, though, often stand somewhere in between these two opposite approaches. As Gries noted, "[e]ven self-proclaimed corpus-driven studies are often not as corpus-driven as they could/claim to be" (2010a, p. 328). However, using both approaches is not necessarily a limitation. Sometimes no matter how appealing the corpus-driven approach might seem, it can provide an unmanageable amount of data, which becomes impossible to generalize (Biber, 1993). Therefore, a combination of the two approaches can actually be more feasible.

3.1.2 Automatic extraction of FSs: issues with corpus statistics

When analysing FSs, target sequences have to be extracted from corpora. While the identification and extraction of some sequences (such as lexical bundles) is rather easy, other sequences (e.g., collocations) have generated a lot of debate. Various methods of collocation extraction have been used, both in corpus linguistics and in computational linguistics, and they will be briefly described in this section.

The most basic method is looking at raw frequencies (Manning & Schütze, 1999; Wei & Li, 2013). This method assumes that a frequency of words occurring together indicates that they are used as a sequence. However, the extraction based solely on frequency leaves a researcher with a lot of combinations of the most frequent function words that are not very interesting from the linguistic perspective (e.g., *of the, and a)*. Hence, even if setting certain frequency cut-off points is a common practice, frequency information as the only criteria is rarely used.

A way of improving the lists of FSs (especially collocations) extracted based on the frequency information is filtering the results according to the part of speech of the words. Manning and Schütze (1999) reported a study by Justeson and Katz, where they filtered the results of frequency counts based on certain patterns of parts of speech (various combinations of nouns, adjectives and prepositions). This simple filtering considerably improved the output. However, this method entails pre-selecting patterns of parts of speech. The patterns used in the aforementioned study, for example, are rather limited as they did not include verbs at all. A more extensive list of patterns, though, could be a useful way to extract more comprehensive lists of FSs.

Another, more common, way of identifying collocations is using certain statistical measures. According to Gries (2010b), more than 20 different measures have been proposed for identifying collocations. However, most of them can be divided into two main groups: measures based on null hypothesis testing and measures that evaluate association between two words (Manning & Schütze, 1999). Both of them will be briefly discussed.

Null hypothesis testing. Various measures are based on the idea, that collocations are words occurring together more often than predicted by chance. They adopt a null hypothesis ‘two words are used together by chance’ and if the data allows discarding this null hypothesis, the words are extracted as collocates. The most widely used measures for null hypothesis testing are *t* scores, *z* scores, chi-square, and log-likelihood ratios. However, the idea underlying the null hypothesis testing becomes problematic in corpus research. Any type of null hypothesis testing assumes a random distribution of data and this is not the case with language (Kilgarriff, 2005; Sinclair, 2008). Also, *p* values are affected by the sample size (Baayen, 2008). Therefore, if we had a large enough corpus, we could prove any sequence of words to be a significant collocation (Barnbrook, 1996; Dunning, 1993).

Association measures. While the null hypothesis testing gives us some indication of the confidence level for rejecting the hypothesis that the words are used together by chance, association measures give us information about the strength of the connection between the words (Clear & Tognini-Bonelli, 1993). For measuring association, Church and Hanks (1990) proposed a measure of a point-wise mutual information (MI), which is widely used in corpus linguistics. MI score can be calculated using the following formula (Church & Hanks, 1990):

$$\mathrm{MI} = \log \frac{\mathrm{P(word\ 1, word\ 2)}}{\mathrm{P(word\ 1) * P(word\ 2)}} = \log \frac{\mathrm{N * Frequency(word\ 1\ word\ 2)}}{\mathrm{Frequency(word\ 1) * Frequency(word\ 2)}}$$

Association measures have also been criticized, though. First of all, they do not have clear cut-off points and it remains an arbitrary choice what to interpret as collocation and what to leave out. The psychological validity of these measures has also been questioned and still requires more research (see Durrant & Doherty, 2010). Furthermore, most of the association measures are applicable only to the sequences of two words, they do not take directionality into account, and they tend to be affected by the frequency of the constituent words (Biber, 2009). So even if association testing is not based on flawed assumptions, it has its own problems.

Various studies have tried comparing different association and null hypothesis testing measures. For example, Barnbrook (1996) compared *z* score, *t* score, and MI for extracting collocations. He concluded that *z* score and MI are rather similar, while *t* score gives slightly different results. Manning and Schütze (1999) tried to rank collocations with same raw frequency (N = 20) based on their MIs and *t* scores. They only looked at ten collocations, but those ten were ranked in an exactly same order. Hence, it seems that the different methods are not drastically different (even if 20 items is a very small sample to draw strong conclusions on). However, certain differences between different the measures are known. For example, MI score disfavours collocations with very high frequency words. Therefore, it tends to exclude function words and favours lexical collocations (Biber, 2009; Gries, 2010b). T score, on the other hand, tends to give more information about the grammatical behaviour of words, extracts collocations with grammatical words, and favours high frequency words (Gries, 2010b). Also, MI is less affected by the corpus size than *t* scores; therefore, MI can be compared across corpora (Hunston, 2002). To sum up, there are certainly differences between various measures, but as Barnbrook (1996) suggested, it is impossible to make any claims about which one is the best measure.

The issues with existing measures were addressed by various researchers. For example, recently Gries (2013) tried to solve some of the problems with the MI by introducing his new measure (ΔP), which takes the directionality of a collocation into account. Wei and Li (2013) suggested a new MI score that can extract n-grams which are longer than two words and less-arbitrary in terms of grammar and semantics than lexical bundles. Daudaravičius and Marcinkevičienė (2004) also suggested a method of *Gravity counts* to identify boundaries of sequences more precisely. However, all these new methods address only certain problems of the extraction of FSs and for now there is no universally accepted way to reliably extract collocations.

3.1.3 Distribution of formulaic language

There have already been various attempts to quantify FSs in language. However, as FSs were defined based on different criteria and extracted using different measures (manual analysis, various statistical methods), the results of these estimations vary considerably. Widely cited figures are those of Erman and Warren (2000), who manually analysed 19 texts (100–800 words long) and estimated an average number of FSs in language to be around 55%. However, their study is rather problematic, as the amount of the data the results are based on is very limited. Also, their criteria for counting a word sequence as a FS can be questioned. For example, they excluded transparent collocations (e.g., *dark night*), but counted contractions (e.g., *I'm*). Hence while the study revealed a large proportion of language to be formulaic, the suggested numbers should be interpreted only as approximations.

More recently, Wei and Li (2013) proposed a new statistical method to extract contiguous n-grams from corpus and found that 58.75% of their corpus was made of FSs, which is a very similar estimate as in Erman and Warren's (2000) study. However, Altenberg (1998), who defined FSs more broadly (as any contiguous sequences occurring in a corpus at least twice), found that these sequences covered up to 80% of his corpus. Not surprisingly the criteria and the method of a study largely affect the results. However, a general tendency seems to be clear: a considerable part of language, both spoken and written, is formulaic in nature.

One issue with these studies, though, is the fact that they only looked at FL overall. However, as already discussed, sequences that are very different in nature can be formulaic. Some of them might have a distinct meaning or be defined by their opacity; others might simply be very frequent. Therefore, in order to achieve a better understanding of formulaicity, it is not enough to conceptualize formulaicity as a phenomenon opposite to the open-choice principle (Sinclair, 2004b). Taking into the account that FL consists of different categories of FSs is also important.

There have been some studies that tried looking at separate categories of FL and estimating how many of certain sequences there are in discourse. For example, Biber et al. (1999) estimated that lexical bundles make up to 21% of academic written discourse and about 30% of conversations, while phrasal verbs (together with prepositional verbs) account for 1% of academic discourse and 2% of conversations. Liu (2011) suggested that phrasal verbs in spoken language and fiction account for about ~0.6% of words (the difference between Biber's and Liu's estimates might be due to slightly different definitions of phrasal verbs adopted). Simpson and Mendis (2003) tried to estimate the number of idioms in academic spoken English and reported a result of about 0.03%, which seems to be a very low percentage. However, they used rather narrow criteria to define idioms and their method of idiom identification could also be challenged (they manually searched a part of the study corpus for idioms, and then used this list of idioms for an automatic analysis of the whole corpus). Thus even if there have been studies looking at separate categories of

FSs, they are not fully comparable and cannot simply be combined together in order to obtain an overall picture of the distribution of various categories of FSs in language.

3.2 Study 1: Distribution of collocations, idiomatic phrases, lexical bundles, and phrasal verbs

The present study looked at the distribution of four categories of FSs (collocations, idiomatic phrases, lexical bundles, and phrasal verbs) in English. It focused on general tendencies of use of different categories of FL rather than on the use of individual FSs.

As register seems to be one of the factors influencing the use of FSs (Biber, 2009), the study also took the register information into account. In this thesis, register is understood as "any variety of language that is systematically related to its situation of use" (Biber & Finegan, 1988, p. 83). Biber (2006) noted that terms 'genre' and 'register' are sometimes used interchangeably and most scholars simply adopt one of them. However, he also noted that when researchers made an explicit distinction between genre and register, studies that adopted the term register usually focused on lexico-grammatical features (words, word types, and grammatical features), while studies focusing on genre looked at socio-cultural aspects of language use. As the present study analysed the use of FL, the term register was adopted. Four registers were analysed: academic prose, fiction, newspaper language, and spoken conversations. As previous studies that focused on lexico-grammatical variation in language used specifically these four registers (Biber et al., 2004, 1999; Conrad & Biber, 2004), this study will add on the existing knowledge of linguistic variation between these registers.

Different registers could potentially differ in two ways: different individual sequences could be used in different registers or there could be differences of the extent to which different categories are used in each register. Both these issues seem to be the case: "[t]here seem to be decided genre preferences for phrasal lexemes in general, as well as for individual expressions" (Moon, 1998b, p. 100). Therefore, register information seems to be an important factor to consider when analysing the use of FSs.

In the present study, the following research questions are addressed:

1. What is the distribution of collocations, idiomatic phrases, lexical bundles, and phrasal verbs in language?
2. Does this distribution depend on the register?

3.2.1 Methodology

The study adopted a distributional approach to formulaicity: sequences that are frequent become entrenched in speakers' mental lexicon and thus formulaic. When adopting the frequency-based approach, the corpus-driven method seems to be the first choice. However, for identifying some categories of FSs (e.g., idioms that are defined by their meaning opacity), the corpus-driven approach could not have been used. Therefore, the study used both corpus-based and corpus-driven approaches. The identification of lexical bundles and collocations (to a great extent) was corpus-driven, while the identification of idiomatic phrases and phrasal verbs was corpus-based. In the following sections, the procedure of extracting FSs will be outlined and the main issues with the extraction will be discussed.

3.2.1.1 Corpus used in the study

The BNC Baby corpus, which is a 4 million word subset of the BNC, was chosen for the study for several reasons. First of all, it is a general balanced corpus, covering academic prose, fiction, newspaper language, and spoken conversations. Therefore, it allowed comparing different registers. Secondly, it had recently become freely available. Certainly, a larger corpus (e.g., the entire BNC) would have allowed a more comprehensive analysis, but it was impossible to use a larger corpus for several reasons. Firstly (and most importantly), some of the parts of the study required manual checking of the data. Secondly, the *Antconc* (Anthony, 2011) software used to extract n-grams sometimes struggled to support even this amount of data. Therefore, a four-million word corpus seemed to be a good starting point considering the limitations.

The BNC is sampled to represent British English from 1980–1993. While a corpus collected more than two decades ago might not represent certain lexical aspects of a current language use, for this particular study it did not seem to be a major concern. The study did not look at the use of individual FSs, but rather at the overall use of different categories of FSs. Hence, even if individual frequent phrases might have changed, the tendencies of language use should arguably remain the same. Table 3.1 presents a short description of the BNC Baby.

Table 3.1 Overview of the corpus used for the study

Register	No of texts	No of words	Content
Academic prose	30	1,002,230	• different subject areas covered
Fiction	25	1,015,884	• imaginative texts for adult readers • no more than one text of an author
Newspaper language	97	960,4130	• 60% national papers, 40% local papers • wide range of topics
Spoken conversations	30	1,013,5407	• informal conversations • various age of speakers • balanced gender distribution

Note: based on Burnard (2008)

3.2.1.2 Extraction of various types of formulaic sequences

Collocations. A distributional approach to collocations was chosen for the study. Three criteria were adopted to extract a word sequence as a collocation:

1. Raw frequency ≥ 5 (in the whole corpus of 4 million running words). This cut-off point was chosen as the MI was reported to tend towards overestimation when the frequency counts are lower than 5 (Dunning, 1993; Kilgarriff, 2005);
2. MI ≥ 3;
3. Composed of lexical words.

The MI score has been criticised for excluding high frequency collocations with function words (Biber, 2009). For the purpose of the study, though, this was actually an advantage, because I was interested in lexical collocations only. As the MI score has no clear threshold, but rather signals the strength of association, different cut-off points can be and have been adopted. For example, Sonbul (2012) used a cut-off of one and Kennedy (2003) used a cut-off point of two. However, three seems to be the most frequently adopted cut-off point in corpus linguistics (Schmitt, 2010, Hunston, 2002).

Despite the fact that collocations are usually extracted from corpora using a span of ± 4 words from the node word, for this particular study, a rather limited span of co-occurrences was chosen. The two collocates had to either be adjacent, or occur within a two-word span. This was chosen to capture both adjacent collocations (such as *immediate effect*) and collocations with one word in between the collocates (such as *make a mistake*). Extracting collocations with a larger span was impossible due to technical limitations.

A raw frequency of five in a four million running word corpus is a low frequency, as it means a collocation had to occur in the corpus a little bit more than once in a million words. The lowest possible frequency cut-off was chosen on purpose, though, so that as many as possible collocations would be captured. That is also the reason why the collocations were initially identified in the entire corpus and this list was then used as a starting point to identify them in the sub-corpora. Inevitably, some collocations were missing from this list even if a low cut-off point was chosen, as there would always be items used only a few times in a particular corpus. However, this is an inevitable limitation of any corpus study when a certain frequency limit is set.

The *Antconc* (Anthony, 2011) software was used for collocation extraction. The procedure is explained in Table 3.2.

Table 3.2 Procedure of extracting collocations

1.	All bigrams and trigrams with a minimum frequency of 5 extracted from the BNC Baby. A list of 99,338 bigrams and 67,888 trigrams obtained
2.	Lists of nouns, verbs, adjectives, adverbs, and proper names extracted from the corpus, based on the part of speech tags
3.	A list of possible patterns of lexical collocation compiled e.g., *noun-noun; verb-noun; adjective-noun, verb * noun, etc.*
4.	All bigrams and trigrams filtered according to these part of speech patterns; only the ones that matched the patterns retained
5.	Frequency lists of all the words obtained
6.	MI scores calculated for all the potential collocations. Sequences with MI scores of ≥ 3 retained
7.	Resulting list of collocations checked against the list of proper names. Collocations with proper names (e.g., *Lewis said*) listed separately and checked manually, if they needed to be excluded
8.	Final list of collocations obtained for the entire corpus: 9,877 adjacent and 3,262 non-adjacent collocations
9.	All bigrams and trigrams with no minimum frequency extracted from each sub-corpus
10.	Extracted sequences checked against the list of collocations
11.	Final lists of collocations and their frequencies obtained for each sub-corpus

When the list of collocations from the entire corpus was obtained (Step 8 in Table 3.2), there were still some decisions to be made. Collocations that had an inserted word could have been classified into various categories. Some binomials were listed, such as *animals and plants, alive and well, black and white*. These binomials were counted as collocations for the purpose of the present study, as they met the criteria for collocations and they can be argued to be a certain type of collocation. Also, some of the collocations had a function word in the middle, like *get a drink, made a face, have a word*. These types of collocations were counted as three-word sequences in the final estimation of words in FSs, in order to avoid underestimating the amount of formulaicity in the corpus.

There were two more problematic cases. Firstly, there were 88 collocations which had an *s* as the inserted word. This *s* was usually a marker of possession, such as *people's lives*. Technically, it should have not been considered a separate word. However, the software treated contractions (hence these *s*-es as well) as separate words and they were included into the word counts as individual words. Therefore, I decided to count these sequences as three word sequences, to make sure that all the contractions in the corpus are treated consistently.

The final issue was the three word sequences consisting of three lexical words such as *indexed sequential file* or *colliding plane wave*. For these trigrams, the MI was calculated as an association between the first and the third words, and the word in the middle was interpreted as an insertion. However, in these cases the association was not between the first and the third word, but rather between the modifier (first word) and the collocation (second and third words). So these sequences were captured both in bigrams list (e.g., *sequential file*) and trigrams list (e.g., *indexed sequential file*), with MI scores ≥ 3 in both cases. These units in their nature are very similar to sequences that Biber (2009) defined as multi-word collocations. As these sequences seemed to be extended collocations, they were counted as three word units for the purpose of this study. The frequencies of the two word sequences were adjusted accordingly, in order not to count same sequences twice.

Issues with collocation extraction. As part of speech tagging is automatic, some errors inevitably remain. Some errors were noticed when compiling the lists of different part of speech words for the study as well. As these lists were too long to fully check them manually, only some adjustments were made. To start with, words consisting of only one letter or two letters (potentially spelling mistakes or letters used to enumerate lists) were checked manually and deleted if they were non-words. Also, words containing punctuation marks as part of a word (e.g., *is{, is.a*) were deleted. Then, articles, which were obviously erroneous in the lists, were deleted (i.e., *the* and *a* in the lists of nouns, verbs, etc.). These adjustments cleaned the results a little bit, but some issues with part of speech lists probably still remained. However, part of speech pattern was only one of the criteria for the collocation extraction. As the sequence had to reach a certain frequency and MI score to be listed as a collation, the final list should contain no random sequences.

One other limitation of the collocation extraction was that it did not take into account collocations with more than one word in between the collocates. To investigate how problematic this might have been, an additional small scale post-hoc corpus analysis was carried out. It is reported in Section 3.3.3.

Idiomatic phrases. As idioms were defined as two or more word sequences that have a figurative meaning, identifying idioms automatically was impossible. Various researchers claimed that idioms are rare in discourse (e.g., Liu, 2003; Moon, 1998a), but at the same time idiom dictionaries list thousands of phrases. So it seems that even if individual idioms might not be frequent, together as a category they might account for an important part of FL.

Ideally, a comprehensive dictionary of idioms could have been selected and all the items from that dictionary searched for in the corpus, making sure that the potential form variations are taken into account. This is obviously a very time-consuming task. However, this task had already been undertaken by Liu (2003), who proposed a list of the most frequent idioms. Liu manually searched for 9,683 idioms compiled from seven idiom dictionaries. He adopted a frequency criterion of two occurrences per million words and this criterion reduced a list of idioms to only 302 items, suggesting that individual idioms are not frequent. Hence this list seemed to be an obvious starting point for idiom extraction.

However, Liu's list (2003) had an important limitation for this particular study: it was compiled based on spoken American English. Liu himself claimed that the idioms in the list might be different, if written or British English corpora were used. To compensate for this limitation an additional list of idiomatic phrases was added (Martinez & Schmitt's PHRASE list (2012)). As the PHRASE list was created based on the BNC, it seemed to be the best option to capture frequent idiomatic language in the BNC Baby.

These two lists are very different in terms of the methodology used to compile them. Table 3.3 briefly describes the methodology behind each of them. Despite being quite different, both of those lists included idioms based on a certain frequency criterion. As the present study looks at FL as a phenomenon largely driven by frequency, these two lists together seemed to be a reasonable starting point for idiom extraction.

Table 3.3 Comparison of lists of idiomatic phrases used in the study

	PHRASE list	**Most Frequently used Spoken American idioms**
Corpus	BNC	Corpus of Spoken, Professional American English Michigan Corpus of Academic Spoken English Corpus of spoken American media English
Mode	written and spoken	spoken
Approach	Corpus-driven: n-grams extracted from the corpus and then manually checked for transparency	Corpus-based: 9,683 candidate idioms chosen from idiom dictionaries and searched for in the corpora
Frequency cut-off	~ 7.87 per million	2 per million
Main criteria	• frequency • morpheme equivalence • non-transparency	• frequency • occurrence in at least two corpora
Number of items	505	302

After comparing the lists, it turned out that they overlapped considerably. This overlap was reassuring, showing that even if the lists were compiled using different corpora and different methodologies, they both yielded similar results. The stages of preparing the lists for the study are summarized in Table 3.4.

Table 3.4 Procedure of extraction of idiomatic phrases

1.	Both lists combined and duplicate items deleted
2.	Phrasal verbs deleted from the list (and added to the list of phrasal verbs)
3.	Nine items deleted from the list as they were judged to be too literal: *add to, amount to, care to, heard to, look like, look to, this stage, those who, to blame*
4.	Items that seemed to be the most problematic (very often used literally) identified for manual analysis: this yielded a list of 29 items to be searched for manually (e.g., *a good, a little, they say;* the full list presented in the Appendix 2). These items were searched for individually. If every sub-corpus yielded less than 200 concordance lines, all of them were checked manually. Else, two random samples of 100 lines were checked to estimate the percentage of the cases where the sequence was used figuratively.
5.	Items that could be different in length were identified (e.g., *(up) to date, in addition (to), as well (as))* Their frequencies were manually adjusted not to overestimate the counts
6.	All possible morphological forms of the items were listed (e.g., *have a clue, has a clue, having a clue)*
7.	Wildcards were added where there were variable slots (e.g., *loose * head).* The results for the items that included wildcards were checked manually to delete any noise

After preparing the final version of the list, the *WordSmith Tools* (Scott, 2012) software was used to search for the items in each sub-corpus. This software was chosen instead of the *Antconc* (Anthony, 2011), as it allowed both searching for a list of items simultaneously, and easily checking the occurrences to exclude. This procedure led to an estimation of the counts of idiomatic phrases in each sub-corpus separately which were then added up to estimate the percentage of idioms in the entire corpus.

As the use of pre-compiled lists implies that the study will not cover all the instances of idioms, a term ‘idiom’ will not be used to refer to this category (especially because Martinez and Schmitt (2012) did not define their sequences as idioms but rather ‘phrasemes’). The term ‘idiomatic phrases’ was adopted instead.

Issues with identification of idiomatic phrases. As the counts of idiomatic phrases were obtained based on the lists of the most frequent items only, some of the instances of use of less frequent idiomatic phrases might be not captured. However, this method of using previously compiled list seemed to be more reliable than, for instance, systematically sampling idioms from an idiom dictionary. Considering practical limitations, the method used in the study seemed to be the best attempt to account (at least partially) for the idiomatic phrases in the corpus.

Lexical bundles. As lexical bundles were analysed mostly by Biber and his colleagues (e.g., Biber, Conrad, & Cortes, 2004; Biber et al., 1999; Conrad & Biber, 2004), his methodology for extracting lexical bundles was adopted. Two main criteria were used to identify lexical bundles: frequency and dispersion of the sequences in corpus. The cut-off points were chosen following Conrad and Biber's (2004) study. They are summarized in Table 3.5.

Table 3.5 Cut-off points for lexical bundles

n-gram	Frequency	Dispersion
3-gram 4-gram	10 per million	5 texts
5-gram 6-gram 7–9-gram	5 per million	5 texts

The *Antconc* (Anthony, 2011) software was used for extracting lexical bundles, because it allows identifying n-grams of any length without any prior input, except of the required frequency and the number of texts. Separate lists of n-grams of each length (tri-grams, four-grams, etc.) were automatically extracted from the corpus. It soon became clear that this category of FSs is extremely numerous, with only the academic sub-corpus yielding 3,125 different three-grams, 410 four-grams, 117 five-grams and 7 six-grams.

Previous studies that analysed lexical bundles more qualitatively, were criticized for looking only at the n-grams of a particular length (usually four-grams), even if this length choice was completely arbitrary (Gries, 2010a). The present study did not assume any specific length of lexical bundles and looked at the sequences of various lengths. However, this choice raised two main problems: setting the best maximum length and avoiding counting the same sequences more than once. These two issues will be briefly discussed.

Initially, my intention was to stop at six-grams, as it seemed that longer n-grams were very unlikely to occur. However, when extracting the sequences, it became clear that even longer n-grams can meet the chosen criteria. Hence no maximum length was chosen, making the extraction of lexical bundles purely corpus-driven. For the newspaper discourse the longest sequences were six-grams, for the academic prose and fiction the longest sequences extracted were seven-grams, while for the spoken discourse, even one nine-gram met the criteria (even if it was only a list of numerals for counting, i.e., *one two three four five six seven eight nine*). As the number of texts in each sub-corpus was rather limited (see Table 3.1), the occurrence in five texts might seem a relatively conservative criterion, but it assured that only well dispersed sequences were included. As the lists of n-grams extracted were very extensive, it seems that this criterion was not problematic.

In order not to count the same sequences more than once, certain adjustments had to be made. When extracting n-grams from corpora, it is inevitable that some of the sequences are identified as shorter bundles, even if they are used as parts of longer phrases (Gries, 2010b). To fix this overlap, the lists of n-grams were saved in *Microsoft Excel* spreadsheets and then the lists of the shorter n-grams were compared to the lists of the longer ones. If a longer n-gram included a shorter n-gram, the frequency of the longer one was subtracted from the frequency of the shorter one. The procedure is illustrated in Figure 3.1.

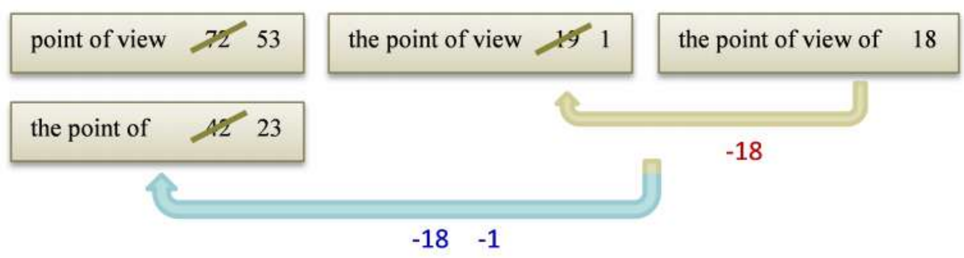


Figure 3.1 Adjustments of frequencies of lexical bundles

Issues with extraction of lexical bundles. Even if identifying lexical bundles seemed to be a straight forward task, a few issues occurred. To start with, the cut-off points (even if they had been used in various studies before) were arbitrary. This led to some problems with frequency adjustment. While the procedure illustrated in Figure 3.1 worked perfectly fine with the academic prose and fiction sub-corpora, when analysing the newspaper language and the spoken conversations this adjustment sometimes led to negative counts. After a closer analysis, it appeared that if longer sequences did not meet the criteria to be included into the lists of lexical bundles, then the adjustment of shorter lexical bundles resulted in negative frequencies. This issue is illustrated in Figure 3.2.

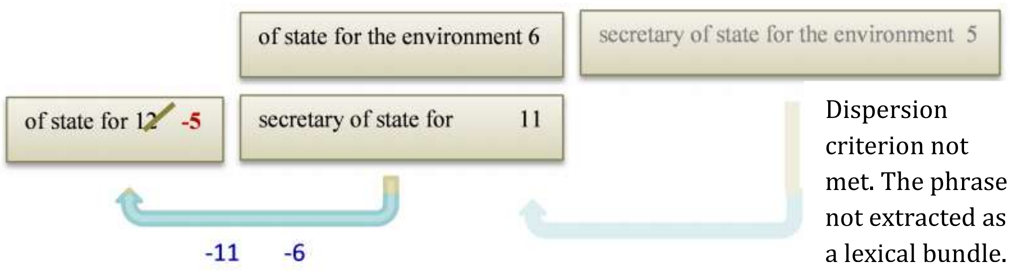


Figure 3.2 Issues with frequency adjustment

However, this was the case only for two sequences in the newspaper sub-corpus. The data of the spoken sub-corpus were a bit messier, but even in the spoken sub-corpus less than one per cent of the sequences ended up with negative counts. Negative frequencies were changed to zero for counting purposes.

Another issue that arose when extracting lexical bundles was a mismatch between the counts when extracting lexical bundles from the entire corpus, as compared to extracting them from each sub-corpus separately and then adding the counts up. This difference was rather considerable: the percentage of lexical bundles in the entire corpus changed from 17% to 31%. After a closer analysis, though, it became obvious that a considerable number of lexical bundles were register-specific. In order to count an n-gram as a lexical bundle in the entire corpus, it had to occur at least 40 times. If an n-gram occurred in one of the registers 20 times, but was infrequent in other registers, it was extracted as a lexical bundle for that sub-corpus but not for the overall corpus. Hence, when extracting lexical bundles from each register separately, register-specific bundles were captured and it increased the number of sequences extracted. Therefore, the number of the lexical bundles in the entire BNC Baby was estimated as a sum of the counts for each sub-corpus.

Phrasal verbs. A ready-made list of the most frequent phrasal verbs was used for the study. The reason for this choice was the fact this study took frequency as one of the criteria for a sequence to be formulaic. Therefore, the use of the list of the most frequent phrasal verbs allowed applying the frequency criterion for phrasal verbs as well for other categories of FSs.

Liu's (2011) list of phrasal verbs was chosen. It was preferred over the Gardner and Davies's list (2007) as it was compiled based on both the BNC and the Corpus of Contemporary American English (COCA) and therefore it is more comprehensive. The 150 phrasal verbs in the list accounted for 62.95% of all the instances of phrasal verb occurrences in the corpora (Liu, 2011). This means that the remaining 12,358 verb + adverbial particle lemmas (which Liu analysed but did not include in the final list, because they did not meet the frequency criterion) accounted for less than 38% of the corpora. Thus it seems safe to assume that the remaining phrasal verbs individually are infrequent.

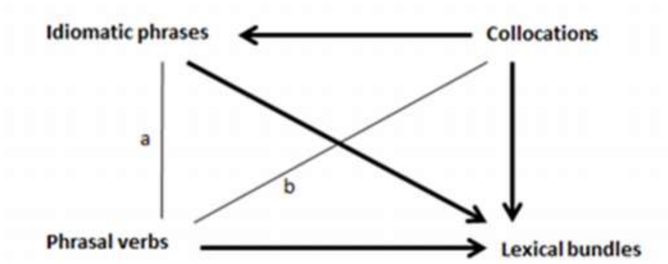
After the analysis of the lists of idiomatic phrases (see the section above), the phrasal verbs that occurred in these lists were added to a list of phrasal verbs, in order to keep the categories of phrasal verbs and idioms as separate as possible. This led to the final list of 211 phrasal verbs (see Appendix 1 for the full list). Then all the inflectional forms of these phrasal verbs were listed. The Cambridge Dictionary (2014) was consulted to check whether a phrasal verb allowed an insertion between the verb and the particle. The phrasal verbs that could be used with insertions were listed separately with wildcards (e.g., *check * out* to capture instances such as *check it out* or *check this out*). These forms with wildcards were searched for separately and the results were checked manually, as the insertion might have led to unwanted noise (e.g., *take part in* counted as an instance of *take in*). The *WordSmith tools* software (Scott, 2012) was used to search for the list of phrasal verbs in each sub-corpus separately and then the percentages were added together for the whole corpus.

Issues with the category of phrasal verbs. It could be argued that phrasal verbs are essentially idioms, as they also have a figurative meaning. The fact that the lists of idiomatic phrases contained some phrasal verbs, also supports this idea. In spite of this, I decided to separate phrasal verbs from idioms for the purpose of this study because phrasal verbs have a very clear structure, which makes them easily distinguishable from other idiomatic phrases.

3.2.2 Adjustments

When the estimates of the number of words of each category were obtained, it was necessary to make sure that the sequences in these categories, and hence the counts, would not overlap. That is, if a sequence *word1 word2* was counted as an idiom, it would not be counted as a collocation, or as a part of a lexical bundle *word1 word2 word3 word4*. Therefore, the counts of each category needed to be adjusted. Two main criteria were adopted for this adjustment:

- if exactly the same sequence fell into two categories, it was counted as a member of the category that was more strictly defined (e.g., a sequence *take advantage* was listed both as an idiomatic phrase and as a collocation, but it was counted as an idiomatic phrase);
- if a shorter sequence was a part of a longer sequence (lexical bundle), it was counted as the longer sequence and deleted from the shorter ones. Figure 3.3 illustrates the process of adjustment.



a. Lists of idiomatic phrases and phrasal verbs compared beforehand. No need to adjust
b. Collocations filtered based on the part of speech. No need to adjust

Figure 3.3 Adjustments in counts

The adjustments were made using *Microsoft Excel* spreadsheets. They led to a decrease of the overall percentage of FL by about 5.2%.

The method used to extract FSs is obviously not without its limitations. However, even if the sequences were identified manually certain inconsistencies, problematic cases, or borderline items would remain. As Erman and Warren claimed "the identification of 'all and only' prefabs [= formulaic sequences] in a text is in practice impossible" (2000, p. 33). Hence the counts of FSs in this study are also only approximations. However, considering the criteria adopted, these approximations should be reliable and the percentages presented in the Results section should give quite a good picture of the distribution of FSs.

3.3 Results

3.3.1 Distribution of formulaic sequences in the entire corpus

The results of the study showed that the four categories of FSs covered about 40% of the BNC Baby corpus. The summary of the results is presented in Table 3.6. The table shows that lexical bundles were by far the most frequently used category, followed by collocations, idiomatic phrases and phrasal verbs.

Table 3.6 Formulaic language in the whole corpus

Part of the corpus	Lexical bundles	Collocations	Idiomatic phrases	Phrasal verbs	Formulaic sequences
Academic prose	193,092 (19.27%)	97,406 (9.72%)	29,288 (2.92%)	5,321 (0.53%)	325,107 (32.44%)
Fiction	261,601 (25.75%)	61,375 (6.04%)	26,314 (2.59%)	12,405 (1.22%)	361,695 (35.60%)
Newspaper language	132,157 (13.76%)	72,261 (7.52%)	21,999 (2.29%)	8,513 (0.89%)	234,930 (24.46%)
Spoken conversations	590,770 (58.29%)	69,102 (6.82%)	24,216 (2.39%)	13,319 (1.31%)	697,407 (68.81%)
BNC Baby	1,177,620 (29.49%)	300,144 (7.52%)	101,817 (2.55%)	39,558 (0.99%)	1,619,139 (40.55%)

Note: the table presents raw counts of the words in each category of FSs. Percentages are presented in parenthesis.

These results can be more easily interpreted graphically. Figure 3.4 presents the distribution of FSs in the entire BNC Baby corpus.

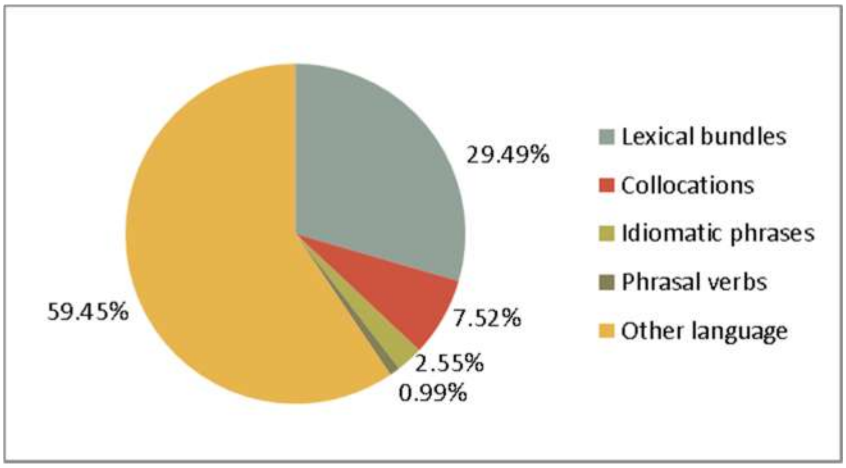


Figure 3.4 Distribution of formulaic sequences in the BNC Baby

As we can see from Figure 3.4, there was a clear difference between the frequencies of the different categories of FSs in the corpus. Lexical bundles were by far the most frequently used category covering almost one third of the corpus. Collocations were the second most frequent category, but they occurred considerably less frequently than lexical bundles and covered about 7.52% of the corpus. Idiomatic phrases covered about 2.55% of the corpus, showing that even if individually idioms might not be particularly frequent, the overall percentage of the phrases that have non-transparent meanings is not negligible.

Finally, phrasal verbs were the least frequent category, covering about 1% of the corpus.

It has to be taken into account, though, that the 59.45% of the corpus, labelled as ‘Other language’ in Figure 3.4, do not necessarily consist of novel language only. Other types of FSs (e.g., proverbs, variable expressions, collocational frameworks) could have occurred in the corpus. There might also be some infrequent idioms or phrasal verbs, which were not included in the lists used for the study. All in all, the results suggest that *at least* 40% of the language of the BNC Baby seems to be formulaic.

3.3.2 Distributions in the registers

To answer the second research question whether the distributions of different categories of FSs depend on the register; the four registers were analysed separately. Figure 3.5 illustrates the results for each sub-corpus.

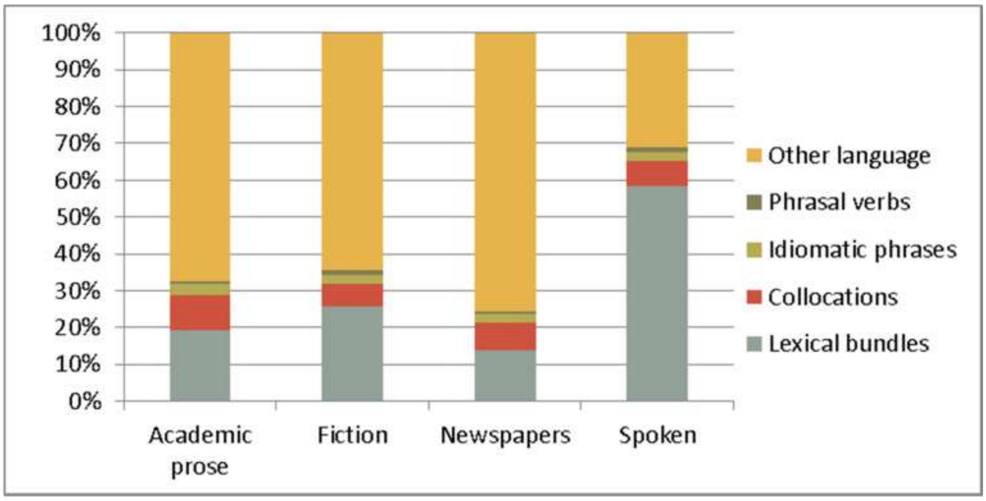


Figure 3.5 Distribution of formulaic sequences in different registers

As becomes clear from Figure 3.5, spoken conversations were the most formulaic: more than 68% of spoken language was covered by FSs, leaving only about 32% of the discourse not covered by the four categories of FSs addressed in this study.

Fiction was the second most formulaic part of the corpus, even if it was considerably less formulaic than spoken discourse. About 35.6% of the fiction sub-corpus was covered by the FSs addressed in this study. The largest difference between the spoken conversations and fiction lied in the use of lexical bundles, which seemed to be driving the difference of the use of FL in these two registers.

Academic prose was only slightly less formulaic than fiction, with FSs covering 32.44% of the sub-corpus. Newspaper language showed the least usage of FSs: only about 24.46% of the texts were covered by the four categories of FSs analysed. Lexical bundles, even if remaining the most frequent category, were used much less in newspaper language than in other sub-corpora.

As shown in Figure 3.5 and Table 3.6 the tendencies of use of FSs remained the same in all the registers, with lexical bundles being the most frequent category, followed by colocations, idiomatic phrases and phrasal verbs (in the same sequence for both the sub-corpora and the entire corpus). However, the estimates of the sequences in each category seem to differ considerably.

In order to reject the null hypothesis that there was no significant difference of the distribution of FSs across the registers, the chi-square analysis was carried out. The chi-square analysis was chosen as it does not assume the normal distribution of the data (Manning & Schütze, 1999). It was carried out using *R studio* software (R Core Team, 2013). The statistical analysis was performed following Gries (2014), as he provided clear step-by-step guidelines for comparing frequencies in different corpora. Table 3.7 provides the observed and expected counts of each category of FL.

Table 3.7 Register and type cross-tabulation

		Academic	Fiction	Newspaper	Spoken
Lexical bundles	Observed count	193,092.00	261,601.00	132157.00	590,770.00
	Expected count	236,454.40	263,065.30	170,867.50	507,232.80
	Residual	-88.17	-2.85	-93.64	117.29
Collocations	Observed count	97,406.00	61,375.00	72,261.00	69,102.00
	Expected count	60,265.93	67,048.34	43,549.58	129,280.15
	Residual	151.28	-21.91	137.58	-167.37
Idiomatic phrases	Observed count	29,288.00	26,314.00	21,999.00	24,216.00
	Expected count	20,443.84	22,744.62	14,773.20	43,855.34
	Residual	61.85	23.67	59.45	-93.78
Phrasal verbs	Observed count	5,321.00	12,405.00	8,513.00	13,319.00
	Expected count	7,942.85	8,836.75	5,739.69	17,038.70
	Residual	-29.41	37.96	36.61	-28.49

A 4 x 4 chi-square analysis revealed that there was a significant relationship between the register and the distribution of different categories of FSs ($\chi^2 = 121{,}971.10$, $df = 9$, $p < .00$, $V = .16$). The Crammer's V showed a weak effect[3] of register on the distribution of different categories of FSs.

Then the data were analysed closer, comparing the residual values to identify which cells deviate from the expected values the most (see Table 3.7). It appeared that the residuals for the spoken part of the corpus were higher than for other registers for all the categories except from phrasal verbs. Therefore, it was decided to check whether the analysis could be simplified by adding all the written registers together and calculating the chi-square values only for spoken versus written parts of the BNC Baby.

A 2 x 4 chi-square analysis was carried out. The chi-square remained significant as in the previous analysis ($\chi^2 = 90{,}250.14$, $df = 3$, $p < .000$), but the effect size increased showing that there was a moderate effect of the mode of language production on the distribution of FSs (Cramer's $V = .24$).

Statistical comparisons showed that the distributions of FSs were significantly different from register to register. The question remained, though, whether the counts of each category were significantly different between the registers or whether there was one specific category (e.g., lexical bundles) that

[3] The magnitude of the effect is interpreted based on the Cramer's V value intervals suggested by Rea and Parker (2005).

drove this difference between registers. Also, it seemed worth analysing whether the counts of the categories were simply proportional to the amount of FL in a specific register (e.g., as spoken conversations were the most formulaic, the counts of all categories of FSs could be expected to be significantly higher in the spoken sub-corpus than in the other parts of the corpus) or whether this relationship between the register and formulaicity was more complicated.

To answer those questions, the counts of each type of FSs were compared across registers using the log-likelihood ratios. All the differences appeared to be significant (see Table 3.8). This result, though, is not very interesting per se, as corpus counts in general tend to be very high, inevitably leading to significant values.

Therefore, in order to compare the frequencies in two corpora in a more meaningful way, it was suggested to calculate the percentage difference (Gabrielatos & Marchi, 2012). Gabrielatos and Marchi proposed this measure (which, as the name suggests, shows what the percentage difference between the counts in two corpora is) to identify and rank key words. However, as this measure allows easily comparing two corpora, it was adopted for the present study. The formula used to evaluate the percentage difference (% DIFF) was the following:

$$\%\ \text{DIFF} = \frac{\text{normalized frequency in Corpus 1} - \text{normalized frequency in Corpus 2}}{\text{normalized frequency in Corpus 2}} * 100$$

(Gabrielatos & Marchi, 2012)

Table 3.8 presents the percentage differences of various categories of FSs in the sub-corpora. The categories are ranked based on their absolute percentage difference. If the percentage difference is negative, it shows that there were fewer sequences in the first corpus than in the second one. The corpus listed first was taken as a Corpus 1; the one listed second served as a Corpus 2. This distinction was purely arbitrary and therefore the ranking is not definitive: if the corpora were changed around, it would lead to somewhat different ranking. However, this tentative ranking seems to be a useful illustration of general trends of which sequences show the most variability in different registers and which follow similar trends in all the registers.

Table 3.8 Percentage differences between the sub-corpora

Corpora compared	Sequences	log-likelihood	*p*	%DIFF
Fiction vs Newspaper	lexical bundles	36,396.98	.000	87.14
Newspaper vs Spoken	lexical bundles	290,333.76	.000	-76.39
Academic vs Spoken	lexical bundles	207,006.62	.000	-66.95
Academic vs Fiction	collocations	8,742.73	.000	60.87
Academic vs Spoken	phrasal verbs	3,456.46	.000	-59.60
Academic vs Fiction	phrasal verbs	2,816.63	.000	-56.52
Fiction vs Spoken	lexical bundles	131,242.62	.000	-55.82
Academic vs Spoken	collocations	5,157.78	.000	42.55
Academic vs Newspaper	phrasal verbs	885.51	.000	-40.10
Academic vs Newspaper	lexical bundles	9,034.51	.000	40.01
Fiction vs Newspaper	phrasal verbs	526.33	.000	37.76
Newspaper vs Spoken	phrasal verbs	823.72	.000	-32.55
Academic vs Newspaper	collocations	2,745.68	.000	29.17
Academic vs Newspaper	idioms	752.08	.000	27.58
Academic vs Fiction	lexical bundles	9,455.50	.000	-25.18
Academic vs Spoken	idioms	540.17	.000	22.31
Fiction vs Newspaper	collocations	1,604.34	.000	-19.70
Fiction vs Newspaper	idioms	181.69	.000	13.08
Academic vs Fiction	idioms	201.94	.000	12.82
Fiction vs Spoken	collocations	475.84	.000	-11.39
Newspaper vs Spoken	collocations	343.17	.000	10.36
Fiction vs Spoken	idioms	82.37	.000	8.41
Fiction vs Spoken	phrasal verbs	34.62	.000	-7.08
Newspaper vs Spoken	idioms	20.51	.000	-4.13

From Table 3.8 it becomes clear that the largest difference existed between the frequencies of lexical bundles in various register. This tendency was already obvious by looking at the Figure 3.5: spoken conversations clearly stood out in terms of the use of lexical bundles. On the other hand, the other trend, which was not as clear from the Figure above, was the difference in the use of phrasal verbs. The use of collocations and idiomatic phrases varied a bit less from register to register.

In order to interpret the percentage differences easier, Table 3.9 presents a summary of the tendencies of use of each category of FSs in the different registers. The registers are ranked based on the estimates of the proportion of the relative sequences in that register.

Table 3.9 Ranking of registers for each category

Collocations
Academic > Newspaper > Spoken > Fiction
Phrasal verbs
Spoken > Fiction > Newspaper > Academic
Idiomatic phrases
Academic > Fiction > Spoken > Newspaper
Lexical bundles
Spoken > Fiction > Academic > Newspaper

The rankings of the register in Table 3.9 did not follow the same pattern as the one for the frequency of FSs in general (i.e., spoken conversations > academic prose > fiction > newspaper language). Instead of being simply the most frequent in spoken conversations and the least frequent in newspaper language, each type of FSs seemed to have specific tendencies of use in different registers. For example, phrasal verbs were the most frequent in spoken conversations, but the least frequent in academic prose, while idiomatic phrases were more frequent in academic prose and fiction, than in spoken conversations and in newspaper language.

3.3.3 Analysis of the variation in collocations

As already briefly mentioned in Section 3.2.2, after all the adjustments applied for extracting various categories of FSs, one important limitation remained. The study only focused on adjacent collocations and collocations with one intervening word. In order to start investigating how much variation there is in collocational use, the following post-hoc analysis was carried out. This was a very small scale exploratory analysis, based on a few items only, but it proved to be informative to at least start looking at the issue of variation in collocational use.

One hundred collocations were sampled from the list of the collocations extracted for the study. As the aim was to cover different types of collocations and to do so in a more or less balanced way (including verb-noun, noun-noun, adjective-noun, and adverb-adjective collocations), the selection of the items was not random. Rather it was attempted to include collocations with various frequencies (adjacent frequency of the selected collocations in the full BNC ranged from 6 to 1,969 occurrences).

Those 100 items (about 0.8% of all the collocations extracted in Study 1) were searched for in the entire BNC (using Davies (2004) interface). The full corpus was chosen for the post-hoc analysis in order to look at collocational variation at a bit larger scale.

The frequencies of adjacent forms, non-adjacent forms and forms with positional variations were listed. The results are summarized in Table 3.10. The first column indicates how much (in percentages) the number of occurrences increased when allowing for non-adjacency (up to 3 words in between the collocates: *word1 (***) word2*) as compared to the frequency of adjacent forms (*word1 word2*) only. The second column indicates how much this difference further increased, allowing for positional variation as well (*word1 (***) word2* and *word2 (***) word1*).

Table 3.10 Effects of the positional variation and adjacency

	Increase of occurrences accounting for non-adjacency (%)	Increase of occurrences accounting for positional variation (%)
Adjective-noun	7.93	5.77
Adverb-adjective/participle	3.22	3.51
Noun-noun	8.21	10.80
Verb-noun	9,119.84	22.07

While the findings presented in Table 3.10 are by no means conclusive, they seem to suggest some clearly emerging trends. First of all, both adjacency and positional variation seem to be rather common. Also, it seems that different types of collocations behave differently. For adjective-noun, noun-noun or adverb-adjective collocations, accounting for positional variation or non-adjacency changed the frequencies only a little bit. The non-adjacency was more or less as common as the positional variation for all these collocations. For the verb-noun collocations, though, the change in frequencies when accounting for non-adjacency was massive (it increased by more than 9,000%). Thus, it seems that verb-noun collocations are the most variable in their forms. One obvious reason for this can be the following. Nouns in the verb-noun collocation can be preceded by a modifier or a determiner and this modifier or determiner is used in between the verb and the noun. In all the other types of collocations, the first word of the collocation is a modifier itself so it precedes the word it modifies.

Sinclair and his colleagues have already tried looking at the variation of collocation forms and introduced the terms of 'position-dependent' and 'position-free' collocates (Krishnamurthy et al., 2004), depending on if a collocation has a preferred configuration. Their analysis suggested that there are more position dependent collocations than position-free collocations and that grammatical collocations seem to be more position-dependent than the lexical ones. While they only looked at a relatively small sample of collocations (N = 384), their results are in line with the present small scale analysis: a considerable number of collocations were position dependent. However, the present analysis extends this finding by suggesting that verb-noun collocations are much more position free than other types of collocations, at least in terms of adjacency.

The variation in form seems not to be limited to collocations only. Moon (1998b), for example, looked at fixed expressions and idioms and showed that about 40% of the phrases he studied did not have a fixed canonical form. Philip (2008) also looked at idioms, proverbs and sayings in corpus data and concluded that the canonical (fixed) form is outnumbered by variations.

In the present analysis, no insertions were allowed when searching for adjacent collocations across all the types of collocations to make the analysis consistent. However, in Study 1 data, when extracting collocations, one word between the collocates was allowed, so a part of the variation in verb-noun collocations was actually captured and the numbers presented in the Results section (3.3.1) are not as underestimated as Table 3.10 would suggest. If instead of comparing adjacent verb-noun collocations to the non-adjacent ones, we compared collocations allowing for one intervening word

(*word1 (*) word2*) to the one allowing three intervening words (*word1 (***) word2*), the increase would be much more moderate (35.75%). It is still larger than for the other types of collocations, but it is not thousands of times larger.

Overall, this small-scale investigation of the use of collocations tentatively suggests that form variation is a very important issue when extracting collocations from corpora and should be accounted for in any attempts to automatically extract collocations.

3.4 Discussion

3.4.1 Comparison of estimates of formulaic sequences

The study shows that there are differences in distribution of different categories of FSs both in English language and in different registers. There have already been claims that various categories of FSs vary in terms of frequency and this study quantified these claims providing estimates of the percentages for each category using exclusive definitions. It seems that overall lexical bundles are about four times more frequent than collocations, ten times more frequent than idioms and as much as 30 times more frequent than phrasal verbs.

The overall percentage of FSs in this study seems to once more suggest that FL is used very frequently. Firstly, it is interesting to discuss the overall percentage of FSs in English in light of the previous research. The overall percentage of FL that the current study suggests is 40.45%. However, it cannot be directly compared to the previous studies due to different methodologies and definitions used. For example, Erman and Warren's (2000) widely cited study reports ~55% of language being formulaic, based on a manual analysis of relatively small number of texts, while Altenberg's (1998) reports a large percentage of ~80%, based on the criteria that were much more inclusive than the ones used for the present study. However, a direct comparison was not the purpose of this study. Nonetheless, the percentage of FL in the BNC Baby (40.45%) is in line with the suggestion that FL makes up from one third to a half of our language (Conklin & Schmitt, 2012).

When comparing estimates of separate categories with the previous studies, the percentage of phrasal verbs in this study (0.99%) fell between those suggested by Biber et al. (1999): 1–2% in academic discourse and conversations; and Liu (2011): ~0.6% in spoken language and fiction. The trend was the same as reported in these previous studies: phrasal verbs were more frequent in spoken conversations and in fiction and less frequent in academic prose and newspaper language.

As for the lexical bundles, their extraction was based on the same criteria as in previous studies. When comparing the results of the present study with Conrad and Biber (2004) results, the counts for academic prose remained almost exactly the same (even if Conrad and Biber used the Longman Spoken and Written Corpus for their research). Conrad and Biber suggested that lexical bundles cover about 20% of the academic discourse, while the result of the present study is slightly lower: 19.3%. As for the spoken language, there was quite a substantial difference between Conrad and Biber's estimate (28%) and the result of the present study (59%).

After looking at the data more closely, it appeared that there was one difference in methodology: Conrad and Biber counted contractions (e.g., *don't*) as one word, while the present study treated them as separate words. This would seem to explain why there was no difference for the academic prose (where contractions tend to be avoided), but the difference appeared in the spoken conversations. In the present study, contractions were counted as separate words for two main reasons. Firstly, the software treated them as two words both in extraction of FSs and in all the other counts (such as overall word count of the corpus). Secondly, it seemed to be more consistent to count contracted forms and their non-contracted counter forms in the same way (e.g., counting both *I don't know* and *I do not know* as 4 word sequences). To investigate to what extent this different treatment of contractions affected differences in the counts, the results were checked again, counting contractions as single words. It turned out that the percentage of lexical bundles in spoken corpus would drop to 52.51% if contractions were treated as single words. However, this number was still considerably higher than the one reported in Conrad and Biber's (2004) study. Therefore, the remaining explanation for this difference seems to be the differences between the corpora used in the two studies. This finding suggests that generalizations across different corpora have to be made with caution. Also, it seems to suggest that the usage patterns of lexical bundles in academic prose are a bit more consistent than in spoken conversations.

Summarizing the overall estimates of frequency of FSs in the BNC Baby, it can be claimed that there are significant differences in the frequencies of use of collocations, idiomatic phrases, lexical bundles, and phrasal verbs in language. This finding suggests that looking at FL as a whole and reporting the overall frequency of FSs in language hides different frequencies of use of different categories of FSs.

3.4.2 Variation between registers

The study showed that register affects the tendencies of use of FSs, which is in line with claims that register has an influence on the use of lexical and lexico-grammatical features (Allwood, 2012; Biber et al., 1999). The results suggest that although the trend of the frequency of use of different sequences remains the same in all the registers (lexical bundles are the most frequent sequences, followed by collocations, idiomatic phrases, and phrasal verbs), each category of FSs has its own profile of use in different registers. For example, lexical bundles were the most frequent in spoken conversations, while collocations were the most frequent in academic prose.

Looking at the overall tendencies of the use of FSs in different registers, spoken conversations were the most formulaic register. As spoken sub-corpus consisted of spontaneous conversations which put time pressure on speakers, this might have led to the adoption of FL in order to speed up language production as well as to aid the hearer and decrease the chances of misunderstanding (Wray, 2002). Importantly, such a high formulaicity of spoken discourse seemed to be driven mostly by a very frequent use of lexical bundles.

Fiction was the second most formulaic register, which is interesting as the language of literature is by definition expected to be creative. In terms of use of some of the categories of FSs (such as phrasal verbs), fiction approached spoken conversations. There might be various explanations for this finding.

One the one hand, fictional texts might contain dialogues that try to imitate spoken language. On the other hand, the high proportion of FSs in fiction seems to emphasize the fact that formulaic use of language is a default way of constructing a linguistic message (Wray, 2008).

Academic prose followed the fiction sub-corpus very closely. The newspaper sub-corpus, though, turned out to be the least formulaic part of the BNC Baby. This result was rather surprising, because one would probably expect newspaper language to be more formulaic than, for instance, fiction. However, this result might be due to the nature of the newspapers sampled for the BNC. The corpus sampled both national and local papers, but it did not cover tabloids or online information portals (not surprisingly, considering the years the BNC covers), which might use more FL. Also, it might be that journalists tend to use more variable sequences, which allow open slots and certain variation and therefore were not captured in this study. However, these claims are purely speculative and would require further empirical research.

The fact that spoken language was distinctively more formulaic than other registers might have been the reason why even if register differences were significant, the mode of language production (spoken or written) had a stronger effect on the distribution of FSs. Spoken language was more formulaic than written language (in line with the findings of Erman & Warren, 2000). However, it is worth noting that spoken conversations exceeded written discourse in the use of lexical bundles and phrasal verbs only, while collocations and idiomatic phrases were used more frequently in at least one of the sub-corpora of written discourse.

It has to be kept in mind that different registers may possibly have even larger effect on the choice of individual FSs, which was not addressed in this study. This claim is supported by the differences of counts of lexical bundles when extracted from the whole BNC Baby compared to the estimates obtained from the separate sub-corpora (discussed in Section 3.2.1.2). In any case, it is clear that even when not analysing the use of individual sequences, different tendencies of use of different categories of FSs appear.

3.4.3 Categorization of formulaic sequences

The present study focused on four categories of FSs, which seemed to be carefully defined not to overlap. However, the analysis showed that drawing a clear line between one category and the other is not always possible. The same sequence can meet the criteria to be included into different categories and shorter sequences can be used as a part of longer sequences. So the boundaries between categories become rather blurry.

This fact is noticeable in previous research as well. For example, in the present study, there was some overlap between idiomatic sequences and collocations (e.g., *large scale* was listed both as an idiomatic phrase and as a collocation). For the purpose of this study, these FSs were considered to be idiomatic phrases because of their non-transparent meaning. Figurative meaning was chosen as the main criteria for identifying idioms in previous research as well (Moon, 1998a). However, these items can also be classified as collocations with figurative meaning, as argued by Macis and Schmitt (2016). Hence the decision to classify them as idioms, or as collocations for that matter, is pretty arbitrary and depends on the extraction criteria adopted. In spite of checking for the overlap between the categories, there might still be non-transparent sequences listed as collocations because they reached the cut-off points for collocations and were not included in the list of idiomatic phrases.

As far as the use of the lexical bundles is concerned, it was noted in the literature that shorter lexical bundles tend to be used as a part of longer lexical bundles (e.g., a four-gram comprising a three-gram) (Biber et al., 1999). However, one other tendency seems not to be discussed that much in the literature: lexical bundles (especially the longer ones) can comprise other types of FSs. That is, idioms, collocations or phrasal verbs can be used as a part of a lexical bundle. Therefore, even if lexical bundles are usually claimed to be semantically transparent, once they contain non-transparent sequences, they become non-transparent as well. For example, a sequence *in spite of* was listed as an idiomatic phrase in this study. This sequence was also extracted as a part of the lexical bundles *in spite of the* and *in spite of all the*. Therefore, it seems that lexical bundles can contain not only shorter n-grams but also other types of

FS, or, to put it the other way around, certain sequences, which have a distinct, meaning can be used with a certain pattern of function words attached to them.

Looking at those overlapping cases between the different categories, it seems that most of them were due to the different approaches to extracting these FSs. While collocations and lexical bundles were mostly defined by their frequency and extracted following a corpus-driven approach, idioms and phrasal verbs were pre-listed based on the semantic criteria. In cases when an idiom or a phrasal verb was frequent in the corpus, they could have been captured by both procedures of the extraction.

Having discussed all these issues with categorization, two main questions remain: 1) Is it worth trying to categorize FSs? and 2) Is the classification adopted in this study valid? I would argue that despite the problems with any attempts to classify FSs, there are practical and theoretical reasons for dividing FSs into different categories. Firstly, these various categories have distinct features that make the classification feasible in the first place. In the present study, certain overlap existed only for a minor part of the sequences. Furthermore, we cannot assume that all the FSs are stored in the mental lexicon and processed in the same way. So looking at different categories separately in a principled way seems to be important. Finally, separating different categories can be useful for teaching purposes: some of the categories could create more problems receptively, others productively and they should probably be approached differently in classrooms. Hence even if a clear-cut distinction between the categories is unviable, the fact that FL is heterogeneous cannot be ignored.

3.4.4 Pedagogical implications

Obviously, this study cannot make any direct suggestions about how to teach various FSs, but its findings could be useful for choosing which categories of FSs to focus on. Table 3.11 provides a summary of how many individual items of each category were extracted/used for the present study and the percentage of the corpus that their covered.

Table 3.11 Number of items considered in each category

Type of sequences	Number of items	Percentage covered
lexical bundles	17,460	29.49
collocations	13,139	7.52
idiomatic phrases	590	2.55
phrasal verbs	211	0.99

Based on the present corpus analysis, lexical bundles provide the largest coverage. Therefore, presumably, if one wants to address FSs in classroom, the most frequent type of FSs should be somehow dealt with. Nevertheless, I would not necessarily advocate extensively focusing on lexical bundles in classroom for several reasons. On the one hand, as far as language reception is concerned, most of the lexical bundles are not problematic, as they tend to be transparent. On the other hand, it seems that it is rather difficult to teach productive use of lexical bundles even during extended periods of time. For example, Cortes (2006) tried teaching lexical bundles to NSs of English during a university level History course. She found after 10 weeks of treatment her students became more aware of lexical bundles, but they were still not producing the same number of lexical bundles as established academics. Considering how numerous the category is, it seems that it is only worth focusing on few sequences selected in a principled way. This is not an easy task, though, as there is no list of lexical bundles of general language to guide the selection. One option could be relying of some statistical measures to select the most useful sequences to teach (e.g., formula teaching worth, Simpson-Vlach & Ellis, 2010). However, considering Cortes (2006) findings, it remains to be established how teachable the items suggested by the formula are.

Collocations also seem to be a very numerous category and they cover at least 7.5% of the corpus. The situation for collocations seems to be very similar to the lexical bundles: they are usually transparent and should not be an issue for language reception. When it comes to production, though, collocations might be more useful than lexical bundles: they are a conventional way of conveying certain meanings. So a failure to choose a right collocate can impede communication (Millar, 2011). It seems that collocations pose serious problems even for advanced learners of language (e.g., Bahns & Eldaw, 1993;

Laufer & Waldman, 2011), which would imply that teachers should offer some help with learning collocations. However, the number of collocations extracted from the BNC Baby (> 13,000) was already too high to directly teach them and if a larger corpus was used, this number would probably increase even more. Therefore, it seems that collocational knowledge (at least a large part of it) remains to be learned incidentally (see Chapter 6 for further discussion on incidental acquisition of collocations). If a teacher decides to address collocations in classroom, there are several lists that attempted to guide the selection of the items. Shin and Nation (2008), for example, created a list of the most frequent collocations that should be included in the most frequent 2,000 words list. However, this list is based on the spoken part of the BNC only and it only covers the most frequent collocations, so it is not clear what to teach to the learners beyond the beginners' level.

Phrasal verbs were the least frequent category. However, the fact that only 211 phrasal verbs covered about 1% of the language, suggests that by teaching phrasal verbs, a relatively high coverage can be achieved by learning a relatively low number of sequences. Also, phrasal verbs have a figurative meaning, so they can cause difficulties even receptively. Luckily, some pedagogically oriented lists of phrasal verbs have already been suggested, not only providing the most frequently used phrasal verbs, but also listing the most frequent meanings, in such a way narrowing down the number of the phrasal verbs to learn considerably (Garnier & Schmitt, 2015).

Regarding the idiomatic phrases, the suggestions are not that obvious. The study did not look at all the idioms, which are very numerous and presumably not very frequent. However, the idiomatic phrases addressed in the study were relatively frequent and non-transparent. Hence, a failure to understand them would decrease the understanding of texts by about 2–3%. However, the list of idioms was not that short and could probably not be taught as quickly as phrasal verbs. There were attempts to list the most useful idioms for students. For example, Grant and Nation (2006) selected a list of 'core idioms' and checked their frequency, but they found that only seven of their selected idioms reached the frequency required to be included into a list of the 7,000 most frequent lexical items. Therefore, the researchers concluded that "[c]ore idioms are not frequent and do not deserve classroom attention"

(Grant & Nation, 2006, p. 15). The idiomatic phrases that were analysed in this study, though, are much more frequent and hence more useful for learners. Therefore, the two lists that were used for their identification (Liu 2003; Martinez & Schmitt, 2012) could be a potential starting point for pedagogy to give a relatively high percentage of coverage.

Furthermore, when teaching FSs, register information has to be taken into account at least to some extent. The difference in the use of various types of FSs across registers could inform the decisions of which sequences to focus on when teaching language in various contexts. For example, in conversational classes, providing learners with ready-made lists of the most frequent lexical bundles might be a good starting point, considering how frequent they are in spoken conversations. When learning academic writing, on the other hand, learners might benefit more from noticing collocational use of words.

3.4.5 Limitations

There are obviously a number of limitations of the study. Some specific limitations, concerning the methodology of extracting each type of FSs, were already discussed in the Methodology section (3.2.1). Therefore, this section will only discuss some general limitations.

To start with, only four categories of FSs were considered. Hence, there are still a number of FSs, which were not accounted for, such as clichés, similes, proverbs, and variable expressions to name just a few. The categories considered in the study were probably the easiest ones to search for in corpora and including other types of FSs would require much more manual work or advanced programming. Allowing for much more variation in sequence use would also be desirable. While it is questionable if it would ever be possible to extract all the FSs (simply because for doing that we would need very clear and indisputable criteria of what is formulaic) looking at more sequences and accounting for variability should be a next step for any study looking at the distribution of FSs in corpora.

Secondly, the categories addressed in the study were only extracted based on the selected definitions. So, for example, grammatical collocations were not counted as FSs, even if they might be even more frequent than lexical collocations (see the findings of Durrant, 2009). Different definitions would have led to different estimates.

Also, the corpus used for the present study (the BNC Baby) is relatively small by the standards of modern corpus linguistics. However, the study could not be fully automated and required some manual checking. Also, the software had technical issues coping with larger corpora. Therefore, four million words seemed to be an adequate starting point to provide at least initial insights about the distribution of various FSs. Further analysis with a larger corpus using more advanced computational techniques could definitely be useful to validate and refine the current findings.

Having in mind all these limitations, it has to be admitted that the counts provided cannot be taken as precise estimates, but rather as approximations giving relative ranking of the categories.

3.5 Concluding remarks and directions for further research

Overall, even if the study somewhat underestimated the extent of formulaicity in language (as it only looked at the four categories), it provided some initial estimates of the percentage of each category of FSs in language. However, the main issue with the extraction of FSs (both in the present study and overall) seems to be the variability of FSs. In the present study, an attempt to account for some variability was made by manually checking the phrasal verbs for potential insertions and trying to list all morphological forms of the idiomatic expressions used. For the phrasal verbs, which allow no positional variations (*get up* but not **up get*), variability was rather easy to account for. The frequent idiomatic phrases used in the present study were also rather fixed (though, if all idioms were considered, the variation would increase considerably). Lexical bundles are defined by their fixedness, so the variability was not an issue for this category. However, when it comes to collocations, not accounting for their form variation might have seriously affected the estimates. This seems to be the case especially for verb-noun collocations (as discussed in Section 3.3.3), as allowing for non-adjacency would potentially lead to a considerable increase in the frequency of this category. However, to the best of my knowledge, there are no ready-made tools that would allow extracting collocations with positional and adjacency variation without any initial input: not focusing on the collocations of a specific word, but rather on all the lexical items in a particular corpus.

Even if variation of the form of FSs is not a new idea (e.g., Langlotz, 2006; Moon, 1998a), there is not much research on how the variability of a form affects processing and acquisition of FL. To my best knowledge, there have not been any published studies which would systematically address the questions of morphological variation, positional variation or adjacency of collocations. So on the one hand, the present corpus study has a limitation of not looking at collocations within a usually adopted span of ± 4 words. On the other hand, it seems that the research is not yet in the position of answering the question whether collocations with, for example, three intervening words are psychologically real. Therefore, we cannot assume that collocation effect extends to items beyond the core adjacent form of collocation. In that sense, the corpus analysis in the present study can be seen as adopting a more conservative approach to collocations, by focusing only on the items that we know show some facilitative effect.

The question of how variation affects collocation processing is an empirical question. Therefore, it seemed to be a logical next step for this thesis to investigate the effect of variability on processing FSs. If FSs were shown to be psychologically real units despite the variation in their form, then the criterion of fixedness for defining FSs should be reconsidered and much more variability should be systematically allowed when extracting the items. On the other hand, if the evidence showed that there was no processing advantage for collocations once they were non-adjacent, then we should reconsider corpus analysis techniques that adopt a ±4 words span for co-occurrence.

With this in mind, an eye-tracking study was designed to look at the processing of non-adjacent collocations. It was carried out with NSs of English first and then replicated with language learners. The two studies are presented in Chapter 4 and Chapter 5.

Chapter 4

Processing of adjacent and non-adjacent collocations by native speakers

FL is interesting not only because it is so widely used (as evidenced by the study reported on in Chapter 3 as well), but also because it has been demonstrated to be cognitively real, and these two features seem to be inextricably connected. Frequently used FSs tend to be processed faster, which leads us to believe that they have some sort of a special status in the mental lexicon. While this result has been reliably replicated, there are still questions to be addressed. One of them is the effect of variability of FSs. As already briefly discussed, FSs can be used in different forms in terms of the position, morphology and adjacency of their elements. The present chapter reports on an eye-tracking study on processing one type of such variation, namely, non-adjacency (i.e., FSs with other words intervening between their elements).

4.1 Background of the study

4.1.1 Processing advantages for formulaic sequences

It is widely accepted that FSs are different from novel phrases and show certain processing advantages (Siyanova-Chanturia, 2013). As discussed in Chapter 2, this finding was replicated for idioms (e.g., Rommers et al., 2013; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Tabossi et al., 2009; Vespignani et al., 2009), lexical bundles (e.g., Nekrasova, 2009; Reali & Christiansen, 2007; Schmitt et al., 2004; Tremblay et al., 2011), phrasal verbs (e.g., Kim & Kim, 2012), and collocations (e.g., Durrant & Doherty, 2010; Sonbul, 2015).

However, most of the studies on processing FSs did not take into account the fact that FSs are not necessarily fixed and in many cases allow some variation. Thus, there is still not enough evidence to know whether processing advantages can be extended to FSs when they are modified. This is an important question for a number of reasons. First of all, for some FSs, the most frequently used forms might not be the fixed ones (see, for example, the analysis of collocations in Section 3.3.3). Also, the question of whether modified FSs retain processing advantages can inform the discussion on holistic storage (see Section 2.2.2). If the FSs lost their processing once modified, this would suggest that there is indeed a core form stored holistically and retrieved directly during language processing. Otherwise, if modified FSs showed processing advantages, it would be more likely due to faster activation of the elements of the FS, rather than its holistic storage.

4.1.2 Studies on modified formulaic sequences

There have been a few studies that looked at processing of modified FSs. In this case *modified* does not mean changed to be novel, but rather not presented in their fixed core form. These studies are reviewed in this section.

In order to look at the positional variation of collocations, Bonk and Healy (2005) carried out a primed word naming task. Target collocations were presented either forwards (*bend – rules*) or backwards (*rules – bend*). Both backwards and forwards conditions showed a facilitative effect. This led the authors to suggest that as positional variation does not disrupt processing, collocations are not stored as holistic units. Rather the links between their elements are entrenched in memory due to their frequent co-retrieval. It would be interesting to expand this study, though, adding collocational directionality measures (used in corpus linguistics, such as ΔP, suggested by Gries (2013)) to the analysis. It could shed more light on the role of preferred position of words in collocations.

Insertions in FSs have also been addressed. Molinaro, Canal, Vespignani, Pesciarelli, and Cacciari (2013) looked at modified complex collocational prepositions (e.g., *in the <u>capable</u> hands of*) in Italian. They compared the core form of the preposition to the modified form. In a self-paced reading task, they found that prepositions with insertion were read slower, probably because of more information to be integrated. The ERP results, though, showed that the insertion did not disrupt processing, with indeed a smaller N400 effect for the noun following the inserted adjective. These results suggest that FSs allow modification without losing their formulaic status, despite the fact that they have their preferred structure.

Taken together these studies seem to suggest that frequent phrases retain their processing advantage even when some variation is introduced. However, not all of the evidence suggests so. For example, Carrol (2015) analysed reading times of various types of FSs when there are words intervening between the elements of FSs and when FSs are reversed. He looked at idioms and binomials as well as associated and non-associated collocations. The only type of phrases that showed a clear facilitative effect when modified was idioms. All of the other types of FSs seemed to lose their formulaic advantage once they were split apart.

There was also a study that looked at the processing of phrasal verbs with a direct object either following the phrasal verb or intervening between the verb and the particle (Gonnerman & Hayes, 2005). This self-paced reading study found that words in the non-adjacent phrasal verb condition were read slower than in adjacent ones. When looking at the reading times of the adverbial particles, the researchers found that the disruption depended on the length of the noun phrase inserted (up to nine words) and the dependency between the verb and the adverbial particle: the more opaque the meaning of the phrasal verb was, the more non-adjacency disrupted processing.

Overall, the previous studies on processing modified FSs suggest mixed results. This is not surprising, considering that different types of sequences were analysed and different methodologies used.

The studies on learning non-adjacent dependencies also seem to add something to the discussion. They show that humans are able to track and learn non-adjacent dependencies from a very early age (van Heugten & Shi, 2010). However, it was also suggested that adjacent dependencies are easier to learn than non-adjacent ones (Gómez, 2002; Newport & Aslin, 2004), even if this finding was later questioned by Vuong et al. (2011) who claimed that adjacent and non-adjacent dependencies are learned simultaneously. These statistical learning studies would suggest that when encountering a FS, we can strengthen the links between its elements no matter if they directly follow each other or not. So we could expect that processing advantages should also hold no matter the adjacency.

The present study sets out to test if adjacent and non-adjacent collocations show similar processing advantages. To do so, an eye-tracking experiment analysing reading time of collocations in adjacent and non-adjacent conditions was designed. Before moving to the design of the present study, the following section presents the eye-tracking technique and the main features of lexical items to consider when designing an eye-tracking experiment.

4.1.3 Eye-movement research

An eye-tracking methodology was adopted for the present study for numerous reasons. First of all, during eye-tracking experiments participants are presented with a very natural task: they are asked to read a text for comprehension, so they are oblivious of the exact purpose of the study. Also, this technique is much more ecologically valid compared to other techniques such as word naming, lexical decision tasks or self-paced reading studies, because language processing during the eye-tracking is not disrupted by any secondary tasks (Sereno & Rayner, 2003).

The following sections briefly describe the nature of eye-movements when reading, the well-known factors that affect eye-movements, and finally the main measures of eye-movements, usually employed in lexical studies.

4.1.3.1 Saccades and fixations

We move our eyes to fixate objects of interest (Wade & Tatler, 2011). When reading, we have to move our eyes to fixate new regions in order to be able to process them. Eye-movement research shows that eye movements are not smooth and continuous but rather consist of a series of fixations and saccades (Wade & Tatler, 2011). Vitu describes saccades as “brief movements of our eyes which bring poorly-resolved peripheral input onto the central, foveal part of our retina” (Vitu, 2011, p. 731). About half of saccadic movements move the eyes to the next word in the text (n + 1); about 20% skip the following word and fixate the word after it (n + 2); another 20% of saccades re-fixate the same word (n); finally, 10% of saccades are regressions: fixations of previous words (Engbert & Kliegl, 2011). Saccades are followed by fixations: periods when an eye is relatively still. However, this is a simplistic explanation, as eyes never stay completely still (Gilchrist, 2011; Martinez-Conde & Macknik, 2011). When reading, we fixate the majority of words (Rayner, 1998).

The main assumption underlying eye-movement research is that eye-movements reflect our cognitive processes. Therefore, by understanding and modelling eye-movements we can arrive at better insights into human cognition. However, there is a debate whether eye movements are generated by high-level cognitive processes or whether they are influenced by low-level oculomotor processes as well (Rayner & Liversedge, 2011; Vitu, 2011). When reading, we have to make two kinds of decisions: *where* to move the eyes next (saccadic movement) and *when* to move the eyes (how long to fixate the currently fixated region). While it seems to be generally accepted that the *when* decision is controlled by cognitive processes, the nature of saccadic movements is less clear (Rayner & Liversedge, 2011). Saccadic movements seem to at least somewhat depend on the basic properties of oculomotor processes (Vitu, 2011). Because of the fact that visual acuity is maximal in fovea, words have to be fixated so that the maximum information can be extracted. This leads to the conclusion "the duration of fixation is controlled by the ongoing comprehension processes, while saccadic targeting is primarily determined by oculomotor constraints and lower-level visual features of the text" (Hyönä, 2011, p. 819). However, even saccadic eye-movements seems to be at least partially influenced by cognitive processes as, for example, predictable words are skipped more often than non-predictable words (Rayner, 1998; Starr & Rayner, 2001). So oculomotor processes alone could not account for these results. In reading research, though, analysing fixation durations is more informative than analysing saccadic patterns.

4.1.3.2 Factors that affect eye-movements

Various factors, such as word frequency, length, predictability, age of acquisition, concreteness, and contextual constraint seem to have an effect on its processing time and hence on the decision *when* to move your eyes (Frisson et al., 2005; Juhasz & Rayner, 2003; Kliegl et al., 2004; Rayner & Liversedge, 2011). As the present study focuses only on lexical processing; only the features that were demonstrated to affect lexical processing will be discussed.

Word frequency. It has been shown on multiple occasions that word frequency has an effect on how long the word is fixated with high frequency words being fixated shorter than low frequency words (Rayner, 1998; Rayner, Ashby, Pollatsek, & Reichle, 2004). The effects of frequency were shown to hold even if the fixated word disappeared after 60 ms: high-frequency words were still fixated shorter, which once again suggests that cognitive processes rather than oculomotor mechanisms determine fixation durations (Rayner, Liversedge, White, & Vergilino-Perez, 2003).

Word length. Kliegl et al. (2004) showed that word length has a larger effect on processing that word than its frequency or predictability. In their data, word length accounted for about 25% of the variance in the reading times. They showed that word length decreased skipping probability and increased re-fixation probability.

Age of Acquisition. Juhasz and Rayner (2003) showed that an age of acquisition has an effect on word processing. However, their experiment used self-reported estimates of age of acquisition, which means that they evaluated perceived rather than actual age of acquisition. Thus this measure could have been confounded by word familiarity.

Concreteness. Juhasz and Rayner (2003) found that concreteness of the word had an effect on reading times, just as it has been shown to be an important factor for other laboratory studies, such as lexical decision or word naming tasks. Concrete words are read faster than abstract words.

Contextual predictability. Predictable words are fixated shorter and they are more likely to be skipped (Rayner, 1998; Rayner et al., 2004; Starr & Rayner, 2001; Vainio, Hyönä, & Pajunen, 2009). This seems to be due to the fact that readers can extract some parafoveal information about the n + 1 word (Starr & Rayner, 2001). Rayner and Well (1996), for example, carried out an eye-tracking study to establish how different levels of contextual constraint affect reading. They looked at words in three conditions: high contextual predictability, medium contextual predictability, and low contextual predictability (as established by cloze tests). The study showed that high-predictability words were skipped more often and fixated shorter. Medium-predictability words were still fixated shorter with no higher skipping rates.

Transitional probability. McDonald and Shillcock (2003a, 2003b) investigated the effect of transitional probabilities on word reading times. Their studies showed that words with higher transitional probability are read faster than words with low transitional probability, suggesting that when reading we draw on our expectations about which words tend to occur together. However, Frisson et al. (2005) critiqued McDonald and Shillcock's studies claiming that transitional probabilities do not have a significant effect on processing if contextual predictability is controlled for. They argued that contextual predictability is more important than transitional predictability even if it is very difficult to disentangle those two effects.

While it has been extensively demonstrated that all the discussed factors have an effect on word processing, which is reflected in eye-movements, it has to be born in mind that most of these factors are interrelated. For example, word length and word frequency tend to correlate and it is not always easy to establish the effects of these factors individually (Kliegl et al., 2004). Correlations can be found between frequency, length, concreteness, and age of acquisition as well (Juhasz & Rayner, 2003). Researchers who tried to disentangle these effects put forward some suggestions. For example, Juhasz and Rayner (2003) suggested that frequency and familiarity show early but lasting effects on fixation durations, while word length significantly predicts fixations only when re-fixations are taken into account. In any case, all these factors are potential confounders in experiments and have to be controlled for.

It has also been demonstrated that not only the word that is currently fixated and its characteristics, but also the preceding and following words, have an influence on reading. Kliegl, Nuthmann, and Engbert (2006), for example, analysed an eye-tracking corpus of 144 sentences read by 222 participants and showed that lexical characteristics of the previous word (n - 1) have a very strong influence on reading times of the current word (n). As for the following word, the effect was still significant albeit smaller. It is worth keeping in mind that features of other words surrounding the word or area of interest can have a large effect on the processing of that word.

Finally, Rayner (1998) warned that considerable inter-reader differences exist: children show somewhat different reading patterns than adult readers, skilled readers show lower fixation times than slower readers, people reading in L2 tend to show more regressions and lower skipping rates. Therefore, when analysing eye-tracking data, it is very important to take individual differences between readers into account.

4.1.3.3 Measures of eye-movement

There are various measures that can be adopted for eye-movement analysis. Rayner (1998) argues that no one of these measures can give a full reflection of cognitive processing when reading. Therefore, he advocates analysing several measures that can complement each other. Table 4.1 summarizes the measures frequently used in eye-movement research (especially in studies on FL) and adopted for the present study.

Table 4.1 Eye-tracking measures

Measure	Definition
First fixation duration	Duration of the first fixation on the area of interest
Gaze duration	All fixation durations prior to moving outside the area of interest
Total reading time	Total time spent reading the area of interest
Fixation count	A number of fixations within the area of interest

Note: based on Roberts and Siyanova-Chanturia (2013)

As lexical identification of a word is necessary before processing syntactic or discourse level information, lexical processing should precede syntactic or contextual processing. Thus early measures of eye-movements (e.g., first fixation duration, gaze duration) are associated with lexical processing, while late measures (e.g., total reading time) with semantic and syntactic processing and phrase level integration (Rayner & Liversedge, 2011). It has been suggested that comparing early and late measures of processing is one of the key advantages of the eye-tracking methodology (Roberts & Siyanova-Chanturia, 2013).

Carrol and Conklin (2014) noted that most of the eye-movement studies on reading so far have focused on individual words, which makes it challenging to directly apply the same measures for analysing multiword areas of interest, as in case of FSs. The suggestion they put forwards is to use a hybrid approach: to analyse both the final word of the phrase and the whole phrase reading times. In such a way, the whole phrase reading captures the effect even if some individual words are skipped and excluded from the individual word analysis. Also, the whole phrase analysis shows an effect if its locus is not the final word. Therefore, this hybrid approach is adopted for the analysis in the present chapter.

4.2 Study 2: Processing non-adjacent collocations in the L1

The present study sets out to test if non-adjacent collocations show the same processing advantage as adjacent ones. To start with, it aims to replicate the collocation processing advantage for adjacent items, consistently shown in previous research (e.g., Durrant & Doherty, 2010; Sonbul, 2015), using an eye-tracking technique and controlling for the contextual predictability. More importantly, the study investigates if this advantage extends to non-adjacent collocations.

If non-adjacent collocations were processed faster than control phrases, it would be in line with the findings of Molinaro et al. (2013), as their study showed that modifications of multiword sequences do not disrupt their processing. Also, it would suggest that non-adjacent collocations extracted from corpora with a certain span of co-occurrence are cognitively real FSs.

On the other hand, there have been suggestions that statistical probabilities of word co-occurrences would not have an effect on processing when words are separated. For example, Rayner, Warren, Juhasz, and Liversedge (2004) claimed that “it is somewhat unlikely that transitional probability effects would survive intervening words” (p. 1298). If there was no processing advantage for non-adjacent collocations, this would give some initial evidence that transitional probability effects disappear if words do not directly follow each other.

4.2.1 Target items

In order to compare the processing of collocations to the processing of novel phrases, four groups of stimuli were created (see Figure 4.1)

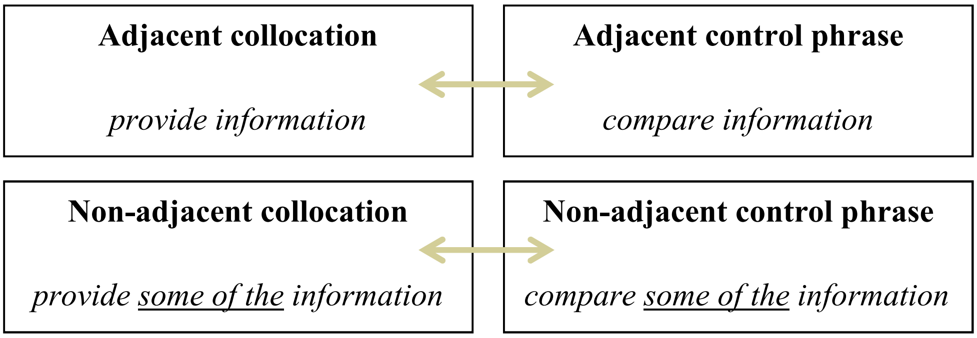


Figure 4.1 Groups of stimuli

Having these four groups of items allowed comparing the processing of adjacent collocations to novel phrases as well as non-adjacent collocations to relative novel phrases. This design allowed me to write sentences that were matched between collocation and control conditions both in terms of syntax and in terms of lexis used. It also allowed me to analyse the effect of both collocation status and adjacency of collocates on the processing speed of different phrases.

A statistical approach to collocations was adopted in this study. Collocations were considered to be two words with an MI ≥ 3. Only verb-noun collocations were used for the study for two reasons. Firstly, a tight experimental control was needed and different syntactic relations between the collocates might have led to differences in processing. This seemed to be the case for Wolter and Yamashita's study (2014), where verb-noun collocations were processed differently than adjective-noun collocations. Also, after the small-scale analysis in the previous study (see Section 3.3.3), it appeared that non-adjacency is most common for verb-noun collocations.

One additional criterion for choosing target collocations was the frequency of their component parts. As collocation frequencies are always lower than frequencies of individual words, this restriction allowed me to select relatively frequent collocations. Therefore, the participants were expected to be familiar with target collocations and have encountered in their everyday language use.

The selection of the target items started from obtaining the individual words to use as node words for collocation selection. A list of the first 2,000 most frequent lemmas of the BNC was obtained (Kilgarriff, n.d.); and separate lists of nouns and verbs were filtered from it. As collocations had to be used without any words in between their elements for the adjacent condition (and a determiner preceding a noun would be required for most nouns), I obtained the list of non-countable nouns from the Kilgariff's list of nouns and used these nouns as a starting point for collocation selection. This list of nouns was searched for in the BNC using the Brigham Young University interface (Davies, 2004). This interface allowed me to search for verb collocates of noun lemmas within the span of 4 words. If a noun had frequent verb collocate (collocation lemma frequency > 50), which would also allow intervening words between the verb and the noun, this collocation was listed as a potential target collocation. The frequencies of the collocations and their component parts were noted, as well as the MI scores.

Control verbs were chosen trying to find verbs that could be naturally used with each noun. All the control verbs were taken from the 2,000 most frequent lemmas as well. Also, control verbs were checked not to be collocates of the noun (their MI < 1). The length of the control verb was matched to the length of the target verb.

Association norms of both collocations and control phrases were checked in two databases (Kiss, Milroy, & Piper, 1973; Nelson, McEvoy, & Schreiber, 1998). To make sure that any observed processing advantage was due to the collocation status of the words rather than their semantic associations, word pairs that were semantically associated were excluded from the study. In 11 cases (6.9% of all the word-pairs), one of the words in the pair was not listed as a key word in the association databases. However, if the second word was on the list and did not yield the first one as its association, the pair was still used for the study.

Three words were always inserted between the verb and the noun in the non-adjacent condition. A number of three was chosen because of the tendency in corpus linguistics to use the span of ± 4. Intervening words were checked in the BNC not to be strong collocates of the target nouns (the MI < 2). This restriction was chosen to make sure that if there was an effect in the data, it would not be driven by the collocation strength between the inserted words and the noun.

Fifty-one blocks of sentences, each containing four sentences (one per condition), were originally written to use in the pilot stage of the study. Care was taken to make sure that the collocations would not appear in the first two positions or in the last position of the sentence. The beginnings of the sentences in all four conditions were always the same up to the target sequence. At least two words after the target sequence were also always retained, to make sure that any potential parafoveal preview of the words after the target sequence would not affect the processing of the target phrases. When possible, the sentences were kept identical for both collocations and control items. When the changes were needed, care was taken to ensure the changes are minimal (see Table 4.2 for an example). For the non-adjacent conditions, the sentences were always the same as in the adjacent conditions with the only difference of the inserted words. All the sentences used in the study are presented in Appendix 5.

Table 4.2 Examples of sentences in experimental conditions

Adjacent collocation	**Adjacent control**
John was asked to **provide information** about the studies to local newspapers.	John was asked to **compare information** about the studies for local newspapers.
Non-adjacent collocation	**Non-adjacent control**
John was asked to **provide** some of the **information** about the studies to local newspapers.	John was asked to **compare** some of the **information** about the studies for local newspapers.

4.2.2 Norming study

Frisson et al. (2005) claimed that collocation status does not have an effect on processing, once contextual predictability is controlled for. Therefore, in order not to overestimate collocational effect in the present study, I had to control for predictability of the collocations. I was aiming to write sentences where the context would not allow guessing the second element of the collocation. However, the very definition of a collocation presupposes that the words have a certain mutual expectancy, so words in collocations are very likely to be more predictable than words in the control phrases. Frisson et al. (2005) also admitted that contextual predictability and transitional probability are intrinsically connected as participants draw on their experience of word co-occurrences to make predictions. In order to disentangle these two effects (collocation status and contextual predictability), a norming study for predictability was carried out and predictability scores were used a as a covariate in the statistical analysis.

4.2.2.1 Design of the norming study

The norming study consisted of two parts. The first part tested the predictability of the items and the second part evaluated the naturalness of all of the sentences. The aim of the norming study was to make sure that any processing advantage for collocations can be attributed to their collocation status, rather than to their contextual predictability or non-naturalness of the control sentences.

Four lists of items were created for the norming study. Items were divided into the lists based on the Latin square design. Each list contained all the target nouns used in one of the four conditions.

To establish the predictability scores, the participants were presented with a pseudo-cloze test with sentence fragments up to the noun and were asked to guess the following word (see Figure 4.2). The participants were informed that their task is not to complete the sentence but only to provide one word that could follow the verb.

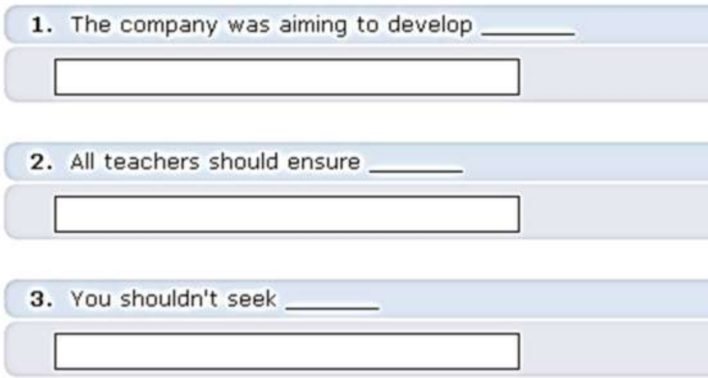


Figure 4.2 Items for the predictability norming

For the naturalness norming, 30 filler items were written. The aim of the filler items was to present the participants with some truly unnatural sentences. As the participants were given the task to rate the naturalness of the sentences, if all the sentences sounded more or less natural, presumably the participants would have tried to classify them to the different categories of naturalness just because of the task effect. All the filler sentences used were grammatically correct. They were checked by a native speaker of English before the norming study.

Each participant read one of the four lists, each containing 81 sentences in a randomized order (same for all the participants). The filler sentences were the same in all four lists. Participants were presented with a Likert scale (from 1 to 5) and were asked to judge how naturally the sentences sounded (as illustrate in Figure 4.3).

Figure 4.3 Items for the naturalness norming

The participants saw the predictability test first, followed by the naturalness rating. The lists were combined in such a way that the same participant would never see the sentence in the same condition in both parts of the norming study. The sentences were presented to the participants online, using the *Bristol Online Surveys* interface. The tasks took approximately 20–30 minutes to complete.

4.2.2.2 Participants of the norming study

Twenty participants took part in the norming study. They were all NSs of English, studying for a major in English. All the participants were from the same population as participants of the main study. Each participant completed only one of the four versions of the study.

4.2.2.3 Results of the norming study

All 20 students saw the same 30 filler items in the naturalness rating task, which allowed me to estimate their inter-rater reliability. Intra-class correlation was calculated using the *psych* package (Ravelle, 2014) in R. An average correlation coefficient was calculated, because the normality and predictability scores used later in the study are based on the average responses of the raters as a group rather than individual rater judgements. Intra-class correlation coefficient (for average measures, random raters) was .91 (p = .00). It allows me to assume that raters that saw different target items were consistent in judging their naturalness.

The items that were predicted by more than two participants in the norming study (predictability > .4) in at least in one of the four conditions were excluded from the study. This led to excluding four items. The items that had an average naturalness rating of below three (in the scale of one to five) in at least one of the four conditions were excluded from the further analysis as well. This led to the removal of two items. As I wanted to have 40 items for the actual experiment, additional five worst performing items were excluded. Further analysis will focus on the 40 items that were retained for the actual study (see Appendix 4 for the full list of collocations and control phrases).

Table 4.3 presents a summary of the predictability scores for the remaining items. As it shows, the mean predictability scores were very low in all conditions.

Table 4.3 Predictability of nouns in different conditions

Condition	Mean predictability (range)
Adjacent collocations	0.04 (0–0.2)
Adjacent controls	0
Non-adjacent collocations	0.07 (0–0.4)
Non-adjacent controls	0.01 (0–0.2)

The summary of the average naturalness scores of the target items is presented in Table 4.4. As the rating of five on the Likert scale was defined as 'very natural', the averages of above four show that sentences in all the conditions were rated as being natural.

Table 4.4 Naturalness of sentences in different conditions

Condition	Mean naturalness (range)
Adjacent collocations	4.46 (3.4–5.0)
Adjacent controls	4.08 (3.2–4.8)
Non-adjacent collocations	4.42 (3.4–5.0)
Non-adjacent controls	4.11 (3.0–5.0)

For the experimental purposes, collocation conditions should have been perfectly matched with their relative control conditions, both in terms of predictability and in terms of naturalness. To check for the differences between the conditions, the following statistical analysis was carried out.

A Shapiro-Wilk test was used to check for normality of the distributions of the predictability scores and naturalness ratings and none of the variables was normally distributed (for the predictability scores: $W = 0.37$, $p = .00$; for the naturalness ratings: $W = .95$, $p = .00$). Non-normal distributions were expected, though, as I was aiming for non-predictable and natural sentences only.

A Kruskal-Wallis test showed a significant difference between the naturalness scores in the four experimental conditions ($\chi^2 = 13.90$, $p = .00$). To further investigate which conditions were different, a pairwise comparison using a post-hoc Pairwise Wilcoxon Rank Sum Tests with Bonferroni correction was carried out. The post-hoc test showed a significant difference (and a moderate effect size) between the adjacent collocations and the adjacent controls ($p = .01$, $r = .36$) and between the adjacent collocations and the non-adjacent controls ($p = .03$, $r = .31$). Thus, the non-adjacent collocations were matched with the non-adjacent controls in terms of the sentences naturalness, but the sentences with the adjacent collocations were more natural than the sentences with the adjacent controls. However, the mean difference of the naturalness ratings for the adjacent collocations and the adjacent controls was only 0.38. As the naturalness rating scale was from 1 to 5, the difference of 0.38 does not seem to be too worrying. It rather reflects the nature of collocations: they sound more natural to NSs than non-collocations.

In terms of predictability, a Kruskal-Wallis test also showed a significant difference between the scores for the four sentence groups ($\chi^2 = 13.52$, $p = .00$). To further investigate which groups were different, a pairwise comparison using the Pairwise Wilcoxon Rank Sum Tests with Bonferroni correction was carried out. It turned out that there was a significant difference between the adjacent collocations and the adjacent controls ($p = .04$, $r = .31$) and also between the non-adjacent collocations and the adjacent controls ($p = .01$, $r = .32$). Hence the adjacent collocations were again more predictable than the adjacent controls. The mean difference in this case was even smaller (.04).

In sum, this norming study allowed me to exclude items that were more predictable or less natural than expected. However, even if all items were judged to be natural and non-predictable, the adjacent collocations were still judged to be more natural and turned out to be more predictable than the control phrases. Therefore, the naturalness and predictability scores of the items were included in the statistical models to account for any differences in different experimental conditions.

4.2.3 Final set of items

After the target collocations were selected, a further analysis of the items was carried out to check if the sentences with collocations and control phrases remained matched well enough after excluding some of the items during the piloting stage. Table 4.5 presents the descriptive statistics of all the collocations and control phrases.

Table 4.5 Descriptive statistics for final set of items

Measure	**Mean**	***SD***	**Range**
MI (collocations)	4.10	0.98	3.02–6.28
MI (controls)	-0.94	1.13	-3.21–0.58
Form frequency (collocations)	138.40	169.00	14–922
Form frequency (controls)	6.23	6.71	0–27
Lemma frequency (collocations)	366.80	491.33	59–2,653
Lemma frequency (controls)	16.20	19.36	0–66
Verb length (collocations)	5.98	1.35	4–9
Verb length (controls)	6.03	1.25	4–9
Verb form frequency (collocations)	6,782.00	4,662.52	1,318–21,980
Verb form frequency (controls)	6,606.00	20,362.20	1,164–54,141
Verb lemma frequency (collocations)	18,220.00	12,012.89	5,222–49,570
Verb lemma frequency (controls)	22,358.00	11,899.40	5,447–86,579

A Shapiro-Wilk test was used to check for normality of the distributions of each variable. The only normally distributed variable as was the MI score for the control items ($W = .94$; $p = .07$). Because of the non-normal distributions, non-parametrical tests were used to check if the target collocations were matched with the control items. An Exact Wilcoxon Rank Sum test was used because this test can cope with tied observations, which were present in the dataset. The analysis was carried out using an R package *exactRankTests* (Hothorn & Hornik, 2013).

The comparisons showed no significant difference for length of verbs in collocation and control condition ($W = 800$, $p = 1$), frequency of verb forms in collocation and control conditions ($W = 752$, $p = .65$), and frequency of verb lemmas in both conditions ($W = 752$, $p = .65$). Therefore, it can be concluded that the verbs in both collocations and control conditions were matched well for their frequency and length.

As for the phrase level comparisons, the collocations were expected to be significantly different from the control phrases in terms of their MI scores and frequencies. This was confirmed by the statistical analysis. An Exact Wilcoxon Rank Sum tests showed that the MI scores for the collocations were significantly higher than those for the control phrase ($W = 1200$, $p = .00$). The same was true for the collocation form frequencies ($W = 1596.5$, $p = .00$) and the lemma frequencies ($W = 1594.5$, $p = .00$).

This statistical analysis confirmed that the target items are well-matched in terms of the verbs used but significantly different in terms of collocation status. So, all the remaining sentences were used for the experiment.

Forty five filler items were also written for the main study. Five of them were presented in the practice trials and another 40 were randomly mixed with the target items for the actual experiment. The aim of using the fillers was to make sure that the participants would not figure out the purpose of the study. The fillers were all natural sounding sentences in English. Care was taken to make sure that the words in these items were from the 3,000 most frequent lemmas list (Kilgarriff, n.d.). Fillers were checked by a native speaker of English to make sure they all sound natural. Thirty of these fillers were followed by *Yes/No* comprehension questions, to make sure the participants paid attention while reading. The filler sentences and questions are presented in Appendix 6.

4.2.4 Eye-tracking set-up

An eye-tracking experiment was set up using the *Experiment Builder* software and an *Eye-Link 1000 Plus* eye-tracker. The experiment consisted of three blocks. The first block contained the instructions and five practice items to allow the participants familiarise themselves with the task. Blocks 2 and 3 both consisted of 20 target sentences and 20 filler sentences, presented in a randomized order for each participant. The target items were divided into four lists using the Latin square design and counter-balancing for the length and the frequency of the items in each list. Each participant was randomly assigned to read one of the lists.

A 9-point calibration was performed before each block of items. For each trial, the participants first saw a dot in the upper left corner of the screen for a drift correction; then a sentence appeared in one line in the middle of the screen. The participants were instructed to read the sentence at their own pace, while their eye-movements being recorded. After reading the sentence they pressed the *Spacebar* and either a comprehension question or a prompt *Ready?*

appeared on the screen. They had to answer the question to move to the next sentence.

The areas of interest for the analysis were pre-set beforehand. For each sentence, two areas of interest were chosen: the final word of the collocation (the same noun in each of the four conditions) and the whole sequence (collocation or collocation with insertions depending on the condition), as illustrated in Figure 4.4. In this sentence, the first area of interest was the word *information*, while the second area of interest was the whole phrase: *provide some of the information*.

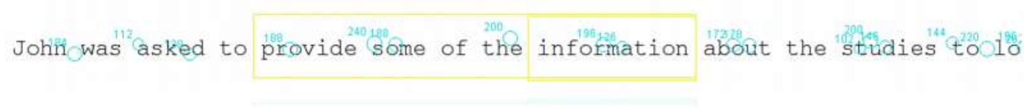


Figure 4.4 Areas of interest

4.2.5 Participants

Initially, 34 participants took part in the study. However, some of them had to be excluded from the analysis for the reasons discussed below. Twenty eight NSs of English (23 females and 5 males) remained for the analysis. They were all studying for a degree in English at the University of Nottingham. Their mean age was 18.89 ($SD = 0.96$). They all reported having normal or corrected to normal vision

The students volunteered to participate in a session in Psycholinguistics lab during their modules undertaken as a part of their degree. Each of them was tested individually. They were first told about the procedure of the study and asked to sign an informed consent form (see Appendix 3 for the Participant information sheet[4]). They were instructed to read the sentences for comprehension and answer the comprehension questions.

The study took about 20–25 minutes to complete. Afterwards the participants were asked if they thought they could guess what the aim of the

[4] All the ethics approval forms and signed consent forms for this and the following studies are available upon request.

study was. None of them reported to have understood the correct purpose of the study.

4.2.6 Data cleaning

To start with, the answers to the comprehension questions were checked for all the participants. The threshold of adequate comprehension was set at answering 80% of the questions correctly, as the questions seemed to be very easy. One participant was excluded from the further data analysis because of not reaching the expected comprehension level (she/he answered only 20 out of 27 questions correctly), as it was assumed she/he did not pay enough attention to the task. The average comprehensions score of the remaining participants was 95%.

Five more participants were excluded from the analysis because of the technical problems with calibration: the data of their recordings were too messy to interpret confidently. The data of the remaining participants were checked manually and when needed a drift correction was performed manually in the eye-movement recordings. Then the recordings were cleaned using the standard procedure of the automatic cleaning function of the *Eyelink DataViewer*. The procedure follows four stages:

1. fixations shorter than 80ms are merged to the nearest fixations if they are less than 0.5 degrees away from each other;
2. the same procedure is repeated for fixations shorter than 40 ms at less than 1.25 degrees of distance;
3. if at least 3 fixations in the area are shorter than 140 ms and there are no longer fixations, they are merged together;
4. fixations shorter than 100 ms or longer than 800 ms are deleted.

This cleaning led to the loss of 2.97% of the data (calculated as the percentage of the total reading time for the areas of interest). Then interest area reports were obtained for each participant. Following the suggestions of Carrol and Conklin (2014), two areas of interest were analysed: the final word and the whole phrase. As there is no one measure of eye-movements that would be the most informative, it seems to be an established practice to analyse a number of eye-movement measures (see, for example, Siyanova-Chanturia, Conklin, & van Heuven, 2011). This approach was adopted in the current study as well.

Figure 4.5 shows a hypothetical fixation path (in order to explain each measure of eye movements analysed). For the final word area of interest, four different measures (two early measures and two late measures) were exported for further analysis: first fixation duration (fixation 4 in Figure 4.5), gaze duration (fixation 4), fixation count (3), and total reading time (fixations 4 + 6 + 7).

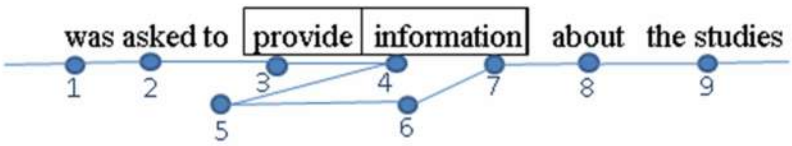


Figure 4.5. Hypothetical fixation path

For the whole phrase area of interest, three eye-movement measures were chosen: first pass reading time (fixations 3 + 4), fixation count (4), and total reading time (fixations 3 + 4 + 6 + 7).

4.2.7 Data analysis

Firstly, the data were entered into an *Excel* spreadsheet. Skipped items were excluded from the final word analysis. Then all the dependent variables (except for the fixation counts) were log-transformed for further analysis in order to reduce the skewedness of their distributions.

The four study conditions were coded as two two-level categorical variables (collocation status and adjacency), specifying custom contrasts to -0.5 and 0.5. This coding was chosen to make sure that the models with interactions evaluate the main effects of collocation status and adjacency across all levels of factors.

Word and phrase frequencies were log-transformed. All the frequencies used were lemma frequencies. The phrase frequencies were increased by one before the transformation, as log transformation is impossible for zero values and some control phrases in the dataset did not occur in the BNC. As no corpus provides all examples of potential language use and one occurrence in a large corpus makes a very small difference. So increasing all the frequencies by one solved a technical problem and did not seem to introduce any theoretical problems.

MI scores are already log-transformed, so no further transformations were undertaken. However, in cases when the phrase did not occur in the corpus and hence the MI score could not be estimated, zero values was inserted as its MI score.

Finally, all the continuous predictor variables were centered in order to avoid multicollinearity in the models (Siyanova-Chanturia, Conklin, & van Heuven, 2011). Table 4.6 presents the range of each variable before the transformations and the median and the standard deviation after the transformation (as the data were centered, the means were always zero; therefore the medians are reported instead).

Table 4.6 Description of the covariates

Variable	Range	Median	*SD*
Age	18–21	0.14	0.95
Trial number	6–85	0.64	22.92
Noun frequency	4,398–50,109	-0.01	0.26
Noun length	4–12	-0.38	2.09
Verb frequency	5,222–86,579	0.02	0.30
Verb length	4–9	0	1.29
Phrase frequency	59–2,653	0.10	0.80
Phrase length	9–40	-2.81	8.23
Insertion length	0–21	-2.94	7.40
MI	-3.21–6.28	0.10	2.61
Predictability	0–0.4	-0.03	0.08
Naturalness	3.2–5	0.13	0.47

The data were analysed using the R-Studio software (R Core Team, 2013), version 3.1.3. Linear mixed-effects models were fitted using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2014, p. 4), version 1.1-7. A linear mixed-effects model analysis was chosen for a number of reasons. Firstly, this analysis has numerous advantages over the traditional ANOVA analysis as it does not make any assumptions of homoscedasticity or sphericity of the data and it is robust against missing data (at least assuming that the data are missing at random). Also, mixed-effects models allows testing effects and interactions of both discrete and continuous predictors in the same model

(Cunnings, 2012; Quené & van den Bergh, 2008). Furthermore, this type of analysis allows including both by-subject and by-item variation in the same model, instead of calculating separate by-subject and by-item *F* scores (Quené & van den Bergh, 2008). Also, these models allow accounting for individual differences, as well as for certain experimental effects such as learning or fatigue (Baayen, Davidson, & Bates, 2008). Importantly, including random slopes and intercepts for subjects and items, allows generalizing results of an experiment to other subjects and items in the population.

To check if any further transformations of the predictors was required to avoid potential multicollinearity, *collin.fnc()* function was used. Kappa values were estimated for all the potential predictors to be used in the models. Before centering the data, it was problematically high ($K = 118.91$), but after centering all the predictors, the kappa value was low and showed no multicollinearity ($K = 5.86$). Therefore, no further transformation was undertaken.

Seven separate models (one for each eye-tracking measure) were fitted. Model fitting always started from a core model, where the outcome variable was predicted by the collocation status and adjacency of the phrase and also the interaction between those two variables. The core model also included the predictability score, to make sure that any effect of collocation status in the models is beyond the effect of predictability. These main predictors were retained in the models regardless of their significance, as they were important in order to answer the research questions.

Then all the potential covariates (see Table 4.6) were added to the model. For the final word reading models, the covariates included trial number of the sentence, list number (one of the 4 experimental lists that the participant was assigned to), noun frequency, noun length, phrase frequency, MI score, insertion length (in letters), sentence naturalness score and participant's age, and gender. For the whole phrase reading models, the covariates included the same predictors as for the final word, together with verb frequency, verb length, and phrase length as potential covariates. Starting from this full model, a backwards step-by-step model selection procedure was applied: the covariates which did not reach *t* values of two were dropped one by one, starting from the ones with the lowest *t* scores. Each new model was then explicitly compared to the previous one using likelihood ratio tests, to check if dropping the covariate does not significantly change the model. Significance level chosen for all the analysis was $p < .05$.

The models included random intercepts for subjects and items and random slopes for collocation status and adjacency. The models with the significant interaction between these two predictors also included a random slope for the interaction. The structure of random effects was kept maximal, following the suggestions of (Barr, Levy, Scheepers, & Tily, 2013), who claimed that a failure to include random effects in the model drastically increases the chances of Type 1 error. The inclusion of the random by-subject effects also allowed statistically accounting for the individual differences between different readers, reported in previous research (Rayner, 1998).

If there was an interaction in the model, it was further analysed graphically (plotted using the *effects* package (Fox, 2009)) and statistically (using the *phia* package (Rosario-Martinez, 2015)).

To estimate *p*-values for the coefficients in the models, the package lmerTest (version 2.0–20; Kuznetsova, Brockhoff, & Bojesen Christensen, 2015) was used.

4.3 Results and interpretation

This section presents the results for the final word reading models first, followed by the models for the whole phrase reading. The models will be briefly discussed, as far as the significant predictors are concerned, leaving the Discussion section (4.4) for the more general discussion of the findings of the study.

4.3.1 Final word: first fixation duration

First fixation duration is defined as “the duration of the first fixation within the area of interest irrespective of whether it is the only fixation or the first of multiple fixations within this region” (Roberts & Siyanova-Chanturia, 2013, p. 219). The first fixation is an early measure of processing indicating early processes of lexical access.

As in all four conditions the final nouns were the same, the reading times between those conditions can be directly compared. The boxplots of the distributions of first fixation duration are presented in Figure 4.6.

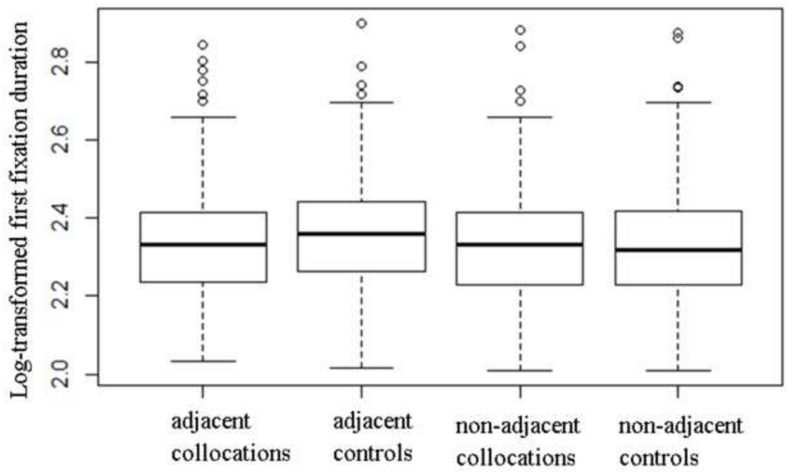


Figure 4.6 First fixation duration. NS data

As Figure 4.6 shows, there seems to be a small difference between first fixation duration for the adjacent collocations and controls, but no difference for the non-adjacent phrases. Table 4.7 presents the means of first fixation duration across the four experimental conditions (the means of non-transformed data are presented for the ease of interpretation). The mean differences seem to confirm the observation that the difference between the adjacent conditions was small and between the non-adjacent conditions was negligible.

Table 4.7 Mean first fixation durations. NS data

Group	Mean first fixation duration (ms)	*SD*	Mean difference (ms)
Adjacent collocations	232.83	92.60	12.71
Adjacent controls	245.54	94.93	
Non-adjacent collocations	226.27	87.70	3.07
Non-adjacent controls	229.34	89.73	

Table 4.8 presents the model of first fixation duration data[5].

Table 4.8 Model of first fixation duration. NS data

	First fixation duration			
Fixed effects	***β***	***SE***	***t***	***p***
Intercept	2.340	0.012	203.00	***
Collocation status	0.005	0.010	0.52	.61
Adjacency	-0.096	0.028	-3.46	***
Noun length	0.007	0.003	2.42	*
Insertion length	0.006	0.002	3.17	**
Predictability	-0.222	0.059	-3.73	***
Collocation*Adjacency	-0.023	0.017	-1.36	.17
Random effects	**Variance**	***SD***		
Item	0.001	0.029		
Collocation | Item	0.000	0.012		
Adjacency | Item	0.000	0.009		
Subject	0.003	0.051		
Collocation | Subject	0.000	0.014		
Adjacency | Subject	0.000	0.021		
Residual	0.018	0.132		

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

As Table 4.8 shows, collocation status did not come out as a significant predictor in the model. Adjacency, on the other hand, was a significant predictor, with the nouns in the non-adjacent conditions fixated shorter compared to the nouns in the adjacent conditions. The fact that the nouns in the non-adjacent phrases were read faster might be due to several reasons. Possibly, after reading a verb and then the modifiers following it, the readers built up some kind of expectations for the noun to follow. Also, as the inserted words very often were shorter (and at least some of them were function words), the readers might have had a chance to benefit from the preview of the noun in the non-adjacent conditions. The fact that the insertion length had an opposite effect than the adjacency (the longer the insertion, the longer the fixation on the noun) was probably due to the same reason. Longer insertions included longer

[5] When interpreting this and the following models, a negative β coefficient shows decreasing reading times, while a positive β indicates increasing reading times. Because of the coding scheme used, the coefficient for collocation status shows the direction for the control phrases. The coefficient for adjacency shows the direction for the non-adjacent phrases.

content words to process. Thus the readers had more information to integrate and potentially had less preview benefit.

The model also shows that the longer the word (noun length), the longer it was fixated; and the more predictable it was, the faster it was processed. These findings are fully in line with the previous research showing that word length and predictability affects processing (Rayner, 1998). It is interesting to note that predictability had a very strong effect (its coefficient is the largest in the model and clearly significant). So it seems that predictability is very important for the first fixation duration, that is, for early lexical access.

4.3.2 Final word: gaze duration

Gaze duration is a "sum of all fixation durations made within a region of interest until the gaze exited either to the left or to the right" (Roberts & Siyanova-Chanturia, 2013, p. 219). Gaze duration, as well as first fixation duration, is an early processing measure. Therefore, the patterns of the effects are usually similar for both first fixation duration and gaze duration, only the effects tend to be larger for gaze duration (Rayner & Liversedge, 2011).

In this final word reading analysis, the area of interest consisted of only one word. Therefore, gaze duration time was different from first fixation time only in the cases when the final word was fixated more than once during the first pass reading (before the eyes leave the area of interest). In this experiment, final words in the phrases were fixated more than once during the first pass reading in 27% of the trials.

The boxplots in Figure 4.7 illustrate the distributions of the gaze duration on the final word.

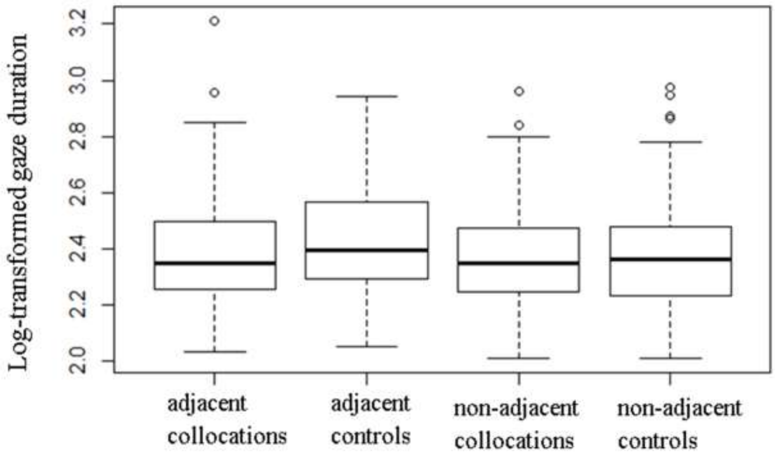


Figure 4.7 Gaze duration. NS data

As Figure 4.7 shows, the adjacent collocations were read faster than the control phrases, but this collocation effect is unclear fort the non-adjacent phrases. The mean differences between the groups are presented in Table 4.9.

Table 4.9 Mean gaze duration times. NS data

Group	Mean gaze duration (ms)	*SD*	Mean difference (ms)
Adjacent collocations	270.83	154.11	21.18
Adjacent controls	292.02	133.54	
Non-adjacent collocations	253.65	114.78	5.91
Non-adjacent controls	259.55	123.87	

From Table 4.9 it becomes clear that the differences between mean gaze duration times for the collocations and the control phrases were larger for gaze duration than they were for first fixation duration. For the non-adjacent phrases the mean difference remained very small, though. The statistical analysis (see Table 4.10) largely confirms this observation.

Table 4.10 Model of gaze duration. NS data

	Gaze duration			
Fixed effects	***β***	***SE***	***t***	***p***
Intercept	2.385	0.013	182.52	***
Collocation status	0.017	0.013	1.30	.20
Adjacency	-0.131	0.036	-3.67	***
Noun length	0.017	0.004	4.27	***
Insertion length	0.007	0.002	3.04	**
Predictability	-0.182	0.075	-2.44	*
Collocation * Adjacency	-0.038	0.021	-1.78	.08
Random effects	**Variance**	***SD***		
Item	0.002	0.040		
Collocation | Item	0.001	0.031		
Adjacency | Item	0.000	0.019		
Collocation * Adjacency | Item	0.001	0.027		
Subject	0.003	0.054		
Collocation | Subject	0.001	0.026		
Adjacency | Subject	0.001	0.033		
Collocation * Adjacency | Subject	0.001	0.027		
Residual	0.026	0.161		

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

Table 4.10 shows that the model of gaze duration time includes exactly the same predictors as the model of first fixation duration. This is not a surprising result, as both of these measures are tapping into the lexical access of the word.

The only difference between the models is the fact that the interaction between collocation status and adjacency in this model almost reached significance ($p = .08$). Even if it was not significant, it was analysed further to see the trend of how collocation status and adjacency interact in the dataset. To visualize the interaction, it was plotted using the *effects* package (see Figure 4.8)

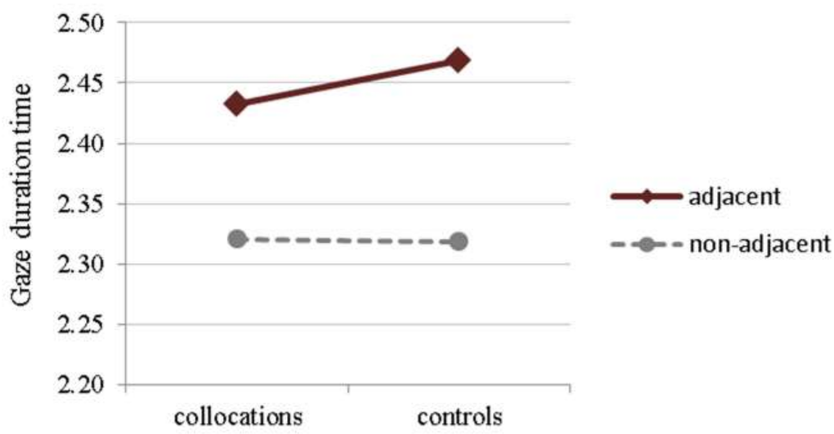


Figure 4.8 Interaction between collocation status and adjacency in the gaze duration model. NSs data

Figure 4.8 suggests a difference between the reading behaviour for the adjacent and non-adjacent phrases. It seems that the collocations were read faster in the adjacent condition. In the non-adjacent condition, there was virtually no difference between the collocations and the control phrases. To go beyond a visual comparison, pairwise comparisons of different groups were carried out. They showed that for gaze duration the effect of collocation status did not reach significance for the adjacent collocations (χ^2 (1, 1013) = 4.29, p = .08); and it was clearly not significant for the non-adjacent phrases (χ^2 (1, 1013) = .02, p = .88).

Even if the interaction between collocation status and adjacency did not reach significance in this dataset, the data seem to suggest that there might be certain differences between the processing of adjacent and non-adjacent collocations even in early stages of processing.

4.3.3 Final word: fixation count

Fixation count is different from all the other measures, as it does not measure reading *time*, but rather a *number* of fixations in the area of interest: "[f]ixation count captures the number of all fixations made within a given region of interest, a single word, or a longer stretch of language" (Roberts & Siyanova-Chanturia, 2013, p. 219). As it is a count variable, truncated at zero, a Poisson distribution was used and a generalized linear mixed model was fitted.

Figure 4.9 presents the boxplots for the distributions of the fixation count. There seems to be a clear difference between the collocations and the control phrases in the adjacent conditions but not in the non-adjacent conditions.

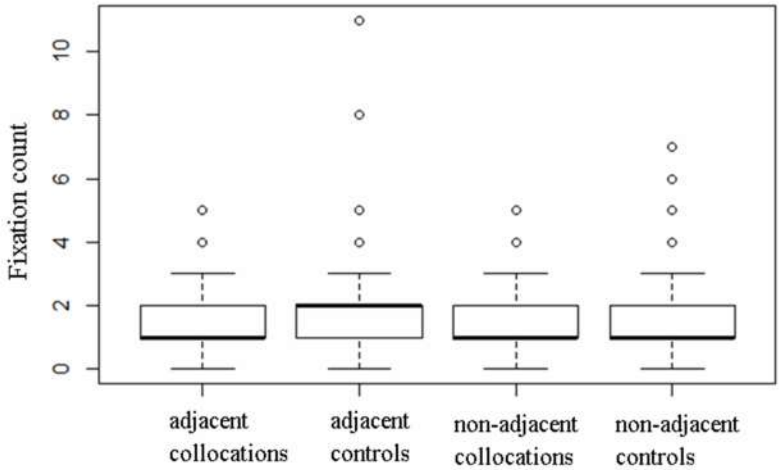


Figure 4.9 Fixation counts of the final word. NS data

The means of fixation counts for all four experimental conditions are presented in Table 4.11.

Table 4.11 Mean fixation counts of the final word. NS data

Group	Mean fixation count	*SD*	Mean difference
Adjacent collocations	1.38	0.85	0.40
Adjacent controls	1.78	1.16	
Non-adjacent collocations	1.26	0.85	0.24
Non-adjacent controls	1.50	1.06	

The mean differences also suggest a larger difference between the collocations and the control phrases in the adjacent condition. The statistical analysis of the data is presented in Table 4.12.

Table 4.12 Model of fixation count of the final word. NS data

	Fixation count			
Fixed effects	***β***	***SE***	***z***	***p***
Intercept	0.351	0.045	7.71	***
Collocation status	0.196	0.054	3.60	***
Adjacency	-0.486	0.149	-3.27	**
Noun length	0.071	0.012	6.13	***
Insertion length	0.025	0.010	2.56	*
Predictability	0.032	0.345	0.09	.93
Collocation * Adjacency	-0.082	0.099	-0.83	.41
Random effects	**Variance**	***SD***		
Item	0.000	0.000		
Collocation | Item	0.000	0.000		
Adjacency | Item	0.000	0.000		
Subject	0.040	0.199		
Collocation | Subject	0.005	0.073		
Adjacency | Subject	0.000	0.019		

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

Table 4.12 shows that both collocation status and adjacency came out as significant predictors, but the interaction between those two variables was not significant. This suggests that the final words in the non-adjacent phrases were fixated fewer times in general, no matter the collocation status. However, at the same time, the words in the collocations were fixated fewer times than the words in the control phrases despite their adjacency.

Noun length had an effect on fixation count with longer words fixated more times and insertion length had the same effect: longer insertions led to more fixations. Interestingly, predictability, which was a significant predictor for the early measures of eye-movements, did not come out as a significant predictor in the fixation count model. As fixation count takes into account re-reading behaviour, it does not show early processing anymore. Therefore, it seems that predictability is more important for early than for late processing.

4.3.4 Final word: total reading time

Total reading time is defined as "all fixations that landed on the target and indicates how much time the participant spent reading the target" (Roberts & Siyanova-Chanturia, 2013, p. 219). It is a measure of late processes as it includes re-reading of the area of interest. In the present study, the participants re-read the area of interest in 33% of the trials.

The visual analysis of the data (Figure 4.10) seems to suggest that there was a difference in total reading time between the collocations and the control phrases for both the adjacent phrases and the non-adjacent ones.

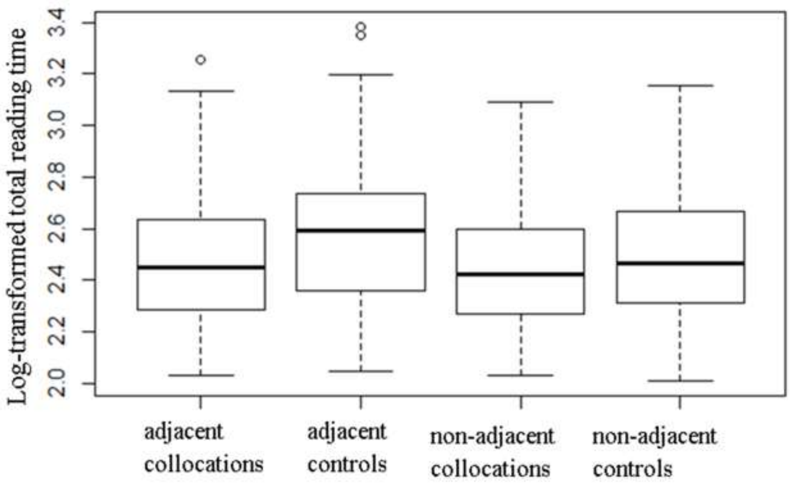


Figure 4.10 Total reading time of the final word. NS data

The mean total reading times for each condition (presented in Table 4.13) show that the mean difference was larger for the adjacent than for the non-adjacent collocations. The model of total reading time is presented in Table 4.14.

Table 4.13 Mean total reading times of the final word. NS data

Group	Mean total reading time (ms)	*SD*	Mean difference (ms)
Adjacent collocations	315.79	229.03	92.55
Adjacent controls	408.34	291.63	
Non-adjacent collocations	275.15	203.76	58.84
Non-adjacent controls	333.99	251.99	

Table 4.14 Model of total reading time of the final word. NS data

	Total reading time			
Fixed effects	***β***	***SE***	***t***	***p***
Intercept	2.489	0.019	132.99	***
Collocation status	0.067	0.017	3.96	***
Adjacency	-0.186	0.052	-3.59	***
Noun length	0.025	0.004	6.01	***
Insertion length	0.011	0.003	3.02	**
Predictability	-0.182	0.097	-1.88	.06
Collocation * Adjacency	-0.055	0.027	-2.02	*
Random effects	**Variance**	***SD***		
Item	0.002	0.041		
Collocation | Item	0.003	0.051		
Adjacency | Item	0.003	0.050		
Collocation * Adjacency | Item	0.003	0.055		
Subject	0.007	0.086		
Collocation | Subject	0.001	0.031		
Adjacency | Subject	0.002	0.044		
Collocation * Adjacency | Subject	0.000	0.021		
Residual	0.041	0.203		

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

Table 4.14 shows that collocation status, adjacency, and their interaction were significant predictors of the total reading time of the final word. Noun length and insertion length were also significant predictors, with reading time increasing when noun length or insertion length increased (as in the previously discussed models). Predictability, as in the fixation count model, was not a significant predictor.

When the interaction between collocation status and adjacency was investigated closer, it confirmed the prediction from the visual inspection of the data: facilitative effect was larger for the adjacent collocations compared to the non-adjacent ones (see Figure 4.11).

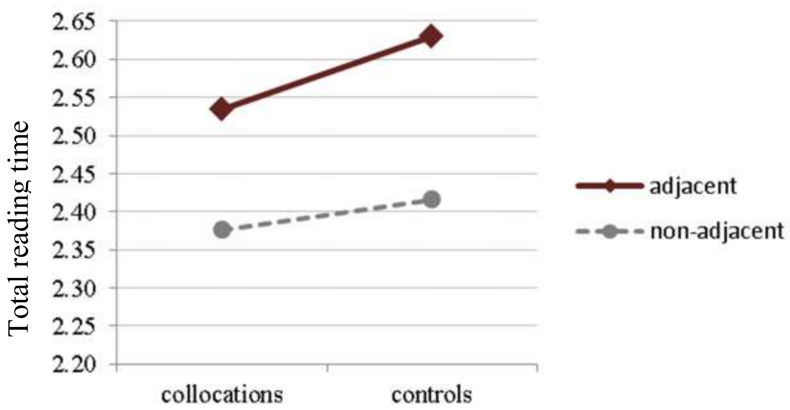


Figure 4.11 Interaction between collocation status and adjacency in the final word total reading time model. NS data

Pairwise comparisons revealed that there was a significant facilitation for the adjacent collocations compared to the adjacent controls: χ^2 (1, 1013) = 17.72, p = .00. For the non-adjacent phrases, on the other hand, the difference between the collocations and the control phrases only approached significance χ^2 (1, 1013) = 3.45, p = .06. This result, taken together with the results of the previous models, suggests that the facilitative effect for collocations is larger for the adjacent phrases than for the non-adjacent ones.

4.3.5 Whole phrase: first pass reading time

First pass reading time for a phrase is an equivalent measure to gaze duration for one word. As a phrase is longer than one word, analysing first fixation duration is not meaningful: it would simply correspond to the first fixation duration on the first word of the phrase and it would not provide much information about the whole phrase reading.

Figure 4.12 illustrates the distributions of the first pass reading time for the phrases in the four experimental conditions. In this case (as for the other whole phrase reading measures as well), the adjacent and non-adjacent conditions cannot be directly compared, as they differ considerably in length.

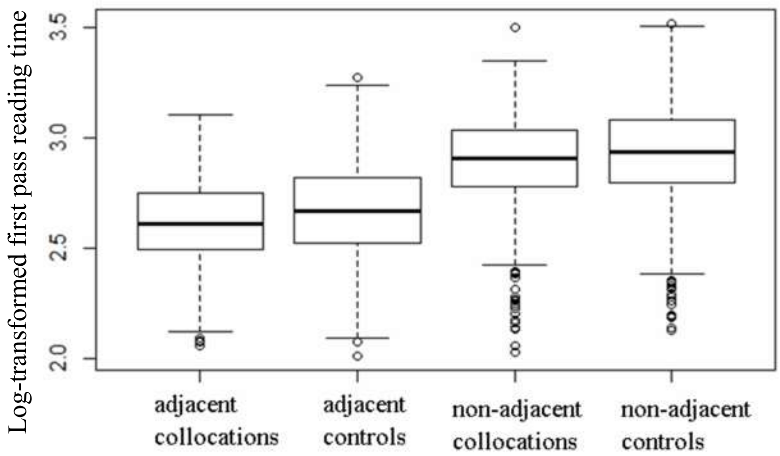


Figure 4.12 First pass reading time. NS data

From the graph and from Table 4.15 it seems that there was a difference between the collocations and the control phrases and this difference was very similar both for the adjacent and for the non-adjacent phrases.

Table 4.15 Mean first pass reading times. NS data

Group	Mean first pass reading time (ms)	*SD*	Mean difference (ms)
Adjacent collocations	444.01	196.39	77.23
Adjacent controls	521.24	283.57	
Non-adjacent collocations	852.41	424.75	76.05
Non-adjacent controls	928.46	493.56	

The model of first pass reading time for the whole phrase area of interest is presented in Table 4.16.

Table 4.16 Model of first pass reading time. NS data

	First pass reading time			
Fixed effects	***β***	***SE***	***t***	***p***
Intercept	2.756	0.018	155.89	***
Collocation status	0.040	0.016	2.48	*
Adjacency	-0.077	0.056	-1.36	.18
Noun length	-0.013	0.006	-2.24	*
Predictability	-0.002	0.097	-0.02	.98
Phrase length	0.022	0.003	6.18	***
Collocation * Adjacency	-0.016	0.026	-0.64	.52
Random effects	**Variance**	***SD***		
Item	0.002	0.050		
Collocation | Item	0.003	0.051		
Adjacency | Item	0.004	0.061		
Subject	0.006	0.077		
Collocation | Subject	0.000	0.014		
Adjacency | Subject	0.000	0.018		
Residual	0.046	0.214		

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

In this model of first pass reading time, collocation status came out as a significant predictor and the collocations were read faster than the control phrases. Adjacency, on the other hand, was not a significant predictor. There was also no significant interaction between collocation status and adjacency, suggesting that the collocations were read faster than the control phrases overall.

Noun length and phrase length were significant predictors as in the final word reading models, making this model for the whole phrase reading very similar to the models for the final word reading. The fact that predictability did not come out as a significant predictor is not surprising, as predictability was estimated for the final word and not for the entire phrase.

4.3.6 Whole phrase: fixation count

Fixation count for a word sequence is exactly the same measure as fixation count for one word. Figure 4.13 illustrates the distributions of the fixation counts. It shows that for the whole phrase reading, there seem to be overall differences between the fixation counts on the collocations and the control phrases.

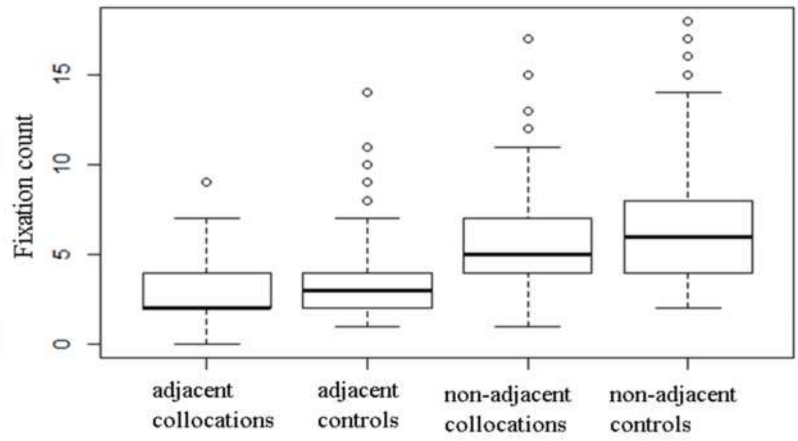


Figure 4.13 Fixation counts of the whole phrase. NS data

Table 4.17 presents the mean fixation counts for the experimental conditions. The differences in this case are even larger for the non-adjacent phrases, but this might be due to their length. As the areas of interest of the non-adjacent phrases comprised the inserted words, they were longer than for the adjacent conditions, and consequently they were fixated more times.

Table 4.17 Mean fixation counts of the whole phrase. NS data

Group	Mean fixation count	*SD*	Mean difference
Adjacent collocations	2.76	1.30	.69
Adjacent controls	3.45	1.75	
Non-adjacent collocations	5.40	2.55	.80
Non-adjacent controls	6.2	3.00	

The model of fixation counts for the whole phrase is presented in Table 4.18. This model is a generalized linear mixed model fit using a Poisson distribution.

Table 4.18 Model of fixation count of the whole phrase. NS data

	Fixation count			
Fixed effects	***β***	***SE***	*z*	***p***
Intercept	1.410	0.041	34.33	***
Collocation status	0.177	0.033	5.38	***
Adjacency	-0.050	0.111	-0.45	.65
Noun length	-0.028	0.012	-2.38	*
Predictability	0.167	0.208	0.81	.42
Phrase length	0.044	0.007	6.35	***
Collocation * Adjacency	-0.086	0.060	-1.43	.15
Random effects	**Variance**	***SD***		
Item	0.006	0.079		
Collocation | Item	0.002	0.041		
Adjacency | Item	0.011	0.104		
Subject	0.036	0.190		
Collocation | Subject	0.000	0.009		
Adjacency | Subject	0.001	0.030		

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

The model reported in Table 4.18 shows that collocation status and length of the words in the phrase were significant predictors for the fixation count. Adjacency of the phrases did not affect the fixation count and there was no interaction between collocation status and adjacency. Overall, the phrases containing collocations were consistently fixated fewer times than the control phrases.

4.3.7 Whole phrase: total reading time

Total reading time for a phrase is an equivalent measure as total reading time for one word: it is a sum of the durations of all the fixations in the area of interest. The boxplots in Figure 4.14 present the distributions of total reading time across the conditions.

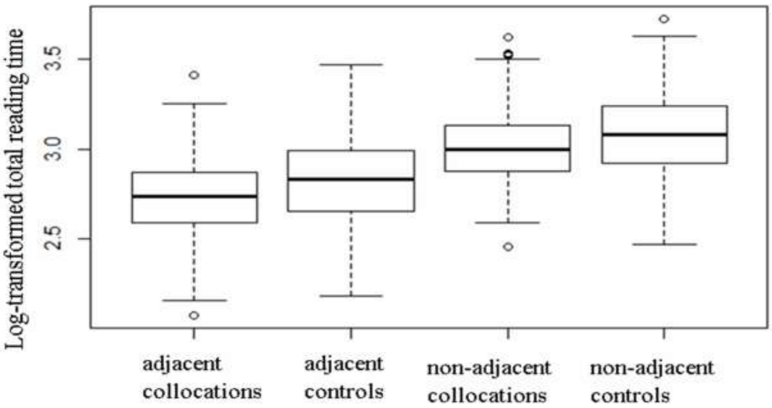


Figure 4.14 Total reading time of the whole phrase. NS data

Figure 4.14 shows that the collocations were read faster than the control phrases across the conditions. This observation is supported by the mean differences of non-transformed total reading times, presented in Table 4.19.

Table 4.19 Mean total reading times of the whole phrase. NS data

Group	Mean total reading time (ms)	*SD*	Mean difference (ms)
Adjacent collocations	611.91	326.45	165.53
Adjacent controls	777.44	450.17	
Non-adjacent collocations	1167.7	632.06	153.60
Non-adjacent controls	1389.06	785.66	

The final model of total reading time of the whole phrase is summarized in Table 4.20.

Table 4.20 Model of total reading time of the whole phrase. NS data

	Total reading time			
Fixed effects	***β***	***SE***	***t***	***p***
Intercept	2.915	0.022	134.21	***
Collocation status	0.080	0.011	6.97	***
Adjacency	-0.052	0.047	-1.09	.28
Noun length	-0.014	0.006	-2.47	*
Predictability	-0.012	0.082	-0.14	.89
Phrase length	0.022	0.003	7.19	***
Collocation * Adjacency	-0.026	0.021	-1.22	.22
Random effects	**Variance**	***SD***		
Item	0.003	0.051		
Collocation | Item	0.000	0.010		
Adjacency | Item	0.003	0.058		
Subject	0.011	0.103		
Collocation | Subject	0.000	0.013		
Adjacency | Subject	0.000	0.021		
Residual	0.031	0.175		

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

The model shows that the total reading time was shorter for the collocations than for the control phrases. As adjacency did not come out as a significant predictor and neither did the interaction between collocation status and adjacency, it can be claimed that the collocations were read faster, no matter if the collocates were adjacent to each other. Noun length and phrase length, as expected from the previous models, came out as significant predictors as well.

4.3.8 Summary of the results

Table 4.21 summarizes the results of the present study showing the significance of the main predictors (collocation status, adjacency, and the interaction between them) in the models.

Table 4.21 Summary of the results. NS data

	Final word reading models			
Fixed effects	First fixation	Gaze duration	Fixation count	Total reading time
Collocation status	–	–	+	+
Adjacency	+	+	+	+
Interaction	–	almost	–	+
Effect for adjacent collocations		almost		+
Effect for non-adjacent collocations		–		almost
		Whole phrase reading models		
		First pass	Fixation count	Total reading time
Collocation status		+	+	+
Adjacency		–	–	–
Interaction		–	–	–

Note: Plus signs (+) indicate a significant effect, minus signs (–) show no significant effect. The word *almost* indicates that the effect was significant at .1 level.

Looking at Table 4.21, it becomes clear that collocation status came out as a significant predictor for the late measures of the final word reading, but not for the early measures, suggesting that collocation effect is more important for semantic and discourse level processing rather than early lexical access.

As for the phrase reading, collocation status consistently remained a significant predictor in all three models. This finding supports the conclusion that collocation status plays a role in phrase processing, no matter the adjacency of the collocation.

4.4 Discussions

4.4.1 Significant predictors in the models

Looking at the models for each of the eye-movement measures, it becomes clear that some of the variables (such as noun length, phrase length, collocation status, and adjacency) came out as significant predictors in most of the models, while some were always discarded. This was the case even if all the considered predictors were chosen based on the suggestions of the previous research. Therefore, it seems to be important to consider why some of these variables did not predict reading times in the present study.

To start with, some of the variables that were non-significant simply confirmed that an adequate experimental control was taken when designing the study. For example, the fact that the naturalness ratings did not came out as a significant predictor in any of the models, confirms that the efforts to make the sentences sound as natural as possible were successful. The fact that the list (a categorical variable coding four different lists of items) was also not a significant predictor suggests that items have been counter-balanced between the lists well. The fact that the trial number was non-significant also shows that the participants experienced no fatigue during the experiment or did not need time to learn the experimental procedure.

Some of the variables non-related to the experimental design but rather characteristic to the target phrases (e.g., noun frequency or verb frequency) were consistently non-significant in the models as well. This result has to be interpreted with caution, though. This finding by no means argues against the previous research suggesting the importance of a word's frequency for its processing. Rather, the experiment was not set up to test the effect of individual word frequency. Therefore, there was no wide range of frequencies of the target items: all words were taken from the 2,000 frequency level. Also, in case of noun frequency, the same nouns were used across all four conditions, which also reduced the variation in noun frequencies. It has also been suggested that word length accounts for much more variation in processing time than word frequency (Kliegl et al., 2004). This claim is in line with the findings of the present study, as the word length was always a significant predictor.

As far as the two variables associated with collocation status (MI and phrase frequency) were concerned, they did not come out as significant predictors either. This might be the case simply because they were both accounted for when the categorical variable of collocation status was introduced to the model.

Predictability score was significant in the models of early measures of final word reading, but not in the other models. This brings two conclusions. Firstly, predictability seems to be important for the early lexical access of the word (in line with Vainio et al., 2009) but not for its later integration. Secondly, collocation status seems to have an effect that goes beyond simple contextual predictability. Even if predictability scores were included in the models, collocation status remained a significant predictor for the late measures of the final word reading times and in all the models for the whole phrase. This finding argues against the suggestion that transitional probabilities do not play a role in processing once the contextual predictability is controlled for (Frisson et al., 2005). It seems that while the predictability is a significant predictor for the early lexical access, collocation status is more important for the discourse level integration and at the phrase level of processing.

4.4.2 Processing adjacent collocations

The first aim of the present study was to replicate the finding that adjacent collocations are processed faster than matched control phrases. The results of the present study are in line with the previous research showing a processing advantage for adjacent collocations (e.g., Durrant & Doherty, 2010; Sonbul, 2015). As the present study controlled for contextual predictability of the collocates, this study supports the findings of McDonald and Shillcock (2003a, 2003b), who suggested the effect of transitional probability (and hence collocation status) on word processing.

For the final word processing, collocation status was only a significant predictor for the late measures, which indicate integration, rather than lexical access. Possibly collocational knowledge becomes important when the phrase integration processes start.

The fact that collocation status was a significant predictor in all the phrase reading models clearly shows that collocations are read faster than matched control phrases. While no one measure can fully represent processing, the fact that all the measures point in the same direction suggest a robust effect of collocation status, especially for adjacent collocations. Collocations in this study were not semantically associated and their predictability was controlled for, suggesting that the processing advantage probably occurred because of the phrase frequency and mutual expectancy of the words in collocations. Control items in the present study, even if being semantically plausible, had very low phrase frequencies in the BNC, while the collocations by definition were frequently used phrases. As all of the participants in the study were NSs of British English, arguably they all should have had enough exposure to those collocations composed of high-frequency words and these phrases were entrenched in their memory facilitating their processing. This result is in line with usage-based theories that claim that our linguistic experience shapes our knowledge about the language (Barlow & Kemmer, 2000).

It is interesting to compare the facilitative effect for collocations found in Sonbul's (2015) study to the results of the present study, as these two studies used the same methodology. In her study, Sonbul only looked at the processing of whole phrases rather than individual words, so only the models for the whole phrase reading times can be compared. She found a processing advantage for adjective-noun collocations while the present study extended the same finding to verb-noun collocations, suggesting that the facilitative collocation effect holds for various types of word sequences. On the other hand, there are also some differences between the results of the studies. While in Sonbul's study, phrase frequency (used as an indication of collocation status) was a significant predictor only in the models of early measures of eye-movement, the present study showed collocation effects on the late measures for the whole phrase reading as well. It is important to note, though, that the present study only looked at two groups of phrases (collocations and controls), while in the previous study, frequency was used as a continuous variable. Also, the MI threshold for extracting collocations was much lower in Sonbul's study (MI > 1), which means that the collocations in the present study were stronger (MI ≥ 3). The frequency threshold in the present study (> 50) was the same as the cut-off for the high-frequency collocation group in Sonbul's study, again suggesting that the present study looked at stronger collocations. The fact that target collocations in the present study were stronger (especially when compared to the ones in Sonbul's mid-frequency condition), might have been the reason why the effect of collocation status was observable in late measures as well as in early ones in the present study.

While Sonbul's study did not show the effect of collocation frequency for late measures, the results of the current study are in line with the results of Siyanova et al.'s (2011) binomial study, which also found the phrase frequency and the phrase type (binomial or not) to be consistently significant in both early and late measures of eye-movements.

4.4.3 Processing non-adjacent collocations

The main finding of the present study, though, is that a processing advantage for collocations holds despite three words intervening between the collocates. For the whole phrase reading models, collocation status was a consistently significant predictor, while adjacency was not, suggesting that both adjacent and non-adjacent collocations are read faster than control phrases. As for the final word reading times, collocation status had a clear effect on fixation counts. For total reading time, only the adjacent collocations were read significantly faster. However, for the non-adjacent phrases the difference between the collocations and the control phrases approached significance. So overall, even the final word reading models suggest some, albeit smaller, effect for non-adjacent collocations.

The difference between the whole phrase and final word reading models might indicate that the final word is not the only locus of facilitation. Possibly, the verbs were re-read more often in control phrases, as they might have been easier to integrate in the context for collocations. Also, the nouns that were skipped were excluded from the final word analysis. This means that some cases of the clearest facilitation for collocations were excluded from the final word analysis, but their whole phrase reading was still analysed. This might have led to clearer trends of facilitation in the whole phrase reading analysis.

The finding that collocation processing is not disrupted even if a collocation is modified is in line with the studies of Molinaro et al. (2013) and Bonk and Healey (2005). Both those studies suggested that even modified FSs retain their processing advantage. However, this result is in contrast with the Carrol's (2015) study. He found that collocations show no processing advantage when they are non-adjacent. On the contrary, the reading of the first element of the collocation not followed by the expected second word was disruptive in his study. One reason that could account for this difference, though, is different types of collocations used. Carrol (2015) analysed noun-noun and adjective-noun collocations, while I used verb-noun collocations. Wolter and Yamashita (2014) have already suggested that there is a processing difference between adjective-noun and verb-noun collocations. Also, as indicated in the Section 3.3.3, verb-noun collocations seem to be used non-adjacently more frequently than other types of collocations. Therefore, noun-noun or adjective-noun collocations might be learned as adjacent dependencies, creating an expectation for the words to co-occur contiguously. Verb-noun collocations, on the other hand, might be learned as non-adjacent dependencies, allowing for variation, both in terms of position of the elements (as indicated in Bonk and Healey, 2005) and in terms of adjacency.

It is also interesting to compare the findings of the present study to the findings of research on phrasal verbs. Gonnerman and Hayes (2005) found that non-adjacent phrasal verbs take longer to read than adjacent ones, and this effect is moderated by the length of the inserted noun phrase and by the opaqueness of the phrasal verb meaning. There are two potential issues to think about when comparing the current study to the study on phrasal verbs. Firstly, the reading times of phrasal verbs were not compared to matched control phrases, the only comparison was between adjacent and non-adjacent conditions. In the present study, non-adjacent collocations also showed a somewhat smaller facilitative effect that adjacent ones (at least for the final word reading models). The fact that the insertion length came out as a significant predictor in all the final word reading models of the present study suggests that the findings of both studies are very much in line.

However, it might also be that there are differences in reading non-adjacent collocations and non-adjacent phrasal verbs because of different syntactic relationships between the elements of the phrases. For collocations in the present study, the insertions always modified the second element (the noun) of the phrase, while for the phrasal verbs a direct object was inserted between the elements of the FS. This difference between syntactic relationships might have led to different processing effort of inserted elements. As the studies adopted different methodologies, the potential differences between non-adjacent collocations and non-adjacent phrasal verbs (driven by different syntactic structures, different semantics or both) remain an empirical question.

4.4.4 Psychological reality of collocations

The results of the present study show that collocations (at least verb-noun collocations), extracted form corpora using a span window of ± 4 are psychologically real. This finding suggest that collocations have some sort of special status in the mental lexicon: we are able to use the information about our previous encounters of the elements in the phrase when processing language.

Even if the study did not look at the activation of the component parts of collocations, the results can still be interpreted as evidence against the holistic storage of collocations. If collocations were stored and retrieved as wholes, once their core form was modified in any way, the processing advantage should disappear, but this was not the case. Wray (2012) noted that it is not clear if the facilitation in processing of FSs occurs due to the holistic processing or faster mapping of the elements of the phrase. The results of the present study seem to argue against the holistic processing hypothesis and support the idea that collocates are somehow linked to each other and can pre-activate each other. This explanation is in line with Hoey's suggestions of lexical priming (2005, 2012). It seems that each time we encounter a collocation, the information about the co-occurrence of the words is stored in our memory. This information later can be used to pre-activate (or prime) the second element of the collocation after encountering the first one. From the results of the present study, it seems that the activation of the elements of a collocation is strong enough to hold even when there are elements inserted between the collocates.

The processing advantage for both adjacent and non-adjacent collocations is also predictable based on research on learning non-adjacent dependencies (e.g., Vuong, Meyer, & Christiansen, 2011). As verb-noun collocations generally seem to be used in their non-adjacent form, potentially, they are learned as non-adjacent dependencies or at least their non-adjacent occurrences seem to be accumulated as a part of co-occurrence information. Thus, when processing collocations, the facilitation occurs for both adjacent and non-adjacent forms.

One other point to make is that collocations extracted from a corpus with an MI cut-off point of ≥3 showed a clear facilitative effect. This finding goes against the findings of Durrant and Doherty (2010), who suggested that a much higher MI scores are needed in order to extract psychologically real collocations. This difference of results might be due to different methodologies used: Durrant and Doherty (2010) used a lexical decision task. As Ellis (2012b) noted, MI score is not a test of significance, it offers a scale of collocational strength. Different experimental techniques might be able to capture facilitation at different levels of strength and eye-tracking seems to be more sensitive than lexical priming. Taken together, Durrant and Doherty's study and the present study suggest that collocations extracted from corpora are psychologically real, but their degree of facilitation depends on their association strength.

4.4.5 Limitations and directions for further research

Inevitably, the study has some limitations. First of all, it only looked at verb-noun collocations. In naturally occurring language, verb-noun collocations are frequently used in non-adjacent form, which made it easier to write naturally sounding experimental stimuli. However, the fact that these items are frequently encountered in their non-adjacent forms might be the reason why their processing advantage holds for non-adjacent forms. Therefore, this finding cannot be extrapolated to other types of collocations (neither other lexical collocations nor grammatical ones) or other categories of FSs without further research.

Also, in order to tease apart the effects of semantic association and collocation status, only non-associated collocations were used as stimuli. It would be interesting to investigate if associated collocations show the same, or maybe even stronger, trend.

As a tight experimental control was needed, the insertion length (in words) was kept constant throughout the experiment. Thus it remains a question how far from each other the collocates can occur and still retain the processing advantage. Equally, if there were fewer words inserted, the effect for non-adjacent collocations might have been even stronger. The research on learning non-adjacent dependencies suggests that the number of intervening elements makes a difference in the chances of learning non-adjacent dependencies (Santelmann & Jusczyk, 1998). The insertion length in the present study was also a significant predictor. Therefore, in future studies, it would be interesting to manipulate this variable to understand its effect better.

Also, the study suggests that collocations are not fixed phrases and allow variation still retaining their formulaic status. However, the study looked only at adjacency, and did not analyse the effect of positional variation (*spend time* as opposed to *time well spent*) or morphological variation (*spend time* versus *spending time*). It would be interesting to combine the questions of adjacency, positional and morphological variation, and directionality of collocational strength in one study in order to better understand how the variability of collocation form affects processing.

Finally, it seemed interesting to investigate whether the findings with NSs would extend to the NNS populations. This will be the question for the following chapter.

Chapter 5

Processing of adjacent and non-adjacent collocations by non-native speakers

Study 2 discussed in the previous chapter showed that non-adjacent collocations have a facilitative effect during reading. However, the question remains whether the same kind of processing advantage holds for NNSs. This chapter starts by reviewing the literature on factors that affect processing of FL in an L2, and then reports on an eye-tracking experiment on processing non-adjacent collocations, which was carried out with NNSs of English.

5.1. Background of the study

5.1.1 Formulaic language in L2 acquisition

Even if we have a linguistic capacity to use language creatively, we do not normally use all the possible word combinations. This leaves NNSs with an enormous task not only to learn lexis and rules of possible linguistic sequences, but also to gain knowledge of what sounds 'native-like. This implies memorizing hundreds of thousands of chunks that NSs use (Pawley & Syder, 1983; Wray, 2002).

While children start acquiring their L1 from chunks rather than from individual words (Tomasello, 2000; Wray, 2002), an L2 is usually learned in a very different way. One usually learns their L2 later in life so they have different cognitive abilities than children. Also, the kind of exposure that an L2 learner receives is very different from the exposure infants rely on when learning their mother tongue: a learner is usually exposed to less language input, which tends to be written rather than spoken, and the language is not necessarily adapted to him. Hence it might take longer and it might be more difficult for an L2 learner to reach an adequate level of knowledge of FL.

Furthermore, research suggests that when learning their L1, infants commit the brain's neural networks to the patterns of the L1 (Kuhl, 2004), which later hinders learning a new linguistic system – an L2. Therefore, when analysing language processing in an L2, there are additional factors to take into account, such as the amount of exposure to the L2, proficiency, and L1 influence. This makes the picture of language processing more complicated for L2 learners than for NSs.

5.1.2 Processing formulaic sequences in an L2

There have been a lot of studies on FL processing in one's L2. It looks like while NSs process FL faster than novel language, for NNSs the effects of formulaicity are rather mixed (Conklin & Schmitt, 2012). These differences might be due to two reasons. Firstly, all the factors discussed above (different amount of exposure, age of acquisition, differences in language learning situation) might play a role. On the other hand, it might be that NNSs have a slightly different set of memorized FSs than NSs. These sequences are formulaic for them, but they are very idiosyncratic and not recognizable as formulaic by NSs (Foster, 2001). If this is the case, then NNSs may have processing advantages for those sequences, but they are simply not targeted in any experiments. Both of these reasons are probably playing a role to some extent: there are certainly differences in the ways of how NSs and NNSs acquire language, but NNSs might also show more individual variation when it comes to the specific sequences being formulaic or not.

The following section discusses research studies that looked at collocation processing by NNSs of English. Some of these studies were already presented in the previous chapter as they had NS participants as well, but they will be briefly discussed here again, focusing of the results for NNS. The discussion of the studies will be based on three factors that were shown to affect processing of FSs: frequency, congruency between the L1 and an L2, and transparency.

5.1.2.1. Frequency effects

Frequency is a cornerstone of language acquisition in usage-based theories. It has been consistently shown to have an effect on one word processing (see, for example, van Heuven, Mandera, Keuleers, & Brysbaert, 2014). Its importance for processing FSs in the L1 has been demonstrated as well (Siyanova-Chanturia, Conklin, & van Heuven, 2011; Sonbul, 2015; Wolter & Gyllstad, 2013). However, the effect of frequency in an L2 acquisition and processing is a bit more problematic. As Wolter and Gyllstad (2013) noted, learners might approach learning of an L2 more analytically and this might diminish the role of frequency in an L2. Also, L2 learners are exposed to the language less and therefore they do not accumulate as much frequency information as L1 speakers do.

There have been a number of studies that tried to analyse the effect of frequency on processing collocations in an L2. Siyanova and Schmitt (2008), for example, carried out two studies on collocation processing looking at the effect of frequency. In the first one, their participants were asked to rate collocations of different frequencies, and in the second one they had to perform the same rating with their response times being recorded. In the first study, while NSs rated frequent collocations higher that infrequent ones, NNSs rated them with smaller differences and did not discriminate between mid-frequency and high-frequency collocations. This finding suggests that NNSs are less sensitive to collocation frequency than NSs. In the second study, NSs were faster overall, but both NSs and NNSs judged frequent collocations faster than non-frequent ones. However, NSs were again more sensitive to the differences between mid-frequency and high-frequency collocations than NNSs. These studies together suggest that NNSs are able to develop sensitivity to native-like collocation frequency even if they do not perform at the NS level.

In their acceptability judgement task, Wolter and Gyllstad (2013) also investigated frequency effects on processing congruent and incongruent collocations in an L2. The NNSs processed frequent collocations faster than infrequent ones, as did the NSs. This finding supports the role of frequency in the L2 acquisition. However, for the NNSs, collocation congruency also played a role in processing, showing that even advanced NNSs still rely on their L1 when processing collocations.

Sonbul (2015), in her eye-tracking study, also found that NNSs were sensitive to collocation frequency. So the frequency effect was replicated using rating tasks, reaction time experiments, and eye-tracking. However, the participants of all these studies were university students, that is, relatively advanced learners who must have already accumulated considerable amount of exposure to English. It seems, though, that the frequency effect is not consistent for learners of different proficiency levels. For example, Siyanova-Chanturia, Conklin and van Heuven (2011), in their binomial study, found that phrase frequency effects held for proficient but not for less proficient NNSs. Thus overall, studies of collocation processing support the effect of frequency for advanced NNSs, but it remains less clear what happens in less advanced or intermediate levels of proficiency, when learners have had less exposure to the target language.

5.1.2.2 Congruency effect

A number of studies tried investigating the effect of congruency for collocation processing. Congruent collocations were defined as collocations that have a literal translation equivalent in one's L1 (Nesselhauf, 2003; E. Peters, 2016). Overall, it seems that congruency plays a role in processing. Yamashita and Jiang (2010), for example, carried out a phrase acceptability judgement task with Japanese English as foreign language (EFL) learners, Japanese English as second language learners (ESL), and NSs of English. Their ESL learners lived in the USA at the time of the experiment and hence presumably had much more exposure to English than the EFL students who lived in Japan. The study showed that the ESL learners processed both congruent and incongruent collocations equally fast, while the EFL learners processed incongruent collocations slower. However, both groups of learners made significantly more errors when processing incongruent collocations compared to the congruent ones. This finding suggests that incongruent collocations are difficult to acquire even for advanced learners who live in an L2 speaking environment. Unfortunately, the study provides no clear indications about how the collocations were defined and extracted, and the target collocations vary considerably in terms of their transparency. Also, the authors acknowledge that the acceptability judgement task is problematic when used with NNSs, as a lot of trials have to be excluded from the response time analysis due to numerous errors.

Wolter and Gyllstad (2011) carried out a primed lexical decision task and looked at congruency effects on collocation processing in an L2. Together with a priming task, they also tested participants' knowledge of the target collocations, which allowed them to take individual familiarity differences into account. The results showed that when an L2-only item (i.e., incongruent collocation) is known to a learner, processing facilitation occurs; when not, there is no priming for incongruent items. This result is not unexpected: it seems rather intuitive that one needs to learn a FS first in order to benefit from its formulaic nature. However, this study underlines the importance of controlling for familiarity of the items to the language learner. The results of the study suggest that while congruency leads to faster processing, once a non-congruent collocation is learned, it can be processed faster as well.

The effect of congruency was replicated in the already-discussed Wolter and Gyllstad's (2013) study as well. Even if the study was mostly concerned with the effect of frequency, congruent collocations were processed faster than incongruent ones, showing that even advanced NNSs still rely on their L1 when processing the L2. Wolter and Yamashita (2014) found that incongruent items do not facilitate processing, also suggesting the importance of congruency when retrieving collocations in the L2.

Overall, looking at all the previous studies together, it seems to that congruent collocations are more likely to show processing advantages in an L2 than incongruent collocations.

5.1.2.3 Meaning transparency

Collocations, depending on definitions and means of identification, can vary considerably in terms of their meaning transparency. Gyllstad and Wolter (2015) tried to address this question. They combined the distributional approach with the phraseological approach (see Chapter 2) and compared processing of free combinations, collocations (items restricted in their compositionality and therefore not fully transparent), and baseline items. It is important to note, though, that both free combinations and collocations in their study were frequent phrases that would be described as collocations by researchers adopting the distributional approach. Gyllstad and Wolter used a semantic judgement task and found that collocations (restricted compositionality) were read slower than free combinations, but both those groups of items were read faster when compared to baseline items (non-related random word combinations). Basically, this study replicated the result that frequent phrases are processed faster than novel phrases. However, it also seems that non-transparent meaning requires more processing time than literal language. This study showed the same processing tendencies both for NSs and for NNSs of English. However, NNSs of English were advanced leaners; hence it remains questionable whether experiments with less proficient L2 speakers would yield the same results.

Overall, it seems that advanced NNSs (as the ones who participated in the discussed studies) show similar processing advantages as NSs when processing collocations. However, their sensitivity to collocation frequency seems to be slightly lower. It also seems that their L1 plays an important role in processing collocations. However, all the discussed studies focused on processing of adjacent collocations only. As already mentioned in the previous chapter, collocations are very frequently used non-adjacently in naturally-occurring language. As demonstrated by Study 2 (see Chapter 4), for NSs non-adjacent collocations retain a similar albeit slightly smaller processing advantage compared to adjacent collocations. The aim of the present study is to explore whether advanced NNSs of English process non-adjacent collocations the same way as NSs.

5.2 Study 3: Processing non-adjacent collocations in an L2

The study aimed to answer two research questions:

1. Do NNSs show a facilitation effect when processing non-adjacent collocations in the same way that NSs do?
2. What role does prior vocabulary knowledge play in processing collocations in an L2?

5.2.1 Materials

As the aim of the study was to compare the reading behaviour of NNSs to NSs, the previous eye-tracking study (Study 2) was replicated using exactly the same methodology and materials. Therefore, for a detailed discussion of the selection and preparation of the target items, see Section 4.2. As all target items were within the 2,000 most frequent words and the words used in the filler sentences were within the 3,000 frequency band, these materials should not have caused any comprehension problems for NNSs participating in the study: relatively advanced speakers living in England and studying at a British university (this speculation was confirmed by the data analysis). Also, all the target collocations were transparent, so understanding the meaning should not have been an issue.

The only difference in the study design was that some additional information about the participants was collected. Firstly, a background questionnaire was added (see Appendix 8). It included questions on everyday language use, other languages known, and self-reported proficiency of English. However, it can be argued that self-ratings are affected by all sorts of psychological and cultural factors, so they can be problematic. Also, they do not give a spread of participants' proficiency, as with a group of fairly advanced language users only the higher end of the scale tends to be used. Therefore, even though the self-reported language proficiency ratings were collected, they were only used for the participants' description but not for the statistical analysis.

Also, I was interested in having a measure of participants' proficiency. Due to the time limits, it would have been impossible to administer an entire proficiency test as a part of the study. One option would have been asking the participants to report their scores of an English exam (e.g., IELTS or TOEFL) which they had to take in order to enter the university. However, this would have been problematic: they might have taken the exam a couple of years before and the results might not reflect their actual proficiency. Also, they might have taken different exams, making the scores difficult to compare.

Therefore, a time efficient measure had to be chosen. As vocabulary knowledge correlates with general language knowledge (see Milton, 2013), a vocabulary test was administered as a part of the study. As collocations are a part of one's mental lexicon, testing learners' lexical proficiency seemed to be a more important than testing their grammatical knowledge.

Unfortunately, there is no one accepted measure of vocabulary knowledge. For this study a part of the Vocabulary Levels Test (VLT) (Schmitt, Schmitt, & Clapham, 2001) was used. It is a multiple matching test, which covers four different word frequency levels (2,000, 3,000, 5,000, and 10,000) and the Academic Word List. An advantage of this test for the present study is that it has been designed to be used in parts, that is, the researchers or practitioners do not need to administer the entire test, rather they can choose the parts of the test that suit their needs. For this particular study, a vocabulary test was used to discriminate between relatively advanced students, hence giving them the 2,000 frequency level seemed to be unnecessary. Therefore, only 3,000, 5,000, and 10,000 parts of the VLT were used.

5.2.2 Participants

Forty participants (24 females and 16 males) took part in the study. They were all students at British university (undergraduates N = 16; postgraduates N = 22). Their age mean was 25.63 years (*SD* = 4.96). They had spent on average 28.88 months in the UK, but there was a lot of individual differences in terms of the length of stay in the UK (*SD* = 31.29). Their self-reported language skills were: speaking – 3.75, reading – 4.3, writing – 4.08, listening – 4.05 (on the Likert scale from one to five, where five means very proficient). Their mean VLT score (for all three levels administered, max = 90) was 66.63 (*SD* = 11.44.) All participants had normal or corrected to normal vision.

They spoke 16 different first languages: Arabic (2), Bengali (1), Cantonese (3), Dutch (1), French (3), German (4), Greek (1), Hungarian (1), Italian (2), Lithuanian (1), Malay (1), Mandarin (5), Polish (4), Portuguese (6), Spanish (4), and Tegulu (1). Care was taken to make sure that speakers of languages that use non-Latin alphabet (e.g., Mandarin, Greek) would be spread equally across the four different lists of the study.

All the participants were paid £6 for their participation. Each of them was tested individually. After explaining the study, the participants signed the informed consent form (see Appendix 7 for the Participant information sheet). Then they were given the instructions to read the sentences and answer the comprehension questions. Afterwards they were asked to fill in the language background questionnaire. Finally, they took the VLT. The whole study took about 45–55 minutes to complete.

5.2.3 Data cleaning

The answers to the comprehension questions were checked for each participant. The participants showed no difficulty in answering comprehension questions: on average they answered 93.6% of the questions correctly (*SD* = 6.37). The results of the 3,000 level of the VLT showed that the participants had mastered the 3,000 frequency band and hence should not have had comprehension difficulties (*M* = 27.9; max = 30, *SD* = 2.44).

The eye-tracking data were checked manually. In some cases drift corrections were manually performed. Then the recordings were cleaned using the standard procedure of the automatic cleaning function of the *Eyelink DataViewer* as in the previous study (see Section 4.2.6). This cleaning led to the loss of 1.61% of the data (calculated as the percentage of total reading time for the areas of interest).

Then reports for two areas of interest (the final word and the whole phrase) were obtained for each participant. The same eye-tracking measures as in Study 2 were used. For the final word area of interest, first fixation duration, gaze duration, fixation count, and total reading time were analysed. For the whole phrase area, first pass reading time, fixation count, and total reading time measures were analysed.

5.2.4 Data analysis

Firstly, the data were transformed using exactly the same procedures as in the previous study. All the dependent variables, except for fixation counts, were log-transformed. All the continuous predictors were centered.

Model fitting always started from a core model with collocation status, adjacency, and interaction between those two variables predicting the outcome variable. As one of the research questions was the effect of the prior vocabulary knowledge on collocation processing, a vocabulary score was also included in this core model. The core model also included a three way interaction between collocation status, adjacency, and vocabulary score.

Starting from this core model, all the covariates were added and a backwards step-by-step model selection procedure was applied. The covariates considered in the models were: trial number, list of presentation, noun length, noun frequency, verb length, verb frequency, MI, phrase length, phrase frequency, insertion length, phrase naturalness, noun predictability, participant's age, gender, and L1. The final selected models are reported for each measure. The models that included a three-way interaction were investigated further by fitting two separate models: one for the adjacent phrases only and one for the non-adjacent phrases.

The structure of the random effects included random intercepts for subjects and items as well as random slopes for collocation status, adjacency, and the interaction between those two variables.

Initially, I intended to include the self-reported language usage information (e.g., answers to the questions about how many hours a week the participants spent reading in English, socializing in English) to the models as well, to be able to see how the amount of language use affects collocation processing. However, the fact that some participants reported reading for up to 90 hours per week for academic purposes only and for up to 28 hours for pleasure, suggested at least some of them exaggerated the numbers when reporting their language usage figures. Therefore, these measures did not seem to be reliable and were not used in the study.

I have also tried using the overall self-rated language proficiency score as a proficiency measure, but it correlated with the VLT score ($r = .58$; $p = .00$) and the latter seemed to be a more objective measure. Therefore, the study only looked at the effect of the prior vocabulary knowledge on processing collocations in an L2.

5.3 Results and interpretation

This section presents the final word reading models first, followed by the whole phrase reading models. The models are briefly discussed, as far as the significant predictors are concerned, leaving the Discussion section (5.4) for the more general discussion of the findings

5.3.1 Final word: first fixation duration

Figure 5.1 provides a visual comparison of the distributions of the distributions of first fixation duration for the final word in all four experimental conditions.

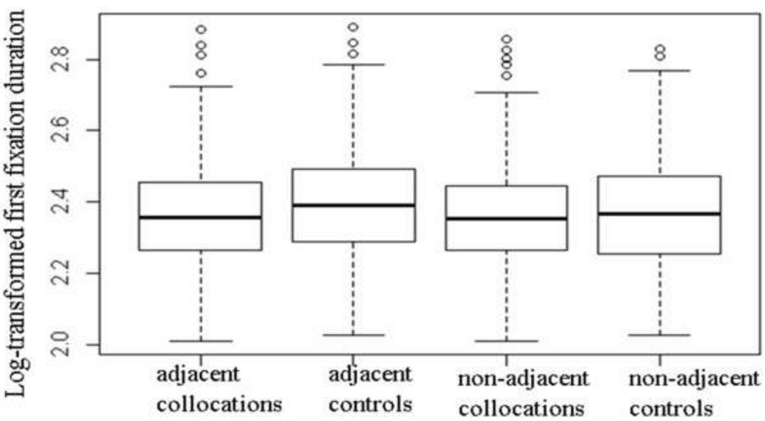


Figure 5.1 First fixation duration. NNS data

When comparing the four conditions visually only, it seems that there were potentially differences between the collocations and the control phrases in both the adjacent and non-adjacent conditions, even if these differences seem to be very small. To make these differences easier to interpret, Table 5.1 presents the means of non-transformed first fixation duration across the four experimental conditions.

Table 5.1 Mean first fixation durations. NNS data

Group	**Mean first fixation duration (ms)**	***SD***	**Mean difference (ms)**
Adjacent collocations	245.52	90.64	18.25
Adjacent controls	263.77	97.73	
Non-adjacent collocations	244.03	88.14	13.77
Non-adjacent controls	251.80	95.70	

From Table 5.1, it seems that there was a similar difference between the collocations and control phrases in both the adjacent and non-adjacent conditions. The final model of the first fixation duration is presented in Table 5.2.

Table 5.2 Model of the first fixation duration. NNS data

	First fixation duration			
Fixed effects	***β***	***SE***	***t***	***p***
Intercept	2.373	0.010	243.57	***
Collocation status	0.023	0.007	3.15	**
Adjacency	-0.075	0.021	-3.56	***
VLT	-0.001	0.001	-0.65	.52
Verb frequency	-0.025	0.012	-2.12	*
Insertion length	0.005	0.001	3.19	**
Age	0.005	0.002	2.79	**
Collocation * Adjacency	-0.022	0.018	-1.23	.23
Collocation * VLT	0.001	0.001	1.07	.29
Adjacency * VLT	0.000	0.001	0.09	.93
Collocation * Adjacency * VLT	-0.002	0.001	-1.06	.30
Random effects	**Variance**		***SD***	
Item (intercept)	0.000		0.007	
Collocation | Item	0.000		0.007	
Adjacency | Item	0.000		0.014	
Collocation * Adjacency | Item	0.002		0.044	
Subject (intercept)	0.003		0.058	
Collocation | Subject	0.000		0.015	
Adjacency | Subject	0.000		0.015	
Collocation * Adjacency | Subject	0.003		0.058	
Residual	0.018		0.133	

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

As Table 5.2 shows, the collocations were read faster than the control phrases. As there was no significant interaction between collocation status and adjacency, it would seem that first fixation duration times on the collocations were shorter than on the control phrases despite the condition (adjacent or not). This finding is somewhat surprising as the NS data did not show collocation effect in the first fixation duration model.

As the negative coefficient for adjacency shows, the final words in the non-adjacent phrases were read faster than in the adjacent phrases. This is in line with the NS data and might be explained by the same reasons, such as preview benefit for nouns when they follow function words or higher expectancy of nouns when they are preceded by modifiers. Insertion length came out as a significant predictor the same way as in the NS data as well, suggesting that the noun reading time increased as the length of the phrase inserted in between the verb and the noun increased.

In contrast to the NS data, age came out as a significant predictor in this model. The model suggests that older participants read the words slightly slower. Age was not a significant predictor for NS data. However, this difference is not surprising. There was much more variation in age in the present experiment: mean age for the NSs in the previous study was 18.89; *SD* = 0.96; mean age for the NNSs in the current study was 25.63; *SD* = 4.96. So it might be that in the previous dataset there was simply not enough variation in age, for it to be a significant predictor.

Predictability was a significant predictor in the NS model, but when added to this model, its coefficient was not significant ($\beta = -0.031$; $SE = 0.046$; $t = -0.67$; $p = .51$) and removing it from the model did not change the model significantly (χ^2 (1, 1539) = 0.47; $p = .49$). Therefore, it was removed from the final reported model. However, the predictability scores were established in a norming study with NSs and even for them predictability proportions were very low. Thus it is not surprising that predictability did not play a role for NNSs.

In the previous study, predictability was always kept in the models even if it was non-significant. This choice was made in order to make sure that collocation status adds something beyond the predictability effect. However, NS data already confirmed this. Also, even for NSs, predictability was only significant in the early measures of the final word reading. For NNSs predictability did not come out as a significant predictor even for the first fixation duration. Therefore, predictability was not kept it in the other models when it was non-significant (even though it was always tested as a covariate).

Vocabulary size, contrary to the predictions, did not come out as a significant predictor for the first fixation duration neither as a main factor, nor in the interaction.

5.3.2 Final word: gaze duration

Gaze duration could be expected to show similar trends as first fixation duration (as they are both early measures of eye-movements). However, the participants fixated 24.5% of the final words more than once during their first pass reading, thus there might be some differences between the models. Figure 5.2 presents the boxplots of the distributions of the gaze duration times in the four experimental conditions.

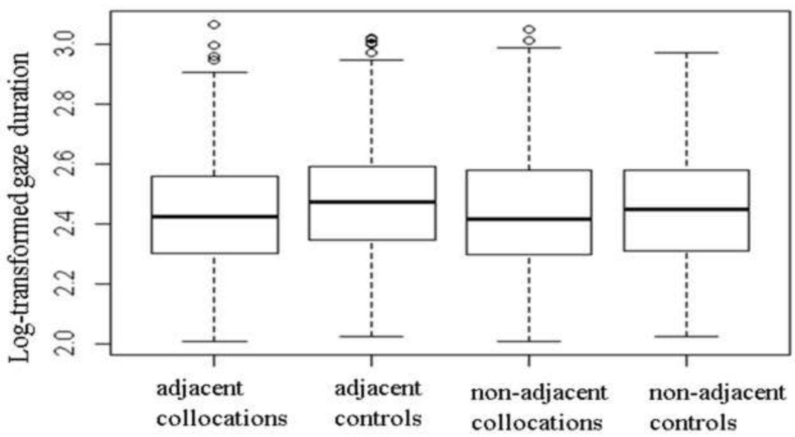


Figure 5.2 Gaze duration. NNS data

From Figure 5.2 it seems that there was a difference in gaze duration times for the adjacent collocations, but for the non-adjacent ones the difference does not seem to be clear. Looking at the mean differences between the groups (see Table 5.3) it appears that while there was a difference between the group means for the adjacent phrases, there was virtually no difference for the non-adjacent phrases. The nouns in the non-adjacent control phrases were even read a little bit faster than in the non-adjacent collocations.

Table 5.3 Mean gaze duration times. NNS data

Group	**Mean gaze duration (ms)**	***SD***	**Mean difference (ms)**
Adjacent collocations	303.87	151.60	30.70
Adjacent controls	334.57	162.33	
Non-adjacent collocations	310.20	159.93	0.39
Non-adjacent controls	309.81	141.20	

The model of the gaze duration times is summarized in Table 5.4. Overall, this model is similar to the NS model (see Section 4.3.2).

Table 5.4 Model of the gaze duration. NNS data

		Gaze duration		
Fixed effects	***β***	***SE***	***t***	***p***
Intercept	2.452	0.013	191.80	***
Collocation status	0.026	0.009	2.81	**
Adjacency	-0.067	0.026	-2.55	*
VLT score	-0.003	0.001	-3.09	**
Noun length	0.013	0.002	5.35	***
Insertion length	0.004	0.002	2.17	*
Age	0.007	0.002	2.70	*
Collocation * Adjacency	-0.040	0.017	-2.27	*
Collocation * VLT	0.000	0.001	0.33	.75
Adjacency * VLT	0.000	0.001	0.56	.58
Adjacency * Collocation * VLT	-0.000	0.001	-0.08	.93
Random effects	**Variance**		***SD***	
Subject (intercept)	0.006		0.074	
Collocation | Subject	0.000		0.018	
Adjacency | Subject	0.001		0.034	
Collocation * Adjacency | Subject	0.000		0.009	
Item (intercept)	0.000		0.017	
Collocation | Item	0.000		0.018	
Adjacency | Item	0.000		0.011	
Collocation * Adjacency | Item	0.001		0.023	
Residual	0.027		0.166	

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

In the present model, collocation status and adjacency came out as significant predictors together with the interaction between them. While for the

NSs, this interaction only approached significance; in the NNS data it was clearly significant.

Significant covariates in both NS and NNS models were noun length and insertion length. However, for the NSs, predictability was also a significant predictor, while for the NNSs, it was not significant. In the NNS data, age came out as significant predictor (as well as in the model of first fixation duration), suggesting that older participants read slightly slower. Also, the VLT score came out as a significant predictor, with higher VLT scores leading to faster reading. This result is expected: people who have more lexical knowledge should be more fluent in reading. There was no interaction between the VLT score and collocation status of the items or their adjacency, though.

To investigate the interaction between collocation status and adjacency further, it has been plotted in Figure 5.3.

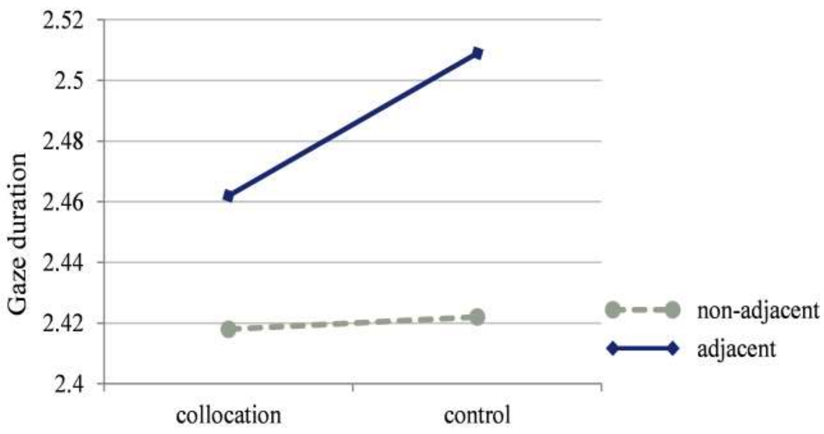


Figure 5.3 Interaction between collocation status and adjacency in the gaze duration model. NNSs data

Figure 5.3 suggests that there was a facilitative effect for the adjacent collocations but not for the non-adjacent ones. This interaction is very similar indeed to the one in the NS data (see Section 4.3.2). The interpretation of the visual inspection of the interaction was confirmed by the statistical analysis. For the adjacent phrases, the difference between gaze duration for the collocations and the control phrases was significant (χ^2 (1) = 12.66, p = .00), while for the non-adjacent phrases, there was no effect of collocation status (χ^2 (1) = 0.27, p = .61).

Overall, the results of the gaze duration analysis suggest that collocations are fixated shorter only when they are adjacent.

5.3.3 Final word: fixation count

To compare the differences in fixation counts between the experimental conditions, the boxplots of their distributions were plotted (see Figure 5.4).

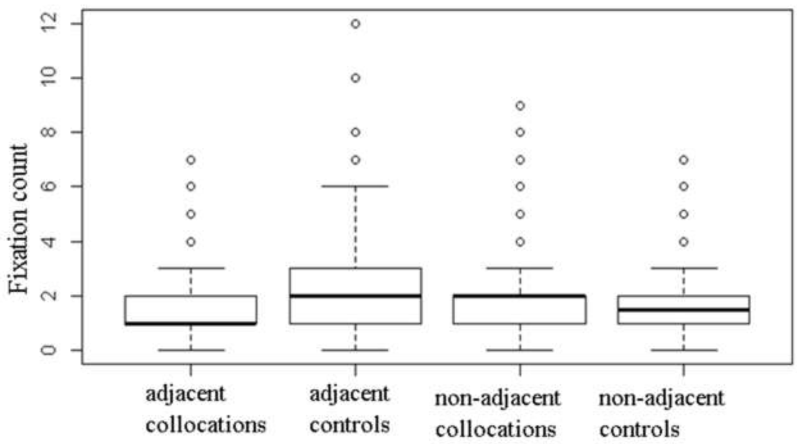


Figure 5.4 Fixation counts of the final word. NNS data

Similarly as for the gaze duration, there seemed to be a difference in fixation counts for the adjacent collocations and control phrases but this difference was not as clear for the non-adjacent phrases. This observation was confirmed by the mean differences presented in Table 5.5.

Table 5.5 Mean fixation counts of the final word. NNS data

Group	Mean fixation count	*SD*	Mean difference
Adjacent collocations	1.79	1.21	0.30
Adjacent controls	2.09	1.41	
Non-adjacent collocations	1.87	1.33	0.04
Non-adjacent controls	1.83	1.23	

Table 5.6 presents the model of fixation count (a Poisson distribution used as the outcome variable is a count).

Table 5.6 Model of fixation count of the final word. NNS data

	Fixation count[a]			
Fixed effects	***β***	***SE***	***z***	***p***
Intercept	0.582	0.048	12.07	***
Collocation status	0.071	0.038	1.86	.06
Adjacency	-0.056	0.039	-1.44	.15
VLT score	-0.011	0.004	-2.66	**
Noun length	0.057	0.010	5.62	***
Collocation * Adjacency	-0.146	0.078	-1.87	.06
Collocation * VLT	0.003	0.003	1.07	.28
Adjacency * VLT	0.002	0.003	0.71	.48
Collocation * Adjacency * VLT	-0.008	0.007	-1.18	.23
Random effects	**Variance**		***SD***	
Subject (intercept)	0.074		0.272	
Collocation | Subject	0.000		0.012	
Adjacency | Subject	0.001		0.039	
Collocation * Adjacency | Subject	0.012		0.108	
Item (intercept)	0.004		0.064	
Collocation | Item	0.000		0.000	
Adjacency | Item	0.000		0.017	
Collocation * Adjacency | Item	0.000		0.006	

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

[a] The model failed to converge

This fixation count model did not converge, so the results have to be interpreted with caution. It is important to note, though, that the model which did not include such a complex random effects structure (no interaction between collocation status and adjacency) converged with significant coefficients both for collocation status and for the interaction between collocation status and adjacency. However, in order not to overestimate the collocation effects, the model with the full random effects structure is reported, following the suggestion of Barr et al. (2013).

In this model, only VLT scores and noun length came out as significant predictors. However, both collocation status and the interaction between collocation status and adjacency almost reached significance (p = .06). Therefore, this interaction was investigated further.

Figure 5.5 shows the interaction plotted graphically. It once again suggests a facilitative effect for the adjacent collocations but not for the non-adjacent ones. Statistical analysis of the interaction confirmed a significant difference between the adjacent collocations and control phrases ($\chi^2 = 7.09$; $p = .02$) and no significant difference for the non-adjacent collocations and controls ($\chi^2 = 0.00$; $p = .97$).

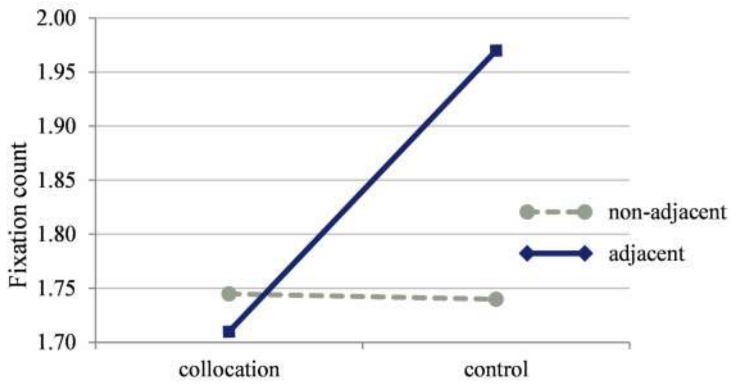


Figure 5.5 Interaction between collocation status and adjacency in the final word fixation count model. NNS data

It is interesting to note that in the fixation count model for the NSs (Section 4.3.3): there was no interaction between collocation status and adjacency, and the collocations showed a facilitative effect in both conditions. Hence it seems that as far as the fixation counts are concerned, NSs and NNSs behave differently: NSs fixate collocations fewer times despite their adjacency, while NNSs fixate collocations fewer times only if they are adjacent.

5.3.4 Final word: total reading time

Total reading measure is different from gaze duration only for the words that are re-read. This was the case for 40.88% of the words in this dataset. The distributions of the total reading time for the four experimental conditions are presented in Figure 5.6.

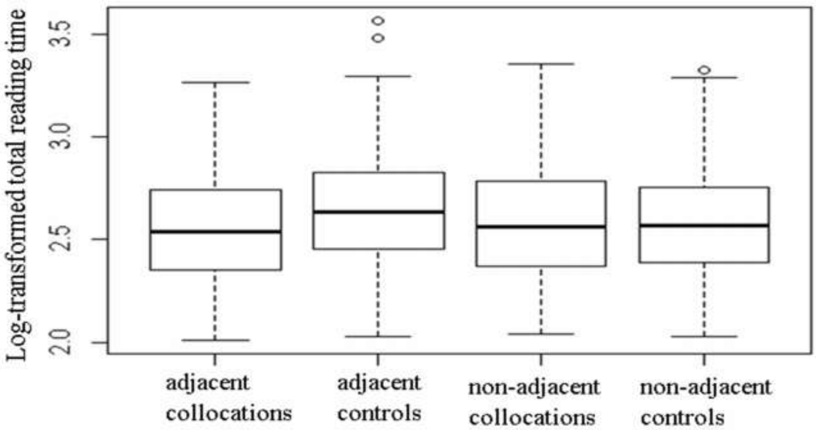


Figure 5.6. Total reading time of the final word. NNS data

For the total reading time, the difference between the collocations and the control phrases in the adjacent conditions was even more obvious than in the previously discussed models. However, there seemed to be no difference for the non-adjacent phrases (see Table 5.7).

Table 5.7 Mean total reading times of the final word. NNS data

Condition	Mean total reading time (ms)	*SD*	Mean difference (ms)
Adjacent collocations	415.18	301.37	104.75
Adjacent controls	519.93	382.07	
Non-adjacent collocations	437.95	327.73	2.98
Non-adjacent controls	440.93	319.20	

Table 5.8 presents the model of final word total reading time.

Table 5.8 Model of total reading time of the final word. NNS data

	Total reading time			
Fixed effects	***β***	***SE***	***t***	***p***
Intercept	2.587	0.022	116.18	***
Collocation status	0.047	0.011	4.29	***
Adjacency	-0.021	0.014	-1.53	.13
VLT	-0.005	0.002	-2.63	*
Noun length	0.016	0.004	3.99	***
Age	0.012	0.004	3.02	**
Collocation * Adjacency	-0.080	0.024	-3.36	**
Collocation * VLT	0.001	0.001	1.32	.19
Adjacency * VLT	0.000	0.001	0.71	.48
Collocation * Adjacency * VLT	-0.002	0.002	-0.95	.34
Random effects	**Variance**		***SD***	
Subject (intercept)	0.017		0.131	
Collocation | Subject	0.000		0.012	
Adjacency | Subject	0.001		0.036	
Collocation * Adjacency | Subject	0.003		0.051	
Item (intercept)	0.002		0.041	
Collocation | Item	0.000		0.009	
Adjacency | Item	0.002		0.041	
Collocation * Adjacency | Item	0.002		0.046	
Residual	0.044		0.209	

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

The model presented in Table 5.8 is very similar to the one in the previous study (Section 4.3.4). The only difference is that insertion length did not come out as a significant predictor, while age and VLT scores did, as in the previous models with the NNS data. While the effects of age and vocabulary scores were already discussed, a potential reason why the insertion length was not a significant predictor might be the fact that there was no facilitation for non-adjacent collocations (as illustrated by the analysis of the interaction below). Therefore, as there was not facilitation for the non-adjacent phrases, the length of the insertion did not play a role any more.

The interaction between collocation status and adjacency is plotted in Figure 5.7.

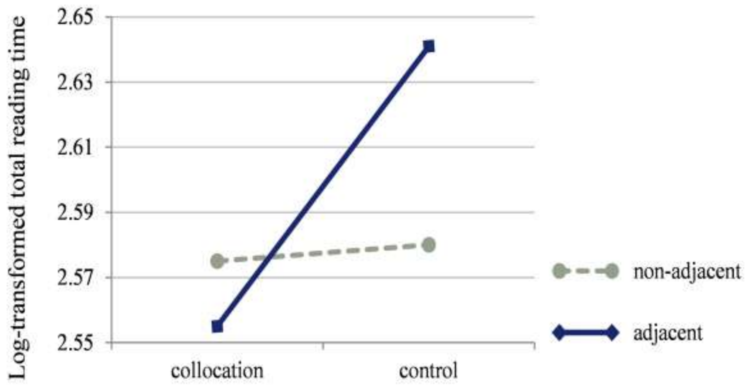


Figure 5.7 Interaction between collocation status and adjacency in the final word total reading time model. NNS data

After further analysing this interaction statistically, it turned out that there was a significant difference between the adjacent collocations and controls (χ^2 (1) = 29.41, p = .00), but not between the non-adjacent collocations and controls (χ^2 (1) = 0.16, p = .68). This finding is in line with the results of the previous models.

Summarizing all the models for the final word reading data, they mostly point in the same direction: NNSs process adjacent collocations faster than control phrases, which is in line with NS data. However, the facilitative effect does not hold for non-adjacent collocations. It seems that once there are words intervening in between the collocates the facilitative effect disappears.

5.3.5 Whole phrase: first pass reading time

When looking at the whole phrase reading behaviour of the NSs, the consistent finding was that collocations were read faster than control phrases and adjacency was not a significant predictor in any of the models. In the present study, though, the final word reading analysis showed that NNSs behave differently than NSs. Therefore, it can be predicted that there would be differences in the whole phrase reading behaviour between the NS and the NNSs as well.

Figure 5.8 compares the distributions of the first pass reading time for the whole phrase in all four experimental conditions. The boxplots show that the distribution of the data was left-skewed. No clear differences between the collocations and the control phrases can be identified from the boxpot. What has to be noted, though, is that the adjacent and non-adjacent conditions cannot be compared directly at this stage because of the very different length of the phrases.

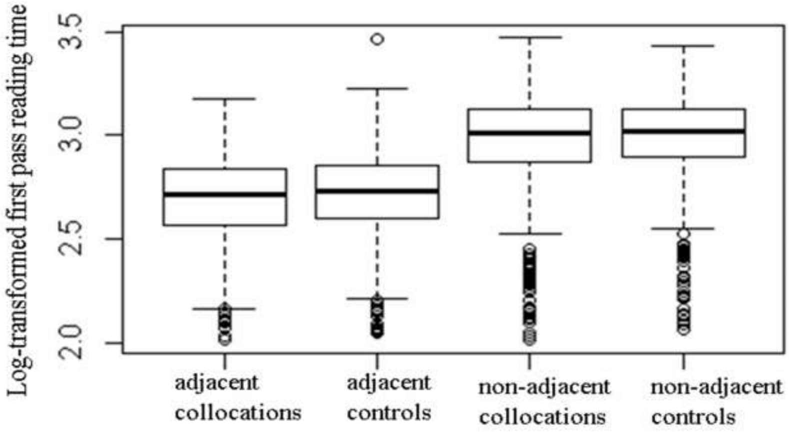


Figure 5.8 First pass reading time. NNS data

Table 5.9 presents the means for all four experimental conditions. The mean differences between the collocations and the control phrases suggest very similar, almost identical, facilitation for the collocations both in the adjacent and in the non-adjacent conditions.

Table 5.9 Mean first pass reading times. NNS data

Condition	Mean first pass reading time (ms)	*SD*	Mean difference (ms)
Adjacent collocations	544.61	255.34	35.65
Adjacent controls	580.26	299.89	
Non-adjacent collocations	1037.19	505.23	30.41
Non-adjacent controls	1067.60	473.69	

The model for the first pass reading time of the whole phrase is presented in Table 5.10.

Table 5.10 Model of first pass reading time. NNS data

	First pass reading time			
Fixed effects	***β***	***SE***	***t***	***p***
Intercept	2.827	0.015	192.21	***
Collocation status	0.025	0.012	2.04	.05
Adjacency	0.059	0.038	1.57	.12
VLT score	-0.004	0.001	-2.89	**
Noun length	-0.017	0.004	-4.24	***
Phrase length	0.015	0.003	5.81	***
Age	0.008	0.003	2.91	**
Collocation * Adjacency	0.001	0.022	0.03	.97
Collocation * VLT	0.000	0.001	0.15	.89
Adjacency * VLT	0.002	0.001	1.99	*
Collocation * Adjacency * VLT	-0.001	0.002	-0.86	.39
Random effects	**Variance**		***SD***	
Subject (intercept)	0.007		0.084	
Collocation | Subject	0.000		0.019	
Adjacency | Subject	0.000		0.011	
Item (intercept)	0.000		0.020	
Collocation | Item	0.001		0.029	
Adjacency | Item	0.002		0.047	
Residual	0.049		0.222	

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

It has to be noted that this model did not include the interaction in the random effect structure because the model with the interaction failed to converge. However, the interaction was also not significant as a fixed effect and therefore including it as a random effect might have made the model too complex for the data.

This model is very similar to the NS model in the previous study (Section 4.3.5). Noun length and phrase length came out as significant predictors in both models, with the same directionality and similar strength of the effect. Also, adjacency did not come out as a significant predictor and neither did the interaction between adjacency and collocation status. While in the NS model collocation status was a significant predictor, in this model it almost reached significance ($p = .05$). This suggests that for the first pass reading time, NNSs (at least relatively advanced ones as in this study) behave very similarly to NSs.

VLT score came out as a significant predictor in this model of the whole phrase reading. The coefficient of this predictor was negative, suggesting that participants with higher VLT scores read faster, as was the case for the final word reading models as well. This finding consistently suggests an effect of vocabulary knowledge on reading.

5.3.6 Whole phrase: fixation count

The boxplots in Figure 5.9 present the distributions of fixation count on the whole phrase in the four experimental conditions. They show that the distributions were right-skewed for all the four groups. From the boxplots there seems to be some difference between the collocations and the control phrases in both the adjacent and non-adjacent conditions, but the significance of this difference cannot be easily predicted.

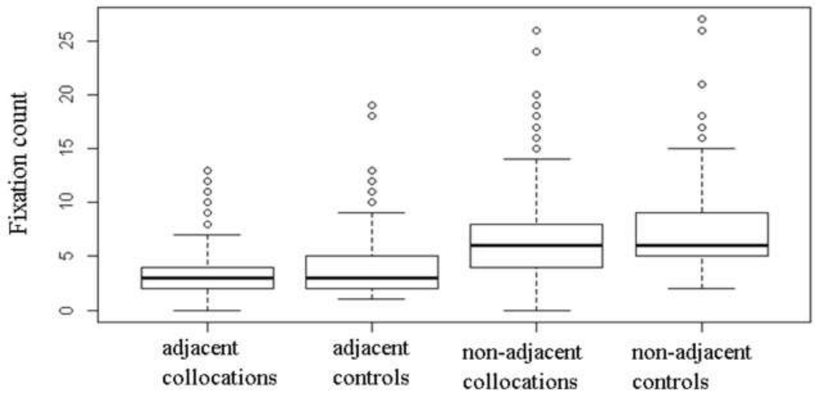


Figure 5.9 Fixation counts of the whole phrase. NS data.

Table 5.11 summarizes mean fixation counts for all four conditions. As it can be seen, there was some difference between the means of fixation counts on the collocations and the control phrases for both the adjacent and the non-adjacent conditions.

Table 5.11 Mean fixation counts of the whole phrase. NNS data

Group	Mean fixation count	*SD*	Mean difference
Adjacent collocations	3.47	2.02	0.56
Adjacent controls	4.03	2.43	
Non-adjacent collocations	6.94	3.71	0.24
Non-adjacent controls	7.18	3.86	

Table 5.12 Model of fixation count of the whole phrase. NNS data

	Fixation count[a]			
Fixed effects	***β***	***SE***	***z***	***p***
Intercept	1.581	0.475	33.30	***
Collocation status	0.084	0.025	3.38	***
Adjacency	0.264	0.075	3.53	***
VLT score	-0.009	0.004	-2.37	*
Phrase length	0.026	0.005	5.33	***
Collocation * Adjacency	-0.010	0.054	-1.85	.06
Collocation * VLT score	0.004	0.002	2.01	*
Adjacency * VLT score	0.002	0.002	1.01	.31
Collocation * Adjacency * VLT	-0.009	0.004	-2.15	*
Random effects	**Variance**	***SD***		
Subject (intercept)	0.076	0.276		
Collocation | Subject	0.001	0.032		
Adjacency | Subject	0.000	0.019		
Collocation * Adjacency | Subject	0.005	0.070		
Item (intercept)	0.009	0.093		
Collocation | Item	0.001	0.029		
Adjacency | Item	0.013	0.113		
Collocation * Adjacency | Item	0.021	0.144		

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

[a] The model failed to converge

The model presented in Table 5.12 was problematic as it failed to converge (same as the model for fixation counts for the final word, see Section 5.3.3). The significant predictors in the model were length of the phrase, collocation status, adjacency, and VLT score. Interestingly, in this model, a three-way interaction between collocation status, adjacency, and VLT score also came out as significant predictors. To investigate this three-way interaction further, two separate models were fitted: one for the adjacent phrases and one for the non-adjacent phrases only. They are presented in Table 5.13.

Table 5.13 Fixation count models for the adjacent and non-adjacent phrases. NNS data

	Fixation count							
	Adjacent phrases only				Non-adjacent phrases only			
Fixed effects	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*
Intercept	1.268	0.049	25.67	***	1.677	0.062	27.06	***
Collocation status	0.138	0.040	3.42	***	0.032	0.032	1.00	.32
VLT score	-0.010	0.004	-2.46	*	-0.008	0.004	-2.01	*
Collocation : VLT	0.008	0.003	2.46	*	-0.000	0.002	-0.03	.98
Phrase length					0.031	0.005	6.00	***
Random effects	**Variance**		***SD***		**Variance**		***SD***	
Subject (intercept)	0.072		0.268		0.078		0.279	
Collocation| Subject	0.004		0.060		0.000		0.002	
Item (intercept)	0.011		0.105		0.012		0.111	
Collocation | Item	0.002		0.045		0.010		0.099	

The models presented in Table 5.13 suggest that the adjacent and non-adjacent phrases were fixated somewhat differently. For the non-adjacent phrases, there was no facilitative effect for the collocations (collocation status did not come out as a significant predictor). For the adjacent collocations, though, there was a significant effect of collocation status and it also interacted with vocabulary size. For the participants with the lowest vocabulary scores, there was almost no facilitative effect for the collocations, while with an increase in the vocabulary size, the facilitation also increased. This is illustrated in Figure 5.10, which plots the interaction between collocation status and VLT score for the adjacent phrases.

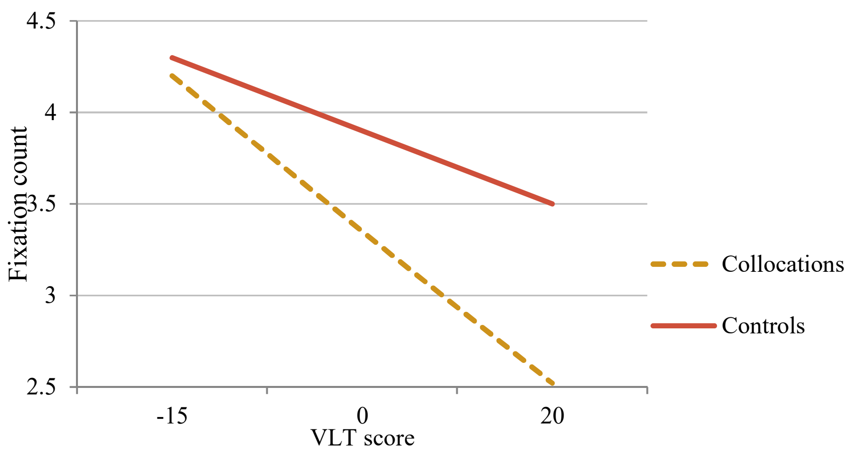


Figure 5.10 Interaction between vocabulary score and collocation status in the adjacent phrases fixation count model. NNS data

5.3.7 Whole phrase: total reading time

Total reading time of a phrase is different from first pass reading time only if the phrase is re-read. This was the case for 51.31% of the phrases in the present study. The boxplots in Figure 5.11 present the distributions of total reading time in the four experimental conditions. From the boxplots it seems that there was a difference between the reading times for the collocations and the control phrases in the adjacent condition, but no clear difference in the non-adjacent condition. The mean differences across the conditions (see Table 5.14 for the comparisons) show that the difference in the adjacent condition was larger than the difference in the non-adjacent condition.

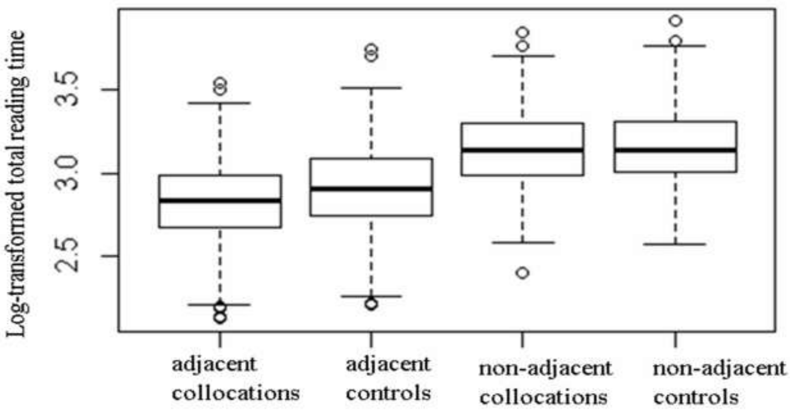


Figure 5.11 Total reading time of the whole phrase. NNS data

Table 5.14 Mean total reading times of the whole phrase. NNS data

Group	Mean total reading time (ms)	*SD*	Mean difference (ms)
Adjacent collocations	802.76	506.79	168.00
Adjacent controls	970.76	641.86	
Non-adjacent collocations	1598.42	909.28	68.84
Non-adjacent controls	1667.26	996.97	

Table 5.15 presents the model of total reading time of the whole phrase.

Table 5.15 Model of total reading time of the whole phrase. NNS data

	Total reading time			
Fixed effects	***β***	***SE***	***t***	***p***
Intercept	3.013	0.023	133.24	***
Collocation status	0.049	0.010	4.88	***
Adjacency	0.059	0.036	1.63	.11
VLT	-0.005	0.002	-2.61	*
Noun length	-0.014	0.004	-3.15	**
Phrase length	0.016	0.002	6.64	***
Age	0.011	0.004	2.53	*
Collocation * Adjacency	-0.061	0.021	-2.91	**
Collocation * VLT score	0.001	0.001	1.46	.15
Adjacency * VLT score	0.001	0.001	0.68	.50
Collocation * Adjacency * VLT	-0.004	0.002	-2.28	*
Random effects	**Variance**		***SD***	
Subject (intercept)	0.018		0.135	
Collocation | Subject	0.000		0.011	
Adjacency | Subject	0.000		0.020	
Collocation * Adjacency | Subject	0.003		0.052	
Item (intercept)	0.001		0.037	
Collocation | Item	0.001		0.027	
Adjacency | Item	0.003		0.057	
Collocation * Adjacency | Item	0.003		0.052	
Residual	0.031		0.175	

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

This model of total reading time is similar to the model of fixation count (it contains a three way interaction). This finding is not surprising, considering that both those measures take into account re-reading behaviour. The three-way interaction was investigated further by fitting two separate models: one for the data of the adjacent phrases only and one for the non-adjacent phrases only. Both of them are presented in Table 5.16.

Table 5.16 Total reading time models for the adjacent and non-adjacent phrases. NNS data

	Total reading time							
	Adjacent phrases only				Non-adjacent phrases only			
Fixed effects	***β***	***SE***	***t***	***p***	***β***	***SE***	***t***	***p***
Intercept	2.872	0.023	124.11	***	3.032	0.030	101.02	***
Collocation status	0.079	0.016	5.09	***	0.021	0.011	1.85	.07
VLT	-0.005	0.002	-2.71	*	-0.005	0.002	-2.38	*
Age	0.011	0.004	2.38	*	0.010	0.005	2.31	*
Collocation * VLT	0.003	0.001	2.37	*	-0.001	0.001	-0.76	.45
Verb length	0.015	0.007	2.09	*				
Verb frequency					-0.046	0.023	-2.03	*
Insertion length					0.018	0.003	6.68	***
Random effects	**Variance**	***SD***			**Variance**	***SD***		
Subject (intercept)	0.018	0.133			0.019	0.138		
Collocation | Subject	0.001	0.023			0.000	0.021		
Item (intercept)	0.002	0.043			0.003	0.050		
Collocation | Item	0.001	0.034			0.000	0.009		
Residual	0.040	0.200			0.022	0.149		

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

As the models in Table 5.16 indicate, for the non-adjacent phrases, collocation status was not a significant predictor. That is, the collocations were not read significantly faster than the control phrases. However, it can be noted that the significance of the coefficient for collocation status approached significance ($p = .07$).

In the model for the adjacent phrases, collocation status was a significant predictor, showing facilitation for the adjacent collocations over the adjacent controls. There was also a significant interaction between VLT score and collocation status, confirming that the facilitative effect for collocations was moderated by one's vocabulary knowledge (see Figure 5.12).

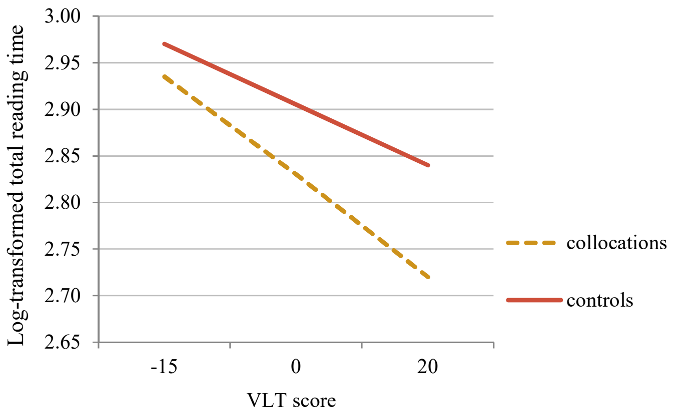


Figure 5.12 Interaction between VLT score and collocation status in the total reading time model. NNS data

Figure 5.12 shows a very similar interaction as in the fixation count model. For the participants with the lowest VLT scores, there was almost no facilitation for the collocations and the degree of facilitation increased with an increase in vocabulary knowledge.

5.3.8 Summary of the results

Table 5.17 summarizes the results of the NS models, as far as the effects of collocation status and adjacency are concerned.

Table 5.17 Summary of the results for non-native speakers

	Final word reading models			
Effects	First fixation	Gaze duration	Fixation count	Total reading time
Collocation status	+	+	almost	+
Adjacency	+	+	–	–
Interaction	–	+	almost	+
Effect for adjacent collocations		+	+	+
Effect for non-adjacent collocations		–	–	–

	Whole phrase reading models		
Effects	First pass	Fixation count	Total reading time
Collocation status	almost	+	+
Adjacency	–	+	–
Interaction	–	+	+
Effect for adjacent collocations		+	+
Effect for non-adjacent collocations		–	almost

Note: plus signs (+) indicate a significant effect, minus signs (–) show no significant effect. The word *almost* indicates that the effect was significant at .1 level.

Table 5.17 shows that the gaze duration, the fixation count and the total reading time models all pointed in the same direction: the collocations were read faster when they were adjacent, but this effect disappeared when the collocations were presented with words intervening in between the collocates. The early measures (first fixation duration and first pass reading time), though, suggested a facilitative effect for all the collocations despite their adjacency.

5.4 Discussion

As this study is based on the same materials as the Study 2, the discussion will focus on both interpreting the findings of the present study and comparing them with the results of the NS data.

5.4.1 Collocation effect

The study aimed to investigate the processing advantage for non-adjacent collocations. Therefore, the main interest of the analysis is the effects of collocation status and adjacency in the models. As Table 5.17 shows, most of the models suggest a facilitative effect for adjacent collocations but not for non-adjacent ones. As exactly the same items were used in both adjacent and non-adjacent conditions, this difference between the conditions cannot be explained by the differences in collocations themselves. This particular group of NNSs read the adjacent collocations in the study faster, so we can assume they were familiar with these lexical items and they had enough exposure to them in order to establish some sort of links between the elements of the collocations in their mental lexicon. However, these links were not strong enough to aid the processing once there were words intervening between the collocates.

However, the results of the study do not allow concluding that there was no effect for non-adjacent collocations. First of all, the models for the early measures of processing (first fixation duration and first pass reading time) showed a main effect of collocation status and no interaction between collocation status and adjacency, suggesting shorter fixations for all collocations. Also, for total reading time for the whole phrase, in the model for the non-adjacent phrases only, collocation status was almost significant. These results taken together suggest that even non-adjacent collocations show an effect, but it is much weaker than and not as clear as for adjacent items. This effect is probably too weak to actually benefit processing as the differences in reading times between the collocations and the controls in the non-adjacent condition were very small. However, it seems that the collocation effect is emergent and it may be that with more exposure to collocations NNSs could reach similar level of processing facilitation as NSs.

5.4.2 Comparison of NS and NNS data

Table 5.18 compares the results from the previous study (Study 2) to the results of the present study. The effects of collocation status and adjacency are summarized.

Table 5.18 Comparison of NS and NNS results

	Final word reading models							
	First fixation		Gaze duration		Fixation count		Total reading time	
Effects	**NS**	**NNS**	**NS**	**NNS**	**NS**	**NNS**	**NS**	**NNS**
Collocation status	–	+	–	+	+	almost	+	+
Adjacency	+	+	+	+	+	–	+	–
Interaction	–	–	almost	+	–	almost	+	+
Effect for adjacent collocations			almost	+		+	+	+
Effect for non-adjacent collocations			–	–		–	almost	–

	Whole phrase reading models					
	First pass		Fixation count		Total reading time	
Effects	**NS**	**NNS**	**NS**	**NNS**	**NS**	**NNS**
Collocation status	+	+	+	+	+	+
Adjacency	–	–	–	+	–	–
Interaction	–	–	–	+	–	+
Effect for adjacent collocations				+		+
Effect for non-adjacent collocations				–		almost

Note: Plus signs (+) indicate a significant effect, minus signs (–) show no significant effect. The word *almost* indicates that the effect was significant at .1 level.

The first difference between NSs and NNSs, which becomes clear from Table 5.18, is the difference between the early measures for the final word. For the first fixation duration, there was a facilitative effect for collocations in the NNS data that was not there for the NSs. In the gaze duration models, NNS data show a clear collocation effect at least for the adjacent items, while for the NSs there was no such effect. From these two measures, it seems that NNSs are more sensitive to collocation status of the phrases than NSs, but this explanation does not seem to be very likely. One potential explanation of these results is that the NSs skipped more nouns when reading collocations and these items were excluded for the word level analysis (3.81% of the nouns skipped overall by the NNSs compared to 9.55% skipped by the NSs). This might have led to a non-significant effect of collocation status for NSs, because the most obvious cases of facilitation were excluded from the analysis. If the NNSs skipped fewer items but fixated the nouns in collocations slightly shorter, this might have been the reason why their data show an effect that is not there for the NS data.

Another potential explanation could be the effect of predictability. Both for the first fixation duration and for the gaze duration models, predictability of the noun was a significant predictor in the NS models but not in the NNS models. As predictability and collocation status are intertwined, it might be that the predictability effect for NSs is actually a very similar effect as the collocation effect for NNSs. If this is the case, then it leads to the conclusion that in early measures of processing, NSs and NNSs behave very similarly (especially if we look at the first-pass reading time models as well).

In the models for the late processing measures, the NSs showed a facilitation for collocations in both conditions (especially for the whole phrase), while the NNSs showed a facilitative effect for adjacent collocations only. So it seems that when we take into account re-reading behaviour, only NSs benefit from collocation status of the words when the elements of the collocation are non-adjacent. However, the data do not allow concluding that there is no facilitation for non-adjacent collocations for NNSs at all. It suggests, though, that the facilitative effect for non-adjacent collocations is much weaker for NNSs.

There might be a number of potential explanations for this difference between NSs and NNSs. Firstly NNSs might be slower and less fluent in integrating inserted modifiers into the context when reading non-adjacent collocations. Therefore, the pre-activation of the second element of a collocation might simply not be strong enough (or not last for long enough) to benefit processing.

Also, the difference between NSs and NNSs can be explained by the differences in their language learning situations (see discussion in Wray, 2002). Frequency seems to be an established factor leading to faster processing of collocations for NSs who learn collocations from exposure. Usage-based theories suggest that frequent patterns become entrenched in memory and can be activated faster (Barlow & Kemmer, 2000; Bybee, 2006; Tomasello, 2005). While this seems to be a good explanation when accounting for the NS behaviour, for the NNSs the effect of frequency might only be a part of an explanation. Obviously, NNSs can and do learn collocations from exposure. However, they might also start from learning them deliberately (potentially as adjacent phrases), and only later benefitting from exposure. Even if we assume that NNSs acquire collocational knowledge from exposure (at least in case language learners living in an L2 speaking country, as the participants of the present study), this process of acquisition can be slightly different for NSs and NNSs. The BNC indicates that all the target collocations of the present study are more frequently used in their non-adjacent forms rather than their adjacent forms. Therefore, it might be that NSs learn them as non-adjacent dependencies from the very beginning. However, as non-adjacent dependencies seem to be somewhat more difficult to acquire than adjacent ones (Gómez, 2002; Newport & Aslin, 2004), NNSs might initially not notice them in the language stream and start from learning these collocations as adjacent phrases, only later allowing them to vary. Therefore, it might be that NNSs have not accumulated enough exposure to know those phrases and to benefit from their collocation status when they are non-adjacent. Hence, the results of the present study do not necessarily suggest that NNSs process collocations qualitatively differently than control phrases, but rather that NNSs did not have enough exposure to these collocations for their elements to be linked strongly enough, in order to facilitate their processing in non-adjacent condition. This explanation would be in line with the usage-based models of language acquisition. More experience with language and more exposure to language lead to greater entrenchment and hence larger facilitation. As NNSs do not have the same amount of exposure to language, even if the cognitive processes underlying the processing of FSs were the same as for NSs (as indicated by the same significant predictors in the models for NS and NNS data), NNSs would not enjoy the same level of facilitation.

5.4.3 Effect of prior vocabulary knowledge

The second research question of the study was whether prior vocabulary knowledge has an effect on collocation processing. The analysis of the significance of the vocabulary scores in the models allows addressing this question. To start with, the results show an overall effect of vocabulary knowledge on reading behaviour. The coefficients for the VLT scores as the main effect were always negative, suggesting that reading times decreased with an increase in vocabulary scores. This finding is not surprising, as an increase in one's vocabulary is related to an increase one's language proficiency, especially receptively (Milton, 2013). So the participants with larger vocabularies were presumably able to access and understand words more fluently.

Also, the VLT score was a part of significant three way interactions for the total reading time and fixation count models in the whole phrase data. Analysed further, these interactions showed that for the adjacent collocations, the facilitative effect depended on learner's vocabulary size. It seems that for the participants with the lowest vocabulary scores, there was almost no facilitation for collocations, but the facilitation increased with an increase in the vocabulary scores. The finding that vocabulary knowledge moderates the collocation effect in reading could indirectly suggest that an increase in vocabulary knowledge leads to an increase in knowledge of collocational patterns of words. It seems that for NNSs with larger vocabulary sizes these collocational patterns in the mental lexicon become more similar to native-like ones.

5.4.4 Comparison to other studies on collocation processing

As already mentioned, most of the previous studies on collocation processing focused on adjacent collocations only. The results of the present study are in line with the studies that showed a processing advantage for adjacent collocations for advanced language learners (e.g., Siyanova & Schmitt, 2008; Sonbul, 2015; Wolter & Gyllstad, 2011, 2013). The reason of this advantage can be the same as for the NSs: familiarity and frequency of the items leads to the special status of phrases in the mental lexicon. This explanation would not necessarily hold for all the language learners, but the participants of the present study (and the previous studies) were all advanced learners living in the L2 speaking country, therefore, presumably receiving a lot of exposure to L2.

It is interesting to compare the findings of the present study to Sonbul's (2015) study as they both used eye-tracking methodology. Sonbul found a facilitative effect for collocations in early but not in late measures of eye-movements (note that she reports whole phrase reading times only). In the present study, the model of the first pass reading time is in line with these results: there was an effect for collocations. However, the present study also shows the effect for total reading time and fixation count of the whole phrase. This difference might be due to the fact that collocations with stronger mutual expectances and larger frequencies were used in the present study (as already discussed in Section 4.4.2).

Sonbul (2015) also found an effect of vocabulary knowledge, using the same vocabulary test. In her study, there was an interaction between collocation frequency and VLT scores, however only for the off-line rating task. In the present study this interaction was present in the eye-tracking data as well. Taken together these two studies suggest that the effect of collocation status is moderated by one's vocabulary knowledge.

It seems important to briefly discuss the findings of the present study focusing on the factors that were shown to be important for processing collocations in an L2: frequency, congruency, and transparency (see Section 5.1.2). As far as frequency is concerned, the target collocations were all relatively frequent. The components of the collocations were all in the 2,000 most frequent words, hence they should have been well-known for the participants (as confirmed by their scores on the VLT). The fact that collocation status was a significant predictor in most of the models is also in line with the findings of frequency effect for NNSs: the frequent items (collocations) were read faster than the non-frequent items (control phrases).

All target collocations were transparent. Therefore, there must have been no confounding effect of transparency. Congruency, on the other hand, remains an issue. Previous studies suggested the effect of congruency (e.g., Wolter & Gyllstad, 2011). However, when the participants have different L1s, it is impossible to control for congruency of the target items for each participant. Wolter and Gyllstad (2011), though, showed that once a collocation is known, its congruency does not matter anymore and participants process it faster. Therefore, maximizing the chances of the target items being known to the participants (selecting frequent and transparent target collocations) provides a partial solution to the problem of congruency. On the other hand, to control for a fact that the target collocations could have included more congruent items for some participants than for others (depending on their L1), I have tried adding an L1 as a potential covariate to all of the models. However, it was never a significant predictor of the reading speed, and hence it was never reported in the models. Possibly, the L1 did not come out as a significant predictor because of the participants' familiarity with the items. On the other hand, all the L1s were represented only by a few speakers. Therefore, the differences between individual readers (as captured in *by-subject* random effects) might have captured potential differences between participants better than their L1s.

5.4.5 Limitations and directions for further research

There were inevitably some limitations of the study and there are numerous ways to extend the current findings.

First of all, the same limitations as in the Study 2 (looking only at one type of collocations and only at constant insertion length) obviously apply for this study as well.

Secondly, the participants came from different linguistic backgrounds. The issue of congruency has already been discussed. However, another problem with a mixed L1 group was that there were participants who do not use Latin alphabet in their L1s. It could be expected that there might be some difference in reading behaviour between the participants whose L1 share the same Latin alphabet with English and those who have different alphabets (e.g., Greek or Arabic). To account for this potential issue two steps were taken. Firstly, during the data collection stage, care was taken to ensure that speakers of languages from other alphabets were equally spread across the four experimental lists. When analysing the data, a binary categorical variable (Latin or non-Latin alphabet) was created and added to the models to see whether it could explain any further variation in the data. The variable did not come out as a significant predictor and did not improve the fit of any of the models. This leads me to conclude that for advanced NNSs (as the ones in the present study) writing system differences do not play a crucial role for processing. As these students were all studying at university in their L2, they seem to have reached a high enough fluency in reading in their L2. With lower proficiency students, though, different writing system might have a larger effect. Therefore, a more rigorous control of the L1 of the participants would be beneficial for the future studies.

Finally, all the NNSs in the present study (as in most of the previous studies) were advanced users of English. It is still unclear how less advanced learners would behave. It might be that they would show no processing advantage even for adjacent collocations. Thus it would be interesting to see at what stage of language learning learners start to become sensitive to collocation status of words and start processing adjacent collocations faster. There has already been a study trying to investigate this question. Sonbul (2012) tried teaching collocations and testing if it leads to their faster processing (using a reaction times experiment). She showed that the amount of input a learner receives in a classroom session is not enough to reach any changes in processing. However, a more sensitive experimental technique might show effects even with beginner learners. Eye-tracking could be a valuable technique as it looks at online processing and requires no recall knowledge or no ability to verbalize it. Therefore, as soon as the learner has a memory trace for a collocation, it could be enough to facilitate processing. Hence it would be interesting to see how much exposure is needed until a facilitative effect occurs using the eye-tracking methodology.

Chapter 6

Incidental acquisition of collocations: Does adjacency matter?

As there are a large number of collocations in language, L2 learners inevitably have to acquire at least a part of them incidentally from exposure. Studies on incidental collocation acquisition (e.g., Pellicer-Sánchez, 2015; Webb et al., 2013) showed that collocations can be learned incidentally, even if the learning rate is quite low. However, while collocations allow insertions and modifications, the previous studies did not take that into account and looked mostly at adjacent collocations.

The results of the Study 3 suggest that NNSs do not read non-adjacent collocations with a native-like facilitative effect. This raises a question of what the effect of adjacency on incidental collocation acquisition is. Study 4 was designed to investigate this question. This chapter briefly presents the main issues with learning collocations in an L2 and reviews studies on incidental acquisition of lexical items. Then it reports on a classroom study (Study 4) carried out to look at the effect of adjacency on incidental collocation acquisition.

6.1 Background of the study

6.1.1 Learning collocations in an L2

Various studies have looked at how L2 learners use collocations and their results vary. Foster (2001), for example, suggested that learners underuse collocations. On the other hand, Durrant and Schmitt (2009) found that NNSs overuse high frequency FSs, but fail to use strongly associated, but less frequent, collocations in their writings. There seems to be an agreement, though, that learners struggle to use collocations in a native-like way. After reviewing numerous studies on how learners from different L1s use collocations, Paquot and Granger (2012) concluded that regardless of the proficiency level, around one third of the collocations learners use are unacceptable. McCarthy (1990) also claimed that "even very advanced learners often make inappropriate or unacceptable collocations" (p. 13). Even though collocations are transparent and do not cause problems for reception, encoding collocations is a challenge for NNSs (Erman, 2007).

There have been different suggestions of why NNSs struggle with collocations. Boers et al. (2014) summarized the difficulties of learning collocations as follows:

- if they are composed of known words, learners do not attend to them;
- the verb in the collocation very often does not carry a lot of meaning and hence remains unnoticed;
- learners do not encounter the same collocation several times in their reading within a short enough span.

Thus one reason for learners' failure to pick up collocations can be their transparency: learners know the meanings of the parts of a collocation, so they understand it when reading but fail to notice and remember it.

In addition, the L1 seems to influence the use of collocations considerably. Bahns (1993) claimed that collocations are usually transferred from the L1 more willingly than idioms, and this transfer accounts for many mistakes that learners make. Nesselhauf (2003) showed that the L1 of the learner can account for about 45% of mistakes in collocational use. Laufer and Waldman (2011) also showed that the persistent errors are usually found in collocations that are transferred from the learners' L1.

Considering the difficulties of acquiring collocations from exposure, there have been suggestions that they need to be taught explicitly in classroom (e.g., Bahns & Eldaw, 1993). However, one of the main problems of teaching collocations seems to be the fact that their number is simply overwhelming. For example, ~13,000 different collocations were extracted for the Study 1 from only a four million word corpus. Thus even if learning collocations from exposure might be a long and slow process, at least part of collocational knowledge has to be acquired incidentally. The following section will define incidental learning and summarize some of the studies on learning individual words incidentally.

6.1.2 Incidental learning

There is an active discussion about implicit/explicit learning, including numerous definitions and theories. Ellis (1994) summarized the distinction as follows:

> Implicit learning is acquisition of knowledge about the underlying structure of a complex stimulus environment by a process which takes place naturally, simply and without conscious operations. Explicit learning is a more conscious operation where the individual makes and tests hypotheses in a search for structure (p. 1).

So implicit learning is unconscious and effortless, while explicit learning is effortful and conscious. This distinction between implicit and explicit learning does not suggest that any one form of learning is better than the other. Quite the contrary: after reviewing the research in psychology, Ellis (2008a) claimed that in situations where the material to learn is relatively simple (or complex but governed by a finite number of critical and salient features), explicit learning is more effective. However, when the material to learn is a bit more randomly structured or influenced by a number of variables with no clear relationships, implicit learning becomes more effective.

Hulstijn (2003) noted that in an SLA context, intentional and incidental learning are the terms that are sometimes practically indistinguishable from the explicit and implicit learning. He notes that the idea of incidental learning was mostly used in SLA vocabulary studies, when carrying out experiments where the participants were not informed about the aim of the experiment. Therefore, the term *incidental* is less theoretically-loaded than *implicit*. It simply implies that participants were not asked to learn target features, without making any strong claims about their consciousness and noticing. His following definition of incidental learning is adopted in the present study:

> a convenient informal, non-theoretical term, referring 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).

There have been a number of studies that looked at incidental acquisition[6] of words. To start with, they tried to answer the question if incidental learning from reading was possible at all. Then the researchers moved on to looking at the effects of frequency and context on learning new words. Later studies extended the scope of research questions and looked at the different aspects of vocabulary knowledge being acquired. As studies on incidental acquisition of collocations build on this body of research, the studies on incidental acquisition of individual words will be briefly discussed here.

The studies on incidental learning of words looked at incidental learning from reading (see Table 6.1), as well as learning from listening (e.g., van Zeeland & Schmitt, 2013) or watching videos (e.g., Montero Perez, Peters, & Desmet, 2015; Rodgers, 2013). However, as the study reported on in this chapter investigated incidental learning from reading, only the research on reading is presented here. While the list in Table 6.1 does not include all of the studies on this topic, it gives a broad overview of the most important findings on incidental word acquisition from reading.

[6] In this thesis, I am not making a distinction between learning and acquisition.

Table 6.1 Reading studies on incidental acquisition of individual words

Studies with NSs	Participants	Treatment	Items	Frequency	Measures	No of words learned (%)	Conclusion
Saragi, Nation, and Meister (1978)	20	A. Burgess's novel *A clockwork orange*	90 *nadsat* words	1–209	MCT	~ 68 (76%)	Considerable amount of learning can occur from reading and repeated exposure
Nagy, Herman, Anderson, and Pearson (1984)	57	1000 word narrative/ 960 word expository text	30 words	1+ (23 words occurred only once)	MCT and interview	~ 3 (10%)	Learning from context is slow but possible even after only 1 exposure
Studies with NNSs							
Day, Omura, and Hiramatsu (1991)	588	30 minutes of reading a short story	17 words	not reported	MCT	not reported	Learners can acquire vocabulary from reading
Horst, Cobb, and Meara (1998)	34	109 pages of simplified novels; reading-while-listening	23 words	2–17	MCT and a word association test	5 (22%)	eight repetitions of the word are required for learning; some initial evidence that stronger learners learn more words
Zahar, Cobb, and Spada (2001)	144	short story; reading-while-listening	30 words	1–15	Multiple matching	2.16 (7.2%)	Context facilitates reading, but it does not necessarily facilitate word learning; effects of frequency are not clear: largest frequency effect for the least proficient learners

Studies with NNSs	Participants	Treatment	Items	Frequency	Measures	No of words learned (%)	Conclusion
Horst (2005)	21	~ 10 graded readers	35 words	not reported	Checklist and Vocabulary Knowledge Scale	17 words (49%)	About half of the words acquired
Pigada and Schmitt (2006)	1	4 graded readers	133 words	1–20+	One-to-one interviews	65% of words enhanced in some ways	Reading can lead to the improvements in the vocabulary depth; spelling knowledge improved most, followed by the grammatical knowledge. 10+ occurrences needed for a sizeable learning
Kweon and Kim (2008)	12	authentic texts; 4–6 hours per day for 5 weeks	367 words	1–20	Self-report test	~ 40%	Words that were learned were retained; nouns learned better than adjectives, verbs learned the worst
Pellicer-Sánchez and Schmitt (2010)	20	Ch. Achebe's novel *Things fall apart*	34 Ibo words	1–28+	One-to-one interviews	Meaning recognition: 43%; orthography: 34%; word class: 20% ; Meaning recall: 14%	Highest gains for meaning recognition; 10+ exposures for learning; a lot of individual variation

Studies with NNSs	Participants	Treatment	Items	Frequency	Measures	No of words learned (%)	Conclusion
Hu (2013)	1	4 graded readers in 6 weeks	63 words	1–18+	Dictation, translation, and sentence writing tasks	~ 46 words improved in some aspects (73%)	Word frequency mostly affected orthographic knowledge; contextual richness had an impact on form-meaning acquisition. Form-meaning link was the most difficult to acquire.
Pellicer-Sánchez (2016)	25 NSs; 23 NNSs	2,300 word story	6 pseudo-words	8	Eye-tracking measures; vocabulary tests for different aspects of knowledge	Form recognition: 86%; meaning recognition: 75%; Meaning recall: 55%	No difference in learning processes between NSs and NNSs; more time spent reading led to better chances for meaning recall; After eight encounters, pseudowords were read in a similar way to known real words
Webb and Chang (2015)	60	20 graded readers; reading-while-listening	200 words	1–70	matching task	High-proficiency group: 63%; Mid-proficiency: 45%; low-proficiency: 28%	Prior vocabulary knowledge has a large influence on incidental vocabulary learning: stronger students show larger gains

Summarizing Table 6.1, the general finding seems to be that incidental vocabulary acquisition occurs, even if it is rather slow, and that different aspects of word knowledge are enhanced to different extents. However, even if all these studies show some learning, directly comparing the number of items learned is rather problematic because the designs of the studies varied considerably. To start with, the duration of the treatment varied from 30 minutes to months, the number of items tested varied from six to 367 and the number of occurrences in the treatment materials varied from one to more than 200. Hence it is not surprising that the learning rates were between 10% and more than 80%. Also, comparing the relative gains (as expressed in the percentages) is rather problematic as learning the same number of words in the study can be reported as a different percentage, depending on the total number of words tested.

Overall, it seems that some factors that play a role in acquisition are frequency of occurrence of the target words and prior vocabulary knowledge. This seems to be a non-controversial finding. At the same time, it has to be noted that reported frequencies in practical terms differ considerably based on the length of the study. Spacing has been reporting to play a role in acquisition and distributed practice has been reliably shown to be more effective than massed practice (Ellis, 1995). So, if a target item occurred eight times in the study materials, there is a difference if it occurred eight times during one reading session or eight times in total during a treatment of four weeks.

Also, even if frequency plays a role, its effect is non-linear and it is not an 'all-or-nothing' effect. Despite the attempts to establish a required frequency threshold, there seems to be none. Webb (2012), who summarized the research on vocabulary acquisition, claimed that there is no specific number of encounters that could guarantee the acquisition even if about 10 encounters are usually suggested. For some aspects of word knowledge, even one encounter is useful.

Overall, the conclusion by Schmitt (2008) seems to hold: "the research confirms that worthwhile vocabulary learning does occur from reading. However, the pick-up rate is relatively low, and it seems to be difficult to gain a productive level of mastery from just exposure" (p. 348). However, he suggests that despite low pick-up rates, incidental learning can be a good way to enhance already existing vocabulary knowledge and increase vocabulary depth. In this sense, it seems that incidental learning might be even more useful for collocations than for individual words, as traditionally collocations would be seen as an aspect of word knowledge (see for example Nation, 2001). Therefore, unsurprisingly, there has already been quite a lot of research on incidental acquisition of collocations.

6.1.3 Incidental learning of collocations

Incidental learning of collocations is different from incidental learning of individual words. First of all, a learner has to notice a sequence of words rather than one word. Also, if a collocation is transparent, its learning does not require inferencing its meaning but rather acquiring knowledge about the two words being used together – form. This might be even more difficult, though, as when the elements of a collocation are known, learners might not pay attention at their partnership (Boers et al., 2014). There might be a lot of different situations of incidental collocation acquisition: both words of a collocation can be known; only one of the words can be known and the other one has to be inferred, and so on. These differences might lead to different chances of learning a collocation, and they complicate the research on incidental collocation acquisition by introducing more potentially confounding variables.

A number of studies looked at acquiring collocations incidentally. For example, Sonbul (2012), in her thesis, compared the incidental acquisition of collocations to the acquisition of enhanced collocations, and to deliberate learning. In her first two studies, she looked at NSs and NNSs of English acquiring medical collocations. For the NSs she found no difference between learning enhanced and non-enhanced conditions, but deliberate learning (where the participants were given a task to learn a list of collocations) was more effective. For the NNSs, though, surprisingly, the only difference she found was between enhanced and incidental condition, suggesting that enhancing collocations in the text might lead to better learning. Her findings suggest that NSs noticed unfamiliar terms more successfully even without the enhancement, while NNSs needed some help. However, in those two studies, the students had to learn two-word medical terms. It is not clear if the collocational knowledge in this case is learning a partnership of two words (adjective and noun), or actually learning a new two-word term. Hence learning transparent collocations that that are tendencies of word use rather than necessary elements of a technical term might be quite different from this kind of learning.

Sonbul (2012) also carried out a classroom study with frequent general purpose collocations (e.g., *familiar face*). In this study, the incidental learning group showed no higher gains than the control group. There might be a couple of reasons for that, though. First of all, her students were rather weak, as shown by the fact that they had problems coping with 2,000 frequency level on the VLT (Schmitt et al., 2001). So they might not have reached enough proficiency to be able to benefit from reading. Also, she only used a three week delayed post-test with no immediate post-test. While this kind of testing is a great opportunity to show durable learning, the delayed test was probably not sensitive enough to capture the potential small gains of the students or, in other words, the treatment was not effective enough to lead to such durable learning.

Szudarski (2012) also looked at incidental acquisition of collocations but focusing on verb-noun collocations. He compared learning in two conditions: 1) reading texts with target collocations embedded and also completing exercises with those target collocations; 2) incidental acquisition of collocations. Reading and performing exercises led to significant gains in both productive and receptive tests. Incidental learning group, on the other hand, was not significantly different from the control group. This finding suggests that learning only from exposure is not effective. However, the study used delexicalized verbs (e.g., *give, take, make*), which might be the reason why the target collocations were difficult to acquire. Also, the target collocations were very frequent, so the control group might have encountered them outside the classroom.

Webb, Newton, and Chang (2013) looked at incidental acquisition of collocations from reading-while-listening. They had 18 target collocations and analysed the effect of repetition (zero, one, five, ten, and 15 occurrences). The collocations they chose were relatively semantically opaque. They concluded that collocations can be learned incidentally and that the number of encounters has an influence on learning. Fifteen encounters, not surprisingly, led to most learning. They concluded that more than five encounters are needed for mastering a collocation. They suggested, though, that hearing a collocation as a phrasal unit might have aided learning, so the results might be not generalizable to reading-only situations. One important issue with this study is that there was no delayed post-test to show the durable gains. Reading-while-listening is also somewhat problematic as it does not seem to be a task that learners are often faced with outside the classroom.

Szudarski and Carter (2014) looked at how input enhancement and input flood can improve the knowledge of collocations. They presented their learners of English with two stories featuring 20 collocations each week for three weeks. They looked at collocations presented six and 12 times over the study period. Their results were rather mixed. While Szudarsky and Carter concluded that input enhancement and input flood can be useful for teaching collocations, it remains unclear why some of the collocations were learned better after six exposures than others were after 12. One of the issues with this study was the type of target items used. The verb-noun collocations all contained delexicalized verbs. As for the adjective-noun collocations, the nouns were infrequent and less likely to be known, while the adjectives were frequent. Therefore, in case of the adjective-noun collocations, it remains questionable if the study investigates the acquisition of collocation as phrases or acquisition of a new unknown word (nouns) and their collocates.

Pellicer-Sánchez (2015) carried out a classroom study where she looked at how the meaning of a new word and its collocational patterns are learned incidentally. Her participants read a short story (~2000 words) with collocations embedded in it for either four or eight times. She administered five different post-tests looking at various aspects of word knowledge. The results showed that collocations: 1) were learned; 2) were learned at the very similar rate as word meanings; and 3) were learned with a larger degree of certainty than most of the other aspects of word knowledge. Interestingly the frequency of the exposure did not play a role in acquisition; at least there was no difference between the four and the eight exposures. However, one limitation of this study, in terms of looking at collocation acquisition, is the use of pseudowords. While pseudowords allow controlling for the previous knowledge, in this study they replaced frequent words, such as *ring* or *bowl*. The collocations used in the study were the most frequent collocations of that replaced real word. This means that there is no way to distinguish whether the readers learned new collocations from reading or whether they simply mapped the meaning on the pseudowords and then transferred their existing collocational knowledge from the real words (e.g., *ring*) to the pseudowords. The latter explanation seems to be somewhat supported by the results of the study: the rate of learning the meaning of the word was not different from the rate of learning its collocation.

Taken together, the studies on incidental collocation learning show rather similar results as research on incidental word learning: learning occurs, but it is rather slow. Also, there is an effect of frequency, but there are no universally accepted thresholds of frequency for incidental acquisition of collocations.

6.1.4 Rationale for the study

To my best knowledge, none of the studies so far have addressed the question of adjacency in incidental collocation acquisition. However, adjacency of the items might actually play an important role as adjacent items can be easier to notice and remember.

Study 3 (see Chapter 5) showed that NNSs benefit for more facilitation when reading adjacent collocations than non-adjacent ones. On the one hand, this might be only an issue with processing that has little to do with learners' ability to acquire non-adjacent collocations. On the other hand, it might indicate that non-adjacent collocations are more difficult to learn as well. This question is important to address as quite a lot of collocations are used non-adjacently in language. Therefore, learners are very likely to encounter them in non-adjacent forms and teachers and materials writers are also very likely to use them non-adjacently. If non-adjacency was shown to hinder learning, teachers or materials writers should aim to maximize the chances of learners encountering the collocations in their adjacent form to facilitate learning those items.

The previous research did not look at adjacency specifically. However, some studies (such as Webb et al., 2013) used a mix of sentences where the target collocations were used both adjacently and non-adjacently (e.g., *broke the silence* and *broke his usual silence*). While this is an ecologically valid approach, it remains unclear if adjacency affects acquisition and if the results of the study would change, had all the items been presented adjacently.

The present study looked at the effect of adjacency and also tried to address some of the limitations of the previous studies on incidental collocation acquisition. First of all, the previous research mostly focused on easy stories consisting of high frequency words (e.g., Pellicer-Sánchez, 2015; Webb et al., 2013). While the use of this kind of stories might be common in language classroom, they are not common in the situations where the learners are already reading authentic texts or learning the subject materials in English, for example, at the university level (paradoxically, as the participants of the studies are usually university students). However, it can be argued that when learners are already using English for their studies, they are left with incidental acquisition as the main source of new collocational knowledge. Therefore, it is important to understand how much incidental learning occurs when the reading materials are challenging in terms of content. To address this limitation, the present study used the materials that could be used for subject learning at a university level.

Also, some studies used pseudowords to look at acquisition (e.g., Pellicer-Sánchez, 2015). While the use of pseudowords ensures perfect control for previous or partial knowledge, it also brings some limitations. Firstly pseudowords might attract more attention. Moreover, as already discussed, when a collocation of a pseudoword is an actual collocations of a word replaced, it is not clear if a learner is learning a collocation or only the meaning of the pseudoword. Also, learning collocations with pseudowords is very different from learning a transparent collocation of two known words. Therefore, the present study investigated the acquisition of real collocations.

Three research questions were addressed in the study:

1. Can transparent collocations be learned incidentally in a naturalistic classroom situation, while reading a text which is challenging content-wise?
2. Is there any difference between learning adjacent and non-adjacent verb-noun collocations?
3. Does prior vocabulary knowledge have an effect on incidental collocation acquisition?

6.2 Study 4: Incidental acquisition of non-adjacent collocations

To answer the research questions, a classroom study following a pre-test, treatment, post-test design was carried out. A between-subjects design was used, with half of the students reading texts with adjacent collocations and half of them being exposed to texts with non-adjacent collocations. The study included both an immediate post-test to capture any immediate gains and a delayed post-test to be able to track any long-term gains. Both receptive and productive knowledge was tested. The following sections present a detailed description of the study materials, piloting stages, and procedure.

6.2.1 Methodology

6.2.1.1 Selection of the target collocations

For the purpose of the study, the target collocations had to be unknown to the participants prior to the treatment. The use of pseudowords was considered, as it is quite common in incidental acquisition studies and it has been suggested as a good practice by Webb et al. (2013). However, as already discussed, the use of pseudowords seems to be rather problematic. Also, as the participants of the present study were university students participating in the experiment during their actual class time, it seemed to be sensible to introduce them to collocations that they can benefit from later in their actual language use.

For this reason, collocations with words from the Academic Word List were used (Coxhead, 1998). As the students had been at the university for only two months before the time of the study, they had not had a lot of exposure to academic English. The item selection started from listing all the verbs in the Academic Word List (Coxhead, 1998). Then, only the verbs that were in the 3,000 most frequent lemmas in the Corpus of Contemporary American English (COCA) (Davies, 2008) list were retained for further consideration. The 3,000 frequency threshold was chosen in order to select collocations that would be comprehensible while reading (individual words known) but may cause problems when trying to produce them. Prior conversations with the teachers of the participants suggested that first year students usually have a receptive vocabulary size of about four to five thousand words.

The COCA (Davies, 2008) was chosen as a reference corpus for a couple of reasons. First of all, it is a large up-to-date corpus of contemporary English. Also, learners of English in Lithuania seem to receive most of their exposure to English outside the classroom from the American TV and music. Therefore, the COCA seemed to be a slightly better representation of the English they might have been exposed to than, for instance, the BNC.

Once the list of verbs was obtained, the second step was to select collocations for these verbs. To do so, the COCA interface online (Davies, 2008) was used. Collocations were defined as phrases that have an MI ≥ 3 and raw frequency ≥ 50. The MI estimates depend on the structure of the collocation you are searching for. The parameters used for extracting the collocations for the study are presented in Figure 6.1 (the words ‘noun’ and ‘verb’ were replaced with the target words). It shows that the MI scores were estimated for the lemmas, allowing up to three word insertions and no positional variation of the collocates.

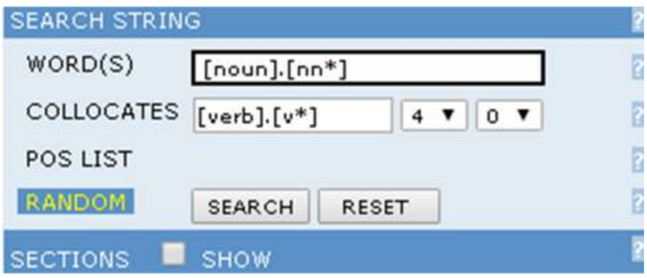


Figure 6.1 Search parameters for collocations

Considering the findings that congruency plays a role in learning collocations (E. Peters, 2016), I decided to focus on incongruent collocations only. In order to select items that were incongruent with the participants' L1 (Lithuanian), I started from discarding congruent items based on my intuitions as a native speaker. When in doubt, a Lithuanian corpus *Donelaitis* (KLC, 2011) was consulted.

After checking the potential noun collocates of the verbs in the list, a list of 98 collocations was obtained. A few collocations from the previous studies (Study 2 and Study 3) were added to this list as well. This final list consisted of 39 different verbs with a number of potential collocates each.

From this list no more than two collocations for each verb were selected for the piloting stage. This led to a pool of 69 potential target collocations. To select the ones that would be included in the actual study, a number of piloting studies were conducted.

6.2.1.2 Congruency piloting

Even if congruency seems to be a rather straight forward concept, in practice, it is not so easy to establish which items are congruent. Following Nesselhauf (2003) and E. Peters (2016), congruent collocations were defined as collocations where the first meaning of *word 1* and the first meaning of *word 2* collocate both in the L2 and in the L1, i.e. a collocation can be translated literally. Only the first meanings of each word in collocation were considered because the subsequent meanings in any dictionary would include more subtle meaning senses and translations, usually based on what words the word collocates with.

To establish the incongruence of the target items, a couple of steps were taken. To start with, the first meanings of both the verbs and the nouns were checked in an English-Lithuanian dictionary (Tildė, 2008) in order to decide if a direct translation is a plausible collocation in Lithuanian. This stage confirmed the intuitions that all the selected items were incongruent.

Afterwards, an additional small-scale piloting study was carried out. Two translators took part in this piloting stage. They both had a Master's degree in Translation studies and a couple of years of experience as freelance translators. Also, they both reported using English every day in their workplaces.

The study consisted of two parts. Firstly, the participants were presented with a list of all 39 verbs and asked to translate them into Lithuanian, using the first translation equivalent that would come to their mind. In the second part, the translators were presented with the list of all 69 collocations, and asked to translate them into Lithuanian. The translators were not allowed to go back and check their previous translations. The study was carried out online using the *Google forms* interface.

If the verb on its own was translated into the same Lithuanian verb as the verb in the collocation, the item was discarded. After this stage, 51 items remained for further investigation.

6.2.1.3 Final selection of items

The next step was to select only 15 collocations for the actual study. To make an informed decision, a teacher of the participants was asked to indicate the items she thought her students would potentially know and these items were discarded. The remaining 31 collocations were kept for the further piloting.

A multiple choice test (MCT) was developed for the remaining items. The noun of the collocation was given in the stem of the question and the participants were asked to select the verb that sounds the most natural with that noun. MCTs were reported to be problematic in vocabulary testing, especially because they tend to overestimate one's knowledge (Gyllstad, Vilkaitė, & Schmitt, 2015; Stewart, 2014). To keep the effects of guessing to a minimum, three steps were taken. Firstly, in the instructions the participants were instructed not to guess blindly. Secondly, an 'I don't know' option was added as the previous research suggests that it can diminish guessing behaviour (Zhang, 2013). Finally, as more options to choose from make guessing more difficult (Stewart, 2012), four distractors were included for each item.

The distractors were all non-collocates of the nouns in the target collocations, as established by the COCA information. However, they were kept semantically similar to the target verb because of the structure of the post-test. Otherwise, having seen the translation of the item in the recall part, the participants might have been able to choose the correct answer without any knowledge of the collocation.

The test was administered to 25 NSs of Lithuanian, all first year students at the same university as the actual participants, all pursuing a degree in languages. They took a paper and pencil test during their actual class time. While they were instructed not to guess, the analysis of their answers strongly suggested that the concerns about guessing were justified: very few 'I don't know' options were ticked, but many unacceptable collocations were selected.

The collocations that were not known to most of the participants were chosen for the actual study, assuming that they have the best chances of being unknown for the actual participants. From the final list of 15 collocations selected, the participants of this piloting stage did not know on average 13.36 (10–15) collocations.

Concurrently, the test was piloted with NSs of English at the University of Nottingham ($N = 16$). Both first year students ($N = 11$) and people about to finish or already holding a PhD ($N = 5$) participated in this part of the study. The PhD group scored higher than the first year students, unsurprisingly, as the target items were academic collocations and the first year students at the time of testing had spent only a bit more than a month at university.

At this stage of piloting, some of the distractors that most NSs were confused about were changed (though Lithuanian data showed that they did not trick the NNSs). Each test item was answered correctly by at least 70% of the NS participants.

This piloting with the NSs and NNSs of English allowed me to choose the 15 target collocations to be used in the actual study.

6.2.1.4 Measurement instruments

Two types of tests were prepared to use in the study: a form recall test and a form recognition test[7]. As the meanings of the collocations were transparent, testing the knowledge of form rather than the meaning seemed to be more relevant.

The format of the recall test was an L1–L2 translation task. The noun of the translated collocation was given already to make sure that the participants are prompted to provide a verb-noun collocation rather than to paraphrase the English phrase. The L1 translations were selected based on the congruency piloting, as during that stage the translators were asked to translate each collocation. The same task was used both in the immediate and in the delayed post-test, but the items were presented in a different random order. An example of the test item is presented in Figure 6.2.

pradėti eiti pareigas	to	duties

Figure 6.2 Item from the recall test

The recognition task was the MCT described above. The same test was used in the pre-test and in both post-tests, but the items and the distractors of each item were presented in a different random order each time. An example of the test item is presented in Figure 6.3. (Full versions of both recognition and recall tests are presented in Appendix 10).

5. to .... duties
   a) adopt
   b) assist
   c) assume
   d) acquire
   e) admit
   f) I don't know

Figure 6.3 Item from the recognition test

[7] The knowledge measured by the form recognition task will be referred to as receptive knowledge, while the knowledge measured by the form recall test will be referred to as productive knowledge in the analysis. I am aware that this is a simplification and that to be able to use a collocation productively requires much more than being able to translate its component verb from one's L1 to the L2. However, this distinction between the productive and the receptive knowledge is adopted for the sake of clarity of the reporting.

An additional small-scale piloting study was carried out with both tests. The aim of this piloting stage was to check if the participants who learn a collocation can answer the test items correctly. I was aiming to find out if the instructions and the translations in the items were not confusing.

An online test consisting of three parts was created. In Part 1, the participants were presented with 60 sentences with target collocations. The sentences were taken from the actual reading materials of the experiment. However, the collocations for this task were enhanced (presented in blue) and the participants were informed that their task was to learn the blue items. An example sentence is presented in Figure 6.4. In Part 2, the participants were presented with the recall test and in Part 3 they were presented with the recognition test.

Certainly, every researcher can appreciate the importance of precision.

Figure 6.4 Enhancement of collocations in the pilot study

Five NSs of Lithuanian participated in this norming study. They were recent university graduates and they had at least one year of English classes at the university level (C1). In the informal conversation afterwards, they reported that while the recognition test was easy, the recall part came somewhat unexpectedly as when reading collocations they could easily understand them and therefore did not spend too much time on memorizing them.

This was largely confirmed by the results of this piloting study. In the recognition test, the participants scored on average 12.2 out of 15 (7–15). Each collocation item was answered correctly by on average 4 learners, showing that recognition format caused no problems for the participants who noticed and learned the collocations.

For the recall test much lower results were to be expected because learning 15 collocations to a recall level is quite a challenging task. The participants scored on average 8.2 out of 15 (5–14) on this test. I was aiming to receive at least one correct answer for each item. This was the case for all the items, except from *to release details*. After discussing it with the participants its translation was changed slightly in the test to make it clearer. Each item in the recall test was answered correctly by on average 2.73 participants.

6.2.1.5 Reading materials

The study was scheduled to take place during actual classes with the first year university students. Therefore, the reading materials had to be related to their course materials and fit into their curriculum. After discussing it with their teacher, two topics for the readings were selected: language and gender and slang.

After surveying a couple of introductory textbooks, monographs, and handbooks on applied linguistics, two following texts were selected: Baxter's (2011) *Gender* and Michael's (2009) *Slang: The People's Poetry.* The text on language and gender provided a brief overview of the history of the field, while the text on slang was an introduction to the book on the use of slang.

Both texts were re-written trying to keep the original sentences but cutting them shorter and introducing the target collocations. As the texts had to be read and discussed during the actual class time (90 minutes), the authentic texts were too long. After the adaptation, the text on language and gender was 2,012 words long, and the text on slang was 2,181 words long for the adjacent versions (2,090 words and 2,230 words respectively for the non-adjacent versions). Both texts were checked by a native-speaker of English to ensure that they sounded natural after the modifications. Ten comprehension questions as well as a few more general questions for discussion were prepared for both texts (see the texts and the questions in the Appendix 11).

In order to make sure that the texts were not too complicated for the target students, the frequency profile of the texts were checked using the VocabProfile (available online at www.lextutor.ca). While the students should have known about 4–5,000 words receptively, considering that there is always a lot of individual variation in any classroom, it was decided to assume receptive knowledge of 3,000 words for this analysis. In the Language and gender text, the 3,000 most frequent word families (BNC-COCA list) covered 95.54% of the text (with proper nouns discarded before the analysis). In the text on Slang, they covered 88% of the text. It seems that for an optimal comprehension the students should understand at least 98% of texts, and for the minimum comprehension, knowledge of at least 95% of words in a text is required (Laufer & Ravenhorst-Kalovski, 2010). So for the Language and gender text, the students were assumed to have at least an adequate comprehension. For the text on Slang, the coverage figure seemed to be problematic. However, a closer inspection of the words which were outside the list of first 3,000 most frequent words in English revealed that a lot of the lower frequency words should not be problematic for the students because they were either explained in the text (e.g., differences between *jargon, slang, cant,* and *argot*), were very likely to be known for the L2 leaners (e.g., *grammar, vocabulary, dictionary, English*) or were cognates with Lithuanian (e.g., *synonym – sinonimas, terminology – terminologija*). If these words were counted as known, they would have accounted for an additional 6.5% of the text. Therefore, the texts were judged to be appropriate for the target population.

Each collocation was inserted in the texts four times. Collocations were presented in the text without any enhancement. Various numbers of occurrences have been considered when designing the study. On the one hand, repetition is important, especially for lower proficiency students (Zahar et al., 2001). On the other hand, it is not easy to insert the same item into a text, for instance, ten times and still make it sound natural. This is especially the case, when 15 items need to be inserted. Hence practicality was an important issue too. Following the classroom study of Pellicer-Sánchez (2015), which showed that there was no difference between four and eight encounters and the learning occurred at about a 40% rate for the receptive knowledge, four occurrences of the target items were chosen. As Sonbul (2012) also found learning (and durable learning in the delayed post-test) from only three exposures both with NSs and NNSs, it was reassuring that four occurrences should be enough to observe some leaning.

When writing the sentences, if an article was intervening between the collocates, the collocation was still considered to be adjacent (e.g., *illustrate the concepts*). This was allowed to make sure the sentences sounded natural. For the non-adjacent conditions, at least two lexical words were inserted between the elements of the collocations and these insertions were different for each occurrence of the collocations (e.g., *adopting new national policies).*

The target collocations and their frequencies in the texts are summarized in Table 6.2.

Table 6.2 Target collocations and their frequencies

		Frequency	
	Collocation	Gender text	Slang text
1.	abandon efforts	1	3
2.	adopt a policy	3	1
3.	appreciate the importance	2	2
4.	assume duties	2	2
5.	challenge claims	3	1
6.	construct an identity	4	0
7.	design experiments	3	1
8.	establish a connection	3	1
9.	exercise control	4	0
10.	generate debate	4	0
11.	illustrate a concept	1	3
12.	meet demand	1	3
13.	pose a danger	1	3
14.	pursue interests	0	4
15.	release details	3	1

6.2.1.6 Procedure

Pre-test. The pre-test took place one week prior to the treatment. The participants were first introduced to the study and asked to sign an Informed Consent form (see Appendix 9 for the Participant information sheet). In order not to draw their attention to the lexis in the study, they were informed that the study looks at second language acquisition and how readers' proficiency affects their comprehension. Then they were given a test booklet, which included the following tasks:

1. pre-test on the target collocations
2. 3,000 and 5,000 levels on the VLT
3. word formation task
4. language background questionnaire

The VLT levels to administer were chosen after the discussion with the teacher of the students. The 3,000 level was administered in order to ensure that the participants were able understand the reading materials and the 5,000 level was administered in order to have a distribution of scores to use in the statistical analysis.

The word formation task was chosen as a distractor task to make sure that the pre-test did not point the participants to focus on the lexical knowledge only. The task was a gap filling exercise, where the learners had to insert a given word in a correct derivational form. The task was taken from a website of the International House Bristol (*http://www.ihbristol.com/free-english-exercises/test-esol/esol-wf-cctv/submitted*), where it was suggested as a task for preparing for the Cambridge First Certificate in English exam. As morphological knowledge was not of interest in this study, this task was not analysed any further. The language background questionnaire and the distractor task can be found in Appendix 12 and Appendix 13.

Treatment and immediate post-test. The treatment session took place during a normal class time (90 minutes). I delivered the session myself. Half of the groups read the text on slang first, while the others read the text on language and gender first. To start with, the students were given 20 minutes to read the first text and to answer 10 true/false questions. Then they were given 10 minutes to discuss the questions and the text with a person sitting next to them. In the meantime, the texts were collected to make sure the students did not spend more time re-reading them. This was followed by a brief group discussion. I was very careful not to use the target collocations and not to elicit their use from the students. After discussing the first text, the same procedure was repeated with the second one.

When the reading part of the class was over, the students were given the productive test first, followed by the receptive test. An immediate post-test was included in the study design after long consideration. As it was shown that the time of the post-test affects the vocabulary gains (with immediate testing, not surprisingly, showing larger gains) (E. Peters, 2014), a delayed post-test would provide a more realistic picture of learning. However, as gains from incidental reading studies are generally small, there was a risk that a delayed post-test might not show any gains, even if there were some (as seemed to be the case for Sonbul's (2012) study). Therefore, the immediate post-test was included in the study to make sure any effects of learning had the best chances of being captured.

After completing the tests, the students were given feedback on their VLT scores from the pre-test week. The reason of this was twofold. First of all, receiving feedback on their vocabulary knowledge was one of the motivational factors for the students to participate in the study. Also, this served as a distractor to minimize the chances of the students remembering the test items for the following delayed post-test.

Delayed post-test. The delayed post-tests were administered two weeks after the treatment sessions. The students first completed the recall and then the recognition tasks. Afterwards they were given a Debriefing sheet with the explanation of the purpose of the study (see Appendix 14). In an informal discussion during this session, none of the students reported figuring out the aim of the study or studying the target collocations at home.

Scoring. The multiple choice tests were scored by giving each answer either zero or one, making no distinction between the incorrect option and the 'I don't know' option chosen. For the productive test, partial knowledge was not scored: only those participants who supplied an expected verb were given one point, other answers were scored zero. Despite the intentions to ignore minor misspellings, there were no cases of misspelling a correct answer. Probably this was the case because the high frequency verbs used in the collocations were known by the participants.

The participants attempted answering most of the items on the productive test. In most cases of the incorrect answers, they provided word combinations do not collocate and could not be used in a translation, e.g., **to start a policy*, **to further your interests*. In some cases the translations provided were plausible collocations but they were not appropriate for the given meaning: for example, *to conduct experiments* instead of *to design experiments*.

Sometimes the participants provided potential translations, even if the words were not collocates based on the COCA information (this was the case for two items only: *to publish details* instead of *to release details*). These answers were scored as zero. There was one problematic item, though, as some participants gave the answer *to understand the importance* instead of *to appreciate the importance*, and it is a collocation based on COCA information. However, these participants were also given zero points as the aim of the study was to look at how students learn the target collocations they were just exposed to. In the piloting stage, the participants provided the expected translation, suggesting that if the target items were learned from the reading materials, the test item itself was not confusing.

6.2.2 Participants

All the first year students of English Studies and Translation Studies at Vilnius University were invited to participate in the study, which took place during the third month of their university course. As the state exam of English they had to take in order to enter university is targeted to B2 level of English, these students could be classified as upper-intermediate users of English.

Initially, 84 students participated in the pre-test (17 males; 67 females). However, five students reported that their first language was Russian. Although they all knew Lithuanian to a native-like level and had the same level of English as other participants, their data were excluded from the further analysis. The main reason for this decision was the fact that the target collocations were not controlled for congruency with Russian.

All the participants attended their classes in nine different groups (five groups of English studies and four groups of Translation studies). These groups were allocated to either adjacent or non-adjacent treatment group, trying to match those two as closely as possible in terms of the number of students, their English exam scores, their vocabulary scores, and their scores in the pre-test. The study sessions took place separately for each group. Only the students who attended all three sessions of the experiment (pre-test, treatment, and post-test) were included in the final analysis.

Out of 54 remaining students, 12 were males and 42 were females. Their mean age was 19.15 years (SD = 1.34). They reported having started studying English when they were on average 8.04 years old (SD = 1.29) and having studied English for on average 10.89 years (SD = 1.47). Their mean score on the state exam of English was 91.47 (max = 100, *SD* = 7.77). They scored on average 27.22 (max = 30, *SD* = 3.06), on the 3,000 VLT level, showing that most of them had fully mastered this level, and 23.87 (max = 30, *SD* = 4.93) on the 5,000 level. They knew on average 4.12 items (max = 15, *SD* = 3.07) on the pre-test, leaving them with on average 10 new items to learn. Most of them reported either not speaking any other languages or speaking one or two languages they learned at school, but none to a proficient level. Forty nine of them reported not having spent more than a month in an English speaking country. Four students reported spending about one month in the UK or in the USA. One student reported having spent year in the UK, but, as her results did not stand out in any way, her data were left in the analysis.

The two treatment groups (adjacent and non-adjacent) were checked to make sure that they did not become significantly different after a number of participants dropped out. Table 6.3 presents the descriptions of the groups.

Table 6.3 Participants in adjacent and non-adjacent treatment groups

	Adjacent condition	**Non-adjacent condition**
Number of students	29	25
Gender	F = 23; M = 6	F = 19; M = 6
Age	19.17 (18–24)	19.12 (18–25)
Pre-test score	4.14 (0–13)	4.24 (0–10)
3,000 VLT score	27.07 (21–30)	27.40 (21–30)
5,000 VLT score	23.45 (16–30)	24.36 (8–30)
State exam score	92.50 (83–100)	90.40 (69–100)

A Mann-Whitney test showed no significant difference between the groups for the pre-test score ($U = 348.50$, $p = .81$), state exam score ($U = 314.50$; $p = .84$); 3,000 VLT level ($U = 340.50$; $p = .69$); 5,000 VLT level ($U = 311.00$; $p = .37$), and age ($U = 324.00$; $p = .56$). Thus the groups remained matched in terms of their language proficiency, age, and previous knowledge of the target items.

6.3 Results

The results of the present study will be presented by addressing the three research questions first, and then by analysing the answering behaviour of the participants a bit more in detail. The data for the Research Questions 1 and 2 were analysed using SPSS version 22 (IBM Corp., 2013). The data for the Research Question 3 were analysed using R (R Core Team, 2013).

All the tests were checked for internal consistency (number of participants $N = 54$; number of items in each test $N = 15$) and the reliability levels were acceptable for all the tests: for the pre-test Cronbach's $\alpha = .74$; for the receptive post-test $\alpha = .77$; for the productive post-test $\alpha = .73$; for the receptive delayed post-test $\alpha = .82$; for the productive delayed post-test $\alpha = .73$.

The normality of the scores on the pre-tests and post-tests were checked and the data turned out to be non-normally distributed. For this reason, non-parametric tests were used for the statistical analysis.

6.3.1. Can transparent collocations be learned incidentally?

The first part of the analysis will try to answer the first research question: can transparent collocations be learned incidentally in a naturalistic classroom situation, while reading a text which is challenging content-wise? As reading activities were mostly likely to increase participants' receptive knowledge, the analysis will start from looking at the acquisition of collocations at the receptive level (form recognition), followed by the analysis of the productive knowledge (recall task).

6.3.1.1 Receptive knowledge

By-subject analysis

A non-parametric Friedman test was used to compare the scores on the pre-test, immediate post-test, and delayed post-test for the entire dataset and for the adjacent and non-adjacent treatment groups separately. The descriptive statistics of this comparison are summarized in Table 6.4.

Table 6.4 Receptive knowledge results summary by subject

Test	Adjacent condition	Non-adjacent condition	All participants
Pre-test	4.14 (3.18)	4.24 (3.07)	4.19 (3.10)
Post-test	6.10 (3.88)	6.08 (3.16)	6.09 (3.53)
Delayed post-test	5.83 (3.91)	6.00 (3.62)	5.91 (3.53)

Note: standard deviations presented in parenthesis

For the entire dataset, a Friedman test showed a significant difference between the pre-test, the immediate post-test, and the delayed post-test (χ^2 (2) = 39.96; p = .00). A pair-wise post-hoc comparison using a Wilcoxon Signed Ranks Test (Bonferroni adjusted alpha value $p < .017$) showed a significant difference with a medium effect size[8] between the pre-test and the immediate post-test (Z = -5.36; p = .00, r = .52) and the pre-test and the delayed post-test (Z = -5.05; p = .00, r = .49) but not between the immediate post-test and the delayed post-test (Z = -1.18; p = .24).

For the adjacent treatment group, a Friedman test showed a significance difference between the pre-test and the two post-tests (χ^2 (2) = 15.24, p = .00). A Wilcoxon Signed Ranks Tests (Bonferroni adjusted alpha value $p < .017$) showed that there was a significant difference between the pre-test and the immediate post-test (Z = -3.73; p = .00, r = .49) and the pre-test and the delayed post-test (Z = -3.45; p = .00, r = .45) but not between the immediate and the delayed post-tests (Z = -1.14; p = .26).

For the non-adjacent treatment group, a Friedman test once again showed a significance difference between the pre-test and the two post-tests (χ^2 (2) = 22.44, p = .00). A Wilcoxon Signed Ranks Test (Bonferroni adjusted alpha value $p < .017$) showed a significant difference between the pre-test and the immediate post-test (Z = -3.84; p = .00, r = .54) and the pre-test and the delayed post-test (Z = -3.76; p = .00; r = .53). There was no difference between the immediate post-test and the delayed post-test (Z = -0.44; p = .66).

[8] The effect sizes in this Chapter are estimate using the formula $r = \frac{z}{\sqrt{N}}$, as suggested by Field (2009). Plonsky and Oswald (2014) argued for adopting field-specific cut-off points for effect sizes in the SLA context. They suggested interpreting r values as follows: $r > .25$ small effect; $r > .4$ medium effect; $r > .6$ large effect. This interpretation is chosen for the present chapter.

To sum up, these results of the by-subject analysis suggest that participants' knowledge significantly increased (all effect sizes medium) from the pre-test to the immediate post-test and did not significantly decline during the two weeks between the post-tests.

By-item analysis

Table 6.5 presents the results of the pre-test and the post-tests when averaged across items (i.e., how many participants answered each item correctly). It has to be noted, though, that due to some participants dropping out, the groups were not perfectly balanced: there were 29 participants in the adjacent group and 25 in the non-adjacent group. Because of that, the by-item analysis was carried out for the adjacent condition and non-adjacent condition separately, but not for the entire dataset.

Table 6.5 Receptive knowledge results summary by item

Test	Adjacent condition (max = 29)	Non-adjacent condition (max = 25)
Pre-test	8.00 (2.88)	7.00 (3.16)
Immediate post-test	11.80 (3.53)	10.07 (3.97)
Delayed post-test	11.27 (4.71)	9.67 (3.83)

Note: standard deviations presented in parenthesis

A non-parametric Friedman test showed a significant difference between the pre-test, the immediate post-test, and the delayed post-test both for the adjacent collocations (χ^2 (2) = 16.32; p = .00), and for the non-adjacent collocations (χ^2 (2) = 11.61; p = .00).

For the adjacent collocations, a Wilcoxon Signed Ranks Test (Bonferroni adjusted alpha value $p < .017$) showed a significant difference between the pre-test and the immediate post-test (Z = -3.31; p = .00, r = .60) and the pre-test and the delayed post-test (Z = -2.99; p = .00, r = .55) but not between the immediate and the delayed post-tests (Z = -0.71; p = .48).

For the non-adjacent collocations, the results were the same: there was a significant difference between the pre-test and the immediate post-test (Z = -3.06; p = .00; r = .56) and the pre-test and the delayed post-test

($Z = -3.21$; $p = .00$; $r = .59$) but not between the immediate and the delayed post-tests ($Z = -0.40$; $p = .69$).

Overall, for the form recognition knowledge, both by-subject and by-item analysis showed the same result: there was a significant gain in knowledge in both the immediate and delayed post-tests and no significant difference between the two post-tests.

6.3.1.2 Productive knowledge

Productive knowledge in this case is defined as the ability to recall the form of a collocation when a meaning is given in the L1. To analyse the increase in the productive knowledge, the procedure was not as straight forward as for the recognition task because no direct measure of the productive knowledge of the target items was included in the pre-test. Therefore, those participants who did not show any receptive knowledge in the pre-test were assumed not to have productive knowledge of these items as well. Hence to analyse the increase in productive knowledge, only a subset of items that the participants did not know in the pre-test was used (584 items in total, 72% of the original dataset).

By-subject analysis

Figure 6.5 summarizes how many participants learned one, two or three items to a productive level. As it become obvious, most of the participants did not gain any productive knowledge of the target items.

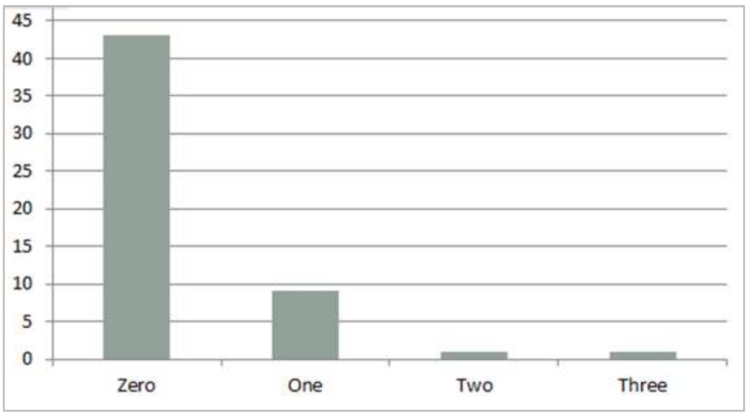


Figure 6.5 Number of items learned

Table 6.6 presents the mean scores in the tests for both participant groups and all participants together.

Table 6.6 Productive gains by subject

Test	Adjacent condition	Non-adjacent condition	All participants
Pre-test	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Immediate post-test	0.17 (0.38)	0.36 (0.76)	0.26 (0.59)
Delayed post-test	0.38 (0.68)	0.40 (0.65)	0.39 (0.66)

Note: standard deviations presented in parenthesis

For the entire dataset, a Friedman test showed a significant difference between the pre-test and the two post-tests (χ^2 (2) = 18.55; p = .00). A post-hoc Wilcoxon Signed Ranks Test (Bonferroni adjusted alpha value $p < .017$) showed that there was a significant difference between the pre-test and the immediate post-test (Z = -3.13; p = .00, r = .30), as well, as between the pre-test and the delayed post-test (Z = -3.67; p = .00; r = .35), but not between the immediate post-test and the delayed post-test (Z = -1.30; p = .19). This result is in line with the result for the receptive knowledge, only the learning gains are much lower and the effect sizes are small for productive knowledge.

A Friedman test also showed a significant difference between the pre-test and the post-tests for both adjacent (χ^2 (2) = 8.91; p = .01) and non-adjacent (χ^2 (2) = 9.75; p = .01) groups separately. A post-hoc Wilcoxon Signed Ranks Test (Bonferroni adjusted alpha value $p < .017$) showed that for the adjacent group there was no difference between the pre-test and immediate post-test (Z = -2.24; p = .03) and the two post-tests (Z = -1.47; p = .14), but there was a significant difference between the pre-test and the delayed post-test (Z = -2.60; p = .01, r = .34).

The same was the case for the non-adjacent group, where there was no difference between the pre-test and the immediate post-test (Z = -2.26; p = .02) and the two post-tests (Z = -0.30; p = .76), but there was a significant difference between the pre-test and the delayed post-test (Z = -2.64; p = .01; r = .37).

This analysis suggests that the productive gains were very small, even if statistically significant. The difference in results for the whole dataset when compared to the results from the adjacent and non-adjacent groups separately might be due to the group size differences. Two separate groups simply had fewer participants, leading to less statistical power. But at the same time the result that only the delayed post-tests scores were higher than the pre-test scores might be indicative of some learning occurring during the testing stage. The immediate translation test was followed by the receptive test. It might be that the receptive test in some cases helped the learners to consolidate their initial knowledge and led to an increase of the score in the delayed post-test

By-item analysis

The same analysis was performed looking at the increase in the test scores by item. The mean numbers of participants who learned each collocation productively are presented in Table 6.7.

Table 6.7 Productive gains by item

Test	Adjacent condition	Non-adjacent condition
Pre-test	0.00 (0.00)	0.00 (0.00)
Immediate post-test	0.33 (0.49)	0.60 (0.83)
Delayed post-test	0.73 (1.03)	0.67 (1.23)

Note: standard deviations presented in parenthesis

A Friedman test showed a significant difference for adjacent items between the pre-test and the two post-tests (χ^2 (2) = 8.00; p =.02) but no significant difference for the non-adjacent items (χ^2 (2) = 5.85; p = .05).

A Wilcoxon Signed Ranks Test (Bonferroni adjusted Alpha value $p < .017$) showed that the only significant difference was between the pre-test and the delayed post-test ($Z = -2.46$; $p = .01$, $r = .33$). Neither the pre-test and the immediate post-test ($Z = -2.24$; p = .03), nor the delayed and the immediate post-tests ($Z = -1.31$; $p = .19$) were significantly different.

By-item analysis further supports the claim that very little productive learning occurred and significant learning occurred only in the adjacent group. Even there, though, the learning only occurred in the delayed post-test and this result, as already discussed, might have been influenced by the additional exposure to the target items in the receptive test during the immediate post-test.

6.3.1.3 Control group

While there was no control group intended in this experiment (as it aimed at analysing differences of acquisition of adjacent and non-adjacent collocations only), there were 11 students who participated in the pre-test, skipped the treatment session and then took part in the delayed post-test. Even if this is a relatively low number of students, it was still interesting to have a look at how their scores changed across the two testing times.

Their mean score in the pre-test was 5.09 ($SD = 3.08$) and in the post-test $M = 5.91$ ($SD = 4.01$). A Wilcoxon Signed Ranks test showed no difference between their scores in the two tests ($Z = -1.15$; $p = .25$). This finding supports the conclusion that the observed learning for the treatment groups was due to the reading task.

Summarizing the answer to the first research question, it can be said that there was clearly some acquisition at a receptive level and it was significant both in by-subject and by-item analysis. The students who missed the treatment session, on the other hand, showed no increase in their test scores. As for the productive knowledge, the gains seem to be very small, suggesting that the exposure to four occurrences of collocations leads to no clear gains in productive knowledge.

6.3.2 Is there any difference between learning adjacent and non-adjacent verb-noun collocations?

To find the answer to the second research question, the gains for the adjacent and non-adjacent conditions were compared. Gains were calculated as the difference between the pre-test score and the post-test score. For this research question, only the data of the receptive test was analysed, as the productive data showed no clear learning effect.

By-subject analysis

Table 6.8 summarizes the gains for the adjacent and non-adjacent treatment groups. Relative gains were estimated as the percentage of items learned, taking into account how many items were already known for each participant.

Table 6.8 Summary of gains by subject

Gains	**Adjacent condition**	**Non-adjacent condition**
Gain in immediate test	1.97 (2.16)	1.84 (1.54)
Relative gains (%)	22.25 (25.09)	18.21 (17.43)
Gain in delayed test	1.69 (2.09)	1.76 (1.81)
Relative delayed gains (%)	19.66 (26.55)	17.93 (19.34)

Note: standard deviations presented in parenthesis

A Mann-Whitney test showed no significant gain difference between the group of students who read the texts with adjacent collocations and those who read the texts with non-adjacent collocations neither in the immediate post-test ($U = 345.00$; $p = .76$) nor in the delayed post-test ($U = 352.00$; $p = .85$). This finding suggests that the treatment conditions made no difference for incidentally learning collocations to the receptive level.

Having in mind that real collocations were used in the study, some students knew more collocations already and therefore had fewer collocations to learn. To account for these differences, relative gains were compared as well. A Mann-Whitney test showed no significant difference in the relative gains of the immediate post-test ($U = 332.00$; $p = .60$) and for the delayed post-test ($U = 356.50$; p = .92).

By-item analysis

The by-item analysis was not as straight forward. As there were more people in the adjacent treatment group, inevitably adjacent collocations could have been learnt by more people. Even if the analysis of the two groups showed no significant difference between the gains in the immediate post-test ($Z = -1.11$; $p = .27$) and in delayed post-test ($Z = -0.82$; $p = .41$), it could have been the case that relatively more students learnt non-adjacent collocations to make the mean differences non-significant. To be able to compare the gains for the two treatment groups, the proportions of people who learned each item were calculated (by dividing the raw number of students out of 29 or 25 accordingly). These proportions summarized in Table 6.9.

Table 6.9 Summary of gains by item

Gains	**Adjacent condition**	**Non-adjacent condition**
Gain in immediate post-test	3.80 (2.60)	3.07 (2.79)
Proportion in immediate post-test	.13 (0.10)	.12 (0.11)
Mean in the delayed post-test	3.27 (2.99)	2.67 (2.02)
Proportion in the delayed post-test	.11 (0.10)	.11 (0.08)

Note: standard deviations presented in parenthesis

A Wilcoxon Signed Ranks test showed no difference between the proportion of students learning adjacent and non-adjacent items in the immediate post-test ($Z = 0.35$; $p = .73$) and the proportion of the students who learned adjacent and non-adjacent collocation in the delayed post-test ($Z = -0.11$; $p = .91$).

Summarizing the analysis on the second research question, there seems to be no difference in learning collocations at the recognition level in adjacent or non-adjacent condition. This was confirmed both in the by-subject and in the by-item analysis.

6.3.3 Does prior vocabulary knowledge have an effect on incidental vocabulary acquisition?

To answer the question whether prior vocabulary knowledge of the students had an effect on their incidental collocation acquisition, a linear mixed effects model was fitted. A model using a binomial distribution was chosen, as the outcome variable was binary: collocations were either learned (1) or not (0). Only the items that were not known by the participants in the pre-test were used for the analysis. In total, 584 observations were used.

Model fitting started from a core model with the treatment condition (adjacent or non-adjacent) and the vocabulary score predicting the learning. To explore what other variables might have influenced learning of the collocations, the following covariates were added: participant's age and gender, collocation frequency, MI, frequency of the noun and verb as well as an interaction between the treatment condition and the vocabulary score. This interaction was included in order to analyse if prior vocabulary knowledge has a different effect on adjacent and non-adjacent collocations, as it was the case in Study 3. For the delayed post-test model, the score on the immediate post-test was also added as a covariate. Then a step-by-step backwards model selection procedure was applied until all the covariates remaining in the model were significant.

Both by-subject and by-item random intercepts were used in the model, as well as random by-item slopes for the condition. The score on the 5,000 level on the VLT was used as an indication of prior vocabulary knowledge. The score on the 5,000 level was chosen instead of the score on the 3,000 level, as it discriminated between the participants better.

The models for the immediate post-test and the delayed post-test are summarized in Table 6.10.

Table 6.10 Models predicting learned items in the post-tests

	Immediate post-test				**Delayed post-test**			
Fixed effects	***β***	***SE***	***z***	***p***	***β***	***SE***	***z***	***p***
Intercept	-1.003	0.178	-5.63	***	-2.051	0.277	-7.41	***
Adjacency	-0.299	0.213	-1.40	.16	-0.040	0.281	-0.14	.89
VLT 5,000 score	0.096	0.022	4.37	***	0.130	0.030	4.30	***
MI					0.449	0.178	2.52	*
Immediate post-test (known)					1.956	0.247	7.93	***
Random effects	**Variance**	***SD***			**Variance**	***SD***		
Subject	0.000	0.000			0.214	0.462		
Item	0.208	0.456			0.385	0.621		
Adjacency | Item	0.043	0.209			0.069	0.263		

$p < .001$: ***; $p < .01$: **; $p < .05$: *

The models summarized in Table 6.10 show that the treatment group (adjacent or not) was not a significant predictor of collocation acquisition, which was expected considering the previous analysis. However, the vocabulary score was a clearly significant factor: larger vocabulary scores increased the probability of collocations to be learned. After calculating the odd coefficients, it turned out with an increase of one point in in the VLT score, the odds of learning an item in the immediate post-test increased by 10%, and in the delayed test by 13%. While the nature of the vocabulary measure used does not allow easily extrapolating these odds to a classroom context, the analysis shows that larger prior vocabulary knowledge predicts higher incidental collocation gains.

I considered including the participant's state exam scores as a measure of the general language proficiency in the model as well. However, they correlated with the VLT scores ($r = .58$, $p = .00$) and the scores were also missing for a few participants. Also, adding it to the model with the VLT score resulted in non-significant coefficient for the state exam scores. On the other hand, when replacing the VLT score with the exam score, the coefficients for the exam scores were significant for both models (in the immediate post-test: $\beta = 0.057$; $z = 3.92$, $p = .00$; in the delayed post-test $\beta = 0.062$; $z = 3.14$; $p = .00$). So an increase of one point in the state exam score increased the odds of learning collocations incidentally by 5.9% in the immediate post-test and 6.4% in the delayed post-test. This finding suggests that participants' general language proficiency is also a significant predictor of collocation acquisition. However, the odds increased more for the increase of the vocabulary knowledge. This is not surprising, considering that collocational knowledge makes a part of one's lexical knowledge.

As expected, for the delayed post-test, the score on the immediate post-test was a significant predictor. The collocations that were learned initially were more likely to be known after two weeks and the score on the immediate post-test came out as the strongest predictor in the model (the odds for the known items to be answered correctly in the delayed post tests were 7.07).

Overall, the answer to the third research question seems to be positive: prior vocabulary knowledge had a positive effect on the incidental acquisition of collocational knowledge.

6.3.4 Analysing individual answers

While the analysis of the mean results is definitely informative, it also somewhat simplifies the results. When looking at the test results at an individual participant level, it soon became clear that the picture is rather complicated. For example, some items were 'known' in the pre-test, but the participants no longer showed any knowledge in any of the post-tests, indicating that they had probably guessed them in the pre-test. Some items were known in the pre-test, known in the immediate post-test, and then unknown in the delayed post-test. Some were known in the first and the third time of testing. These two scenarios would probably suggest some partial knowledge. To investigate how problematic these inconsistencies were, the following tables were created. Table 6.11 summarizes answers in the immediate post-test.

Table 6.11 Item knowledge in the immediate post-test

	Known	**Learned**	**Guessed in the pre-test**
Raw score	177	150	49
Percentage	21.85	18.52	6.05

'Known' in Table 6.11 refers to items that were answered correctly in both pre-test and the immediate post-test, indicating some stable knowledge. As Table 6.11 shows, about 22% of the all the items were known to the participants.

'Learned' in Table 6.11 indicates the items that were not-known in the pre-test and known in the post-test. This was the case for about 19% of all the items. These are the items that are of the main interest of the study.

As the test was a MCT, judging from the very low number of 'I don't know' options chosen, there must have been quite a lot of guessing. The last column in Table 6.11 suggests that about 6% of the items answered correctly in the pre-test were not known in the immediate post-test. It could be speculated that these items were guessed successfully in the pre-test. Six percent of items being guessed, though, does not seem to be a very high number. In the previous analysis, when comparing the pre-test scores and the immediate post-test scores, these 6% of correct guesses remained in the pre-test score, thus actually, the significant gains after the treatment should be even a bit larger than the analysis indicates, as the pre-test scores were slightly overestimated.

For the delayed post-test there were even more potential scenarios for answering the item. An attempt to summarize them is presented in Table 6.12.

Table 6.12 Item knowledge in the delayed post-test

	Known	Know but forgotten	Known again	Learned and retained	Learned but forgotten	Guessed
Raw score	161	16	18	81	69	59
Percentage	19.88	1.97	2.22	10	8.52	7.28

In Table 6.12 ‘known’ refers to the items that have been answered correctly in all the tests, that is, pre-test, immediate post-test, and delayed post-test. Table 6.12 suggests that the participants had a stable knowledge of about 20% of items in the test.

‘Known but forgotten’ stands for items that were answered correctly in both pre-test and immediate post-test, but not in the delayed post-test. As it can be seen from Table 6.12, it was the case for about 2% of the items, which is a rather low percentage. The participants probably had at least partial knowledge of these collocations, but it might have not been strong enough to allow them to answer the item correctly on all the tests. In any case, this percentage is not high.

‘Known again’ stands for an interesting category of items: they were known in the pre-test, not known in the immediate post-test and then answered correctly again in the delayed post-test. This would probably also suggest some partial knowledge, similarly to the ‘know but forgotten’ scenario.

Participants learned and retained for at least two weeks about 10% of all the items. If we calculate the relative percentage of the items learned (participants could not have learned the items they already knew), this relative percentage would be about 24% for the immediate post-test and about 12% for the delayed one. Hence there was definitely some learning, even when we discard all the behaviour that is potentially confounding the effect if only the group means are analysed.

The column ‘learned but forgotten’ stands for the items that were learned in the immediate post-test, but then not answered correctly in the delayed post-test. The table shows that this was the case for quite a few items (about 9%). This is obviously not a surprising finding, as most of the vocabulary studies report some decline in knowledge for the delayed post-test.

'Guessed' in this table stands for items that were never answered correctly before, but were answered in the delayed post-test. They have been labelled as 'guessed' as they could be simply guessed successfully in the last test. Having in mind that a few items were guessed correctly in the pre-test, this is not a surprising finding. However, some of the items in this category might show some consolidation of knowledge. Also, there is a possibility that the participants had encountered these collocations somewhere else during those two weeks and learned them outside the study classroom.

This closer analysis of the participants' behaviour cannot provide any statistical comparison, but it can be an interesting supplement for the statistical analysis. On the one hand, it illustrates that by looking at the means only we are ignoring some tendencies (such as guessing behaviour). On the other hand, this analysis seems to support the statistical analysis by showing that there were indeed a number of items learned during the experiment and quite a lot of them retained to the post-test. Thus looking at these results would suggest that there is definitely some learning at the receptive level for both the immediate and the delayed post-test. However, it might be that there is also some larger decline in knowledge between the two tests than the statistical analysis seems to suggest.

6.4 Discussion

First of all, the study replicated the finding that incidental learning of collocations occurs even after four encounters, as shown by Pellicer-Sánchez (2015). Importantly, collocations that were learned during the reading session were mostly retained two weeks later: the scores on the immediate post-test and the delayed post-test were not significantly different. So the incidental learning was durable.

These findings are contrary to the findings of Sonbul (2012), who found no learning for incidental condition in classroom context, and Szudarski (2012), who also showed that collocations were learned only when some explicit activities were introduced. However, Sonbul's learners were much less proficient than the participants of the present study. It seems, thus, that a certain level of proficiency is required for the incidental learning to be successful (or at least for lower proficiency learners, more input might be necessary). The fact that prior vocabulary knowledge played a role in acquisition in the present study further supports this claim.

As far as Szudarski's (2012) study is concerned, he used delexicalized verbs in his study and this might have been the reason why the gains of incidental learning group in his study were not greater than the gains of the control group. Taken together with the present study these findings seem to suggest that findings on one type of collocations cannot easily be generalizable to other types of collocations.

Even if collocations in the present study were learned incidentally, admittedly the gains were rather small: students learned on average 1.9 collocations receptively during the session (about 20% of the unknown items). This might seem to be a very low gain. However, considering that the reading activity during the session took only 40 minutes and the students were tested only on 15 collocations, the findings seem to be rather positive. If the students learned two collocations in every university session, they could considerably improve their collocational knowledge. Also, it has to be considered that even studies directly focusing on collocations in exercises sometimes showed very similar modest gains (e.g., Boers et al., 2014).

It is not easy to compare the gains in the present study (i.e., 20% or on average two items in a session) to the gains in previous studies on incidental collocation acquisition. Studies on incidental acquisition of collocations used pseudowords (Pellicer-Sánchez, 2015), looked at delexicalized verbs (Szudarski, 2012) or used reading-while-listening task as a treatment (Webb et al., 2013). Therefore, these studies are not directly comparable. When looking at the rates of learning individual words, the relative percentages for gains in receptive knowledge were anywhere from 10% (Nagy et al., 1984) to almost 80% (Pellicer-Sánchez, 2016), depending on the treatment and the aspect of vocabulary knowledge assessed. In any case, the present study suggests a rather low uptake rate, which is probably expected, considering that the texts were complex and the reading took place in only one session. However, it has extended the findings of previous studies by showing that collocations can be learned in more realistic conditions, when reading academic texts, rather than short simplified stories.

Even though frequency was shown to be an important factor in previous studies, no strong claims about the four encounters can be made in the present study. All target collocations were presented four times across the two texts in one session. However, some students might have read the texts twice, some might have re-read parts of the texts, and others might have read them only once. Some readers might have skipped certain paragraphs. So the fact the collocations were always repeated four times does not by any means ensure that everyone was exposed to them four times. This is the very realistic situation, though, and this is an issue for any reading study. Even if students are not allowed to go back when reading (e.g., they are given a story on separate cards, as in Pellicer-Sánchez, 2015) there is no way to stop the keen students from re-reading some parts of the story twice before going to the next section. Therefore, the number of occurrences of the word cannot simply be assumed to be the number of encounters each participant receives, unless there is a direct processing measure to show this was the case. This issue, though, seems to be often overlooked in the studies on incidental learning.

The present study also looked at the gains in productive knowledge. Not surprisingly, the gains were very low, with most of the students not learning any collocations productively. The statistical analysis showed different results for the productive knowledge than for the receptive knowledge. When looking at the treatment groups separately, there was no difference between the pre-test score and the immediate post-test score. The significant differences appeared only between the pre-test and the delayed post-test scores. This finding might indicate that the participants learned the target items not only from the reading but also from the receptive test. If the participants reached a receptive level of knowledge of an item when reading the texts, they were then exposed to that item again during the receptive post-test and might have consolidated their knowledge before the delayed productive test. Even if statistical by-subject analysis for the entire dataset showed significant gains for both the immediate and delayed post-tests, these gains in practice were negligible: on average it was less than one item learned per participant. Thus, these results indicate that reading a text with four occurrences of a collocation is not enough to gain productive level of knowledge of that item. This finding is in line with previous suggestions that it is difficult to learn an item to the productive level incidentally from reading (Schmitt, 2008).

As for the main research question of the study, whether there is any difference between learning adjacent and non-adjacent collocations, the answer seems to be clearly negative. Somewhat surprisingly, the non-adjacent collocations were learned equally well as the adjacent ones. This was the case for by-subject and by-item analysis for the receptive knowledge.

This result might look rather controversial. Firstly, the previous eye-tracking study (Study 3) with advanced NNSs of English showed no processing advantage for non-adjacent collocations. Also, the research on learning non-adjacent dependencies suggested that learning non-adjacent dependencies is more difficult than learning adjacent dependencies (Newport & Aslin, 2004). Therefore, this finding will be discussed further.

The results show that learners acquire non-adjacent collocations equally well as adjacent ones. This would indirectly suggest that they are able to track the transitional probabilities for non-adjacent lexical elements as well as for adjacent ones. This is not a particularly surprising result, though, as a recent study on statistical learning showed that adjacent and non-adjacent dependencies can be learned in parallel (Vuong et al., 2011). Also, it has been pointed out that it is easier to learn non-adjacent dependences if the item to learn carries a semantic meaning (this might seem trivial, but this is often not the case in artificial language learning experiments) (Vries et al., 2012, p. 2074). Hence the aspect of meaning might have had an effect on learning the non-adjacent phases in the study.

The fact that there was no difference between learning adjacent and non-adjacent collocations might also be due to the nature of the target collocations. The verb and the noun of the collocations were always used as the elements of the same phrase (the noun was a direct object of the verb, with the insertions only modifying the noun). So the readers needed to attend to both the verb and the noun in order to understand the meaning of the phrase. The fact that both elements of the collocations were used in the same phrase might have led to their co-activation, potentially aiding the entrenchment of the links between them in memory.

Also, in order to ensure that the learners acquired non-adjacent collocations rather than fixed n-grams, the insertions varied for each occurrence. It was shown that variation of insertions make non-adjacent dependencies easier to notice and learn from the input (Gómez, 2002). Hence this variation of the insertion in the present study might have been beneficial for learning.

When comparing the result from this classroom study to the results of the previous eye-tracking study (Study 3), initially they might seem contradictory, but not necessarily so. In the eye-tracking study, the online processing is measured. Thus for benefitting from collocation status of words, one needs to be fluent and use language with a high degree of automaticity. In this study, on the other hand, no online processing was evaluated. It might well be that it took longer for the participants to read nouns in non-adjacent phrases or to answer test items if they were in non-adjacent treatment group. The results simply show that they are able to learn collocations even if there are word intervening in between the collocates.

One important issue to discuss is what is meant by ‘learning’ in the present study. When discussing learning adjacent and non-adjacent collocations, I have referred to an increase of receptive knowledge about the link between the two words. That is, the meaning of the words is assumed to be known beforehand, and it is only the partnership of the words to be acquired. Also, the test was only measuring the ability to choose an answer in a MCT, which is arguably problematic. This was the case in most of the previous studies as well, though. The question remains of what the findings of the study mean in a real life situation. Recognition level of knowledge clearly does not imply recall level of knowledge (as indicated by the study itself as well). Gaining a score on a MCT does not suggest these students would be able to produce the collocation in their own language. However, it can be an initial stage of learning and more exposure over time could potentially lead to stronger collocational knowledge. Recognition level knowledge might help noticing your own mistakes or might accumulate over time to reach a recall level of knowledge.

6.4.1 Pedagogical implications

Considering that collocations seem to be difficult to acquire even at advanced levels of language proficiency (see Section 6.1.1 for a discussion), research focusing on collocation acquisition has important pedagogical implications. The results of the present study can have implications both for teachers and material writers.

There have been suggestions that the best way of learning FSs is to aim for as much native-like exposure as possible (Meunier, 2012). If the adjacency turned out to be an issue in acquisition, natural language exposure would arguably not be very helpful for acquiring collocations, especially verb-noun collocations as they are frequently non-adjacent. This study suggests, though, that adjacency does not have an effect on acquisition. Thus, receiving exposure to naturally occurring collocations (be them adjacent or non-adjacent) seems to be a good way of increasing one's knowledge of FL.

As for materials writers, teachers adapting texts for their students' needs, and researchers designing the materials for their studies, this finding is very positive. It indicates that there is no need to put an enormous effort to keep the collocations adjacent for better learning, they can be used naturally and this should lead to the same gains. My personal experience with writing items for the Studies 2 and 3 showed what a difficult task it is to use verb-noun collocations only adjacently. It is much easier to provide exposure to collocations that are used in both adjacent and non-adjacent forms depending on the context.

Also, it seems that four repetitions of an item might lead to some durable learning. More repetition would be even more effective (e.g., Webb et al., 2013). As Boers et al. (2014) pointed out, though, collocations are usually not encountered enough times for the students to benefit from the exposure. So if we want the learners to incidentally acquire collocations, the reading materials have to be manipulated to ensure there is some repetition (and four encounters are not that much to aim for). However, it has to be noted that the need to provide repetition is not specific for the incidental learning situation. Explicit focus on any vocabulary items would also require some repetition (e.g., Peters, 2014). So the need of providing repetition when writing materials is not a limitation to deal with when creating a favourable situation for incidental vocabulary acquisition, but rather a necessity for any learning (both implicit and explicit) to occur.

Onnis (2012) criticized the field of SLA for not looking at the findings of the studies on statistical learning and not integrating those findings into teaching materials enough. He claimed that any of the studies on statistical learning that use artificial languages with adult participants essentially simulate the process of acquiring an L2. Therefore, he advocated drawing on the findings of statistical learning research when designing teaching materials, that is, optimizing the salience of certain features relevant for statistical learning. In the present study, I was trying to incorporate some of the findings from the research on statistical learning (e.g., making sure the insertions are always different) into my materials. Following the suggestions of Onnis (2012), the task for both language teachers and materials writers becomes creating the materials that would draw on our knowledge about memory and learning processes and to maximize the chances of the sequences to be learned.

6.4.2 Limitations

The study has a number of limitations. First of all, all the participants of the study were upper-intermediate learners of English, so the results cannot easily be generalized to the beginner learners who might find it more difficult to learn collocations incidentally. Weaker learners, who use the language less fluently, might access word meanings slower and thus struggle to keep the entire verb phrase in the memory to benefit from processing verb and noun together. Therefore, they might struggle with non-adjacent collocations more. Also, most of the learners in this study were relatively motivated, as learning English is crucial for their future careers. Had the participants paid less attention to the tasks or failed to do the readings seriously, the gains may well be lower.

Secondly, the study looked only at one type of collocations: transparent verb-noun collocations. If the meanings of the collocations had been non-transparent, the participants might have learned the formal link between the words equally well, but they would have not necessarily work out the meaning of the collocation equally successfully. Therefore, it would be interesting to replicate the study with non-transparent collocations.

Another remaining question is whether grammatical collocations would be learned at the same rate as lexical ones. As there have been suggestions that grammatical collocations are the most commonly used collocations (Durrant, 2009), this question is an important one to consider in the future. Even if adjacency would probably not be the most important variable for grammatical collocations, looking at their incidental acquisition would be interesting.

One other issue to address is the use of both an immediate and a delayed post-test. Webb et al. (2013) warned that using a delayed post-test in incidental learning studies is problematic as there is no way of checking if the learners have studied the items at home after the session. This appeared to be the case with Webb et al.'s (2013) study. While it is true that there is no way the researcher can check that no one studied the target collocations after the treatment or simply encountered them in reading somewhere else, I would argue that a delayed post-test was not very problematic in the present study. In an informal conversation after the present study, none of the students admitted figuring out the purpose of the study or studying the items at home. One reason why the students in the present study might have paid less attention to the individual items might have been the fact that the study emphasized the content of the texts and the discussion was focused on the content. Also, I was an external researcher, so after the treatment session and the immediate post-test (where their teachers were not even present) they did not expect to see me again. Therefore, rather than being a limitation, delayed post-test seems to be a crucial part of the study in order to look at durable learning. However, it has to be kept in mind that there is a possibility that the students might have encountered the target collocations outside the classroom between the immediate and delayed post-tests.

The use of immediate post-test is also somewhat problematic, as it gives the students one more exposure to the collocations and a chance to consolidate their knowledge. So using an immediate post-test, seems to lead to higher gains (E. Peters, 2014). While this is certainly a concern, I would argue that it only warns against taking the numbers reported and expecting them to be easily generalizable to real classroom situations, other contexts, and other items. It does not, however, undermine the finding that there were gains in collocational knowledge and that they were similar for both adjacent and non-adjacent conditions. Ideally, though, in the future studies, I would have half of the students taking an immediate post-test and a delayed post-test and the other half taking the delayed post-test only, to be able to evaluate how much effect on learning the use of the immediate post-test had.

Also, as in any acquisition study, the students might have learned more collocations (rather fully or partially) or new individual words that were not tested during the study. Therefore, potentially the study underestimates the amount of learning to a certain extent.

Chapter 7

General discussion and conclusions

The aim of this chapter is to place the findings of the four studies conducted for this PhD thesis into the context of previous research on FL. The thesis approached the phenomenon of FL from three different perspectives, using three different research methodologies, and looking at the distribution, processing, and incidental acquisition of FSs, mostly focusing on collocations. Therefore, it seemed worth looking at the findings of Study 1, Studies 2 and 3, and Study 4 separately first, and then bringing the findings together in order to discuss some limitations and provide potential directions for future research.

7.1 Summary of the findings and general discussion

7.1.1 Distribution of formulaic sequences

The first study of this thesis investigated the distribution of different categories of FSs in language. Initially, I intended to provide a comprehensive breakdown of the overall percentage of FSs in language into the percentages of each specific category (covering all the potential categories), in order to systematically address the idea of the heterogeneous nature of FL. However, after starting to design the study, it immediately became clear that this would be an enormous (if not an impossible) task. Therefore, Study 1 finally focused on four categories of FSs only: collocations, idiomatic phrases, lexical bundles, and phrasal verbs. The results suggest that the overall percentage of these four categories of FSs in language falls within previously suggested limits (up to half of the language (Conklin & Schmitt, 2012)) and covers about 40% of the BNC Baby. However, the findings also suggest that different categories of FSs vary considerably in their frequency, with lexical bundles being by far the most frequent category, followed by collocations, idiomatic phrases, and phrasal verbs.

These differences in frequencies would not be as interesting per se, but they seem to illustrate (and to be driven by) a different nature of the phrases analysed. For example, idioms, taken as a group, are actually not very frequent. Therefore, frequency cannot be a driving factor of their formulaic nature. On the other hand, idioms are known to NSs and they very reliably show a processing advantage over novel language (Rommers et al., 2013; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Tabossi et al., 2009; Vespignani et al., 2009). As idioms are characterised by their opaque, figurative meaning, they need to have an underlying unitary concept, represented in the mental lexicon in some way (see, for example, the superlemma theory of idiom representation by Sprenger et al. (2006)). This conceptual layer might be the reason they are showing the processing advantage and psychological reality. So from the distributional perspective, idioms might be the peripheral members of FL, while from the phraseological perspective they are the core members.

Lexical bundles, on the other hand, are at the other extreme – they are particularly frequent but very often transparent and incomplete, both semantically and structurally (Biber et al., 2004). However, from the usage-based perspective to language acquisition, frequency is a reason for the phrases getting entrenched in memory (Kemmer & Barlow, 2000). If this is the case, it should not matter whether a phrase is complete or not, or whether it consists of lexical words or mostly function words. As long as a sequence is frequently used, the words of the sequence should become linked to each other in the mental lexicon and start to be predictable, facilitating each other's processing. As we have some empirical evidence for faster processing of lexical bundles (e.g., Nekrasova, 2009; Reali & Christiansen, 2007; Schmitt et al., 2004; Tremblay et al., 2011), lexical bundles (or any other sequences defined by their frequency) can still be defined as formulaic even if their nature of formulaicity is very different from that of idioms.

So there seems to be at least two different reasons for phrases to become formulaic: meaning-based (formulaicity at the concept level), and frequency-based (formulaicity based on recurrence); and these two types of formulaic phrases differ considerably in their frequency distributions. This is not a clear cut distinction, though: some collocations, for example, can be both frequent, and idiomatic. However, this distinction gives a reason to try and look at different categories and their frequencies separately.

Study 1 also suggested that one of the biggest challenges in extracting FSs is accounting for their variability. The small-scale post-hoc analysis of the use of collocations, clearly suggests that form flexibility is more common for some types of collocations than for the others. However, this analysis only looked at collocations. Variation in form might be even more important for other FSs that allow variable slots, such as prepositional frameworks (Renouf & Sinclair, 1991) or potentially certain formulaic conversational routines that were not even covered in the Study 1.

One systematic way to address the variability of forms of FSs would be moving away from the idea of collocations (in a general sense, as words occurring next to each other) to the idea of meaning shift unit, as suggested by Cheng et al. (2009). In this approach a FS is conceptualized as a co-selection of words to convey a specific meaning. Cheng et al. (2009) argued by looking at concgrams (which operationalize meaning shift units), we would be able to uncover the phraseological profile of the language more completely than by looking at collocations. While this seems to be an interesting idea, it has two main limitations. Firstly, as already discussed, semantically incomplete phrases seem to be formulaic as well. And secondly, while conceptually the idea of concgrams is very appealing, it is still not clear how to systematically approach the actual data to investigate this phenomenon.

Overall, Study 1 raised the question of how to best categorize FSs and approach their extraction. Different approaches of extracting FSs led to the extraction of sequences of two different types: those with underlying meaning and those with frequency as a defining feature. The frequencies of use of those two types of FSs differed considerably. Therefore, it might be the case that their storage in the mental lexicon and their processing would differ as well. This conclusion once again warns against generalizing the findings of one type of FSs to FL overall.

7.1.2 Processing of formulaic sequences

Following on the issue of flexibility of FSs, the two processing studies of the thesis investigated if FSs retain their processing advantage when they are presented in a non-fixed form. As Study 1 suggested that the flexibility of form is particularly common for collocations, they were chosen as the FSs to focus on. There are various potential types of form variability in collocations: positional variability (*to spend time* versus *time spent*), morphological variability (*to spend time* versus *spending time*), and non-adjacency (*to spend time* versus *to spend a lot of time*). To my knowledge, none of these aspects of variation have been extensively addressed in processing or acquisition studies.

Studies 2 and 3, therefore, looked at one aspect of variability, namely adjacency, which seems to be particularly common for verb-noun collocations. Collocations have been shown to have a processing advantage over matched control phrases (e.g., Durrant & Doherty, 2010; Sonbul, 2015). Studies 2 and 3 investigated whether the same kind of processing advantage would hold when processing collocations with three intervening words. As such, these studies attempted to bridge a gap between corpus linguistics (where allowing a certain span of the co-occurrence, usually ± 4 words, is a norm) and psycholinguistics (where only the adjacent collocations have mostly been investigated).

Study 2 looked at NSs of English and showed that they read non-adjacent collocations faster than matched control phrases even if the effect of collocation status seems to be stronger for adjacent items than for non-adjacent ones. So it seems that verb-noun collocations are not fixed phrases in the mental lexicon, but rather allow certain flexibility, still retaining their formulaic nature. This finding can be easily accounted for by the usage-based approach to language as verb-noun collocations are very frequently used in their non-adjacent forms. Therefore, non-adjacent collocations could potentially be learned as non-adjacent dependencies, forming some sort of links between words, but not creating a holistic entry of the phrases.

This finding is interesting to discuss in light of the findings on other types of collocations and idiomatic phrases. Adjective-noun and noun-noun collocations seem to lose any processing advantage once their elements are split apart (Carrol, 2015). Hence, there seems to be differences in processing of different collocations. However, this difference seems to be rather predictable: as indicated by the analysis of variation in collocation forms in Chapter 3, verb-noun collocations are much more frequently used non-adjacently than other types of collocations. If for other types of collocations language users do not receive enough exposure to their non-adjacent forms, those non-adjacent forms are not predictable and not facilitated in processing.

As far as flexibility of idioms is concerned, some studies suggested that once an idiom is split apart, its processing advantage disappears (e.g., McGlone, Glucksberg, & Cacciari, 1994), while others seem to show that even when idioms are modified, they retain some of the processing advantage (e.g., Carrol, 2015). However, even if modified idioms lost the processing advantage, this difference between collocations and idioms would seem to be easy to explain. Collocations are items defined by the frequency of co-occurrence of their elements. Therefore, their formulaicity seems to be driven by probabilities and tendencies of use, creating certain expectancies of the words to follow each other (Wray, 2002). Idioms, on the other hand, are defined by their figurative meaning and at least this meaning has to be stored holistically. So it might be that for idioms, both the holistic and the compositional routes of access are available, as predicted by dual route processing models (e.g., Abel, 2013, Van Lancker Sidtis, 2012). The existence of the direct access option for collocations is somewhat questionable, though. First of all, their meanings are more or less compositional. Also, as non-adjacent collocations are processed faster, we would have to assume that a collocation is retrieved directly when it is adjacent, but the elements of the collocation pre-activate each other as well, leading to a computational advantage when the collocation is non-adjacent. While this could be the case, it would raise a question whether the holistic entry is necessary as faster mapping of the elements can account for processing advantages for both adjacent and non-adjacent items. Another option could be considering the idea of a formuleme proposed by Van Lancker Sidtis (2012). According to this idea, collocations, as any other FSs, would have their core form stored in the mental lexicon and then modified in use. It would remain unclear, though, what a core form for collocations would be: their adjacent form (*to spend time*), non-adjacent form with optional open slots (*to spend (xxx) time*) or, depending on usage, an adjacent form for some collocations and a non-adjacent form for others. The design and the results of the current study do not allow choosing one of the discussed options.

Drawing on the discussion above, it seems that it might be reasonable to move away from the question whether FSs are processed faster due to holistic storage or faster mapping of their elements. It seems to be likely that the answer is different for different types of FSs. When it comes to verb-noun collocations, though, I would argue that holistic storage, at least at the lexical level, is not very likely. Rather collocations seem to be words connected by some sort of associative links (which is fully in line with the suggestions of the lexical priming theory (Hoey, 2005, 2012)). These links help a language user anticipate the second element of the collocation when encountering the first one. Predictability, at various levels of language granularity, was shown to be a core aspect of language processing (Kuperberg & Jaeger, 2016) and the elements of verb-noun collocations seem to remain predictable even when there are other words intervening between them.

Study 3 replicated Study 2 with NNSs of English in order to investigate if the same collocation effect for non-adjacent collocations would hold for language learners. The results showed that adjacent collocations are processed faster than control phrases (as already suggested by Sonbul, 2015; Wolter & Gyllstad, 2013), but this processing advantage does not hold for non-adjacent collocations. As the study looked at advance learners of English, this finding suggests that NNSs struggle to reach native-like processing advantage for collocations.

There seems to be two plausible explanations for this finding. One the one hand, the integration of the inserted words in the phrase might be more demanding for NNSs and therefore it might cancel out any facilitative effects for collocations. On the other hand, language learners have inevitably had much less exposure to English than NSs. Therefore, it might be that for them the links between the collocates are simply weaker due to less entrenchment. Therefore, the difference in the results of NSs and NNSs does not necessarily suggest qualitatively different processing of non-adjacent collocations. It might simply show the weaker association between the collocates in the mental lexicon of NNSs.

This difference in findings between NSs and NNSs brings up an important issue that FSs we identify in different corpora are not necessarily formulaic for all language users. As the target collocations in Study 2 and Study 3 were relatively frequent, they might as well have been formulaic for all the NSs in the study. However, this might not be the case for all the FSs: some rarely used items might differ in psychological reality even for NSs. Coming back to the idea of the holistic storage, then, it could be possibly argued, that it might be a fluid concept at the level of individual speaker as well. For instance, if a restrictive collocation is first encountered in its adjacent form only, it can be initially stored as a whole, but later (with subsequent exposures to its non-adjacent form and other variations) it can move on from the holistic representation to a more flexible one. Even if this is admittedly only a speculation, it seems to be important to keep in mind these potential individual variations. Usage-based approach conceptualizes language as a system shaped by experience (Bybee, 2006). Thus, every language user have their individual experiences with language and, consequently, slightly idiosyncratic mental lexicons.

7.1.3. Incidental acquisition of formulaic sequences

Building on the idea that non-adjacency might be difficult master for NNSs; Study 4 was designed to address the question whether non-adjacent items can be learned incidentally from reading. There have been a number of studies suggesting that in general incidental acquisition of collocations is possible (Pellicer-Sánchez, 2015; Webb et al., 2013). Study 4 built on these findings, showing that incidental acquisition of collocations is not only possible but also occurs when learners are reading complex texts and focusing on the academic content. Also, it suggested that non-adjacent verb-noun collocations are learned equally well (or equally slowly, depending on how one wants to interpret the learning rate in incidental acquisition studies) as adjacent collocations.

The finding that non-adjacent collocations are learned from exposure is important, as (at least verb-noun) collocations seem to be used non-adjacently very frequently. If they were not learned when encountered non-adjacently, they would probably have very little chances of being learned incidentally from exposure. It seems, though, that non-adjacency is not a problem for acquisition. So even when encountering a non-adjacent form of a collocation, the link between the collocates is strengthened. Therefore, the lack of processing advantage in Study 3 seems to be due to the link between the elements of the collocations being not strong enough, rather than non-adjacent collocations not being learned.

7.2 Limitations and directions for future research

The limitations of each study are already addressed in the chapters presenting the studies. Therefore, in this section, I will focus on more general limitations that could also be taken as directions for future studies.

First of all, three of the studies focused on the use of only one type of FSs, namely, collocations. As already argued in Study 1, FL is a heterogeneous phenomenon. Therefore, the results of the processing and acquisition studies in this thesis cannot be generalized to all the FSs; rather they are limited to verb-noun collocations only.

For example, Study 2 suggested that collocations are flexible in their forms and do not lose their formulaic nature when used in a non-adjacent form. The research on form variation of idioms suggests that the results for collocations are probably not generalizable to idioms (e.g., McGlone et al., 1994). The items like lexical bundles, that are extracted based on their fixedness, would probably lose their formulaic nature if split up as well, even if research seems to suggest that manipulating the form of a recurrent lexical string does not necessarily hinder its processing (Molinaro et al., 2013). However, further research on FSs that allow variable slots (e.g., collocational frameworks) would be interesting. I would speculate that they should show similar processing advantages as non-adjacent collocations. This remains an empirical question, though.

Even if we are careful not to overgeneralize the findings of the studies to other types of FSs, the question remains whether they are generalizable to all types of collocations. Research on adjective-noun and noun-noun collocations (e.g., Carrol, 2015) seems to suggest that they behave differently from the verb-noun collocations addressed in the thesis. This is the case even if they are still lexical collocations, so generalizing the findings to grammatical collocations would be even more problematic. Overall, more research on different categories of collocations would be needed in order to understand which types of collocations allow flexibility and which ones are more fixed in their forms.

Furthermore, the studies of this thesis only addressed one aspect of variation of collocations: non-adjacency. Positional variation or morphological variation are still to be explored. Positional variation has been addressed a little bit in processing studies (Bonk & Healy, 2005), but the findings are not conclusive. To my knowledge, no studies so far have investigated the morphological variation of the elements of collocations. While in English this might not be such an important issue (Stubbs, 2009), in other languages with more complex morphological systems it could be.

It would also be interesting to consider, how much we can alter a form of a FS for it to still be recognized as formulaic. In idiom research, it was suggested that we do have quite a lot of freedom, but we have to retain enough of the elements of an idiom (at least enough of the underlying concepts) for the idiom to be recognizable (Omazic, 2008). While the collocations can potentially allow lot of various insertions, the length of insertion for the sequence to be recognized as a collocation is probably limited. These could be interesting questions for further investigations.

Finally, and this limitation unfortunately holds for a large part of the studies on FL, only English language data were analysed in this thesis. Therefore, the findings are definitely not easily generalizable for other languages or for Language as a phenomenon. Moon suggested that "[t]he phenomenon of multi-word items and idiomaticity are generally held to be universals in natural languages" (1998c, p. 57). While the central principles of the use of FL would most probably hold for any specific language due to the underlying principles of human cognition, different language systems would very likely lead to different findings on specific aspects of use of FSs. For example, Granger (2014) looked only at lexical bundles in English and French and already showed that different frequency thresholds might be needed for different languages and that sequences can allow different degrees of variation in different languages. This leads me to believe that differences between languages would be even more prominent if less related languages were studied. Some of the main issues addressed in this thesis, such as flexibility of FSs, could become even more central to the languages with more complex morphological systems or more flexible word order. Frequency as the only criterion could hardly capture even relatively fixed sequences and positional variation of words in sequences could be much more common than in English. Therefore, the speakers of the language might be exposed to slightly different patterns, very often not simply recurrent adjacent sequences but some sort of variable expressions. The analysis of flexibility of FSs in languages other than English could add considerably to our understanding of language processing and acquisition.

7.3 Concluding remarks

While the criterion of fixedness has traditionally been an important part of definitions of FSs, it seems to be more important for some FSs than for the others. Although fixed form might define lexical bundles and (at least) a part of idioms, collocations seem to stand out as a category that allows quite a lot of flexibility in use, without losing the formulaic nature and processing facilitation.

This thesis addressed the question of non-adjacency in verb-noun collocations. It showed that non-adjacency is an issue for extracting collocations from corpora, but also that non-adjacent collocations are read faster by the NSs of English and can be learned incidentally from exposure as well as adjacent collocations. However, these findings seem to be only a first step towards better understanding the flexibility of collocations and potentially other formulaic sequences.

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.

Adolphs, S., & Durrow, V. (2004). Social-cultural integration and the development of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use* (pp. 107–126). Amsterdam; Philadelphia: John Benjamins Publishing.

Alali, F. A., & Schmitt, N. (2012). Teaching formulaic sequences: The same as or different from teaching single words? *TESOL Journal*, *3*(2), 153–180.

Allwood, J. (2012). Register. In *The Encyclopaedia of Applied Linguistics*. Blackwell Publishing Ltd. doi: 10.1002/9781405198431.wbeal0469.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology : Theory, Analysis, and Applications* (pp. 101–122). Oxford: Clarendon Press.

Anthony, L. (2011). AntConc (Version 3.2.4.). Japan: Waseda University.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge; New York: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bahns, J. (1993). Lexical collocations: a contrastive view. *ELT Journal*, *47*(1), 56–63.

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., Lieven, E., & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences*, *106*(41), 17284–17289.

Barfield, A., & Gyllstad, H. (2009). Introduction: Researching L2 Collocation Knowledge and Development. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: multiple interpretations* (pp. 1–18). New York: Palgrave Macmillan.

Barfield, P. A., & Gyllstad, D. H. (2009). *Researching collocations in another language: Multiple interpretations*. New York: Palgrave Macmillan.

Barlow, M. (2011). Corpus linguistics and theoretical linguistics. *International Journal of Corpus Linguistics*, *16*(1), 3–44.

Barlow, M., & Kemmer, S. (Eds.). (2000). *Usage-based models of language*. Stanford, California: The Center for the Study of Language and Information Publications.

Barnbrook, G. (1996). *Language and computers: A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.

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., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package (Version 1.1-7). Retrieved from http://CRAN.R-project.org/package=lme4>

Baxter, J. (2011). 23. Gender. In S. James (Ed.), *The Routledge Handbook of Applied Linguistics* (pp. 331–343). London ; New York: Routledge.

Biber, D. (1993). Co-occurrence patterns among collocations: a tool for corpus-based lexical knowledge acquisition. *Computational Linguistics*, *19*(3), 531–538.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam; Philadelphia: John Benjamins Publishing.

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

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*(3), 263–286.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at …: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371–405.

Biber, D., & Finegan, E. (1988). Drift in three English Genres from the 18th to the 20th century. In M. Kytö & O. Ihalainen (Eds.), *Corpus Linguistics, Hard and Soft: Proceedings of the Eighth International Conference on English Language Research on Computerized Corpora* (pp. 83–103). Amsterdam: Rodopi.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Bishop, H. (2004). Noticing formulaic sequences. A problem of measuring the subjective. *LSO Working Papers in Linguistics*, *4*, 15–19.

Blais, M. J., & Gonnerman, L. M. (2013). Explicit and implicit semantic processing of verb–particle constructions by French–English bilinguals. *Bilingualism: Language and Cognition*, *16*(4), 829–846.

Bobrow, S. A., & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, *1*(3), 343–346.

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb–noun collocations. *Language Teaching Research*, *18*(1), 54–74.

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, *32*, 83–110.

Bonk, W., & Healy, A. F. (2005). Priming effects without semantic or associative links through collocation. Presented at the 46th Annual Meeting of the Psychonomic Society, Toronto, Canada.

Brown, J. (2014). Knowledge of collocations. In J. Milton & T. Fitzpatrick (Eds.), *Dimensions of Vocabulary Knowledge* (pp. 123–139). Basingstoke: Palgrave Macmillan.

Burnard, L. (2008). Reference Guide to BNC Baby. Retrieved from http://www.natcorp.ox.ac.uk/corpus/baby/manual.pdf

Butler, C. S. (1998). Collocational Frameworks in Spanish. *International Journal of Corpus Linguistics*, *3*(1), 1–32.

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

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*(6), 668–683.

Cambridge Dictionaries Online. (2014). Retrieved from http://dictionary.cambridge.org/

Carrol, G. (2015). *Found in Translation: a psycholinguistic investigation of idiom processing in native and non-native speakers*. University of Nottingham.

Carrol, G., & Conklin, K. (2014). Eye-tracking multi-word units: some methodological questions. *Journal of Eye Movement Research*, *7*(5), 1-11.

Čermák, F. (2001). Substance of idioms: Perennial problems, lack of data or theory? *International Journal of Lexicography*, *14*(1), 1–20.

Cheng, W. (2008). Concgramming: A corpus-driven approach to learning the phraseology of discipline-specific texts. *CORELL: Computer Resources for Language Learning*, *1*(1), 22–35.

Cheng, W., Greaves, C., Sinclair, J., & Warren, M. (2009). Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied Linguistics*, *30*(2), 236–252.

Cheng, W., Greaves, C., & Warren, M. (2006). From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics*, *11*(4), 411–433.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22–29.

Citron, F. M. M., Oberecker, R., Friederici, A. D., & Mueller, J. L. (2011). Mass counts: ERP correlates of non-adjacent dependency learning under different exposure conditions. *Neuroscience Letters*, *487*(3), 282–286.

Clear, J., & Tognini-Bonelli, E. (1993). Computational Tools for the Study of Collocation. In M. Baker & G. Francis (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 271–292). Amsterdam: John Benjamins Publishing.

Colson, J. P. (2008). Cross-linguistic phraseological studies: An overview. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 191–205). Amsterdam: John Benjamins Publishing.

Columbus, G. (2011). Processing MWUs: Are MWU Subtypes Psycholingustically real? In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 194–210). London ; New York: Continuum.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1), 72–89.

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

Conrad, S., & Biber, D. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, *20*, 56–71.

Cooper, T. C. (1998). Teaching idioms. *Foreign Language Annals*, *31*(2), 255–266.

Corrigan, R., Moravcsik, E., Ouali, H., & Wheatley, K. (2009a). *Formulaic Language: Acquisition, loss, psychological reality, and functional explanations*. Amsterdam; Philadelphia: John Benjamins Publishing.

Corrigan, R., Moravcsik, E., Ouali, H., & Wheatley, K. (2009b). *Formulaic Language: Distribution and historical change*. Amsterdam; Philadelphia: John Benjamins Publishing.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(4), 397–423.

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, *17*(4), 391–406.

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

Coxhead, A. (1998). *An academic word list.* ELI Occasional Publications 18, School of Linguistics and Applied Language Studies, Victoria University of Wellington: Wellington.

Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2015). Assessing Lexical Proficiency Using Analytic Ratings: A Case for Collocation Accuracy. *Applied Linguistics*, *36*(5), 570–590.

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

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: A case for contrastive analysis. *Studies in Second Language Acquisition*, *7*(1), 73–79.

Dahlmann, I., & Adolphs, S. (2007). Pauses as an indicator of psycholinguistically valid multi-word expressions (MWEs)? *MWE '07 Proceedings of the Workshop on a Broader Perspective on Multiword Expressions*, 49–56.

Daudaravičius, V., & Marcinkevičienė, R. (2004). Gravity counts for the boundaries of collocations. *International Journal of Corpus Linguistics*, *9*(2), 321–348.

Davies, M. (2004). BYU-BNC. (Based on the British National Corpus from Oxford University Press). Retrieved from http://corpus.byu.edu/bnc/

Davies, M. (2008). The corpus of contemporary American English: 450 million words, 1990-present. Retrieved from http://corpus.byu.edu/coca/

Day, R., R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, *7*(2), 541–551.

Dörnyei, Z., Durrow, V., & Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use* (pp. 87–106). Amsterdam; Philadelphia: John Benjamins Publishing.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1), 61–74.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, *28*(3), 157–169.

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, *6*(2), 125–155.

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL-International Review of Applied Linguistics in Language Teaching*, *47*(2), 157–177.

Ellis, N. C. (1994). Introduction. Implicit and explicit language learning: An Overview. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 1–32). London; San Diego: Academic Press.

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

Ellis, N. C. (2008a). Implicit and explicit knowledge about language. In J. Cenoz & Hornberger, *Encyclopaedia of Language and Education* (2nd ed., Vol. 6: Knowledge about Language, pp. 1–13). New York: Springer Science+Business Media LLC.

Ellis, N. C. (2008b). Sequencing in SLA. *Studies in Second Language Acquisition*, *18*(1), 91–126.

Ellis, N. C. (2012a). Corpus Study: Cognitive Implications. In *The Encyclopaedia of Applied Linguistics*. Blackwell Publishing Ltd. doi: 10.1002/9781405198431.wbeal0261.

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

Ellis, N. C. (2015). Implicit AND explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and Explicit Learning of Languages* (pp. 1–24). Amsterdam: John Benjamins Publishing Company.

Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistic. Introduction to the special issue. *Applied Linguistics*, *27*(4), 558–589.

Engbert, R., & Kliegl, R. (2011). Parallel graded attention models of reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 787–801). Oxford ; New York: Oxford University Press.

Erman, B. (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, *12*(1), 25–53.

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.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London: SAGE.

Firth, J. R. (1957). A synopsis of linguistic theory 1930-1955. In F. Palmer (Ed.), *Selected papers of Firth 1952-1959* (pp. 168–205). London: Longman.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 75–93). Harlow: Longman.

Fox, J. (2009). Effect displays in R for Generalised Linear Models. *Journal of Statistical Software*, *8*(15), 1–27.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology. Learning, Memory & Cognition*, *31*(5), 862–877.

Gabrielatos, C., & Marchi, A. (2012). Keyness: Appropriate metrics and practical issues. Presented at the CADS International Conference, Bologna.

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly*, *41*(2), 339–359.

Garnier, M., & Schmitt, N. (2015). The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, *19*(6), 645–666.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. Psychonomic Bulletin & Review, 16(3), 486–490.

Gibbs Jr., R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, *28*(5), 576–593.

Gilchrist, I. D. (2011). Saccades. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 85–95). Oxford ; New York: Oxford University Press.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183–206.

Gonnerman, L. M., & Hayes, C. (2005). The professor chewed the students… out: Effects of dependency, length, and adjacency on word order preferences in sentences with verb particle constructions. *Proceedings of the Twenty- Seventh Annual Conference of the Cognitive Science Society*, 785–790.

González Fernández, B., & 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.

Granger, S. (2014). A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast*, *14*(1), 58–72.

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 27–46). Amsterdam; Philadelphia: John Benjamins Publishing.

Grant, L., & Nation, P. (2006). How many idioms are there in English? *ITL International Journal of Applied Linguistics*, *151*, 1–14.

Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, *18*(1), 109–136.

Greaves, C. (2009). *ConcGram*. John Benjamins Publishing.

Gries, S. T. (2008). Phraseology and linguistic theory. A brief survey. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 3–25). Amsterdam; Philadelphia: John Benjamins Publishing.

Gries, S. T. (2010a). Corpus linguistics and theoretical linguistics. A love-hate relationship? Not necessarily... *The Bootcamp Discourse and Beyond. Special Issue of International Journal of Corpus Linguistics*, *15*(3), 327–343.

Gries, S. T. (2010b). Useful statistics for corpus linguistics. In A. Sánchez & M. Almela (Eds.), *A mosaic of corpus linguistics: selected approaches* (pp. 269–291). Frankfurt am Main: Peter Lang.

Gries, S. T. (2013). 50-something years of work on collocations: What is or should be next. *International Journal of Corpus Linguistics*, *18*(1), 137–166.

Gries, S. T. (2014). Frequency tables, effect sizes, and explorations. In D. Glynn & J. Robinson, *Corpus methods for semantics: quantitative studies in polysemy and synonymy* (pp. 365–389). Amsterdam; Philadelphia: John Benjamins Publishing.

Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL - International Journal of Applied Linguistics*, *166*(2), 278–306.

Gyllstad, H., & Wolter, B. (2015). Collocation processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*. doi: 10.1111/lang.12143

Hoey, M. (2004). The textual priming of lexis. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and Language Learners* (pp. 21–42). Amsterdam; Philadelphia: John Benjamins Publishing.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. New York: Routledge.

Hoey, M. (2009). Corpus driven approaches to grammar. In U. Römer & R. Schulze (Eds.), *Exploring the Lexis-grammar Interface* (pp. 34–47). Amsterdam; Philadelphia: John Benjamins Publishing.

Hoey, M. (2012). Lexical Priming. In *The Encyclopaedia of Applied Linguistics*. Blackwell Publishing Ltd. doi: 10.1002/9781405198431.wbeal0694

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *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.

Hothorn, T., & Hornik, K. (2013). *exactRankTests: Exact Distributions for Rank and Permutation Tests. R package version 0.8-27.* Retrieved from http://CRAN.R-project.org/package=exactRankTests

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, *19*(1), 24–44.

Hu, H. M. (2013). The effects of word frequency and contextual types on vocabulary acquisition from extensive reading: A case study. *Journal of Language Teaching and Research*, *4*(3), 487–495.

Hulstijn, J. H. (2003). Incidental and Intentional Learning. In K. J. Doughty & M. H. Long, *In The Handbook of Second Language Acquisition.* (pp. 349–381). Berlin; Victoria: Blackwell Publishing.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, S. (2012). Corpus Linguistics: Historical Development. In *The Encyclopaedia of Applied Linguistics*. John Wiley & Sons, Inc. doi: 10.1002/9781405198431.wbeal0257

Hyönä, J. (2011). Foveal and parafoveal processing during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 819–838). Oxford ; New York: Oxford University Press.

IBM Corp. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Irujo, S. (1986). A piece of cake: learning and teaching idioms. *ELT Journal*, *40*(3), 236–242.

Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, *91*(3), 433–445.

Johnson, J. (2012). Bootstrapping language: Are infant statisticians up to the job? In P. Rebuschat & J. N. Williams (Eds.), *Statistical Learning and Language Acquisition* (pp. 55–90). Boston; Berlin: de Gruyter Mouton.

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use* (pp. 269–301). Amsterdam; Philadelphia: John Benjamins Publishing.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology. Learning, Memory & Cognition*, *29*(6), 1312–1318.

Kemmer, S., & Barlow, M. (2000). Introduction: A Usage-Based Conception of Language. In M. Barlow & S. Kemmer (Eds.), *Usage Based Models of Language* (pp. vii–xxviii). Stanford, California: The Center for the Study of Language and Information Publications.

Kennedy, G. (2003). Amplifier collocations in the British National Corpus: Implications for English language teaching. *TESOL Quarterly*, *37*(3), 467–487.

Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory*, *1*(2), 263–276.

Kilgarriff, A. (2007). Googleology is bad science. *Computational Linguistics*, *33*(1), 147–151.

Kilgarriff, A. (n.d.). BNC database and word frequency lists. Retrieved from http://www.kilgarriff.co.uk/bnc-readme.html#lemmatised

Kim, S. H., & Kim, J. H. (2012). Frequency effects in L2 multiword unit processing: Evidence from self-paced reading. *TESOL Quarterly*, *46*(4), 831–841.

Kiss, G. R., Milroy, C., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The Computer and Literary Studies*. Edinburgh: University Press.

KLC. (2011). *DONELAITIS*. Kaunas: VDU.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35.

Kremmel, B., Brunfaut, T., & Alderson, J. C. (2015). Exploring the role of phraseological knowledge in foreign language reading. *Applied Linguistics*. doi: 10.1093/applin/amv070

Krishnamurthy, R., Sinclair, J., Daley, R., & Jones, S. (2004). *English collocation studies: the OSTI report*. London: Continuum.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843.

Kuiper, K., Van Egmond, M.-E., Kempen, G., & Sprenger, S. (2007). Slipping on superlemmas. *Multi-Word Lexical Items in Speech Production. The Mental Lexicon*, *2*, 313–357.

Kuperberg, G. R., & Jaeger, F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59.

Kuznetsova, A., Brockhoff, P. B., & Bojesen Christensen, B. (2015). lmerTest: Tests in Linear mixed effects models (Version 2.0-29). Retrieved from http://CRAN.R-project.org/package=lme4>

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.

Langacker, R. W. (2000). A Dynamic Usage-Based Model. In M. Barlow & S. Kemmer (Eds.), *Usage-Based Models of Language* (pp. 1–64). Stanford, California: The Center for the Study of Language and Information Publications.

Langlotz, A. (2006). *Idiomatic Creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English* (Vol. 17). Amsterdam: John Benjamins Publishing Company.

Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, *19*(12), 1247–1252.

Laufer, B. (1989). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review*, *6*(1), 10–20.

Laufer, B. (2000). Avoidance of idioms in a second language: The effect of L1-L2 degree of similarity. *Studia Linguistica*, *54*(2), 186–196.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language*, *22*(1), 15–30.

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

Lewis, M. (1997). *Implementing the Lexical Approach: Putting Theory Into Practice* (1st edition). Hove: Cengage Learning.

Lewis, M. (2000). Learning in the lexical approach. In M. Lewis (Ed.), *Teaching collocation: Further developments in the Lexical Approach* (pp. 155–184). Hove: Language Teaching Publications.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, *18*(2), 85–102.

Lin, P. M. S. (2014). Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System*, *42*, 164–176.

Liu, D. (2003). The most frequently used Spoken American English idioms: A corpus analysis and its implications. *TESOL Quarterly*, *37*(4), 671–700.

Liu, D. (2011). The most frequently used English phrasal verbs in American and British English: A multicorpus examination. *TESOL Quarterly*, *45*(4), 661–688.

Macis, M. & Schmitt, N. (2016). Meaning matters: The figurative and polysemous nature of collocations and their place in ELT. *ELT Journal,* doi: 10.1093/elt/ccw044.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, Mass.: MIT Press.

Marco, M. J. L. (2000). Collocational frameworks in medical research papers: A genre-based study. *English for Specific Purposes*, *19*(1), 63–86.

Martinez-Conde, S., & Macknik, S. L. (2011). Microsaccades. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 95–115). Oxford ; New York: Oxford University Press.

Martinez, R. (2011). *The development of a corpus-informed list of formulaic sequences for language pedagogy*. University of Nottingham, Nottingham.

Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, *45*(2), 267–290.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, *33*(3), 299–320.

Matlock, T., & Herediag, R. R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 251–274). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.

McCarthy, M. (1990). *Vocabulary*. Oxford; New York; Toronto: Oxford University Press.

McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. In *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 1619–24).

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.

McGlone, M. S., Glucksberg, S., & Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Processes*, *17*(2), 167–190.

Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics*, *32*, 111–129.

Michael, A. (2009). *Slang: The People's Poetry*. Cary: Oxford University Press.

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

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use* (pp. 57–78). European Second Language Association.

Molinaro, N., Canal, P., Vespignani, F., Pesciarelli, F., & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? A reading time and ERP study of collocational complex prepositions. *Language and Cognitive Processes*, *28*(6), 1–27.

Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology*, *83*(3), 176–190.

Mondini, S., Jarema, G., Luzzatti, C., Burani, C., & Semenza, C. (2002). Why is 'Red Cross' different from 'Yellow Cross'?: A neuropsychological study of noun–adjective agreement within Italian compounds. *Brain and Language*, *81*(1), 621–634.

Mondini, S., Luzzatti, C., Saletta, P., Allamano, N., & Semenza, C. (2005). Mental representation of prepositional compounds: Evidence from Italian grammatical patients. *Brain and Language*, *94*(2), 178–187.

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal*, *99*(2), 308–328.

Moon, R. (1998a). *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford: Clarendon Press.

Moon, R. (1998b). Frequencies and forms of phrasal lexemes in English. In A. P. Cowie (Ed.), *Phraseology : Theory, Analysis, and Applications* (pp. 79–100). New York: Oxford University Press.

Moon, R. (1998c). Vocabulary connections: multi-word items in English. In M. McCarthy & N. Schmitt (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 40–63). Cambridge; New York: Cambridge University Press.

Nagy, W. E., Herman, P. A., Anderson, R. C., & Pearson, P. D. (1984). *Technical Report No. 319: Learning words from context*. Illinois: University of Illinois.

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

Nation, I. S. P., & Webb, S. (2010). *Researching and Analyzing Vocabulary* (1st edition). Boston: Heinle ELT.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford; New York: Oxford University Press.

Nekrasova, T. M. (2009). English L1 and L2 speakers' knowledge of lexical bundles. *Language Learning*, *59*(3), 647–686.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Retrieved from http://w3.usf.edu/FreeAssociation/.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, *11*(3), 283–304.

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; Philadelphia: John Benjamins Publishing.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162.

O'Donnell, M. B., Romer, U., & Ellis, N. C. (2013). The development of formulaic sequences in first and second language writing: Investigating effects of frequency, association, and native norm. *International Journal of Corpus Linguistics*, *18*(1), 83–108.

O'Donnell, M. B., Scott, M., Mahlberg, M., & Hoey, M. (2012). Exploring text-initial words, clusters and concgrams in a newspaper corpus. *Corpus Linguistics and Linguistic Theory*, *8*(1), 73–101.

Omazic, M. (2008). Processing of idioms and idiom modifications. A view from cognitive linguistics. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 67–79). Amsterdam; Philadelphia: John Benjamins Publishing.

Onnis, L. (2012). The potential contribution of statistical learning to second language acquisition. In J. Williams & P. Rebuschat (Eds.), *Statistical learning and language acquisition*. Berlin: Walter de Gruyter.

Pace-Sigge, M. T. L. (2013). The concept of Lexical Priming in the context of language use. *ICAME Journal*, 37, 149–174.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–227). London: Longman.

Pellicer-Sánchez, A. (2015). Learning L2 collocations incidentally from reading. *Language Teaching Research*. doi: 10.1177/1362168815618428.

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

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* (Vol. 198). Cambridge: Cambridge University Press.

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.

Philip, G. (2008). Reassesing the canon. Fixed phrases in general reference corpora. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 95–108). Amsterdam: Philadelphia: John Benjamins Publishing.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*(1), 1–28.

Pinker, S. (1998). Words and rules. *Lingua*, *106*(1–4), 219–242.

Plonsky, L., & Oswald, F. L. (2014). How big is ‘big’? Interpreting effect sizes in L2 Research. *Language Learning*, *64*(4), 878–912.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425.

Ravelle, W. (2014). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, USA. Retrieved from http://CRAN.R-project.org/package=psych Version=1.4.8

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z reader model. *Journal of Experimental Psychology. Human Perception & Performance*, *30*(4), 720–732.

Rayner, K., & Liversedge, S. P. (2011). Linguistic and cognitive influences on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 751–766). Oxford ; New York: Oxford University Press.

Rayner, K., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text. Cognitive control of eye movements. *Psychological Science*, *14*(4), 385–388.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290–1301.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509.

R Core Team. (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reali, F., & Christiansen, M. H. (2007). Word chunk frequencies affect the processing of pronominal object-relative clauses. *The Quarterly Journal of Experimental Psychology*, *60*(2), 161–170.

Rea, L. M., & Parker, R. A. (2005). *Designing and Conducting Survey Research: A Comprehensive Guide* (3rd edition). San Francisco: Jossey-Bass.

Rebuschat, P., & Williams, J. (2012). Introduction: Statistical learning and language acquisition. In P. Rebuschat & J. Williams (Eds.), *Statistical learning and language acquisition* (pp. 1–13). Boston; Berlin: de Gruyter Mouton.

Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer, B. Altenberg, & H. Longman (Eds.), *English Corpus Linguistics* (pp. 128–143). New York: Longman.

Revier, R. L. (2009). Evaluating a New Test of Whole English Collocations. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: multiple interpretations* (pp. 125–138). New York: Palgrave Macmillan.

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, *35*(2), 213–235.

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. Victoria University of Wellington.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, *25*(5), 762–776.

Rosario-Martinez, H. (2015). phia: Post-hoc interaction analysis. R package (Version 0.2-0). Retrieved from http://CRAN.R-project.org/package=phia

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition*, *69*(2), 105–134.

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, *6*(2), 72–78.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, *14*, 357–357.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*(1), 206–226.

Schmitt, N. (Ed.). (2004). *Formulaic Sequences: Acquisition, Processing, and Use*. Amsterdam; Philadelphia: John Benjamins Publishing.

Schmitt, N. (2005). Formulaic language: fixed and varied. *Elia*, *6*(6), 13–39.

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

Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-driven recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic Sequences* (pp.127-152). Amsterdam; Philadelphia: John Benjamins Publishing.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88.

Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 173–189). Amsterdam; Philadelphia: John Benjamins Publishing.

Scott, M. (2012). WordSmith Tools (Version 6). Liverpool: Lexical Analysis Software.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493.

Serrano, R., Stengers, H., & Housen, A. (2015). Acquisition of formulaic sequences in intensive and regular EFL programmes. *Language Teaching Research*, *19*(1), 89–106.

Shin, D., & Nation, P. (2008). Beyond single words: the most frequent collocations in spoken English. *ELT Journal*, *62*(4), 339–348.

Shukla, M., Gervain, J., Mehler, J., & Nespor, M. (2012). Linguistic constraints on statistical learning in early language acquisition. In P. Rebuschat & J. N. Williams (Eds.), *Statistical Learning and Language Acquisition* (pp. 171–202). Boston; Berlin: de Gruyter Mouton.

Simpson, J., & Speake, J. (2008). *A Dictionary of Proverbs* (5th edition). Oxford: Oxford University Press.

Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly*, *37*(3), 419–441.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. Applied Linguistics, 31(4), 487–512.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (1996). The empty lexicon. *International Journal of Corpus Linguistics*, *1*(1), 99–119.

Sinclair, J. (1998). The lexical item. In E. Weigand (Ed.), *Contrastive Lexical Semantics* (pp. 1–25). Amsterdam; Philadelphia: John Benjamins Publishing.

Sinclair, J. (2000). Lexical grammar. *Darbai ir Dienos*, *24*, 191–203.

Sinclair, J. (2004a). *How to use corpora in language teaching*. Amsterdam; Philadelphia: John Benjamins Publishing.

Sinclair, J. (2004b). *Trust the text: language, corpus and discourse*. New York: Routledge.

Sinclair, J. (2007). Defining the Definiendum. In G. M. de Schryver (Ed.), *A Way with Words: Recent Advances in Lexical Theory and Analysis* (pp. 37–48). Kampala: Menha Publishers Ltd.

Sinclair, J. (2008). Borrowed ideas. In A. Gerbig & O. Mason (Eds.), *Language, People, Numbers: Corpus Linguistics and Society.* (pp. 21–41). Kenilworth: Rodopi.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, *64*(3), 429–458.

Siyanova-Chanturia, A. (2013). Eye-tracking and ERPs in multi-word expression research: A state-of-the-art review of the method and findings. *The Mental Lexicon*, *8*(2), 245–268.

Siyanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language : Corpus linguistics and Linguistic theory. *Corpus Linguistics and Linguistic Theory*, *11*(2), 285–301.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*(2), 251–272.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a phrase 'time and again' matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 776–784.

Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, *36*(5), 549–569.

Sonbul, S. (2012). *The interface of explicit / implicit lexical knowledge: Acquisition and processing of collocations by natives and nonnatives*. University of Nottingham, Nottingham.

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocation 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. *English Language Teaching Journal*, *XXXI*(3), 211–221.

Sosa, A. V., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: collocations involving the word of. *Brain and Language*, *83*(2), 227–236.

Spöttl, C., & McCarthy, M. (2004). Comparing knowledge of formulaic sequences across L1, L2, L3, and L4. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 191–225). Amsterdam; Philadelphia: John Benjamins Publishing

Sprenger, S. A., Levelt, W. J. M., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, *54*(2), 161–184.

Starr, M. S., & Rayner, K. (2001). Eye movements during reading: some current controversies. *Trends in Cognitive Sciences*, *5*(4), 156–163.

Stewart, J. (2012). A multiple-choice test of active vocabulary knowledge. *Vocabulary Learning and Instruction, 1*(1), 53–59.

Stewart, J. (2014). Do multiple-choice options inflate estimates of vocabulary size on the VST? *Language Assessment Quarterly*, *11*(3), 271–282.

Stubbs, M. (2009). Memorial Article: John Sinclair (1933–2007) The Search for Units of Meaning: Sinclair on Empirical Semantics. *Applied Linguistics*, *30*(1), 115–137.

Svensson, M. H. (2008). A very complex criterion of fixedness: Non-compositionality. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 81–93). Amsterdam; Philadelphia: John Benjamins Publishing.

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. (2014). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*. doi: 10.1111/ijal.12092.

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, *37*(4), 529–540.

Teliya, V., Bragina, N., Oparina, E., & Sandomirskaya, I. (1998). Phraseology as a language of culture: Its role in the representation of a cultural mentality. In A. P. Cowie (Ed.), *Phraseology : Theory, Analysis, and Applications* (pp. 55–75). Oxford: Oxford University Press.

Tildė. (2008). *Lietuvių-anglų kalbų žodynas*.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam; Philadelphia: John Benjamins Publishing.

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*(1/2), 61–82.

Tomasello, M. (2005). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, Mass.: Harvard University Press.

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.

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: Acquisition, Processing, and Use* (pp. 153–172). Amsterdam; Philadelphia: John Benjamins Publishing.

Vainio, S., Hyönä, J., & Pajunen, A. (2009). Lexical predictability exerts robust effects on fixation duration, but not on initial landing position during reading. *Experimental Psychology*, *56*(1), 66–74.

Van Lancker Sidtis, D. (2009). Formulaic and novel language in a 'dual process' model of language competence. In R. Corrigan, E. Moravcsik, & H. Quali (Eds.), *Formulaic Language, Volume 2 : Acquisition, Loss, Psychological Reality, and Functional Explanations* (pp. 445–470). Amsterdam; Philadelphia: John Benjamins Publishing Company.

Van Lancker Sidtis, D. (2010). Formulaic expressions in mind and brain: Empirical studies and a dual-processing model of language competence. In J. Guendouzi, E. Loncke, & E. Williams (Eds.), *The Handbook of Psycholinguistic and Cognitive Processes: Perspectives in Communication Disorders* (pp. 247–272). London: Taylor and Francis.

van Heugten, M., & Shi, R. (2010). Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *The Journal of the Acoustical Society of America*, *128*(5), 223–228.

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*(6), 1176–1190.

Van Lancker Sidtis, D. (2012). Two-Track Mind: Formulaic and Novel Language Support a Dual-Process Model. In M. Faust (Ed.), *The Handbook of the Neuropsychology of Language* (p. 342–367). Malden; Oxford: Blackwell Publishing Ltd.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*(3), 609–624.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2009). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, *22*(8), 1682–1700.

Vilkaitė, L. (2016). Formulaic language is not all the same: comparing the frequency of idiomatic phrases, collocations, lexical bundles, and phrasal verbs. *Taikomoji Kalbotyra*, *8*, 28–54.

Vilkaitė, L. (in press). Are non-adjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Vitu, F. (2011). On the role of visual and oculomotor processes in reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 731–751). Oxford ; New York: Oxford University Press.

Vries, M. H. de, Petersson, K. M., Geukes, S., Zwitserlood, P., & Christiansen, M. H. (2012). Processing multiple non-adjacent dependencies: evidence from sequence learning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1598), 2065–2076.

Vuong, L., Meyer, A. S., & Christiansen, M. H. (2011). Simultaneous online tracking of adjacent and non-adjacent dependencies in statistical learning. In C. Carlson, C. Hölscher, & T. Shipley, *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (964-969). Austin, TX: Cognitive Science Society.

Wade, N. J., & Tatler, B. W. (2011). Origins and applications of eye movement research. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 17–45). Oxford ; New York: Oxford University Press.

Webb, S. (2012). Repetition in incidental vocabulary learning. In *The Encyclopaedia of Applied Linguistics*. John Wiley & Sons, Inc. doi: 10.1002/9781405198431.wbeal1425.

Webb, S., & Chang, A. C. S. (2015). How does prior vocabulary knowledge affect vocabulary learning in an extensive reading program? *Studies in Second Language Acquisition*, *37*(4), 651–675.

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, *43*(1), 55–77.

Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*, *32*(3), 259–276.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, *63*(1), 91–120.

Wei, N., & Li, J. (2013). A new computing method for extracting contiguous phraseological sequences from academic text corpora. *International Journal of Corpus Linguistics*, *18*(4), 506–535.

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 collocation processing. *Studies in Second Language Acquisition*, *35*(3), 451–482.

Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, *36*(5), 1193–1221.

Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching*, *32*(4), 213–231.

Wray, A. (2002). *Formulaic Language and the Lexicon*. New York: Cambridge University Press.

Wray, A. (2008). *Formulaic language: pushing the boundaries*. Oxford; New York: Oxford University Press.

Wray, A. (2010). 'We've had a wonderful, wonderful thing': Formulaic interaction when an expert has. *Dementia*, *9*(4), 517–534.

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.

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.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: effects of frequency and contextual richness. *Canadian Modern Language Review*, *57*(4), 541–572.

Zhang, X. (2013). The *I don't know option* in the Vocabulary Size Test. *TESOL Quarterly*, *47*(4), 790–811.

Appendices

Appendix 1

Study 1

Phrasal verbs used in the study

Liu's (Liu, 2011) Phrasal verbs list

1. Back up
2. Blow up
3. Break down
4. Break off
5. Break out
6. Break up
7. Bring about
8. Bring back
9. Bring down
10. Bring in
11. Bring out
12. Bring up
13. Build up
14. Call out
15. Carry on
16. Carry out
17. Catch up
18. Check out
19. Clean up
20. Close down
21. Come about
22. Come along
23. Come around
24. Come back
25. Come down
26. Come in
27. Come off
28. Come on
29. Come out
30. Come over
31. Come through
32. Come up
33. Cut off
34. End up
35. Figure out
36. Fill in
37. Fill out
38. Find out
39. Follow up
40. Get back
41. Get down
42. Get in
43. Get off
44. Get on
45. Get out
46. Get through
47. Get up
48. Give back
49. Give in
50. Give out
51. Give up
52. Go ahead
53. Go along
54. Go around
55. Go back
56. Go down
57. Go in
58. Go off
59. Go on
60. Go out
61. Go over
62. Go through
63. Go up
64. Grow up
65. Hand over
66. Hang on
67. Hang out
68. Hang up
69. Hold back
70. Hold on
71. Hold out
72. Hold up
73. Keep on
74. Keep up
75. Lay down
76. Lay out
77. Line up
78. Look around
79. Look back
80. Look down
81. Look out
82. Look up
83. Make out
84. Make up
85. Move back
86. Move in
87. Move on
88. Move out
89. Move up
90. Open up
91. Pass on
92. Pay off
93. Pick out
94. Pick up
95. Play out
96. Point out
97. Pull back
98. Pull out
99. Pull up
100. Put back
101. Put down
102. Put in
103. Put off
104. Put on
105. Put out
106. Put up
107. Reach out
108. Rule out
109. Run out
110. Send out
111. Set about
112. Set down
113. Set off
114. Set out
115. Set up
116. Settle down
117. Show up
118. Shut down
119. Shut up
120. Sit back
121. Sit down
122. Sit up
123. Slow down
124. Sort out
125. Stand out
126. Stand up
127. Start out
128. Step back
129. Sum up
130. Take back
131. Take down
132. Take in
133. Take off
134. Take on
135. Take out
136. Take over
137. Take up
138. Throw out
139. Turn around
140. Turn back
141. Turn down
142. Turn off
143. Turn out
144. Turn over
145. Turn up
146. Wake up
147. Walk out
148. Wind up
149. Work out
150. Write down

Phrasal verbs added from the PRASE list (Martinez & Schmitt, 2012)

1. account for
2. act on
3. call for
4. call on
5. care for
6. come across
7. consist of
8. focus on
9. get away
10. get into
11. get to
12. go for
13. go into
14. look after
15. look for
16. provide for
17. put forward
18. rely on
19. stand for
20. switch on
21. think about
22. turn into
23. turn on
24. work on

Phrasal verbs added from *The most frequently used spoken idioms* list (Liu, 2003)

1. deal with
2. fall apart
3. follow through
4. get across
5. give away
6. go after
7. go with
8. hand out
9. leave out
10. live up
11. live with
12. pass out
13. run into
14. run through
15. screw up
16. sell out
17. sign off
18. stick with
19. stick to
20. throw away
21. touch on
22. turn in

Added 3-word phrasal verbs

1. Come up with
2. Follow up on
3. Get away with
4. Get back to
5. Get out of
6. Go along with
7. Hang out with
8. Hold on to
9. Keep up with
10. Make up for
11. Make up of
12. Put up with
13. Run out of
14. Wind up in
15. Wind up with

Appendix 2

Study 1

Idiomatic phrases used in the study

Phrases used *The most frequently used spoken idioms list* (Liu, 2003)

(a) level playing field
a rule of thumb
according to
after the fact
all along
all of a sudden
all out
all over again
as a fair game
as a matter of fact
at * expense (of)
at issue
at large
at one's disposal
at stake
back and forth
be better off
be in for
be in place
be on * mind
be on track
be open to ideas
be over * head
be under way
beg the question
bits and pieces
bring into effect
by and large
by hand
call * into question
can't get over something
cast light on
catch headlines
chances are
come into effect
come to mind
do * best
do away with
down the road
draw the line
err on the side of
fall short
first and foremost
for * sake
for * sake (of)
for certain
for real
for sure
for that matter
for the time being
from scratch
get a clue
get a handle on
get a say
get a voice
get hands on
get hold of
get in somebody's way
get in the way
get over something
get rid of
get to the point
give * a break
go into effect
go wrong
grab hold of
hang in there
have * in place
have * on * mind
have a clue
have a handle on
have a part (in)
have a role (in)
have a say
have a voice
have hands on
have in mind
have nothing to do with
have something to do
hit headlines
hit home
hits headlines
hold * accountable
in * interest
in * interest of
in * place
in * view (of)
in * way (of)
in a fashion (that)
in bad shape
in control
in due course
in essence
in good faith
in good shape
in keeping with
in light of
in no way
in place of
in private
in public
in respect to
in some fashion (that)
in some sense
in some way
in someone's way
in the eyes of
in the fashion (that)
in the fore of
in the interest (of)
in the long run
in the long term
in the pipeline
in the short run
in the short term
in the wake of
in the works
in the wrong
keep * eyes on
keep an eye on
keep in mind
keep on track
leave * alone
make a difference
make fun of
make good on sth
make headlines
off the top of * head
on and off
on the horizon
on the verge of
on time
once and for all
out of control
play a part (in)
play a role (in)
push the envelope
put * to rest
put on hold
quid pro quo
regardless of
right away
right off the bat
shed light on
so long as
so on and so forth
so to speak
take * for granted
take * toll
take effect
take issue with
take steps
the ball is in somebody's court
the big picture
the other way around
things are up for grabs
to * advantage
to * best knowledge
to * best knowledge of
to * credit
to some extent (of)
to some extent (that)
to that effect of
to that effect that
to the contrary
to the extent (of)
to the extent (that)
to the/this effect that (of)
up for grabs
up front
up in the air
with * eyes on

Phrases used from the PHRASE list (Martinez & Schmitt, 2012)

(at) the outset
(get) rid of
(some) kind of
(up) to date
a bit (of)
a case of
a couple of
a degree of
a few
a further
a go
a good deal
a great deal
a handful of
a long way
a lot
a mere
a number of
a question of
a range of
a single
a variety of
about to
above all
all but
all over
all right
all sorts of
all the time
all the way
all too
along with
and all that
and so on
apart from
are to
as a result
as a whole
as far as
as follows
as for
as good as
as if
as it were
as long as
as of
as opposed to
as soon as
as such
as though
as to
as usual
as well (as)
as yet
at a time
at all
at best
at first
at last
at least
at once
at one time
at present
at risk
at the moment
at the outset
at the time
at this point
at times
backed by
based on
be run by
be expected to
be likely to
bear in mind
better off
bother to
bound to
but then (again)
by contrast
by far
by means of
by no means
by now
by the time
by the way
by virtue of
by way of
come to terms with
common sense
concerned with
consistent with
contrary to
could hardly
‘d better
day to day
dealing with
do so
due to
each other
entitled to
even so
even though
ever since
except that
faced with
feels like
find *self
first of all
owing to
followed by
fond of
for good
for instance
for life
for long
for sale
for some time
for the moment
found to
free from
from time to time
full time
give rise to
given that
go away
good at
greater than
had better
half past
have a look
have got (to)
have to
held that
how about
I’m afraid
I mean
if only
if so
in * own right
in a position to
in a sense
in a way
in accordance with
in addition (to)
in advance
in any case
in case
in charge
in common
in conjunction with
in contrast (to)
in detail
in effect
in fact
in favour
in front of
in full
in itself
in line with
in mind
in need
that sort of thing
that which
the above
in order to
in other words
in part
in particular
in place
in practice
in principle
in question
in respect of
in return (for)
in short
in so far as
in spite of
in terms of
in the absence of
in the course of
in the end
in the event (of)
in the face of
in the first place
in the interest of
in the light of
in the meantime
in the same way
in the sense that
in theory
in this respect
in time
in touch (with)
in turn
in view of
in which case
instead of
is to
it takes
just about
key to
kind of
known to
large scale
last night
let alone
limited to
little more than
long ago
long before
long term
look forward to
lots of
make * way
make sense
make sure
make use of
may well
meant to

might as well
mind you
more and more
more or less
more so
most likely
never mind
next door
next to
no doubt
no good
no idea
no longer
no matter
no more than
no one
no sign of
no such
no wonder
not even
not only
nothing but
of course
of little
oh dear
oh no
oh well
old fashioned
on * own
on average
on behalf of
on board
on the basis (of)
on the grounds
on the market
on the one hand
on the other hand
on the part of
on the road
on the way
on the whole
once again
once more
one another
opposed to
or anything
or so
or something
or whatever
other than
ought to
out there
over the years
over there
over time
part time
point of view
prior to
prove to be
provided that
put together
quite a lot
rather than
reflected in
regard to
right now
said to be
set to
shake * head
short of
short term
shown to
sight of
so as to
so called
so far as
so that
some kind of
some more
something about
something like
(that)
something of a
sort of
straight away
subject to
such a
such as
such that
supposed to
take account of
take advantage
take care of
take into account
take part in
take place
taken account of
taken care of
taken into account
takes account of
takes advantage
takes care of
takes into account
takes part in
taking account of
taking care of
thanks to
that much
that s it
the bulk of
the extent to which
the following
the former
the latter
the lot
the means
the odd
the other day
the outset
the sight of
the whole thing
there are
these days
things like that
think so
third party
those who
thought of
thought of (as)
thought of as
thought so
to death
to do with
to some extent
to the extent
to the point
too many
too much
touch of
under way
up and down
up to (date)
way out
way round
wealth of
well being
were to
what about
what if
when it comes to
whether or not
with a view to
with regard (to)
with regard to
with respect to
worth of
would appear
would say
would you like
yet another
yet to
you see

Phrases from the PHRASE list (Martinez & Schmitt, 2012) checked manually

a good
after all
at the same time
a little
at work
can tell
come to
for all
get to
going to
happen to
heard of
if you like
in hand
is to
just a
or two
out of
put it
so far
that is
the case
they say
to come
to date
to go
to me
up to
used to

Appendix 3

Study 2

Ethics information sheet

Date: 15.11.2014

INFORMATION SHEET

As part of my PhD in the School of English, I am carrying out a study of reading behaviour of native speakers of English. I am going to compare the differences in reading patterns when reading different kinds of phrases.

You will read a list of sentences presented on the screen one by one. Please read them naturally and try to understand every sentence. Some of the sentences will be followed by questions checking your comprehension. You should try to be as accurate as possible when answering these questions. Your eye-movements will be recorded while you are reading and used for data analysis.

You are free to withdraw from the study at any time. At every stage, your name will remain confidential. The data will be anonymized before the analysis and will be kept securely and used for academic purposes only.

Should you have any further queries about the study, please feel free to contact myself or my supervisor, Prof. Norbert Schmitt, who can be reached at norbert.schmitt@nottingham.ac.uk or by phone on +44 (0) 115 951 4847. You may also contact the Head of School, Prof. Josephine Guy, on +44 (0) 115 951 5921.

Signed

Laura Vilkaitė

laura.vilkaite@nottingham.ac.uk

University of Nottingham
School of English
NG7 2RD
United Kingdom
Tel: +44 (0) 115 951 5900
http://www.nottingham.ac.uk/english/index.aspx

Appendix 4

Study 2 and Study 3

Collocations and control verbs used in the study

Collocation	Insertion	Control verb
achieve status	a more secure	ignore
acquire land	a bit of	receive
adopt policies	a number of	choose
allow access	all its members	sell
attract investment	all the needed	obtain
cause damage	any kind of	admit
commit murder	more than one	ignore
confirm findings	all the typical	address
create reality	some aspects of	assess
develop skills	some of the	present
develop software	their own new	present
earn respect	at least some	want
ensure safety	the best possible	expect
express concern	the entire society's	explain
extend knowledge	the degree of	change
gain control	at least some	leave
gain experience	a bit of	want
hold views	their own individual	keep
improve performance	the level of	support
improve quality	the level of	discuss
increase capacity	the average available	examine
influence behaviour	the type of	recommend
lose confidence	some of the	keep
meet demand	the widely discussed	avoid
meet expectation	the almost impossible	stand
offer advice	some much better	expect
perform duties	basic and simple	complete
promote development	new kinds of	control
protect interests	only his own	support
provide information	some of the	compare
provide services	free or cheap	consider
raise cash	a lot of	leave
receive treatment	any form of	arrange
reduce costs	at least some	divide
reduce pressure	the existing academic	accept
require attention	all of their	Involve

Collocation	Insertion	Control verb
require effort	any sort of	imply
save energy	all forms of	sell
seek help	your older brother's	avoid
spread news	all the positive	accept

Appendix 5

Study 2 and Study 3

Sentences used in the four experimental conditions

a – adjacent collocation, **ac** – adjacent control phrase, **b** – non-adjacent collocations, **bc** – non-adjacent control

	Collocation		Sentence
1	provide information	a	John was asked to provide information about the studies to local newspapers.
		ac	John was asked to compare information about the studies for local newspapers.
		b	John was asked to provide some of the information about the studies to local newspapers.
		bc	John was asked to compare some of the information about the studies for local newspapers.
2	promote development	a	Committee members wanted to promote development of cheap houses in rural areas.
		ac	Committee members wanted to control development of cheap houses in rural areas.
		b	Committee members wanted to promote new kinds of development of cheap houses in rural areas.
		bc	Committee members wanted to control new kinds of development of cheap houses in rural areas.
3	gain control	a	Everyone was trying to gain control in the region without any success.
		ac	Everyone was trying to leave control in the region to the local authorities.
		b	Everyone was trying to gain at least some control in the region without any success
		bc	Everyone was trying to leave at least some control in the region to the local authorities.
4	improve quality	a	The managers knew they should improve quality and customer service if they wanted to increase sales.
		ac	The managers knew they should discuss quality and customer service in their next meeting.
		b	The managers knew they should improve the level of quality and customer service if they wanted to increase sales.
		bc	The managers knew they should discuss the level of quality and customer service in their next meeting.

	Collocation		Sentence
5	improve performance	a	New types of training can improve performance without the use of drugs.
		ac	New types of training can support performance without the use of drugs.
		b	New types of training can improve the level of performance without the use of drugs.
		bc	New types of training can support the level of performance without the use of drugs.
6	allow access	a	Everyone thinks that the association will allow access to its website without any fee.
		ac	Everyone thinks that the association will sell access to its website for only 20 pounds.
		b	Everyone thinks that the association will allow all its members access to its website without any fee.
		bc	Everyone thinks that the association will sell all its members access to its website for only 20 pounds.
7	reduce pressure	a	In the end, they will have to reduce pressure at the university committee meetings.
		ac	In the end, they will have to accept pressure at the university committee meetings.
		b	In the end, they will have to reduce the existing academic pressure at the university committee meetings.
		bc	In the end, they will have to accept the existing academic pressure at the university committee meetings.
8	require attention	a	Students knew that learning engineering would require attention and hard work if they wanted to succeed.
		ac	Students knew that learning engineering would involve attention and hard work if they wanted to succeed.
		b	Students knew that learning engineering would require all of their attention and hard work if they wanted to succeed.
		bc	Students knew that learning engineering would involve all of their attention and hard work if they wanted to succeed.
9	attract investment	a	It is not an easy task to attract investment as nobody wants to risk their money at the moment.
		ac	It is not an easy task to obtain investment as nobody wants to risk their money at the moment.
		b	It is not an easy task to attract all the needed investment as nobody wants to risk their money at the moment.
		bc	It is not an easy task to obtain all the needed investment as nobody wants to risk their money at the moment.
10	meet demand	a	The city council cannot meet demand for new primary schools anymore.
		ac	The city council cannot avoid demand for new primary schools anymore.
		b	The city council cannot meet the widely discussed demand for new primary schools anymore.
		bc	The city council cannot avoid the widely discussed demand for new primary schools anymore.

	Collocation		Sentence
11	require effort	a	This choice doesn't require effort or hard work from your side.
		ac	This choice doesn't imply effort or hard work from your side.
		b	This choice doesn't require any sort of effort or hard work from your side.
		bc	This choice doesn't imply any sort of effort or hard work from your side.
12	raise cash	a	Everyone expects that Joe will raise cash for the environment protection fund.
		ac	Everyone expects that Joe will leave cash for the environment protection fund.
		b	Everyone expects that Joe will raise a lot of cash for the environment protection fund.
		bc	Everyone expects that Joe will leave a lot of cash for the environment protection fund.
13	offer advice	a	I'm sure the agency will offer advice on all the aspects of the financial situation.
		ac	I'm sure the agency will expect advice on all the aspects of the financial situation
		b	I'm sure the agency will offer some much better advice on all the aspects of the financial situation.
		bc	I'm sure the agency will expect some much better advice on all the aspects of the financial situation.
14	cause damage	a	They didn't want to cause damage to the machines because this would affect their public image.
		ac	They didn't want to admit damage to the machines because this would affect their public image.
		b	They didn't want to cause any kind of damage to the machines because this would affect their public image.
		bc	They didn't want to admit any kind of damage to the machines because this would affect their public image.
15	save energy	a	The company will try to save energy more effectively because they care about the environment.
		ac	The company will try to sell energy more effectively because they want to earn more money.
		b	The company will try to save all forms of energy more effectively because they care about the environment.
		bc	The company will try to sell all forms of energy more effectively because they want to earn more money.
16	seek help	a	You shouldn't seek help when you can easily do the task yourself
		ac	You shouldn't avoid help when you cannot do the task yourself.
		b	You shouldn't seek your older brother's help when you can easily do the task yourself.
		bc	You shouldn't avoid your older brother's help when you cannot do the task yourself.

	Collocation		Sentence
17	ensure safety	a	All teachers should ensure safety in their schools.
		ac	All teachers should expect safety in their schools.
		b	All teachers should ensure the best possible safety in their schools.
		bc	All teachers should expect the best possible safety in their schools.
18	develop software	a	The company was aiming to develop software for business management by the end of the year.
		ac	The company was aiming to present software for business management by the end of the year.
		b	The company was aiming to develop their own new software for business management by the end of the year.
		bc	The company was aiming to present their own new software for business management by the end of the year.
19	influence behaviour	a	This new expert claims to influence behaviour at work and make companies more successful.
		ac	This new expert claims to recommend behaviour at work that will lead companies to success.
		b	This new expert claims to influence the type of behaviour at work and make companies more successful.
		bc	This new expert claims to recommend the type of behaviour at work that will lead companies to success.
20	spread news	a	Jane is very professional; she will never spread news until the information is confirmed.
		ac	Jane is very professional; she will never accept news until the information is confirmed.
		b	Jane is very professional; she will never spread all the positive news until the information is confirmed.
		bc	Jane is very professional; she will never accept all the positive news until the information is confirmed.
21	increase capacity	a	The government plans to increase capacity and promote improvement of schools next year.
		ac	The government plans to examine capacity and promote improvement of schools next year.
		b	The government plans to increase the average available capacity and promote improvement of schools next year.
		bc	The government plans to examine the average available capacity and promote improvement of schools next year.
22	express concern	a	Yesterday the minister tried to express concern about the changing situation of banks.
		ac	Yesterday the minister tried to explain concern about the changing situation of banks.
		b	Yesterday the minister tried to express an entire society's concern about the changing situation of banks.
		bc	Yesterday the minister tried to explain an entire society's concern about the changing situation of banks.

	Collocation		Sentence
23	lose confidence	a	I think Tom will lose confidence in his abilities when he finds out he was not chosen.
		ac	I think Tom will keep confidence in his abilities even when he finds out he was not chosen.
		b	I think Tom will lose some of the confidence in his abilities when he finds out he was not chosen.
		bc	I think Tom will keep some of the confidence in his abilities when he finds out he was not chosen.
24	perform duty	a	Companies have new employees perform duties that senior employees do not want to do.
		ac	Companies have new employees complete duties that senior employees do not want to do.
		b	Companies have new employees perform basic and simple duties that senior employees do not want to do.
		bc	Companies have new employees complete basic and simple duties that senior employees do not want to do.
25	gain experience	a	They say that if you gain experience in management, you will easily find a job in a big organization.
		ac	They say that if you want experience in management, you should look for a job in a big organization.
		b	They say that if you gain a bit of experience in management, you will easily find a job in a big organization.
		bc	They say that if you want a bit of experience in management, you should look for a job in a big organization.
26	acquire land	a	Jim hopes to acquire land in his old village and become a farmer.
		ac	Jim hopes to receive land in his old village when his father decides to move into town.
		b	Jim hopes to acquire a bit of land in his old village and become a farmer.
		bc	Jim hopes to receive a bit of land in his old village when his father decides to move into town.
27	adopt policies	a	The city wanted to adopt policies which were environmentally friendly.
		ac	The city wanted to choose policies which were environmentally friendly.
		b	The city wanted to adopt a number of policies which were environmentally friendly.
		bc	The city wanted to choose a number of policies which were environmentally friendly.
28	create reality	a	A writer's job is to create reality and make the reader believe in it.
		ac	A writer's job is to assess reality and make the reader think about it.
		b	A writer's job is to create some aspects of reality and make the reader believe in them.
		bc	A writer's job is to assess some aspects of reality and make the reader think about them.

	Collocation		Sentence
29	achieve status	a	Young people in this country want to achieve status in society and improve their lives.
		ac	Young people in this country want to ignore status in society and improve their lives without considering social class.
		b	Young people in this country want to achieve a more secure status in society and improve their lives.
		bc	Young people in this country want to ignore a more secure status in society and improve their lives without considering social class.
30	extend knowledge	a	Their project was not sufficient to extend knowledge and understanding of language acquisition.
		ac	Their project was not sufficient to change knowledge and understanding of language acquisition.
		b	Their project was not sufficient to extend the degree of knowledge and understanding of language acquisition.
		bc	Their project was not sufficient to change the degree of knowledge and understanding of language acquisition.
31	provide services	a	The council promised to provide services for poor citizens free of charge.
		ac	The council promised to consider services for poor citizens becoming free of charge.
		b	The council promised to provide free or cheap services for poor citizens despite their age.
		bc	The council promised to consider free or cheap services for poor citizens despite their age.
32	meet expectations	a	Sue told me that she cannot meet expectations and still complete the project on time.
		ac	Sue told me that she cannot reach expectations and still complete the project on time.
		b	Sue told me that she cannot meet the almost impossible expectations and still complete the project on time.
		bc	Sue told me that she cannot reach the almost impossible expectations and still complete the project on time.
33	earn respect	a	During their life, every person should earn respect and try to contribute to society.
		ac	During their life, every person should want respect and try to contribute to society.
		b	During their life, every person should earn at least some respect and try to contribute to society.
		bc	During their life, every person should want at least some respect and try to contribute to society.
34	develop skills	a	Potential candidates will have to develop skills that would lead them to success in the future.
		ac	Potential candidates will have to present skills that would lead them to success during the interview.
		b	Potential candidates will have to develop some of the skills that would lead them to success in the future.
		bc	Potential candidates will have to present some of the skills that would lead them to success during the interview.

	Collocation		Sentence
35	hold views	a	I respect people who hold views on what is wrong and what is right.
		ac	I respect people who keep views on what is wrong and what is right to themselves.
		b	I respect people who hold their own individual views on what is wrong and what is right.
		bc	I respect people who keep their own individual views on what is wrong and what is right to themselves.
36	reduce costs	a	The organization tried to reduce costs in order to save some money.
		ac	The organization tried to divide costs in order to have each department pay their share.
		b	The organization tried to reduce at least some costs in order to save some money.
		bc	The organization tried to divide at least some costs in order to have each department pay their share.
37	protect interests	a	This lawyer tries to protect interests and ensure that justice is done in all the situations.
		ac	This lawyer tries to support interests and ensure that justice is done in all the situations.
		b	This lawyer tries to protect only his own interests and ensure that he makes a profit.
		bc	This lawyer tries to support only his own interests and ensure that he makes a profit.
38	receive treatment	a	Jane didn't want to receive treatment even if she was feeling very weak.
		ac	Jane didn't want to arrange treatment even if she knew her son was very weak.
		b	Jane didn't want to receive any form of treatment even if she was feeling very weak.
		bc	Jane didn't want to arrange any form of treatment even if she knew her son was very weak.
39	confirm findings	a	The study tried to confirm findings from the previous research studies.
		ac	The study tried to address findings from the previous research studies.
		b	The study tried to confirm all the typical findings from the previous research studies.
		bc	The study tried to address all the typical findings from the previous research studies.
40	commit murder	a	Tom claimed he could commit murder if it would help to protect his family.
		ac	Tom claimed he could ignore murder if it would help to hide his son's crime.
		b	Tom claimed he could commit more than one murder if it would help to protect his family.
		bc	Tom claimed he could ignore more than one murder if it would help to hide his son's crimes.

Appendix 6

Study 2 and Study 3

Filler sentences for the eye-tracking task

	Sentence	Question	Answer
1	I have always loved American culture, but I am afraid of flying so I never went there.	Am I interested in American culture?	Yes
2	I am sure Harry finds Emily attractive, but he is too afraid to ask her out.	Are Harry and Emily a couple?	No
3	My grandparents have been married for more than 50 years and they are still such a lovely couple.	Are my grandparents dead?	No
4	The police found a supermarket assistant who recognised the man who tried to pay for goods with a stolen credit card.	Did an assistant help the police?	Yes
5	I know I will never forget that amazing summer we spent together in Italy.	Did I enjoy my time in Italy?	Yes
6	Jack should have a rest after walking for more than 5 hours today.	Did Jack play football today?	No
7	When Molly found out she was the winner of the largest prize, she knew her life would change completely.	Did Molly win the prize?	Yes
8	My grandmother had a house next to the park, so she used to go for a walk every day.	Did my grandmother spend all day at home?	No
9	During the lesson, we were asked to draw a circle with our eyes closed.	Did we have to draw a house?	No
10	When I was little, we always used to go for weekend trips to the beach with my family.	Did we used to go the sea?	Yes
11	I have an impression that my neighbour never sleeps; every time I come home late she is sitting in the garden.	Do I ever see my neighbour?	Yes
12	My new neighbour is a very nice person; she is always smiling when I see her.	Do I like my new neighbour?	Yes
13	Oliver's new girlfriend is very pretty, but I don't think she is a nice person.	Do I like Oliver's new girlfriend?	No
14	We always buy fresh milk from a local market on Tuesday mornings.	Do we buy milk in supermarkets?	No
15	James has always wanted to work in television and now he is presenting the weather.	Does James work in television?	Yes
16	Jane said she couldn't wait to retire and grow roses in her little garden.	Does Jane want to continue her career?	No
17	Jim usually goes to a pub every Friday, but this week he preferred to stay at home and watch TV.	Does Jim often go to pubs?	Yes

	Sentence	Question	Answer
18	Sue has a very large collection from drawings of modern artists which she hopes to sell one day.	Does Sue want to keep all the drawings?	No
19	Poor Luke hates going to the theatre, but because his wife loves it, he has to see a performance at least once a month.	Is Luke a theatre lover?	No
20	Megan always wanted to travel to unusual places, so after finishing university she went to the Far East.	Is Megan afraid of travelling to new places?	No
21	My father insists that I should save up for a flat instead of getting a loan.	Is my father against me getting a bank loan?	Yes
22	My grandfather drinks green tea every morning because he believes it is good for his heart.	Is my grandfather drinking coffee?	No
23	You are not allowed to smoke in public places, so you should not smoke in this bar.	Is smoking in the bar allowed?	No
24	Adam fell down and hurt his knee so badly that his mother had to take him to the hospital.	Was Adam injured?	Yes
25	When John came home, his father was sitting in the kitchen alone and reading a newspaper.	Was John's mother in the kitchen?	No
26	We couldn't afford to live in a big city any more so we had to move to a small village in a valley.	Was living in the valley cheaper?	Yes
27	In order to pass the test in history, students were told they had to learn everything about World War II.	Were students asked to study ancient history?	No
28	They say you should exercise every day if you want to stay fit.	Ready?	
29	Alice wanted to study drama at the college, but her father thought it was a stupid idea.	Ready?	
30	Everyone knows that the Sun is the closest star to the Earth.	Ready?	
31	Matt thinks that everyone who plays in a rock band becomes famous one day.	Ready?	
32	I am leaving for a holiday tomorrow, so I made a list to make sure that I don't forget anything.	Ready?	
33	Everyone would agree that designing a home for a growing family is quite a challenge.	Ready?	
34	When I was younger, I always took criticism too seriously.	Ready?	
35	When Oscar decided to become a soldier, his father got mad, because he always wanted his son to be a lawyer.	Ready?	
36	Doctors say you should eat at least five vegetables or fruit every day.	Ready?	
37	The business manager estimated the profit of the past year and was very pleased.	Ready?	
38	Mistakes are a natural part of learning, so you should not be afraid of making them.	Ready?	
39	Before starting the lecture, we had a brief introduction to the safety rules.	Ready?	

	Sentence	Question	Answer
40	The institute is carrying out a study on life satisfaction in England.	Ready?	
41	Everyone is perfectly capable of learning a new language if they are ready to work hard.	Ready?	
42	I bought two tickets to a fashion show and I hope Olivia will go with me.	Ready?	
43	When Lucy was 15 years old, she wanted to paint all the walls of her room grey.	Ready?	
44	This book tells a story about the adventures of an old servant named Joe.	Ready?	
45	Thomas is a scientist and he is always looking for new participants for his experiments.	Ready?	

Appendix 7

Study 3

Ethics information sheet

Date: 23.02.2015

INFORMATION SHEET

As part of my PhD in the School of English, I am carrying out a study of reading behaviour of non-native speakers of English. I am going to compare the differences in reading patterns when reading different kinds of phrases.

You will read a list of sentences presented on the screen one by one. Please read them naturally and try to understand every sentence. Some of the sentences will be followed by questions checking your comprehension. You should try to be as accurate as possible when answering these questions. Your eye-movements will be recorded while you are reading and used for data analysis.

After the eye-tracking part of the experiment, you will be asked to complete a short test and a questionnaire about your language background. All of the data collected for the experiment will be strictly anonymous.

You are free to withdraw from the study at any time. At every stage, your name will remain confidential. The data will be anonymized before the analysis and will be kept securely and used for academic purposes only.

Should you have any further queries about the study, please feel free to contact myself or my supervisor, Prof. Norbert Schmitt, who can be reached at norbert.schmitt@nottingham.ac.uk or by phone on +44 (0) 115 951 4847. You may also contact the Head of School, Prof. Josephine Guy, on +44 (0) 115 951 5921.

Signed

Laura Vilkaitė

laura.vilkaite@nottingham.ac.uk

University of Nottingham
School of English
NG7 2RD
United Kingdom
Tel: +44 (0) 115 951 5900
http://www.nottingham.ac.uk/english/index.aspx

Appendix 8

Study 3

Language background questionnaire

Participant's ID:

1. What is your mother tongue? ________________________________
2. Country of origin? ____
3. How old were you, when you started learning English?____________
4. How long have you been living in the UK or other English speaking country? (years, months)

____

5. Do you have a degree in English?
   - yes
   - no

6. Please rate your proficiency in English, when 1 indicates poor and 5 indicates excellent:

	1 (poor)	**2**	**3**	**4**	**5 (excellent)**
speaking					
reading					
writing					
listening					

7. How many hours a week do you:
   - … read in English for your studies? ________________
   - … read in English for pleasure? ________________
   - …watch TV/movies in English? ________________
   - … socialize with your friends in English? ________________
   - … socialize online in English (Twitter, Facebook…)? ________________

8. What other languages do you speak? ____

9. What is your age? ______________

10. What is your gender?
    - female
    - male

11. Are you currently an undergraduate or a postgraduate student?
    - undergraduate
    - postgraduate taught (MA)
    - postgraduate research (PhD)

Appendix 9

Study 4

Ethics information sheet

Date: 07.09.2015

INFORMATION SHEET

You are being asked to take part in a research study on second language acquisition and incidental learning, which I am carrying out as a part of my PhD in the School of English, University of Nottingham.

You will be asked to take a vocabulary test and a few tests of language proficiency to start with. You will get your scores and the estimate of your vocabulary size after completing the study. You will also be asked to give some information about your English learning background.

The second part of the study will take place next week, when you will be asked to read a couple of short texts about linguistics and sociolinguistics and answer some comprehension questions about these texts. You should try to perform to the best of your abilities when answering those questions.

All of the data collected for the experiment will be strictly anonymous. Your name will be associated to the participant number and only this number will be used to analyse the data of the experiment. The data will be anonymized before the analysis and will be kept securely and used for academic purposes only. You are free to withdraw from the study at any time.

Should you have any further queries about the study, please feel free to contact myself or my supervisor, Prof. Norbert Schmitt, who can be reached at norbert.schmitt@nottingham.ac.uk or by phone on +44 (0) 115 951 4847.

Signed

Laura Vilkaitė

laura.vilkaite@nottingham.ac.uk

University of Nottingham
School of English
NG7 2RD
United Kingdom
Tel: +44 (0) 115 951 5900
http://www.nottingham.ac.uk/english/index.aspx

Appendix 10

Study 4

Recall and recognition tests

I. Please translate the phrases in Lithuanian into English. You are already given the noun, you should only write a verb instead of the blank.

pradėti eiti pareigas	to	....	duties
paaiškinti sąvoką	to	....	a concept
užmegzti ryšius	to	....	connections
užsiimti savo pomėgiais	to	....	your interests
patenkinti paklausą	to	....	demand
suprasti *kažko* svarbą	to	....	the importance *of something*
abejoti teiginiais	to	....	claims
parengti eksperimentą	to	....	an experiment
formuoti tapatybę	to	....	an identity
pradėti taikyti politiką	to	....	a policy
kelti diskusijas	to	....	debate
kontroliuoti	to	....	control
būti pavojingam	to	....	a danger
nustoti stengtis	to	....	efforts
paskelbti detales *apie...*	to	....	details

II. Some words in language sound more naturally together than others. In this task you are presented with one noun and five verbs each time. Your task is to choose which verb you think sounds the most natural when used with that noun. Sometimes you will find that more than one verb could work in certain contexts. You should simply choose the one that sounds ***the most natural*** *to you! If you do not know the answer, please* ***do not guess*** *but choose "I don't know" instead.*

Example: 0. to .... a role
a) play
b) do
c) make
d) use
e) gain

1. to .... efforts
a) leave
b) deny
c) abandon
d) neglect
e) quit
f) I don't know

2. to .... the importance of *X*
a) assure
b) appreciate
c) imply
d) respect
e) find
f) I don't know

3. to .... interests
a) pursue
b) keep
c) extend
d) gain
e) hold
f) I don't know

4. to .... details
a) reflect
b) clear
c) expose
d) release
e) allow
f) I don't know

5. to .... demand
a) serve
b) track
c) engage
d) meet
e) reach
f) I don't know

6. to .... a policy
a) retain
b) accept
c) adopt
d) select
e) advance
f) I don't know

7. to .... a danger
a) yield
b) throw
c) pose
d) hold
e) bring
f) I don't know

8. to .... a claim
a) fight
b) challenge
c) negotiate
d) perform
e) suspect
f) I don't know

9. to .... duties
a) assume
b) adopt
c) acquire
d) admit
e) assist
f) I don't know

10. to .... a concept
a) expose
b) reveal
c) interpret
d) display
e) illustrate
f) I don't know

11. to .... connections
a) establish
b) acquire
c) enable
d) build
e) settle
f) I don't know

12. to .... experiments
a) design
b) draw
c) direct
d) adapt
e) construct
f) I don't know

13. to .... an identity
a) found
b) produce
c) design
d) adjust
e) construct
f)I don't know

14. to .... control
a) employ
b) apply
c) exercise
d) handle
e) practice
f) I don't know

15. to .... debate
a) develop
b) bring
c) drive
d) produce
e) generate
f) I don't know

Appendix 11

Study 4

Texts and comprehension questions

Collocations are presented in bold only for the reader of the thesis. In the actual study, the collocations were presented without any enhancement. Insertions used in non-adjacent condition texts are presented in the parenthesis. Each participant read only one version: either with insertions or without them.

Language and Gender

Differences between women and men have **generated (a very intense) debate** for many years now and receive a lot of attention. Talbot (2003: 478) suggests that the talk of women and men is often thought of in the following opposite terms:

women's talk	***men's talk***
sympathy	problem-solving
supportive	oppositional
listening	lecturing
private	public
connection	status
intimacy	independence

Overall, women seem to prefer **establishing (strong personal) connections** between people and being supportive listeners, while men are described as aiming to **exercise (their very traditional) control** and being rather competitive. A stereotypical woman **assumes (traditional housewife's) duties**, cooking and raising kids, while a stereotypical man is a politician **adopting (new national) policies** or a scientist making discoveries. Folk-linguistic evidence has long described the language of women and men to be different and this belief lives on in media texts today. Much of this evidence comes in the form of sayings, literature, diaries, essays, advertising, and newspapers **releasing (newly published) details** about various studies. It usually takes a *prescriptive* perspective (how women's language ought to be) rather than a *descriptive* view (how it actually is) thus revealing deeply rooted ideological assumptions about gender.

Language and Gender (also known as 'Feminist Linguistics') is a relatively new field within sociolinguistics, usually said to be marked by the publication of Lakoff's *Language and Woman's Place* in 1975. The field has since **generated (quite a lot of) debate** among applied linguists. Linguists were keen to analyse the relationship between language and gender. They **designed (various kinds of) experiments** and gathered authentic data to explore folk-linguistic beliefs, and **challenge (the very traditional) claim** that males and females speak and act differently. They showed that language is a primary means of **constructing (one's gender) identity**, and that gender identity sometimes leads to inequalities between men and women. Consequently, two aspects emerged in language and gender research: first, how women and men talked, and second, how women/men/boys/girls were represented in language. Today, these questions appear more integrated in a concern to explore how people **construct (various personal) identities** in general, and gender is only a part of a more complex picture.

While the feminist approach has changed since the 1970s (there have been developments in women's status, governments have **adopted (gender equality) policies** to protect their rights, etc.) there is still an agreement that gender continues to be highly relevant to the way people interact through language, and in the way they are represented by gendered 'discourses'. Within its short history, language and gender scholars have repeatedly contested the terms `gender' and 'sex', which are not regarded as synonyms. 'Gender' has now stabilised as a term to distinguish people in terms of their socio-cultural behaviour, and to represent masculine and feminine behaviours as a continuum rather than as a two clearly opposite categories (Holmes 2001). 'Sex' is used to refer to categories distinguished by biological characteristics (i.e. 'male' and 'female').

The field of language and gender is most strongly associated today with a range of studies, which focus on the distinctively gendered ways in which people interact in various social and professional contexts. Three early but still highly influential theories were *deficit, dominance,* and *difference* (see the following sections that **illustrate (these three important) concepts**). They all emphasised gender opposition. The researchers, however, did not **design (well-controlled linguistic) experiments** or gather large amounts of data to support their theories, but rather used observational techniques. These theories tended to compare men and women as members of two opposite social groups, supporting the popular stereotype that 'men are from Mars and women are from Venus' (Gray 1992). These claims continue to receive attention in media that tries to **release (all sorts of) details** about various studies without necessarily being very precise about them.

Deficit theory

Lakoff's (1975) 'deficit' theory claims that from an early age, girls are taught how to use a separate 'woman's language': they are made to use language in a 'lady-like' way. She suggested that women's lower status in American society in the 1970s was reflected through a basically poorer version of men's language. This language was full of hesitation, indirect, and therefore a more powerless version of men's, trapping the women in an unending double issue:

> If [a girl] refuses to talk like a lady, she is laughed at and criticized as unfeminine. If she does learn [lady-like language], she is ridiculed as unable to think clearly, unable to take part in a serious discussion: in some sense as less than fully human. These two choices which a woman has - to be less than a woman, or less than a person - are highly painful.
>
> (Lakoff 1975: 5)

Lakoff's often cited characteristics of women language are the following:

- Hedges: *sort of, kind of, it seems like*, etc.
- Use (super)polite forms: *Would you mind ..., I'd appreciate it if ..., ... if you don't mind*.
- Use tag questions: *You're going to dinner, **aren't you**?*
- Use empty adjectives: *divine, lovely, adorable*, etc.
- Use hypercorrect grammar and pronunciation
- Apologise more: *I'm sorry, but I think that* ...
- Use modal constructions: *can, would, should, ought to - Should we turn up the heat?*
- Use more intensifiers: especially *so* and *very* - *I am **so** glad you came*!
- Lack a sense of humour: women do not tell jokes well and often don't understand the jokes.

Lakoff argued that this lady-like language made the message less clear and showed that the speaker lacks confidence and does not **exercise (any sort of) control** over the conversation. These findings unsurprisingly **generated (a considerable amount of) debate** and later researchers **challenged (the above**

presented) claims for a variety of reasons. One of them was Lakoff's evidence of use being largely based on her own intuitions, without any real evidence, and her failure to **appreciate (the very great) importance** of different functions of all language forms. However, they remain very popular and widely used in popular media to **meet (reader's and viewer's) demand** for interesting and sensational information.

Dominance theory

Dominance theory had two distinct branches: *language as social interaction*, which considered how gender inequalities were constructed through interactions between men and women, and *language as a system* focusing on 'sexism' within the language.

In terms of *language as social interaction*, dominance theorists viewed everyday conversation as instruments to construct unequal gender relations. Again, early feminist researchers did not **design (any large-scale) experiments** or larger studies to support their claims, but rather conducted numerous small-scale, interactional studies of spoken everyday conversations which examined the nature and frequency of talk, silences, questions, interruptions and `back-channelling' (e.g. the woman's use of responses while the man is talking). For example, a famous study by Fishman (1978) showed that men do most of the talking and **exercise (most of the) control** over the conversation, while women are supportive and encouraging listeners and try to **establish (warm and friendly) connections** between the interlocutors. In addition to this, DeFrancisco (1991) showed that men often respond to the conversation of their female partners with silence, rather than with encouraging responses, with the effect that women often stop talking. Such research showed that women were not the sex which talks more, as stereotyped in folk-linguistics: in fact, men talked far more.

In terms of *language as a system*, Spender (1980) argued that language has evolved over the centuries to serve male needs, to represent male interests, to allow men to express male experiences; in short, it is 'man-made'. Spender was concerned with the way that grammars and dictionaries prescribed the use of masculine terms such as *he or man* as general terms to refer to both males and females, thus reinforcing the male-centered view of the world. She also noted that there are some terms that are very clearly associated with men and marked for women (e.g. if you are told a new minister **assumed (their official political) duties** on Wednesday, would you think the minister is a man or a women?). Also, there are gaps for a womancentred vocabulary to describe certain female experiences in positive ways, such as remaining a single woman (while men are called *bachelors* with no negative meaning to it, a negative term *spinster* is used for women).

However, the problem with dominance research was that it appeared to adopt the very male-centered perspective that it was criticising (Coates 2004). For example, by proposing that a male speech style (assertive, direct, competitive, goal-orientated) was more 'powerful' than a female speech style (supportive, co-operative, process-orientated), theorists were in danger of not **challenging (this very popular) claim** that females use inferior language. Therefore, researchers have to be very careful not to **release (vague and imprecise) details** about their studies, in order not to encourage reproducing stereotypes.

Cultural Difference Theory

While dominance theory helped to **generate (a lot of public) debate** about equality and to reveal the apparent tendencies of males and females for different linguistic styles of interaction, it took an unfairly negative view of women's talk. Researcher from Cultural Difference Theory argued that women and men are different 'sub-cultures' learnt through friendly interactions as children in same sex peer groups. So boys learn how to compete with others, to **exercise (some kind of) control**, to use goal-oriented language, to say things for impact and effect. Girls, on the other hand, learn how to **establish (good personal) connections** between individuals, based on equality and trust, to co-operate with others to get things done, and to express feelings and emotions (Maltz and Booker 1982). These contrasting conversational goals corresponded to differently gendered speech 'styles': while 'women speak and hear a language of connection and intimacy, men speak and hear a language of status and independence' (Tannen 1990: 42). Given these separate conversational goals, Coates (1988) argued that women's talk should be 're-valued' in much more positive ways by feminist linguists as different but equal to men's, not inferior.

Main current issues

Since the 1990s, language and gender research has firmly distanced itself from gender difference theories, and is emphasising the diversity of gender rather than difference. Social constructionist theory (e.g. Bergvall et al. 1996; Butler 1990) suggests that males and females are not born, or even simply socialised into a pre-fixed gender identity, but they **construct (their**

own personal) identity through their interactions. According to this view, individuals don't *have* gender, they *do* gender through repeated behavioural and linguistic interactions. This post-modern perspective argues that there are no special male or female characteristics, only ones that are brought into being through repeated actions and linguistic actions. Any apparent characteristics are the effects we produce by way of particular things we do. Thus, according to Butler (1990), we learn many aspects of our identity, such as being feminine or masculine, by performing certain actions.

In this way, gender has constantly to be affirmed and publicly displayed by repeatedly performing particular acts in accordance with the cultural norms. Language is therefore not just a medium to convey social life and interactions, but an essential, constitutive factor. So particular uses of language become culturally associated with masculinity and femininity; they become symbolically gendered or 'index' a gendered identity (Ochs 1992), rather than being the property or attributes of males and females.

According to the social constructionist perspective, gender can therefore be seen as relational, a process, something that is done, and an important resource for **constructing (one's individual) identities** and gender roles. If gender (and indeed sex) are cultural constructs only, they can be challenged and resisted. Gender has the potential to be redefined in terms of multiple roles and positions for men and women. There is a range of ways in which people can speak and act, some of which may be stereotypically coded 'masculine' and others 'feminine', but they are potentially available to all.

In many ways, feminist linguists have successfully accomplished their mission: they made the public **appreciate (the widely discussed) importance** of gendered language, while simultaneously debating that gender is not nearly as significant as we all once thought. They **abandoned (any sort of) efforts** to describe 'women's language' or separate gender-based speech styles, and started looking at complex social identities and linguistic diversity. They have indeed challenged gender as a category. Yet, scholars continue to be interested in the relationship between language and gender as it is performed in various contexts such as running business, leadership, **adopting (international and national) policies**, entering education, learning a second language, promoting health, and in diverse locations around the world.

Adapted from: *Baxter, J. 23. Gender. In Simpson James (ed.) "The Routledge Handbook of Applied Linguistics", 2011, London and New York Routledge. 331-343*

Language and Gender

I. Please answer these comprehension questions on your own, without discussing them with a person next to you.

1.	Lakoff is the person who is associated with the start of the field.	**True**	**False**
2.	Women are usually described as good sympathetic listeners.	**True**	**False**
3.	Early studies of Language and Gender were based on various linguistic experiments.	**True**	**False**
4.	Early studies depicted women language as a poorer version of men language.	**True**	**False**
5.	A lot of early theories saw gender as a very complex construct, while it is actually rather simple.	**True**	**False**
6.	Dominance theory completely rejected gender stereotypes.	**True**	**False**
7.	Cultural difference theory focuses on gender being learned through language.	**True**	**False**
8.	Word *sex* is used to refer to a biological category, while word *gender* – to socio-cultural category.	**True**	**False**
9.	While initially Language and Gender research focused on women language only, nowadays it focuses on men language.	**True**	**False**
10.	Language and Gender research continues to receive a lot of attention today.	**True**	**False**

II. Please discuss the following questions with a person next to you. Be ready to share your answer with the class.

1. Do you think you use women language/men language?

2. In 1975 Lakoff claimed:

> *If [a girl] refuses to talk like a lady, she is laughed at and criticized as unfeminine. If she does learn [lady-like language], she is ridiculed as unable to think clearly, unable to take part in a serious discussion: in some sense as less than fully human. These two choices which a woman has - to be less than a woman, or less than a person - are highly painful. (5)*

Do you agree? Would you say this still applies today? Has the situation changed?

3. Do you agree that we construct our gender differently in different situations? Can you think of one example of a situation, when you would use language stereotypically associated with men and one when you would use women's language?

SLANG AND JARGON

The essentials of slang

If I were to say to you right now, “Dude, I know where we can grab some awesome mongo,” what would you do? Would you think *mongo* could **pose (any sort of) danger**? Would you call the police? Would you come along to see what I meant by *mongo*? Would you look up *mongo* in the dictionary before we left, so that you know that you were getting into? I mean, what kind of word is *mongo*? It doesn’t sound like an English word. It certainly is not a word you learned by studying for vocabulary quizzes. Is it a bad word for something good, or a good word for something bad? If you looked it up in dictionaries, how would lexicographers **illustrate (this newly popular) concept**? Would they label *mongo*: as *standard* or *nonstandard*, as *slang* or *jargon*, *dialect* or *colloquial*, or even *vulgar* of *taboo*?

Here is how Ted Botha defines *mongo* in his book *Mongo: Adventures in Trash*: “Some [collectors of sidewalk trash] also have a word for what they find, a word that is suitably playful and vague. It sounds like it could be French, Chinese or even African, but it is American slang, used in New York for any discarded item that is picked up, retrieved, rescued. That word is *mongo*”. You can relax now, *mongo* does not **pose (any real life) danger**. All the other dictionaries that enter *mongo* give a similar definition to **illustrate (this simple linguistic) concept** and all of them agree that *mongo* is American slang.

But is it? Does Standard English possess a term for stuff once thrown out and then retrieved? That item “deserves” to be saved from the trash; it has value, as determined by the rescuer. Surely Standard English has no word for ‘appealing trash’ other than *mongo*. According to one way of thinking, we can **challenge (this seemingly obvious) claim** that *mongo* is slang, and think that it is a “standard” term, one that gives a whole new meaning to taking trash.

You might think that the argument sounds like one that lexicographers would have over drinks. Who else would care? And anyway, *mongo* is obviously slang; everyone I ask about it, including lexicographers, agrees that it’s slang. But how do they know? On what bases do they judge it? Do we know what counts as slang and what doesn’t, not by thinking about it, but by the way we feel about it? Should we collect large databases of data and **design (a number of different) experiments** to figure out what is slang and what isn’t? When you encounter the *slang* label in dictionaries, does it reflect reasoned conclusions about words or intuitions about them? Or should we simply **abandon (all of our) efforts** to make clear descriptions?

Slang by definition

Native speakers of American English recognize American slang more or less automatically: the brain sorts the labels much more efficiently than any lexicographer. They might not be able to **illustrate (this widely-used) concept of slang** clearly, but they definitely have an idea of what slang is. We know slang when we hear it, and we know how to use slang in our own speech, because social, aesthetic, and linguistic knowledge guides us. When dictionaries define slang they attempt to identify the social, aesthetic, and linguistic characteristics or properties that distinguish slang from other language. Here is a definition from a prominent dictionary:

> *Encarta World English Dictionary*: 1. **Very casual speech or writing** words, expressions, and usages that are casual, lively, or playful replacements for standard ones, are often short-lived, and are usually considered unsuitable for formal contexts. 2. **Language of an exclusive group** a form of language used by a particular people, often deliberately created and used to exclude people outside the group.

Definitions agree that slang is “in-group” language, the use of which determines who belongs to a group and **pursues (many common) interests** together and who does not.

The American Heritage College Dictionary definition provides *argot* and *jargon* as near synonyms for *slang*; along with *cant*, they have been used as synonyms, not only in everyday speech, but also in thoughtful writing about language. *Jargon* is the oldest of the terms, entering English from French in the fourteenth century to mean ‘unintelligible talk’. *Cant* started as a verb, probably in sixteenth century, and referred to beggar’s whine (cant likely derives from Latin *cantus* ‘song’ and refers to sing-songy speech). *Argot* arose in the nineteenth century as a synonym for *cant*. Choosing among these supposed synonyms is itself an exercise in jargon, the jargon of linguistics and historians of English. Though the arguments over terminology may sound incomprehensible, we should not **abandon (all of our linguistic) efforts** to make clear distinctions. Certainly, every researcher can **appreciate (the very obvious) importance** of precision. The Oxford English Dictionary suggests that *jargon* can be “applied as a negative term to any mode of speech abounding in unfamiliar terms, or peculiar to particular set of persons, as the language of scholars or philosophers, the terminology of science or art, or the *cant* of a class, sect, trade or profession”, which brings the speech of philosophers, rocket scientists, thieves, lawyers, and beggars too close to one another for comfort.

Criminals speak *argot* when they’re doing their nasty business. Anyone **pursuing (any kind of) interests** or vocations, from doctors to model train enthusiasts, employs *jargon* suited to that particular occupation. Any other language that characterizes a group and identifies speakers with that group end up *slang* by default. Is *mongo* slang? If the police **release (some interesting new) details** about a crime where art thieves strip your walls of *mongo*, *mongo* for the thieves is argot; if workers cleaning streets pick up *mongo* at the street, then *mongo* is jargon; when you and I talk about *mongo*, because we just read Ted Botha's book and we're all about the next big thing, then *mongo* is slang. Slang, jargon, and argot aren't essentially the characteristics of a word; one or another of them applies depending on who uses the word, in what situation, for what reasons.

We enjoy slang just for its casual, lively, informal, maybe a bit disrespectful and playful elements, and some combination of those elements is what alerts the ear to lexical trouble: slang **poses (absolutely no real) danger** to linguistic standards, but challenges them a bit, and each synonym it supplies must add some social meaning to the standard alternative's lexical meaning. If you want to be casual, don't *spend time* or, as in Standard English, *relax* with your friends: *hang out* with them; even better, just *hang*, or be cool and *chill out*, or *chillax* - just sit around, *like chillin'* with your friends. Does language follow attitude here, or make it, or perhaps a little of both?

The definitions of slang also disagree about the stamina of lexical items we consider slang. Are they short lasting, or do they persist in the vocabulary? And if they survive, do they survive as slang, or do they "move up" from slang to standard spoken English, relatively informal English that isn't invested in fitting in or standing out, as they lose their restrictive, in-group quality? Some definitions describe slang as "short-lived". But slang's life-span is problematic, and no simple estimate adds up. For instance, most of us are convinced that *like*, as in "It was *like* the best concert ever" or "He was *like*, 'It was the best concert ever,'" originated recently in urban California slang, but Alexandra D'Arcy has recently reported that this is language mythology, because examples of both likes can be found in rural speech in England… in the nineteenth century.

Having said all of this, I have to agree, from intimate experience with slang, that most of it disappears before we notice it. But slang's short half-life isn't the biggest problem. As Eble points out, slang "is widely used without precision, especially to include informal usage and technical jargon, and the social and psychological complexities captured in slang vocabulary make the term difficult to define." You may be surprised to learn that "the aim of using slang is seldom the exchange of information. More often, slang serves social purposes: to identify members of group, to **establish (close in-group) connections**, to change the level of discourse in the direction of informality, [and] to challenge established authority." In the market of social meaning, "slang is the linguistic equivalent of fashion and serves much the same purpose."

Language of purpose versus language of being

Doctors discussing a diagnosis, engineers trying to **meet (their customers' increasing) demand** for the new products, or politicians **adopting (important international) policies**: they all use jargon in their conversations. If you want to **assume (prime minister's) duties**, you should learn political jargon. On the face of it, jargon behaves very much like slang. For instance, if you work as a barman in a national chain restaurant that serves lots of frozen blended drinks, you might end up **assuming (the incredibly boring) duties** of *blender tender* on Friday and Saturday night. Like many slang terms, *blender tender* is playfully constructed and it's casual, too, in the sense that it's a more casual term than Cocktail-Bartender, or whatever official term the restaurant uses to identify the human employee. Blending frozen drinks for eight hours to **meet (thirsty customers') demands** isn't a pleasure, exactly; the humour conveyed in jargon compensates some of the boredom at work. A term like *amateur diner* is used among waiters for the customers who eat out once a year, perhaps on Mother's Day, and who aren't really sure what the menu items are or how to order them effectively. Customers provide waiters with most of their income; one might think they deserve some respect. *Amateur diner*, you'll note, is a little irrelevant.

If both slang and jargon are playful, a bit disrespectful and casual, then why aren't they both the same thing? Should we **abandon (all these unsuccessful) efforts** of separating them and use one term for them both? The distinction lies in this: slang is the language of a group that **pursues (some of their shared) interests** together but not necessarily shares a purpose. It's a language of being. As said before, *hang* is perfect slang: *hanging* is just being, whether alone in your apartment or with a group of friends. Slang can belong to small groups (junior high school groups) or big ones (all Americans under forty). If members of a group have things in common (and if they didn't they wouldn't be a group), then they are bound to speak a common language that separates them in some fashion from the language of those who aren't in the group.

Jargon is language at work; the workgroup is the in-group. If you are a waiter in a restaurant, then you and your fellow waiters do not **pursue (a lot of common) interests**. Rather you have common purposes, namely, to serve well, to **meet (the restaurant customers') demand** and thereby to make money for restaurant and for yourself. Everyone who eats in restaurants knows that waiters are paid *tips* but might not realize that waiters *tip out* (that is, 'pool and share tips') at the end of the *shift*. To make big *tips*, waiters persuade customers to run up big bills: they might suggest that you buy more so that you spend more, or *upsell*. You are not supposed to know about *upselling*, or that the waiters call you a *camper* because you sit too long at your table without ordering anything or that to make money, a waiter needs to *turn tables* (serve the table quickly to get new people in).

But customers need not to worry about the jargon spoken around them while they chat, network, flirt, eat, and drink. It's not for them. Once in a while a waiter will approach my table with a familiar phrase: "Hello, my name is James, and I will be your waiter tonight. Before I take your drink orders, let me tell you about our specials tonight. For *apps*, we have…" Wait a minute, James, my internal voice interrupts. I do not belong to your group; I don't want to understand the mysteries of your profession, tell me about the *appetizers*, not *apps*.

Amateur diner and *camper* are playful synonyms for Standard English phrases ('inexperience customer' and 'customer who sits at a table too long' respectively) and so resemble slang. But the social circumstances that provoke jargon are different from those that provoke slang: in jargon, there is something at stake beyond positioning oneself in the social circumstance. Jargon may be stylish, but it is not, as Eble suggests of slang, analogous to fashion, jargon has to roll up it sleeves and do real work. So jargon is practical: it's lively and unambiguous, and it helps busy people do their jobs efficiently. But its use is also social: social relations within the restaurant depend on it, and efficient work toward the common goal depends on certain social relations and not having good group dynamics might **pose (a very real) danger** to the functioning of the group. As funny as it might seem, we should really **appreciate (the already discussed) importance** of jargon and slang in a professional and personal life.

Adapted from: *Michael, A. (2009) Slang: The People's Poetry. Cary, NC: Oxford University Press*.

Slang and Jargon

III. Please answer these comprehension questions on your own, without discussing them with a person next to you.

1.	*Mongo* is classified as a standard English word.	**True**	**False**
2.	Slang dictionaries cannot define slang precisely.	**True**	**False**
3.	Common language users are usually quite good at knowing which words are slang.	**True**	**False**
4.	Slang shows your belonging to a group of people with common interest.	**True**	**False**
5.	All slang words eventually become words of standard English, if they add something to a meaning.	**True**	**False**
6.	Slang and jargon are essentially the same thing.	**True**	**False**
7.	Jargon is usually associated with your profession.	**True**	**False**
8.	The author claims that customers in restaurants always learn some restaurant jargon.	**True**	**False**
9.	*Amateur diner* means an inexperienced restaurant customer.	**True**	**False**
10.	Slang is language of purpose, while jargon is language of being.	**True**	**False**

IV. Please discuss the following questions with a person next to you. Be ready to share your answer with the class.

1. Do you use *slang*? When? Could you give any example of Lithuanian slang?

2. Can you think of any group you belong to that uses professional *jargon*? Give some examples of the jargon you use.

Appendix 12

Study 4

Language background questionnaire

Please answer the following questions about your language background.

1. What is your first language? ____
2. How old were you, when you first started learning English?_______________
3. How long have you been studying English? ____________________________
4. Have you taken the State School leaving exam of English? If yes, what was your score? ____
5. Have you been to an English speaking country (UK, USA…) for more than a month? If yes, for how long?____
6. What other languages do you speak?__________________________________
7. What is your age? ______________
8. What is your gender?
   - o female
   - o male

Appendix 13

Study 4

Word formation task

Read the text below. To fill the gaps, use the word given at the end of some of the lines to form a word that fits the sentence. There is an example at the beginning (0).

CCTV

0. visitors to the UK are often quite surprised at just how many	**VISIT**
1._________________ cameras there are. Some of these cameras are	**SECURE**
2._______________owned and used to watch over shops, offices	**PRIVATE**
and ***3.***________________ buildings when they are not in use. But	**INDUSTRY**
the ***4.***______________ of CCTV cameras, as they are called, are	**MAJOR**
placed in city centres and public buildings to ***5.***______________ the	**SURE**
safety of the general public. Local governments and the police share	
the ***6.***______________________ for the camera's operation and	**RESPONSIBLE**
maintenance and for paying somebody to sit and watch at the other	
end. What is even more surprising is how ***7.***__________________	**EFFECT**
these cameras are at preventing ***8.***__________________ activity.	**CRIME**
According to one leading ***9.***____________________, only three	**POLITICS**
percent of crimes are solved using evidence from CCTV	
recordings. Many people now think that the government could	
spend the same money more ***10.***_________________ on having a	**USE**
few extra police officers on the street.	

(*http://www.ihbristol.com/free-english-exercises/test-esol/esol-wf-cctv/submitted*)

Appendix 14

Study 4

Debriefing

Incidental acquisition of formulaic sequences. Study debriefing

This study was concerned with incidental acquisition collocations. Collocations are sequences of two words used together very frequently, such as ***meet demand***. This study focuses specifically on acquisition of non-adjacent collocations, that is, collocations with words intervening, e.g. ***meet*** *the growing* ***demand***. This type of collocations is very frequently used in language and this study looks at whether learners pick can them up unintentionally when reading.

How was this tested? In this study, you were asked to read some texts for comprehension. These texts contained some adjacent collocations (where the elements of collocations directly follow each other, as in ***meet demand***) and some non-adjacent collocations. You were not asked to focus on them during the reading stage, but you were later tested on acquired knowledge of these collocations

Hypotheses and main questions. We expect to find that adjacent collocations will be noticed and therefore learned better than non-adjacent collocations. We will examine this hypothesis statistically by comparing your answers to adjacent and non-adjacent collocations in the follow-up test. We will also investigate the effect of your vocabulary size on learning new vocabulary items, i.e. collocations.

Why is this important to study? Recently in second language teaching there was a shift from focusing on single words to focusing on teaching longer chunks of language – multiword sequences. However, there are still a lot of unanswered questions about how the learners acquire those multiword sequences and what the best ways of introducing them in classroom are. As the number of multiword sequences in language is very high, it is important to understand how realistic is to pick them up unintentionally, when engaging in different kinds of language tasks.

What if I want to know more? If you are interested in learning more about the acquisition of collocations or if you would like to receive the summary of the findings of this study, please contact me at laura.vilkaite@nottingham.ac.uk.
If you have concerns about your rights as a participant in this experiment, please contact my supervisor, Professor Norbert Schmitt at norbert.schmitt@nottingham.ac.uk.

Thank you again for your participation!