How many CDs do you have?: The impact of teaching-materials on formulaic language processing in Japanese learners of English

Julian Northbrook

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

April 2020

Abstract

Formulaic sequences are processed more quickly and sound more natural than matched non-formulaic language for native speakers. However, little is known about how the speeded processing and intuitions about naturalness emerge and develop in second language learners, especially at the early stages. The research presented in this thesis explores, from a usage-based perspective, the relationship between the input in teaching materials and the development of a processing advantage in Japanese learners of English for a type of formulaic language: 'lexical bundles', which are frequently occurring sequences of words that may constitute a complete utterance (e.g., *I don't really know*) but often do not (e.g., *in the middle of*).

Using a corpus-based approach, Study 1 compared the most frequent lexical bundles in Japanese junior high school English textbooks with a native-speaker reference corpus and asked whether they were quantitatively and qualitatively representative of those found in conversational English. Overall, lexical bundles were found to be very frequent in the textbooks, but they deviated considerably from native-speaker language in terms of their frequency distribution. Study 2 used a 'phrasal judgment' task to assess the extent that the textbook materials impact upon processing for the students exposed them. Students responded to lexical bundles from their textbooks significantly faster and more accurately compared to control items (e.g., *going to visit* vs *going to say*). They were also sensitive to the frequency of occurrence in their textbooks in a continuous manner – i.e., lower frequency items were responded to slower, and higher frequency items faster, across the frequency continuum. Study 3 explored the perceived naturalness of lexical bundles using an offline naturalness rating task and compared performance by a group of adult Japanese learners of English and native English speakers. Naturalness ratings from both groups were predicted by frequency of occurrence, once again in a continuous manner. However, the Japanese English learners were sensitive to textbook frequency, and the native speakers to frequency counts from the British National Corpus. Study 4 used a repetitive phrasal judgment task, combined with exposure to lexical bundles, either in an 'enhanced' (by underlining) or 'hidden' (i.e., no enhancement) condition, in constructed materials. The study showed that repetition facilitated processing, and

further that there was an initial, but short lived, benefit for textual enhancement. The results also demonstrated that processing behaviour changed after just one exposure to items, and further a post-test conducted one week later indicated that students retained knowledge of the lexical bundles.

Overall the results from the studies presented in this thesis provide a valuable new perspective on how a processing advantage for formulaic sequences emerges in reaction to input; in this case due to teaching materials. The fact that even early-stage EFL learners display a formulaic advantage for sequences extracted from their learning materials, and that this remains years after encountering these materials, suggests that very careful consideration needs to be given to the quality of the materials we present students. Crucially, the findings call into question prior claims that L2 learners process language in a word-by-word, computational manner, or else that formulaic processing only emerges at higher proficiency levels. Notably, the set of findings support a 'usage-based' model of language development, whereby processing develops incrementally, with each and every exposure to a sequence of words in second language speakers.

Contents

Declaration....8
Published and Submitted Manuscripts....9
Acknowledgements....10
Terminology and Typographical Conventions....11
Chapter 1. Introduction....12
Overview of the Thesis....14
Chapter 2. How Nativelike Is Your Processing? Nativelike Fluency, Selection and the Emergence of a 'Processing Advantage'....17
Defining formulaic language: do you know it when you see it?....18
Defining lexical bundles....22
Exploring the processing advantage....24
How do we go from input to optimised processing?....32
Chapter 3. "What Are You Talking About?"....41
Introduction....41
Lexical bundles in second language learning....42
Methodology....46
Analysis and results....50
Overview of lexical bundles....50
Analysis of frequent lexical bundles....51
Discussion....57
Conclusion....61
Chapter 4. Is What You Put in What You Get Out?....63
Introduction....63
Formulaic sequences in non-native speakers....65
How does a processing advantage for lexical bundles emerge?....67
Methodology....71
Experimental stimuli....71
Participants....73
Procedure....74
Analysis and Results....74
Discussion....78

Conclusion 82
Chapter 5. “That Sounds About Right” 83
Introduction 83
Formulaic sequences in second language learning 84
Methodology 87
Participants 87
Materials 88
Procedure 90
Analysis and Results 90
Discussion 94
Conclusion 96
Chapter 6 - “Did You See That?” 98
Introduction 98
Learning formulaic sequences 100
Optimising the Input: Repetition and Enhancement 102
Methodology 106
Experimental stimuli 106
Participants 107
Procedure 108
Analysis and Results 109
Discussion 114
Conclusion 119
Chapter 7. Materials Matter 120
Summary of studies 120
Developing a ‘processing advantage’ for textbook-language 122
Optimising the input 125
Perceived Fluency 126
Optimising teaching materials for exposure 129
Limitations and directions for future work 131
Final Conclusions 134
References 136

Appendices 158

Appendix 1. Chi-square scores testing the difference between JHSETC frequency and SUBTLEXus frequency (ranked in the order they are discussed in the paper). 158

Appendix 2a. LSMEANS statistics for Response Time by Repetition. 159

Appendix 2b. LSMEANS statistics for Response Time by Lexical Bundle Type. 159

Appendix 2c. LSMEANS statistics for Repetition: Lexical Bundle Type interactions by lexical bundle type. 160

Appendix 2d. LSMEANS statistics for Repetition: Lexical Bundle Type interactions by repetition. 161

List of Figures

Figure 1. Experiment Flow for Study 4. 108
Figure 2. Mean response times in milliseconds for experimental conditions across repetition for Study 4......109

List of Tables

Table 1. Number of lexical bundles by type and token for each of the word lengths in the JHSETC and SUBTLEXus. 51
Table 2. The 10 most frequent lexical bundles in the JHSETC and SUBTLEXus corpora at each bundle size......52
Table 3. Comparison of yes/no fragment bundles, do you~ and what do you~ bundles in the JHSETC and the SUBTLEXus corpus......53
Table 4. Summary of non-native speakers' age, years of studying English, and self-rating of English proficiency for participants in Study 2. 73
Table 5. Mean response times in milliseconds and standard deviations in brackets in each of the three experimental conditions for Study 2......75
Table 6. Summary of fixed and random effects for the best-fit model for Study 2. 76
Table 7. Summary of non-native speakers' age and self-rating of English proficiency for participants in Study 3. 88
Table 8. Mean naturalness ratings and standard deviations for lexical bundles and controls for native and non-native participants in Study 3. 91
Table 9. Summary of fixed and random effects for the best-fit model for Study 3. 92
Table 10. Mean response times in milliseconds and standard deviations in brackets in each of the four experimental conditions in Study 4......109
Table 11. Summary of Significant Predictors for Study 4. 111

Declaration

I declare that the work presented here is my own and was conducted during my time as a PhD student at the University of Nottingham. Several parts of this thesis have been published in or submitted to peer reviewed journals: The study outlined in Chapter 3 was published in *International Journal of Corpus Linguistics* (Northbrook & Conklin, 2018b). The study presented in Chapter 4 was published in *Applied Linguistics* (Northbrook & Conklin, 2018a). Chapter 5 has been submitted and is currently under review (Northbrook & Conklin, under review). Chapter 6 has been submitted and is currently under review (Northbrook, Conklin, & Allen, under review).

All chapters within this thesis represent adapted versions of the published papers. Only the published versions should be considered authoritative, and any citations or page references should be taken from the published versions.

Published and Submitted Manuscripts

Northbrook, J., & Conklin, K. (2018). “What are you talking about?” An analysis of lexical bundles in Japanese junior high school textbooks. *International Journal of Corpus Linguistics*, 23(3), 311-334.

Northbrook, J., & Conklin, K. (2018). Is What You Put in What You Get Out? — Textbook-derived Lexical Bundle Processing in Beginner English Learners. *Applied Linguistics, 40*(5), 816-833.

Northbrook, J., & Conklin, K. (in submission). “That sounds about right” – An examination of lexical bundle naturalness intuitions in native and non-native English speakers.

Northbrook, J., Conklin, K., & Allen, D. (in submission). “Did you see that?” – The role of repetition and enhancement on lexical bundle processing in English learning materials.

Acknowledgements

I would like to express my gratitude to Kathy Conklin, my supervisor, for her guidance and patient feedback throughout my doctoral studies. I would also like to express my gratitude to David Allen for his support on numerous occasions, including providing feedback on several studies, and helping with data collection and setting up/running the experiment presented in Chapter 6.

Many thanks also go to Masao Nakamura and the other teachers at the school where research for Chapter 4 was conducted, as well as to the Japanese students who volunteered to participate in the studies presented here.

Terminology and Typographical Conventions

The terms ‘formulaic language’, ‘formulaic sequences’ and ‘multi word units/sequences’ are used interchangeably throughout this thesis. I make no technical distinction between these terms. I also use the term ‘lexical bundle’ to refer specifically to lexical bundles, and also the shortened ‘bundle’ to refer to the same thing. Because lexical bundles are central to this thesis, they are defined in detail in Chapter 2, however it should be noted that many studies use different terms to refer to the same thing. When describing the literature, unless otherwise stated I use the terms used by the original author on the assumption that it refers to the same unit of analysis as lexical bundles.

Direct quotations are presented throughout in double quotation marks: “…”. Terms used in a semi-technical sense are presented in single quotation marks: ‘…’. Examples of phrases are presented in italics, e.g., *in the middle of the*.

Chapter 1. Introduction

A key concern for learners of English as a second language is speaking the language fluently and naturally, whether it be in business situations, social or day to day life. Over the last two and a half decades, mounting evidence has shown that the acquisition and use of formulaic sequences may be a key component of this (e.g., Conklin & Schmitt, 2008; Pawley & Syder, 1983; Siyanova-Chanturia & Pellicer-Sánchez, 2019). However, important questions remain as to how this can be achieved. Notably, very few studies have considered the quality of the teaching materials that are used with learners of English; despite the fact that delivering the contents of textbooks is perhaps the biggest pressure teachers face (e.g., Browne & Wada, 1998; Creese, 2008) and subsequently they exert a powerful effect on teachers' practice (Sato & Kleinsasser, 2004).

In the Japanese context, which is the focus of the empirical studies presented in this thesis (Chapters 3–6), English education has been a hot topic of debate and research for many years (Butler & Iino, 2005; Hashimoto, 2009). English is big business in Japan, both in the private sector—where it constitutes a 670 billion yen industry (Seargeant, 2009)—and in schools where it is studied for a minimum of five, but often at least nine years. Yet, the time and cost put into learning English "is not matched by an equally high level of communicative proficiency among the population" (Seargeant, 2009, p. 3). Indeed, Japan has been extensively criticised for its level of English ability and English education (Aspinall, 2006; Butler & Iino, 2005; Kikuchi & Browne, 2009; Kwan, 2002), and despite attempts at reform, Japan is outperformed by other countries on English tests such as the Test of English as a Foreign Language (TOEFL) (Hagerman, 2009; Tukahara, 2002; Yoshida, 2003).

Many explanations for this poor performance have been put forward. Some accounts point to school entrance examination washback (Aspinall, 2006; Brown, 2000; Guest, 2008; Hagerman, 2009; Yoshida, 2003) which results in lessons that emphasise translation skills (Bailey, 1999; Kikuchi, 2006) and 'yakudoku' (grammar translation) (O'Donnell, 2005; Rapley, 2010) as the source of the problem. Other explanations focus on the 'communicative' approach taken in Japanese textbooks when there is a lack of real communicative need (Browne & Wada, 1998; Nishino,

2008; Nishino & Watanabe, 2008; Rapley, 2010), lack of confidence in teachers with communicative methods (Gorsuch, 2001; Nishino, 2011), students' debilitating shyness and anxiety (Andrade & Williams, 2009; Cutrone, 2009; Doyon, 2000; Kimura, Nakata, & Okumura, 2001 2001) and their disposition towards taciturnity (Anderson, 1993; Cutrone, 2010; King, 2012; Nozaki, 1993). Still other explantions have focused on a failure of teachers to question the appropriateness of what they do in class (Stout, 2003), contrasting cultural values (see for example Loveday, 1986 section 5) and the acceptance of certain myths concerning Japanese people's inability to learn languages (Hinenoya & Gatbonton, 2000).

One area that has received very little attention is that of the teaching materials used in Japanese schools. Usage-based approaches to second language learning (discussed at length in Chapter 2) put a premium on the linguistic input that learners receive, and hold that any sequence of words that learners encounter frequently will, all else being equal, be processed faster than unfamiliar strings of language. In the literature this is generally referred to as a 'processing advantage' – that is, faster processing when compared with controlled 'non-formulaic' sequences of words. The formulaic/ non-formulaic dichotomy is somewhat arbitrary, as will be discussed in the following section, however the processing advantage associated with it has been demonstrated by a significant body of work (e.g., Arnon & Snider, 2010; Conklin & Schmitt, 2008; Jiang & Nekrasova, 2007 among many others). This resesearch will be reviewed at length in the subsequent chapter.

The key concern of this thesis is the relationship between input and the development of a processing advantage for high-frequency sequences of words. This is explored in the context of Japanese secondary school English teaching – a particularly apt population for this kind of research because, in contrast to university students living and studying abroad (with whom most such research is conducted), we are able to control for English input to a relatively high degree. Schools in Japan are required to use one of six sets of Ministry of Education Culture, Sports, Science and Technology (MEXT) mandated textbooks. Therefore, other than students who grew up abroad or went to special schools (for example international or immersion schools, which operate outside of the typical MEXT guidelines) we know that all Japanese students have been exposed to the English in these books.

As was stated at the outset, formulaic language is important for fluency and proficiency (e.g., Conklin & Schmitt, 2008; Pawley & Syder, 1983; Siyanova-Chanturia & Pellicer-Sánchez, 2019). The focus of this thesis is on a subtype of formulaic language, 'lexical bundles'. Formulaic language and lexical bundles will be discussed in detail in Chapter 2. However, in brief, lexical bundles are sequences of words that occur frequently in language, are generally corpus derived and may constitute a complete utterance, as in *how's it going*, but often do not, as in *a lot of the*. Only a handful of papers have argued for more emphasis on formulaic sequences in the Japanese context (Asare, 2006a, 2006b; Okuwaki, 2012, 2014). Crucially, almost no published empirical data exists on school-age learners, in the Japanese context or outside of it, perhaps with the exception of a single study focusing on exemplar-based development of the 'be' verb (Tode, 2008). Nobody, to my knowledge, has studied the MEXT mandated English textbooks in terms of formulaic language or usage-based theories of language development. Therefore, it was necessary to start right at the beginning with a corpus-based study of the textbooks (Chapter 3). The aim of this was to determine whether junior high school learners were exposed to formulaic language – in this case lexical bundles. The findings from this study were used to test the role of input; if input matters, we should see an advantage for lexical bundles encountered in the textbooks (Chapters 4–6). Each empirical study is self-contained and presents its own literature review discussing the issues that are most relevant to that investigation. Because of this, a certain degree of repetition is unavoidable in the general introduction to each study. However, I have tried to ensure that where information is repeated, it is done so because it helps to illuminate the overall picture being created. This also allows me to present each chapter as it was originally designed – as a stand-alone study, with minimal amendments from the published and submitted versions.

Overview of the Thesis

Following this introduction, Chapter 2 begins by defining formulaic language in general, and 'lexical bundles' specifically, as well as what is meant by a 'processing advantage'. In this chapter, I will also outline the theoretical underpinnings of the research: namely that of usage-based accounts of second language learning and the

predictions they make for the development of the aforementioned processing advantage in relation to input.

Chapter 3, as mentioned above, explores lexical bundles in communicative Japanese junior high school textbooks and looks at whether they are representative of conversational English. To do this, I used a corpus-based approach that compares the most frequent lexical bundles in the textbooks to those in an English reference corpus, (the SUBTLEXus – Brysbaert & New, 2009). This chapter provides the basis for materials selection for subsequent chapters (Chapters 4–6), as well as for measures (i.e., frequency) that are used to predict phrasal processing in tasks reported in them (with the exception of Chapter 6, which uses measures from the COCA – Davies, 2008).

Chapter 4 is the first behavioural study. It takes the lexical bundles identified in the English textbooks in Chapter 3 and explores the processing of textbook-language in beginner learners of English in a Japanese secondary school. To do this, I use a phrasal judgment task and analyse the data using mixed-effects modelling. The results are discussed in terms of pedagogical implications and material design.

Following from this, Chapter 5 explores the perceived naturalness of lexical bundles and aims to assess the extent to which prior experience of English modulates intuitions. This was achieved using an offline naturalness rating task, comparing scores from native English speakers and non-native speakers, with Japanese as their first language. This was done in relation to corpus derived frequencies from the BNC and the corpus of Japanese English textbooks reported in Chapter 3. Findings are discussed in terms of usage-based accounts of second language development, particularly the role of early-stage exposure in the development of naturalness intuitions.

Chapter 6 asks whether we can optimise learning materials with lexical bundles in order to facilitate processing, and whether textual enhancement can further aid learning. Adult Japanese learners of English read a series of short stories repeating low-frequency 3-word lexical bundles; half of these were underlined in the text, while the other half were unenhanced. After each story, participants performed a phrasal judgement task on the lexical bundles from the story as well as ones not seen in the story (half of which were repeated in each task and half which were new in each).

Findings are discussed in relation to the literature on formulaic sequence acquisition, and in relation to learning material design.

Finally, Chapter 7 provides a general discussion of the findings of the studies as they relate to the wider literature on formulaic sequences. I discuss the implications of the findings for the creation of quality language learning materials and propose some 'best practices' as far as the instruction of English in the classroom is concerned. I also provide some final conclusions, and some potential areas for future research.

Chapter 2. How Nativelike Is Your Processing? Nativelike Fluency, Selection and the Emergence of a 'Processing Advantage'

The resurgence of corpus linguistics following the 'Chomskyan Era' marked a turning point in our understanding of language as it is used in real communicative situations. Throughout the 50s, 60s and 70s the actual language people produced was, for the most part, considered uninteresting with the focus being on 'competence', or the underlying rule system. Before this there had been considerable interest in the probabilistic nature of language (Shannon, 1949) and 'performance' – or what people *do* with language (Skinner, 1957). Over the last two decades, we have seen a return to this way of thinking and ever-advancing technology has allowed us to build larger and larger corpora and conduct experiments which were never before possible. Following early, pioneering work (Pawley & Syder, 1983; Sinclair, 1991 amongst others), a different way of looking at language emerged – not simply as a set of words and rules to combine into sentences, but as a system that uses larger, multi-word blocks of language known as 'formulaic sequences'.

The focus of this thesis is on formulaic language—and in particular *lexical bundles*—which is central to the debate about words and rules versus usage-based approaches to language acquisition and processing. In order to 'set the scene' for my research on lexical bundles, this chapter lays the foundation for the subsequent empirical ones. I begin this chapter with a broad overview of formulaic language research as it pertains to second-language learning. Following this I define the specific type of formulaic language this thesis uses as its primary unit of analysis—the *lexical bundle*—before stepping back to provide an overview of research which has aimed to assess the psycholinguistic reality of formulaic sequences in terms of what is known as a 'processing advantage'. I will then explore usage-based and cognitive theories which go some ways towards explaining the acquisition and processing of formulaic language.

It should be noted that both the literature on formulaic language and usage-based approaches is extremely broad in its scope. It would be impossible to include everything in this review, nor would it make sense to do so. Therefore, I have limited the discussion here to the literature most relevant to the empirical chapters. Notably, I

focus my discussion on compositional, non-figurative formulaic sequences and only touch briefly upon the rich—and extensive—research concerning idiom processing. While idiomatic language acquisition and processing is certainly interesting and important in the field of formulaic language in general, because of the nature of idioms (which are semantically opaque, figurative phrases), it is largely irrelevant to the majority of the empirical research presented here.

Defining formulaic language: do you know it when you see it?

The expression, '*I know it when I see it*' is a formulaic sequence that refers to something which is hard to categorise but can be intuitively recognised for what it is. This particular expression gained notoriety when, in 1964, Supreme Court Justice Potter Stewart ruled manager Nico Jacobellis as 'not guilty' following a lengthy trial where he was accused of obscenity for showing Louis Malle's film *Les Amants* (The Lovers) at the Heights Art Theatre, in Cleveland. As a result of this ruling, the expression 'I know it when I see it' became one of the most famous—and controversial—phrases in the history of supreme court opinions (Gewirtz, 1996), and a household expression as far as formulaic sequences are concerned. However, does this expression accurately define the linguistic category of which it is a part of? Does one simply, *'know it when I see it'* as far as formulaic language is concerned?

Over the last three decades, there has been a steady increase in the interest of formulaic language across three broad areas: monolingual language representation and processing (e.g., Arnon & Snider, 2010; Wray, 2002, 2008; Wray, 2012); second language learning and bilingualism (e.g., Conklin & Schmitt, 2008; Conklin & Schmitt, 2012; Ellis, 2012; Ellis, Simpson-Vlach, & Maynard, 2008; Siyanova-Chanturia, Conklin, & van Heuven, 2011 amongst others); and language impairment (e.g., Lancker & Kempler, 1987; Sidtis, Canterucci, & Katsnelson, 2009; Van Lancker Sidtis, 2012; Van Lancker-Sidtis, 2004). Yet the question of what exactly formulaic language *is*, still remains a deceptively difficult one to answer. Superficially at least, the term 'formulaic language', and the associated term 'formulaic *sequence*,' point to recurrent sequences of words in language performance. However, there are almost as many definitions as there are angles from which the topic has been explored. Just

some of the names given to formulaic language in the past have included: ‘chunks’ (Ellis, 1996), ‘phrase-chunks’ (Servan-Schreiber & Anderson, 1990), ‘ready-mades’ (Weinert, 1995), ‘prefabricated patterns’ (Hakuta, 1974), ‘lexicalised sentence stems’ (Pawley & Syder, 1983), ‘multi-word units, chunks, phrases, sequences, patterns and expressions’ (e.g., Arnon, McCauley, & Christiansen, 2017; Arnon & Snider, 2010; Biber, 2009; Eskildsen & Cadierno, 2007; Jolsvai, McCauley, & Christiansen; Kim & Kim, 2012; Siyanova, 2010; Siyanova-Chanturia, 2013; Siyanova-Chanturia, Conklin, & Schmitt, 2011), ‘constructions’ (e.g., Ellis, 2003; Goldberg, 2006) and over fifty more terms (Wray, 2002). Further, these have been used to describe a range of item types including, but certainly not limited to collocations (e.g., Ellis, Frey, & Jalkanen, 2009; Sosa & MacFarlane, 2002; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2014), idioms (e.g., Conklin & Schmitt, 2008; Siyanova, 2010; Siyanova-Chanturia, Conklin, & Schmitt, 2011), binomials (e.g., Siyanova-Chanturia, Conklin, & van Heuven, 2011; Siyanova-Chanturia & Janssen, 2018) and lexical bundles (e.g., Tremblay, Derwing, Libben, & Westbury, 2011). While the category of formulaic language can be defined in a variety of ways, the same is true of the subcategories of formulaic language; what one researcher may refer to as collocations could easily be seen as at least partially figurative or idiomatic (e.g., *thick skin* is a collocation meaning ‘ability to ignore verbal attacks and criticism’). Formulaic subtypes exist on a spectrum, rather than being completely separate phenomena. That being said, most would agree that *kick the bucket* is an example of an idiom, while the sequence *a lot of* is not. The expression *kick the bucket*, however, could feasibly be classified as a lexical bundle – a recurring sequence of words that is corpus defined (Biber, Johansson, Leech, Conrad, & Finegan, 1999). In other words, we might know formulaic language and different types of it when we see it, but we also might not.

Leedham (2012) argues that at the heart of this terminological confusion, there is a division between types of formulaic sequence that “hinges on semantic unity” (p. 157). That is, on the one hand, sequences of language are recognised intuitively as whole chunks of meaning. These tend to be conventionally ‘whole’ items such as idioms, proverbs and other ‘institutionalised’ expressions such as *how are you?*, *thanks very much*, and indeed, *I know it when I see it*. On the other hand, researchers use computational methods to extract chunks on the basis of frequency of occurrence

(Biber et al., 1999, being a poignant example) which may lead to 'complete' units, but often does not. This relates closely to what Siyanova-Chanturia and Van Lancker Sidtis (2019) describe when they make a distinction between literal-compositional and figurative non-compositional sequences: the former is comprised of literal sequences including "lexical-bundles, verb-particle constructions, and irreversible binomials" (p. 44), as well as n-grams. The latter category is comprised of non-literal, figurative idioms. It is necessary, then, to keep in mind that when talking about 'formulaic language' we are not dealing with a single phenomenon – rather we are talking about a group of related phenomena that have been studied across varied data sets and in different theoretical contexts and disciplines (see for example Wray, 1999; 2000 for a discussion on this).

For second language learners, formulaic sequences in general are thought to aid communication by freeing up processing resources to increase fluency both in terms of production for the speaker and comprehension for the listener (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Kuiper, 2004; Pawley & Syder, 1983). Indeed formulaic sequences have been discussed at length in terms of fluency in both first and second language speech production and nativeness in second (see e.g., Wood, 2010a for a good overview of these issues) . Empirical evidence for the 'benefits' of formulaic language will be outlined in the subsequent chapter, but in brief Pawley and Syder (1983) roughly divide them into two areas: what they term 'nativelike selection', and 'nativelike fluency'. Put simply, native speakers sound natural because they select nativelike language, and they speak fluently because they default to conventionalised expressions that they store—either has wholes or as routines—and use *as is* without needing to resort to resource-draining computational processes. This serves to reduce the processing load and allows longer stretches of speech to be processed in working memory (Kuiper & Haggo, 1984).

Hymes (1972), in his taxonomy of situational competence outlined four parameters that mark native-speaker communication: that utterances are possible, well-formed, appropriate and that they are *performed*, or put more simply, that they are actually 'done' in the real world. Thus, for second-language learners, this means that the use of formulaic language is a mark of nativelikeness, and therefore proficiency (Boers et al., 2006; Dickinson, 2012; Kuiper, 2004; Wray & Fitzpatrick,

2008). In other words, learners who seek to become accepted members of a speech-community must produce utterances that are not just possible, well-formed, and appropriate—which as Foster (2001) points out has tended to be the focus of classroom teaching, even at the cost of "being slow and odd" (p. 90) in the language—but should be expected and nativelike. Thus, nativelike selection can be considered a sensitivity to collocational preferences (Wray, 2000).

The idea of collocational preferences can be contrasted to generative grammar and the 'words-and-rules' approach where the emphasis is on the potential to create an infinite number of unique sentences. In reality we do not use this creative potential much of the time. This means that although native speakers could say, *Pleasant morning.* or *Could you aid me in this task?* they tend not to do so. This is not because these phrases are ungrammatical, which is clearly not the case, but rather because there are more common, socially institutionalised phrases available – *Good morning.* and *Could you help me with this?.* It is probably safe to say that few people would disagree that these more common alternatives sound more natural and are understood more readily. We must conclude, then, that communicative competence in a language is not simply the ability to produce grammatically correct utterances; rather it is the ability to select frequent and conventional patterns that are institutionalised in the language (Biber et al., 1999; Ellis, 2001).

The use of formulaic sequences also serves to increase fluency (or *nativelike* fluency). The term 'fluency' gets used in different ways and can mean different things (Schmidt, 1992). Studies have explored the relationship between formulaic sequence instruction and perceived oral fluency in second language speakers finding it, on the whole, very positive (e.g., Boers et al., 2006; Dickinson, 2012; Gatbonton & Segalowitz, 2005; Taguchi, 2007). In terms of fluency in the sense of automatic processing, which is closer to the definition used by Pawley and Syder (1983), it refers to the efficient, automated processing of languge. Pawley and Syder (1983) observed that native speakers seem to produce fluent stretches of spontaneous, connected speech that is beyond our computational abilities. They state that "there is … a sizable collection of evidence of several different kinds that the largest unit of novel discourse that can be fully encoded in one encoding operation is a single clause of eight to ten words" (p. 202). Conklin and Schmitt (2012) give the example *You can lead a horse*

to water, but you can't make him drink. pointing out that this is beyond the limits of 8-10 words – yet native speakers have no trouble with this kind of utterance. On the contrary; they seem to process them faster and more efficiently than novel stretches of speech (e.g., Siyanova-Chanturia, Conklin, & Schmitt, 2011; Sprenger, Levelt, & Kempen, 2006; Swinney & Cutler, 1979 among others).

The current study focuses on compositional formulaic sequences broadly and more specifically one specific sub-type of this category: the lexical bundle. The following section will provide an in-depth working definition of lexical bundles, before considering the behavioural research and the role formulaic language plays in processing in subsequent sections.

Defining lexical bundles

As described above, studies have not used terminology consistently when talking about formulaic language. The term 'lexical bundle' is commonly used in corpus-based studies, and studies looking at idiosyncratic differences between text genres. However, generally speaking, behavioural studies have used the term 'formulaic sequence' to refer to what I call 'lexical bundles' (with the exception of perhaps Tremblay et al., 2011 who explicitly use the term lexical bundle), or the term 'n-gram'. This makes discussing the literature somewhat problematic; while studies may be using the same or very similar items, they do not define or label them in the same way. In this thesis, I will describe the literature using the terms used by the authors of each study with the understanding that, unless otherwise stated (for example, in the case of idioms or binomials), we are discussing the what I call 'lexical bundles'. When discussing my own empirical research, I use the term 'lexical bundle'.

Lexical bundles have often been considered a middle-ground between purely creative, syntax-driven language and completely fixed phrases such as conventional expressions and idioms (Skehan, 1998). Lexical bundles can be semantically opaque (for example, *may as well* cannot be understood in terms of its constituent parts) but more often are transparent in that the sequence carries the meanings of its individual parts (for example, *let's go to the*). They can be difficult to identify without the use of corpus software because lexical bundles span traditional boundaries and grammatical

categories. They are often incomplete fragments and are hierarchical in that they exist on multiple levels. For example, a 4-word lexical bundle may contain two 3-word bundles, which in turn also contain two bigrams each (e.g., *in the middle of* contains *in the middle* and *the middle of*, and also *in the*, *the middle* and *middle of*). Biber et al. (1999) define lexical bundles as extended collocations that constitute:

> A recurring sequence of three or more words […] a combination of words must frequently recur in order to be considered a lexical bundle. In most cases, these bundles are not structural units, and they are not expressions that speakers would recognise as idioms or other fixed lexical expressions (p. 990).

And further that:

> To qualify as a lexical bundle, a word combination must frequently recur in a register. […] Lexical sequences are counted as 'recurrent' lexical bundles only if they occur at least ten times per million words in a register. These occurrences must be spread across at least five different texts in the register (to exclude individual speaker/ writer idiosyncrasies). Because 5-word and 6-word bundles are generally less common, a lower cut-off of at least five times per million words is used for those types (p. 993).

I also use this description as a starting point when discussing the literature around lexical bundles, and in the empirical chapters (with some modifications; discussed in the relevant chapters). This view of lexical bundles is based purely on occurrence in a corpus. Phraseological approaches to defining formulaic language differ from this. Phraseological approaches, stemming from the Russian tradition (Cowie, 1998), see multi-word sequences as ranging from free combinations (e.g., *blow the trumpet*), to restricted collocations (e.g., *blow a fuse*), figurative idioms (e.g., *blow your own trumpet*) and finally pure idioms (e.g., *blow the gaff*) (Howarth, 1998, p. 164). These formulaic sequence types relate closely to Siyanova-Chanturia and Van Lancker Sidtis (2019)'s definition of figurative non-compositional items. Importantly, in this tradition there should be a meaning or function associated with a given formulaic sequence.

This means that although corpus tools may be used to assist the identification of items, ultimately it is down to subjective decision—aided by detailed selection criteria (see for example Myles, Hooper, & Mitchell, 1998; Weinert, 1995)—as to whether an item is formulaic or not. Phraseological viewpoints, then, tend to favour items which constitute complete structural phrases. Lexical bundles, on the other hand, are statistically defined items that emerge out of language on account of the way language users construct discourse and do not necessarily represent linguistic units with grammatical, semantic or pragmatic status (Biber, 2009; Stubbs & Barth, 2003). They are, in the words of Biber et al. (1999), "recurrent expressions, regardless of their idiomaticity, and regardless of structural status" (p. 990). However lexical bundles do tend to be both register and discipline specific (Biber, 2009; Stubbs & Barth, 2003) with frequency counts for a given lexical bundle tending to be quite different between, for example, conversation and academic writing (Cortes, 2004; Hyland, 2008). These differences will be discussed at length in Chapter 3.

Exploring the processing advantage

Generative accounts of language, and the minimalist mental-architectures they assume, have given way to a different way of thinking whereby language development is driven by exposure to and usage of language. Broadly speaking, these are known as 'usage-based' accounts, which encompass a wide range of theoretical stances (Barlow & Kemmer, 2000) but posit a maximalist system whereby language development occurs in reaction to language use (e.g., Barlow & Kemmer, 2000; Bybee, 2006; Ellis, O'Donnell, & Römer, 2013; Tyler, 2010). In contrast to generative accounts, which aimed to outline systems which used a minimal number of computational rules to produce infinite numbers of utterances from these, usage-based accounts tend to see language as a largely redundant collection of constructions and patterns of usage. Language development is driven by language experience – or in more extreme accounts, development and experience of language are seen as synonymous (de Bot & Larson-Freeman, 2011).

Usage-based approaches are important in terms of formulaic language research because they not only move away from the traditional view of words being the

primary unit of language, but they also blur the distinction between syntax and the lexicon. A major outcome of usage-based theories is that language is far less creative than was once postulated. That is, although we certainly can produce an infinite number of sentences, and as Pinker (1995) claims, produce utterances that are wholly unique and original, "appearing for the first time in the history of the universe" (p. 22), more often than not we simply do not do so (Pawley & Syder, 1983). Instead, people tend to "cobble together constructions they have registered earlier" (Larsen-Freeman & Cameron, 2008, p. 100), speaking in highly predictable chunks of language. Or put it another way, people tend to produce language in the way they have previously experienced it.

Higham (1997) points out that the common factor for all chunking models and theories—a detailed discussion of which I will save for the following section—is that performance can be predicted by some metric derived from the frequency at which features occur within a given stimulus. And indeed, frequency effects at all levels of processing are amongst some of the most robust found in psycholinguistics (e.g., Bybee & Hopper, 2001; Ellis, 2002a, 2002b). Behavioural studies have looked at reaction times for formulaic sequences compared to matched novel language, and on the whole formulaic sequences have been found to have significantly faster reaction times when compared to non-formulaic sequences. This advantage has been found for various types of multi-word units including compositional sequences such as collocations (e.g., Ellis et al., 2009; Sosa & MacFarlane, 2002; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2014), binomials (Siyanova-Chanturia, Conklin, & van Heuven, 2011) and lexical bundles (e.g., Tremblay et al., 2011), as well as figurative idioms (Conklin & Schmitt, 2008; Siyanova, 2010; Siyanova-Chanturia, Conklin, & van Heuven, 2011). A processing advantage has also been found in different populations including adults (e.g., Conklin & Schmitt, 2008; Jiang & Nekrasova, 2007; Siyanova, 2010), children (Bannard & Matthews, 2008) and second language learners (discussed extensively below) and is posited as a key argument for the inclusion of multi-word units in models of first and second language processing.

A wide literature has demonstrated that native speakers are sensitive to phrasal frequency (Bannard & Matthews, 2008; Edmonds, 2013; Ellis et al., 2008; Hernández, Costa, & Arnon, 2016; Jiang & Nekrasova, 2007; Kapatsinski & Radicke, 2009; Kim

& Kim, 2012; Rammell, Sidtis, & Pisoni, 2017; Schmitt & Underwood, 2004; Shantz, 2017; Siyanova & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Siyanova-Chanturia & Janssen, 2018; Sonbul, 2014; Sonbul & Schmitt, 2013; Sosa & MacFarlane, 2002; Tremblay & Baayen, 2010; Tremblay et al., 2011; Tremblay & Tucker, 2010; Underwood, Schmitt, & Galpin, 2004). Indeed, experimental results are remarkably consistent for native speakers. Studies exploring the processing advantage in second-language speakers, however, have provided less consistent results. As touched on in the opening of this chapter, one explanation is that non-native speakers have not encountered formulaic sequences a sufficient number of times for a processing advantage to emerge. In general, learners have less exposure to language than a native speaker does their first language (Conklin, 2020). Another explanation—and key to this thesis—is that non-native speakers do process formualic sequences in a nativelike manner, however the exact sequences they have been exposed to are different and may not have been tested in existing studies because they have used native-speaker norms to identify them – see Conklin (2020).

In one of the earliest studies with second language speakers, Underwood et al. (2004) used eye-tracking to compare the fixation time and count for terminal words in idioms and non-idiomatic phrases. They found that the terminal words were read faster in idioms by native speakers, while the results for non-native speakers were inconclusive. Schmitt, Grandage, and Adolphs (2004) conducted a dictation task with embedded formulaic language. They predicted that long bursts of speech would overload working memory and force participants to reconstruct what they had heard, while formulaic sequences would be retrieved as wholes, and therefore be produced more accurately. However, this was not the case for the majority of items for both L1 and L2 speakers. In another study making use of a self-paced reading task, Schmitt and Underwood (2004) failed to find evidence of a processing advantage for L2 speakers, and suggested that the task might not be suitable to find a formulaic advantage, although other studies have since shown self-paced reading to be effective for studying lexical bundle processing in native speakers (Tremblay et al., 2011) and in non-native speakers (Kim & Kim, 2012; Shantz, 2017). Babaei, Taleb Najafabadi, and Fotovatnia (2015) found no advantage for lexical bundles over non-lexical

bundles for moderately proficient non-native speakers in an EFL setting, concluding that non-native speakers do not store or process lexial bunldes as “holistic” units (p. 12; this point will be discussed further below).

Siyanova and Schmitt (2008) in a multi-experiment study explored adjective–noun collocations (e.g., *social services*) in native and non-native speakers and found non-natives demonstrated poorer intuition than their native-speaker counterparts regarding the frequency of collocations. Likewise, one of their studies showed that non-native speakers were slower than native speakers in processing collocations. In a later study, Siyanova-Chanturia, Conklin, and Schmitt (2011) used eye-tracking and found that native speakers had an advantage for idioms compared with novel sequences, but that non-native speakers did not. Further, in a more recent study, and one of the few looking at production, Siyanova-Chanturia and Janssen (2018) used a phrase elicitation task to explore the processing of binomial expressions. Similar to Siyanova-Chanturia, Conklin, and van Heuven (2011), they found native speakers were sensitive to both the order (e.g., *bride and groom* vs. *groom and bride*) and the frequency of binomials. However, no such effects were found for non-native speakers. Valsecchi et al. (2013) used eye-tracking to investigate the reading of lexical bundles by native and lower-proficiency non-native speakers. They found that while native speakers had a processing advantage for lexical bundles compared with non-lexical bundles, non-native speakers did not. The authors concluded that this was due to lack of proficiency and that a processing advantage would emerge with increased proficiency.

Other studies have found evidence for a formulaic processing advantage in second-language speakers. The first study to do so in resepect to compositional formualic sequences was by Jiang and Nekrasova (2007). The researchers used a phrasal judgment task where participants were instructed to judge whether the word order was correct for 4-word formulaic sequences. They found that formuliac sequences were responded to significantly faster than novel-control items by both native and non-native speakers. However, because they did not control for word or substring frequency their conclusion that their findings are indicative of ‘holistic storage’ (discussed below) are tenuous. A similar study—and also an important study in that it is one of the few studies to date to look at online processing of formulaic

sequences in lower-level, EFL learners—Isobe (2011) conducted a phrasal judgment task. The author, like Jiang and Nekrasova (2007), found 4-word formulaic sequences were responded to significantly faster and with greater accuracy. Further, in a second experiment, involving a timed familiarity judgement task, the author found familiarity and phrasal frequency to be important in predicting response times for the same items. However, Isobe (2011), like Jiang and Nekrasova (2007), did not control for word or substring frequency which raises questions about the findings.

Ellis et al. (2008) ran a series of three experiments based on the 'academic formulas list' (Simpson-Vlach & Ellis, 2010), and demonstrated various corpus derived metrics to be psycholinguistically valid. Crucially, they found evidence for phrasal processing and frequency effects in advanced non-native speakers. Conklin and Schmitt (2008) likewise found that idiomatic sequences embedded into stories were read faster by both native and non-native participants compared with control sequences. Kim and Kim (2012) conducted a self-paced reading task with items categorised into four frequency bins and found that native speakers and, to a slightly lesser degree, non-native speakers were sensitive to collocational frequency. They took their results as evidence that native speakers chunk and store all items, and non-native speakers chunk and store only high frequency items. Similarly, Shantz (2017) found evidence for phrasal processing in non-native speakers using a self-paced reading task.

Sonbul and Schmitt (2013) conducted a study of collocations designed to test both native and non-native speakers' implicit and explicit knowledge of them. They found comparable results for native and non-native speakers: both groups were sensitive to explicit measurements, but neither group were sensitive to implicit measurements. Hernández et al. (2016) studied 4-word sequences using a phrasal judgement task focused on native speakers as well as upper-intermediate and lower-advanced non-native speakers. They used the same methodology and experimental items as Arnon and Snider (2010), asking whether mode of exposure made a difference: classroom-based environments, compared with learners from immersion environments. They found evidence for nativelike processing that matched that of the native speakers, but no difference between the environments.

As we see, findings from second-language speakers are mixed and seem to

undergo a change from studies which largely find no evidence of a processing advantage in this group, or at least evidence which is non-conclusive (in the case of Siyanova & Schmitt, 2008, for example), to studies which, on the whole, find similar effects for non-native speakers as native speakers published around the 2011 mark (Babaei et al., 2015 and Valsecchi et al., 2013 being important exceptions to this). This may be in part to do with Arnon and Snider (2010)'s highly influential study which criticised Underwood et al. (2004) for not controlling for word frequency, commenting that: "The effects reported so far provide limited evidence for an effect of phrase-frequency on adults; we need more evidence from adults when substring frequency is controlled for." (p70). The same criticism can also be lodged against many of the early studies: Jiang and Nekrasova (2007), Schmitt et al. (2004), Ellis et al. (2008), Conklin and Schmitt (2008) and Isobe (2011).

Failing to control for word frequency makes results very difficult to interpret because there is no way of knowing whether processing advantages are indeed the result of whole string frequency (as authors have claimed) or whether the individual words happen to be more frequent in the 'formulaic' strings, thus giving the illusion of a formulaic sequence-driven processing advantage where in fact none exists. A good example of this would be *going back to* (formulaic) vs. *turn back to* (non-formulaic) taken from Jiang and Nekrasova (2007). In the SUBTLEXus corpus (Brysbaert & New, 2009), part of the 'SUBTLEX' family, which has been shown to accurately predict word processing in various languages (e.g., Cai & Brysbaert, 2010; Cuetos, Glez-Nosti, Barbón, & Brysbaert, 2011; Dimitropoulou, Dunabeitia, Aviles, Corral, & Carreiras, 2010; van Heuven, Mandera, Keuleers, & Brysbaert, 2014), *going back to* appears 897 times; significantly more than *turn back to*, which only appears 16 times. The only difference between these phrases is the headword – *going* vs. *turn*. The word 'going' has a frequency of 104,585 (2,051/million), and 'turn' has a frequency of 15,136 (296/million). This means 'going' is approximately seven times more frequent than 'turn'. Thus, it is hardly surprising that *going back to* is responded to faster than *turn back to* – regardless of the frequency of the string. A similar issue exists in Isobe (2011), where she cites the comparison between *come up with* and *run up with*, where the word 'come' is more than three times more frequent than 'run', making it hard to diferentiate frequency effects driven by the individual words and the sequence.

Arnon and Snider (2010), who did control for word frequency, conducted a phrase judgement study similar to that of Jiang and Nekrasova (2007). They found that higher frequency phrases were processed faster than low frequency ones and argue that this provides evidence that phrasal frequency information is represented in the brain. Similarly, Hernández et al. (2016) controlled for word frequency in a phrasal judgment task, providing more convincing evidence of a formulaic advantage in second language speakers. Valsecchi et al. (2013) and Babaei et al. (2015), whose studies were done with lower proficiency learners, did control for word frequency but did not find evidence of a processing advantage. They took this to mean a processing advantage does not emerge until later in second-language development, with Babaei et al. (2015) stating that lexical bundles "are not processed holistically in the mind of [Persian] EFL learners" (p. 12). Similarly, Valsecchi et al. (2013) say that their results "indicate that the mastery of formulaic sequences comes only later in the acquisition process of a foreign language" (p. 11). The implication of these statements is that formulaic language is treated differently—more compositionally—by non-native speakers than native ones. If this is the case, a key focus in research should be on when and how a chunking mechanism is 'turned on' in second language learning. An alternative explanation for null results in the literature, is that non-native speakers are sensitive to phrasal frequency for sequences they have encountered before; however, they have been tested on sequences they have not had sufficient exposure to. If we test non-native speakers on formulaic language that we know that they have encountered, we will see the typical processing advantage.

As may already be apparent from the discussion of the literature, many of the studies have used a processing advantage as evidence for 'holistic storage'. In other words, faster processing indicates that multi-word items are stored in the lexicon as a single unit, in the same way as an individual word is generally assumed to be (although see Elman, 2009; 2011 for a contrasting view of this assumption). In fact, Wray (2002) defines formulaic language as a sequence that is "stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (p. 9). This definition is echoed by Underwood et al. (2004), who state: "formulaic sequences are stored and processed holistically" (p. 167), and further by Jiang and Nekrasova (2007): "Our findings provide clear and

straightforward support for a holistic view of formula representation and processing" (pp. 441-442). However, as Edmonds (2013) argues, equating fast processing with holistic storage is problematic, and she and criticises Underwood et al. (2004) and Jiang and Nekrasova (2007) for unequivocally interpreting their findings in this way, noting that "there is not an exclusive relationship between facilitation and holistic storage" (Edmonds, 2013, p. 91; see also Siyanova-Chanturia, 2015).

Not only have more recent studies been more carefull in controling for individual word frequecies (as part of item selected and/or statistically in analyses), researchers have been more conservative when interpreting their results. For example, Arnon and Snider (2010) even comment that holistic storage is "unlikely" (p. 72) and rather see the processing advantage as evidence that the brain stores frequency information pertaining to phrases. This view will be discussed at length in the following section on usage-based theories of second language learning. However, at this point it is worth noting that regardless of whether formulaic sequences are stored as single, 'morpheme-like units' (Wray, 2002, 2008) or whether they are represented as routines that facilitate (speed) processing, the result is, ultimately, the same. That is, native speakers at least, and likely second-language speakers, produce and comprehend formulaic sequences in a faster, more efficient manner than non-formulaic sequences.

Finally, before closing this section, it is worth noting a few key points. The vast majority of the studies mentioned above are compehension studies. Studies which have looked at the production of formulaic sequences, specifically in non-native speakers are, in the words of Siyanova-Chanturia and Van Lancker Sidtis (2019) "still a rarity, with some areas having received virtually no attention from L2 resaearchers" (p. 39). Notable exceptions are Ellis et al. (2008) who included a production task in their battery of tests (Experiment 2) and Siyanova-Chanturia and Janssen (2018) who examined the production of binomial expressions in L1 and L2 usesrs finidng that a phrase elicitation task was facilitated by binomial frequency, but only for native speakers. Tremblay and Tucker (2010) used a reading aloud task to explore the processing of 4-word formulaic sequenes. Improtantly, this study is also one of the few that included probabilistic measures such as unigram, bigram, trigram, and quadgram frequency of occurrence as a part of the analysis. Bannard and Matthews (2008), working from a usage-based approach looked at the production of 4-word

units in children, using a sentence repetition task. Crucially, they showed that children were sensitive to the frequency of items derived from a corpus of child-direct speech. To my knowledge, this is the only study which has attempted to match stimuli in a task to the population by utilising specialised corpora. This is an issue that I believe has been overlooked in the literature and one which will be taken up in detail below, as well as in subsequent chapters (in particular Chapters 3 and 4).

How do we go from input to optimised processing?

In this section I will discuss how usage-based theories explain a processing advantage emerging as a result of input. As briefly noted in the previous section, usage-based accounts of language encompass a wide range of theoretical stances (Barlow & Kemmer, 2000) but posit a maximalist system whereby language development occurs in reaction to language use (e.g., Barlow & Kemmer, 2000; Bybee, 2006; Ellis et al., 2013; Tyler, 2010) using general cognitive systems (Ellis & Wulff, 2015; Wulff, 2019). In this section I will limit my discussion to the facets of these approaches most relevant to the current discussion: the input-driven development of a processing advantage for formulaic sequences in general, and lexial bundles in particular.

Essentially, we can think of the 'processing advantage' observed in the literature as the emergent property of language optimisation via the attuning of mental processing systems to real world language (experienced via input). We can think of this process as similar to the way an internet browser 'caches' frequently accessed webpages: the first time a page is visited, the browser must load all data files present on the page. However, by storing this information in a dedicated data-cache, the next time a page is visited the browser can make use of this stored information and loading times are greatly reduced. An important question in the formulaic language literature is whether all data is cached for every instance of language we receive (e.g., Bod, Hay, & Jannedy, 2003; Bybee, 1998, 2006; Bybee & Hopper, 2001; Mclaughlin, Rossman, & McLeod, 1983), or whether only frequently encountered language that surpasses a certain number of instances (akin to setting your browser to cache only after, say, five visits to a website) is stored (e.g., Goldberg, 2006; Wray, 2002, 2008), in other words after a 'threshold' is reached. These two viewpoints will be discussed in relation to

language learning materials; notably, both can be considered usage-based, and therefore run counter to the predictions made by traditional 'words-and-rules' accounts.

In a words-and-rules approach, no frequency effects are predicted for phrases, or at least not in the sense of phrases as multi-word units. That is, although words-and-rules approaches predict no frequency effects for procedural sequences of words, they do not reject the concept of formulaic sequences per se. Rather, in a words-and-rules approach formulaic sequences must be given the status of words – that is, single items symbolically stored in the lexicon. The reason for this is that they make a distinction between stored forms and the grammatical rules that are applied to them, thus positing two qualitatively distinct systems (Pinker, 1999; Pinker & Ullman, 2002; Ullman & Walenski, 2005). Phrasal frequency effects pose a challenge for such models because they blur the distinction between lexis and syntax (or, in more extreme viewpoints do away with it altogether, which is discussed below). In order for formulaic language, and a formulaic advantage, to be incorporated into words-and-rules models, all instances of formulaic sequences need to be stored as single, symbolically represented items (and thus constitute 'long words').

To argue for formulaic language to be represented as 'words' is, however, somewhat problematic. As Arnon and Priva (2013) point out, it undermines the qualitative difference between stored and computed forms proposed by words-and-rules accounts, as well as having a number or problems. First, the number of items that would need to be included in the lexicon would be immense, making search times for an item in the lexical store long (e.g., Baayen, Hendrix, & Ramscar, 2013). This will be discussed in more detail below. However it is worth considering at this point just how large a percentage of every day speech formulaic language may constitute: Erman and Warren (2000) estimate that over 52 to 59 percent of spoken and written language is made up of formulaic sequences; Sorhus (1977) similarly estimated 41 percent. Van Lancker-Sidtis and Rallon (2004) are more conservative, estimating 25 percent. However, even at this lower figure, the number of formulaic sequences required for nativelike speech is immense. To put this into context, the Google Web1T Database (Brants & Franz, 2006) lists approximately 244 million sequences of three words, 328 million of four words, and 294 million of five words. Clearly including formulaic

sequences as ‘words’ is no simple matter.

Not only this, but even for typically ‘formulaic’ sequences such as core-idioms which are considered non-decomposable (Grant, 2005), very few, if any, are indeed truly non-flexible (Kyriacou, Conklin, & Thompson, 2019; Read & Nation, 2004). In a system which requires instances to be stored in memory as unanalysable wholes, this would not be the case. Further, how would we account for processing advantages found for lexical bundles which show not only an influence of phrasal frequency, but also of sub-string frequency (e.g., Tremblay & Baayen, 2010)? It has already been pointed out that formulaic sequences are often hierarchical in nature, and a phrase such as *what’s the weather like?* may consist of *what’s the*, *what’s the weather*, *the weather*, *the weather like* – all of which may receive processing advantages as individual sequences, or as a single amalgamation. This is clearly difficult to account for in a lexical approach and has led to scholars even as early as Bolinger (1975) to look for alternative, or more nuanced ways of explaining formulaic language.

Usage-based theories, in contrast, either blur the distinction between the ‘words’ and the ‘rules’ that operate in cognitive processing, or, at the more extreme end of the scale, do away with the distinction altogether (e.g., Elman, 2009 proposes a 'Simple Recurrent Network' model that sees all linguistic material in terms of a dynamic network and requires no 'lexicon'). One class of usage-based model is that of ‘frequency-threshold’ approaches, which posit a processing system that defaults to pre-fabricated items being stored in long-term memory that ‘switches’ to an analytical, syntax driven, system when needed (e.g., Goldberg, 2006; Skehan, 1998; Van Lancker Sidtis, 2012; Wray, 2002; Wray & Fitzpatrick, 2008). In contrast to the words-and-rules approach, these approaches are emergentist models which integrate memory-based storage (primary processing) and analytical processes (secondary processing) and are, therefore, fundamentally dual-processing systems. Importantly, under this umbrella of models, a sequence of words would need to be encountered a certain number of times for a formulaic processing advantage to emerge.

Exactly how many times a sequence would need to be encountered to reach this theoretical ‘threshold’ is not clear. However, a good analogy might be of a ‘boiling point’ where, just as water that is heated does not turn into vapor until it reaches 100°C, only once a certain number of encounters is achieved will a sequence of words

become stored and processed as such. As mentioned, to account for this qualitative shift in processing style, models of this type require two distinct processing systems: one to deal with analytical, rule-processed language and another to process sequences which have passed the required threshold. Key to the current discussion is that such approaches maintain a frequency-driven distinction between phrases which are stored and those which are not, and this happens in reaction to input. This kind of model is attractive because it allows a language user to make the most of the efficient processing that formulaic sequences afford, while still retaining the ability to use language creatively via analytical processes such as syntax. Skehan (1998) describes these models as such:

> … two systems coexist, the rule-based analytic, on the one hand, and the formulaic, exemplar-based on the other. In the former case, compact storage and powerful generative rules operate together to 'compute' well-formed sentences. In the latter, the central role is occupied by a very large, redundantly structured memory system and (presumably) less powerful rules which operate on chunks much of the time, rather than on individual items (p. 54)

In Wray's 'Needs Only Analysis' (Wray, 2002, 2008), prefabricated retrieval is seen as the default processing mode, with computational systems only kicking in when necessary. Formulaic sequences and language chunks are not analysed unless a need arises, making it a serial processing model. Thus, the processing system is thought to default to a more efficient, routinised system, but allows for flexibility when—and only when—needed. Models such as this can be considered 'chunk-making' and 'chunk-breaking' approaches, in that the system acquires chunks and then breaks down (e.g., *hurryup* is acquired as a chunk and later the components *hurry* and *up* are recognised), and at the same time chunks are formed from high-frequency sequences of individual words over time (e.g., Goldberg, 2006; Wray, 2002, 2008). To extend (and admittedly stretch to its limits) the analogy of a web-browsers' caching system, this might be likened to viewing a page via a screenshot where text and images are merged into a single file, and only loading the original in its constituent parts if needed. Other models posit a system whereby computation and direct access run in

parallel and it is conceived of as a 'race'. Because retrieval of a prefabricated form bypasses computational processes, it is inherently faster. However, if a prefabricated unit does not exist, computation would win the race. Although parallel dual processing models have looked at access of literal and figurative meanings in idioms (Bobrow & Bell, 1973), little has been said about parallel processing in regards to non-compositional sequences such as lexial bundles, although Logan's (1988) 'Instance Theory' makes fairly clear preditions for them.

Instance Theory predicts memory-based storage for utterances which were initially constructed word-by-word and which are replaced—i.e., 'chunked'—with larger constructions over time through use (Logan, 2005; Robinson & Ha, 1993; Skehan, 1998). Instance Theory can explain the formulaic advantage through a process of replacing rule-based, algorithmic performance with memory-based performance. With each exposure to a language string (the original theory considered strings of phonemes to create words, but the concept is equally applicable to formulaic sequences – see Schmidt, 1992 for a good overview), an 'instance' of it is imprinted in memory. With repetition, eventually that instance reaches 'boiling point' and is compounded in memory such that its representation is accessed faster than computational systems can compute the same string. From that moment on, the sequence is retrieved as a complete unit, not created using rules.

The frequency-threshold models described so far, however, are not without problems. First, dual processing assumes a symbolic approach to language storage. In one sense, threshold models, and the dual processing that make up an important part of them, are simply more flexible versions of words-and-rules approaches and come with many of the same problems. Instance Theory, for example, only says that the exact instances that have been encountered are stored in memory (Schmidt, 1992) and therefore we should expect novel or infrequently encountered strings to be processed computationally, not via single-step memory retrieval. However, even Logan's own experiments (Logan, 1988; Logan & Stadler, 1991) provided evidence of speeded processing beyond exact instances. Further, these theories lead to an infeasibly large number of items—many of which would be very similar—which would need to be stored in memory (Baayen et al., 2013). These quantities would grow so large that as experience of language was accumulated, recognition times would become longer, less

efficient and more error-prone (Nelson, 2018). Clearly this is contrary to the processing advantage observed in behavioural studies. Not only this, but as mentioned above, how are we to account for studies which find facilitation not just from complete sequences, but also their sub-strings (e.g., Tremblay & Baayen, 2010)? Further, if formulaic sequences are stored as instances of items on the basis of a frequency theshold (Goldberg, 2006; Wray, 2002, 2008), how are we to define such a threshold? While some studies have collapsed frequency into 'binned' high vs. low metrics (e.g., Jiang & Nekrasova, 2007; Kim & Kim, 2012; Sosa & MacFarlane, 2002, among others), others such as Arnon and Snider (2010) have found processing latencies are better predicted by statistical models that include frequency as a continuous metric.

An alternative approach that does account for frequency of occurrence as a continuous variable are the 'single-system' processing accounts that Arnon and Snider (2010) refer to as 'distributed', or 'continuous' processing models. In these approaches, all linguistic material—whether words or larger sequences—is processed by the same cognitive mechanism (Arnon et al., 2017), and they advocate for a system driven by the probabilistic information present in language resulting in higher frequency items being processed more quickly with computation speeded on a word-by-word and sound-by-sound basis (Jurafsky, 2003). Importantly for our discussion, this means that 'continuous' accounts predict each and every exposure to a sequence of words or sounds via input will affect its subsequent processing (Bod et al., 2003 2003; Bybee, 1998, 2006; Bybee & Hopper, 2001; Mclaughlin et al., 1983). Importantly, while frequency-threshold approaches are essentially based on symbolic representation, and store formulaic sequences as a 'holistic whole', distributed approaches procces language in a sound-by-sound manner, with the processing advantage observed in the literature arising as a result of the degree of activation across the network's constituent parts. To put this more simply, a difference beween higher or lower frequency in these approaches is not a difference of type (stored or not stored), but rather it is a differnce in activation (Arnon & Snider, 2010). The processing advantage is seen as algorithmic "fine-tuning" (in the words of DeKeyser, 2015, p. 95) rather than a qualitative change in processing style (for example Anderson, 1992; Baayen, 2010; Baayen et al., 2013; Elman, 2009, 2011).

These approaches have important implications for the processing studies with

non-native speakers described above. If 'fine-tuning' and a processing advantage result from input, we must ask ourselves what input have the learners received? In other words, an absent processing advantage would indicate that sufficient input had not been received for the system to be fine-tuned. In contrast to this, researchers like Babaei et al. (2015) take the lack of a processing advantage as an indication that lexical bundles "are not processed holistically in the mind" (p. 12), concluding that because they did not find a processing advantage, second-language speakers must process language in a compositional, word-by-word manner. However, their study used lexical bundles from Biber et al. (1999) and extracted frequency scores from the COCA corpus (Davies, 2008). Arguably then, this study might only tell us that this particular group of Persian speakers did not display a processing advantage based on lexical bundles from Biber et al. (1999) or frequency scores form the COCA corpus; however, it does not tell us they do not process lexical bundles in a nativelike manner. It could simply be that these learners were exposed to input different in its statistical distribution to native-speaker language. If we tested them on lexical bundles that we were sure they had encountered, distributed or continuous accounts would say that they would process them in a nativelike way; it is just that they were tested on the 'wrong' lexical bundles (see also Conklin, 2020).

Corpora provide the frequency counts used in processing studies and are generally viewed as a measure of linguistic experience. Thus, if a given unit of language is frequent in a large, representative corpus, it is assumed that speakers will have encountered this unit a number of times aligned with its corpus-frequency, and likewise, for lower-frequency units. For native speakers, and especially for high-frequency items, we are probably justified in assuming generalisability on the basis that most people will likely have quite similar exposure to their language. However, as Gernsbacher (1984) notes, corpora only provide an approximation of actual exposure, which may be particularly problematic for low-frequency words. Conklin (2020) gives the example of the word 'recasts', which, with a frequency of only five in the BNC corpus, can be considered very low. However, in the SLA community this term is likely to be fairly familiar and, in reality, a high-frequency item. Thus, in a lexical decision task, we would expect the word to be responded to quickly: perhaps faster, even, than a higher frequency word in the BNC. In this sense, we must be critical of

the corpora we use, and the measures of frequency derived from them. In the kind of studies described above, for native speakers we are probably, as stated, justified in assuming generalisability on the basis that most people probably have quite similar exposure to their language and, therefore, to high-frequency formulaic sequences. However, can this be said of non-native speakers? To do so arguably goes against everything which we know about second language development from usage-based approaches where, as demonstrated above, experience of language is seen as the key driving force behind acquisition and subsequently processing (e.g., Barlow & Kemmer, 2000; de Bot & Larson-Freeman, 2011). In terms of EFL learners, and in the context of the subsequent studies presented in this thesis, this 'experience of language' generally means classroom instruction and materials (Meunier, 2012; Nunan, 1991).

The textbooks used in many language classrooms are based on the words-and-rules approach (Gilmore, 2015). It is assumed that grammar should be the primary unit of instruction, alongside vocabulary. Therefore, the content of textbooks, examples and dialogues are generally written to exemplify specific grammar points, not necessarily to provide a 'nativelike' sample of discourse. Under a words-and-rules approach, this makes sense because we expect learners to acquire the rules and words they need, then combine them to produce language. Indeed, it still makes sense (albeit to a lesser extent) if we subscribe to a frequency-threshold model which predicts only items either learned (i.e., explicitly taught) as whole units or sequences of words experienced a sufficient number of times will be stored—and thus processed—as wholes. Therefore, we can assume the supporting context has limited influence, unless it is sufficiently repetitive (i.e., passes a 'threshold' and is represented in memory). However, a continuous approach would predict that learners take in all language they are exposed to and process it accordingly. Under this view, the quality of learning materials is extremely important – whether it is items chosen for explicit instruction, or the input learners receive via supporting text. To put it simply, under this viewpoint, if the input is not 'nativelike', neither will be the language developed by the learner.

Clearly this is an important factor to consider in the development of learning materials. However it has not, to my knowledge, been touched upon before with second language learners (the only study to my knowledge that considers input

matched to the participant-group is Bannard and Matthews, 2008, who as mentioned in the previous section, studied formulaic sequence processing in monolingual children). Sadly, very little has been said about formulaic sequences in teaching materials; and the only studies that have looked at this have tended to examine textbooks based on pre-defined lists taken from native-speaker corpora, but have not considered the actual lexical bundles that naturally exist in the textbooks as a result of the way they are written (e.g., Koya, 2004; Mcaleese, 2013; Tsai, 2014 – this will be discussed in detail in the following chapter). For this reason, the starting point for the four empirical studies presented in Chapters 4, 5 and 6 was to ascertain the formulaic nature of the input learners are likely to receive (Chapter 3). This will provide the core data required to explore the usage-based processing of lexical bundles in second language learners which we do in Chapter 4 (asking whether junior high school students have a processing advantage), in Chapter 5 (which explores naturalness-intuitions in Japanese English learners based on frequency metrics taken from their early textbooks), and finally in Chapter 6 which looks at optimising the materials learners are presented with.

Chapter 3. "What Are You Talking About?"

In a communicative approach to language teaching, students are presented with 'authentic' language, which is thought to allow them to produce it in a nativelike way. The current study explores whether the lexical bundles in communicative Japanese junior high school textbooks are representative of conversational English. To do this, we use a corpus-based approach that compares the most frequent lexical bundles in the textbooks to those in an English reference corpus. The study finds that although lexical bundles are very frequent in the textbooks, and conform relatively well to English patterns at shorter lengths (3-word lexical bundles), they deviate considerably at longer ones (4-, 5- and 6-words). This has important implications for the communicative utility of the language in the textbooks. The study presented in this chapter was published in the *International Journal of Corpus Linguistics,* as 'What are you talking about?' An analysis of lexical bundles in Japanese junior high school textbooks' (Northbrook & Conklin, 2018b).

Introduction

Many schools and educational agencies advocate taking a communicative approach in foreign language classrooms. In this approach, students encounter and are asked to produce authentic language in the classroom, which is thought to give them the ability to reproduce it naturally in the real world. The cornerstone of the communicative approach is providing students with authentic language materials. This view of foreign language teaching aligns well with usage-based accounts of language acquisition, which maintain that language develops in reaction to use and exposure (e.g., Barlow & Kemmer, 2000; Bybee, 2006; Ellis et al., 2013; Tyler, 2010). Crucially, this means that the quality of input defines the quality of language learned. In the case of English as a Foreign Language (EFL) students, this input is primarily from classroom instruction and materials (Meunier, 2012; Nunan, 1991). Despite this, studies focusing on secondary school textbooks have found the dialogues in them to be inauthentic, (e.g., Gilmore, 2004; Nguyen & Ishitobi, 2012; Römer, 2004, 2005) in that they do not accurately represent the situations students need to be prepared for. Relevant to the current study, Ogura (2008) finds that although the Japanese Ministry of Education

(MEXT) places an emphasis on communicative competence, and textbook designers for upper-secondary school claimed their books are communicative, the books did little to promote communicative competence.

To ensure that textbooks are providing authentic, real-life opportunities for communicative practise, we need to look at whether students are exposed to frequent, high utility language relevant to general everyday communication. To assess the authenticity of the teaching materials in MEXT approved junior high school textbooks, the present study examines the use of lexical bundles (e.g., *you know what*). More specifically, the study looks at whether the highly frequent lexical bundles that are present in the textbooks occur frequently in real language and similarly whether low-frequency ones appear and occur less frequently. To assess this, we have used a modified version of the framework outlined by (Stubbs & Barth, 2003). The authors created three corpora to represent distinct genres of English – academic language, fiction, and what they label 'learned' language (consisting of texts from sources such as natural and social sciences and humanities). They then extracted lexical bundles from these corpora and analysed the resulting lists for differences between the type and frequency of the items found. They found that each of the text types contained many lexical bundles; however, the lexical bundles were qualitatively different. Lexical bundles have also been used to differentiate between television genres (Bednarek, 2012), legal genres (Breeze, 2013) and academic and conversational genres (e.g., Conrad & Biber, 2004; Gray & Biber, 2013). In the same way, an investigation of lexical bundles should be able to demonstrate how close the language used in communicative textbooks is to the language used in real communicative situations.

Lexical bundles in second language learning

Speakers tend to express the same things again and again, using the same words each time, resulting in repetitive language use (e.g., Biber, 2009; Erman & Warren, 2000; Nattinger & DeCarrucio, 1992; Pawley & Syder, 1983; Wray, 2002). Lexical bundles are, as discussed in the previous chapter, a type of repetitive language described as "recurring sequence[s] of three or more words" (Biber et al., 1999, p. 990) and are

usually defined according to a frequency threshold and criterion for dispersion. For example, Biber et al. (1999) set a minimum of 10 times per million words for 3- and 4- word lexical bundles, and at least five times per million words for 5- and 6-word lexical bundles. However, this number is somewhat arbitrary, and studies tailor the selection criteria to fit their individual needs, particularly when working with small corpora (see Biber, 2006 for a detailed discussion of this). Importantly, unlike some other kinds of formulaic language such as idioms or phrasal verbs, lexical bundles are not necessarily linguistic units with idiomatic, grammatical, semantic or pragmatic status (Biber, 2009; Stubbs & Barth, 2003). That is, they are often incomplete fragments (*in the middle of the*), and often span grammatical categories (*yes I do I*).

Lexical bundles, and formulaic language in general, are important for second language learners for two key reasons. First, by knowing chunks of language, learners can express themselves beyond their current level (Myles et al., 1998; Myles, Mitchell, & Hooper, 1999; Nattinger & DeCarrucio, 1992; Wray, 2002). For example, one does not need to have learned about post-head adverbial modification to use the phrase *thank you very much* in conversation, or to even consider that technically speaking the related phrase *thanks very much* is not even 'grammatical' (although it is very much a standard, accepted expression). Second, as discussed at length in Chapter 2, knowledge of formulaic language helps learners to free up memory resources (Kuiper, 2004; Pawley & Syder, 1983) and sound more natural (Boers et al., 2006; Dickinson, 2012; Pawley & Syder, 1983).

Seminal studies such as those from Pawley and Syder (1983), Biber (2009), Erman and Warren (2000), Nattinger and DeCarrucio (1992), among others, have demonstrated the pervasiveness and importance of formulaic language, including lexical bundles, and a great deal of research has been carried out exploring the phenomenon across multiple disciplines. Studies from corpus linguistics have aimed to categorise the function and structure of lexical bundles (e.g., Biber, Conrad, & Cortes, 2004; Biber et al., 1999; Hyland, 2008). In second language acquisition, studies have tried to identify the best lexical bundles to include in teaching (e.g., Simpson-Vlach & Ellis, 2010), and how best to use lexical bundles to improve fluency (e.g., Boers et al., 2006; Dickinson, 2012; Taguchi, 2007). Others have also looked at lexical bundles in learner output (e.g., Cortes, 2004; Yaoyu & Lei, 2011). Further,

behavioural studies have investigated the psychological reality of lexical bundles, among other kinds of formulaic sequences (see for example Conklin & Schmitt, 2012; Siyanova-Chanturia & Martinez, 2014 for good overviews).

In terms of studies within corpus linguistics which have looked at lexical bundles in textbooks, the majority are in academic and university contexts (Biber, 2009; Biber et al., 1999; Chen, 2010; Simpson-Vlach & Ellis, 2010; Wood, 2010b). Only a few studies have focused specifically on multi-word units in materials for secondary school students, though none of them are *lexial bundle* studies per se, and fewer still on secondary school materials in the Japanese context. Koya (2004) compiled a list of collocations based on collocation dictionaries, and examined their usage in Japanese senior-school textbooks. Only a small number of the collocations appeared in the textbooks, and over half only appeared once with only eight items appearing more than five times. In a similar study, Tsai (2014) compiled a list of collocations based on the prescribed wordlists for Taiwanese, Ministry approved textbooks. Only a small number of collocates were presented in the textbooks, and similar to Koya (2004) these were not repeated often.

Although only a few studies have looked specifically at lexical bundles in secondary school materials, there have been studies looking more generally at formulaic sequences explicitly taught in textbooks (Koprowski, 2005; Mcaleese, 2013), as well as one looking at teacher talk in Korean classrooms (Kwon & Lee, 2014), and several examining lexical bundles in academic teaching materials. Like the studies above, these studies primarily found item selection to be insufficiently representative of naturally produced language. Koprowski (2005) analysed three English as a Second Language course books and found many lexical bundles in the books did not appear in a reference corpus, and many were deemed not useful due to their low-frequency in the reference corpus. Mcaleese (2013) compared formulaic sequences taught in a commercially available course book to the Bank of English corpus. Formulaic sequences were common in the textbook, but a large proportion were very low-frequency. The author concluded that the textbook failed to accurately represent real language usage, and was therefore of limited value.

In academic contexts, Biber et al. (2004) compared lexical bundles in four separate genres of English: conversation, classroom teaching, textbooks and academic

prose. They found lexical bundles to be very common in conversation and university teaching, and to a lesser extent, in textbooks and academic prose. Importantly, it appears that textbooks and academic prose match in terms of the number of lexical bundles (see also Biber, 2009; Biber et al., 1999; Simpson-Vlach & Ellis, 2010). This is an encouraging finding and could be taken as an indication that academic teaching materials are authentic. However, the authors did not compare the exact lexical bundles in each genre. Studies by Wood (2010b), focusing on EAP textbooks, and Chen (2010), focusing on lexical bundles in textbooks for electrical engineering students, however, showed that formulaic language in textbooks was not representative of language found in authentic reading materials. Wood (2010b) extracted and analysed lexical bundles from a corpus of six commercially available textbooks finding that most items were in the instructional components of the material but not in the reading texts, meaning that although students did get exposure to lexical bundles, it was not in the texts that were supposed to represent authentic academic reading. Importantly, the lexical bundles they did encounter were not the same as the ones they could expect to encounter outside of the textbooks. Similarly, Chen (2010) found that only a third of the lexical bundles from actual electrical engineering books appeared in preparatory textbooks and the ones that did appear covered only a narrow range of functional categories.

These studies show that, although students are exposed to lexical bundles and formulaic language, there is a mismatch between the lexical bundles in the input and what students might encounter in the real world. Learners are given the potential to acquire lexical bundles and enjoy the benefits which come with them – however, the lexical bundles presented are not necessarily the ones they need. This is problematic because per usage-based accounts of language learning, students become fluent in the actual language they are exposed to. Take for example a phrase such as *in the middle of the*. The first time this is encountered, it will be processed in a word-by-word manner. However, it will leave an imprint in memory that is strengthened by subsequent exposures (e.g., Logan, 1988 − see also the discussion in Chapter 2). Thus, when reading for the first time there would be no difference between *in the middle of the* and any other sequence of words. However, when reading it for the second time, there should be a 'boost' in processing speed because we already have an imprint

stored in memory (see for example Conklin & Schmitt, 2008; Isobe, 2011; Jiang & Nekrasova, 2007). Where lexical bundles in materials are nativelike, in that frequent items occur frequently and less frequent items less frequently, this boost will match that of native speakers. Where language is not representative of everyday language, students experience a boost in language they are unlikely to use. Further, because of the imprint left in memory, formulaic sequences such as lexical bundles are highly predictable and tend to be considered more natural sounding (Ellis, 2001; Pawley & Syder, 1983). This means that acquiring lexical bundles that are non-nativelike will make it harder to judge the appropriateness of a given phrase, a problem which potentially could impact on a student's confidence using the language (Wray & Fitzpatrick, 2008). This will be the topic of Study 3, presented in Chapter 5. Unfortunately, very little research has looked at materials designed for more general English, such as those used for beginners and/or in primary and secondary school programmes. The current study aims to fill this gap by exploring lexical bundles in a corpus of MEXT-approved junior high school textbooks, to determine if they are representative of naturally produced language. In other words, the study examines the following research questions:

1. Do the textbooks contain lexical bundles and provide opportunities for their acquisition?
2. Are the lexical bundles that appear frequently in the English textbooks also frequent in native-speaker English?

If the answer to these questions is '*yes,'* it will be taken as an indication that the communicative textbooks are representative of naturally produced language, and are preparing students for real-life communication.

Methodology

A corpus of Japanese junior high school English textbooks (referred to as the JHSETC – the Junior High School English Textbook Corpus) was constructed and consists of 152,966 word tokens, and 6,718 word types. Although small, the corpus can be

considered exhaustive because it contains all of MEXT's currently approved textbooks (six series, with three books in each making a total of 18 textbooks). The six series consist of: *Columbus 21 English Course* (2013), 2,279 types and 24,071 tokens; *New Crown English Series* (2013), 2,523 types and 22,142 tokens; *New Horizon English Course* (2013), 2,751 types and 26,598 tokens; *One World English Course* (2013), 2,514 types and 24,845 tokens; *Sunshine English Course* (2013), 2,864 types and 31,481 tokens; and *Total English New Edition* (2013), 2,131 types and 23,829 tokens. It must be noted that transcripts for listening components were not included in the corpus because they were not uniformly available across the different series. This is not considered to be a problem because the listening activities are based on the contents of the textbooks and therefore only add to the existing data, but do not add any new language or lexical bundles[1]. Further, elements such as wordlists, textbook titles and section headings were not included in the corpus.

The language presented in these textbooks is, for the most part, presented as oral language in the form of dialogues and monologues. The SUBTLEXus (Brysbaert & New, 2009) was selected as the reference corpus. It is made up of scripted language from film and television, that although is not spontaneous speech, it has been written to approximate it. Crucially, the SUBTLEXus has a number of advantages over other spoken corpora such as the COCA's spoken component, the Santa Barbara Corpus of Spoken American English, the Switchboard corpus or the Buckeye Speech Corpus: (i) it contains both conversations and monologues; (ii) it is modern; (iii) it includes a very wide range of sources (8,388 files of subtitles from American films and television series); (iv) it is the least idiosyncratic of the alternatives mentioned above; and (v) it

[1] To demonstrate that including listening scripts to the corpus makes little difference to the overall data set, we transcribed the listening exercises from the *New Crown* series. It would have added 5,019 words to the corpus – assuming listening scripts are all similar in size, this would have added around 30,000 words to the overall corpus to take it from around 150,000 words to 180,000 words. We generated a list of lexical bundles from the listening 'texts'. The listening scripts had 41 lexical bundles. Ten did not appear in the textbook version, but three of which are actually part of a larger lexical bundle: *does she play, high will be, the high will, the high will be, an old man, does he play, he doesn't play, kind of bird, something that has, who am I.* We compared the top 10 lexical bundles from the New Crown with and without the listening scripts. There were no differences and therefore we conclude that including the listening scripts does not change the overall findings of our study.

has the highest correlation with word processing variables (Brysbaert & New, 2009). The SUBTLEXus is also a good approximation of the kind of language American native speakers are regularly exposed to; the Statistic Brain Research Institute found that 99% of American households have TVs, 65% of Americans regularly watch TV while eating dinner, and on average American children spend 1,480 minutes per week, or around 3.5 hours a day watching television (Statistic Brain Research Institute 2017).

One drawback of the SUBTLEXus corpus, however, is that the size of the corpus, at 51 million words of English, makes it problematic as a reference for one of just over 150,000 words. Different approaches are used to deal with size differences amongst corpora. One method is to normalise each frequency into a value per million words (see for example McEnery, Xiao, & Tono, 2006) , or alternatively a 'Zipf score' can be used (van Heuven et al., 2014) which ranks items on a logarithmic scale between 1 and 7 (where 7 is considered extremely high-frequency). However, both approaches are problematic when using corpora as small as the JHSETC. For example, taking a 5-word lexical bundle and normalising it would mean that one occurrence in the 150,000-word JHSETC would indicate a frequency of roughly 7/million, or a Zipf score of over 4. However, this is out of line with what we know about 5-word lexical bundles, which are already thought to be frequent if they occur at a rate of more than 5/million (Biber et al., 1999). An alternative method involves considering a similarly sized, representative sample of the larger corpus. Ideally, this would be achieved using stratified random sampling techniques (e.g., Biber, 1993) to take a sample from multiple elements within the corpus (in this case, film subtitles), preventing over-representation of items that happen to be highly frequent in a single text. Unfortunately, the SUBTLEXus is only available as a single file of randomised lines (due to copyright restrictions), so this was not possible. Instead, a random sample of 19,500 lines was taken from the 51-million-word SUBTLEXus to create a corpus of 152,964 tokens. This is somewhat crude as a sampling method, but it should be noted that no special selection criteria were applied to the files making up the SUBTLEXus – rather subtitles were downloaded en masse and anything with an OCR error rate of less than 2.5% (worked out using the *Aspell* spell checker) was included.

Lexical bundles three to six words long, which occurred four or more times in at

least three of the six textbook series were extracted from the JHSETC using *AntConc* (Anthony, 2014). Words containing an apostrophe (e.g., *let's*, *don't* or *isn't*) were considered one word. In the JHSETC the character '/' appeared because the textbooks contain many exercises which include alternatives separated by the forward slash (e.g., *Yes, I am. / No, I'm not*). This naturally leads to bundles such as *am No I'm*, which cannot be considered valid items. Thus, extracted items containing a forward slash were removed from the analysis. As recommended by Biber et al. (1999), we also removed items from the JHSETC that crossed speaker turn boundaries (e.g., *do you have I have*). A list of lexical bundles was extracted from the sampled SUBTLEXus. Because the SUBTLEXus is a single text file (as described above) we were unable to include range in the extraction. However, as a confirmative measure, following Biber et al. (1999), we removed any 3- and 4-word items that did not reach the threshold of 10/million, and any 5- 6-word items that did not reach the threshold of 5/million in the full SUBTLEXus. This ensured that none of the lexical bundles from the smaller version of the SUBTLEXus were over-representative of the entire corpus. We then conducted a Pearson's product-moment correlation test comparing the frequency scores from the sampled SUBTLEXus lexical bundles with the scores from the full corpus. The results were significant (r (1688) = .82, $p < .0001$), so we can assume that lexical bundles from the sampled SUBTLEXus generalise to the full corpus.

The lexical bundles were classified according to the number of words they contain (3-, 4-, 5- and 6-word lexical bundles) and ranked from most to least frequent. In many cases, a longer lexical bundle contains shorter lexical bundles. For example, *in the middle*, and *the middle of* are both 3-word bundles and are contained in the 4-word bundle *in the middle of*. These, in turn, are part of the 5-word lexical bundle *in the middle of the*. For the purposes of the current study, items such as *in the middle* and *in the middle of* are treated as independent lexical bundles. One important issue is that if a shorter lexical bundle does not occur, or is under-represented in the JHSETC, the larger lexical bundles which contain it will also be under-represented. For example, if the 3-word bundle *do you have* was under-represented in the corpus, this would mean that *do you have any*, *do you have a*, *do you have that*, etc. would also be under-represented. When discussing frames and lexical bundles that have multiple continuations or preceding words, we use the '~' notation to indicate that we are

referring to the pattern, and not the discrete lexical bundle (e.g., *do you have~* can be *do you have any, do you have a, do you have that,* and so on).

Analysis and results

Little is known about lexical bundles in textbooks for teenagers and beginner level students, therefore first we provide an overview of the lexical bundles as they appear in the JHSETC and compare them to the SUBTLEXus. We then analysed the lexical bundles by comparing and examining the JHSETC items side by side with the SUBTLEXus lexical bundles. Following Stubbs and Barth (2003), the analysis is limited to the most frequent items from each lexical bundle size in order to make the data set more manageable. We started with the 10 most frequent lexical bundles in the JHSETC and SUBTLEXus corpora at each lexical bundle size and labelled them as follows: (i) items that appear in the top 10 of the other corpus, (ii) items that appear in the top 50 of the other corpus, (iii) items that do not appear in the top 50 of the other corpus, and finally (iv) items that do not appear at all in the complete reference list. We then counted the number of items in each category, and finally examined them closely for differences and similarities.

Overview of lexical bundles

The extraction technique outlined above generated four sets of lexical bundles. The distribution of these across the word lengths is displayed in Table 1. There are significantly more lexical bundles in the JHSETC across all sizes by both type and token. This is also true of each word length for both type and token. This suggests that the language used in textbooks is considerably more formulaic than that of language contained in the SUBTLEXus – especially at longer lexical bundle lengths.[2]

---

[2] We also considered whether there were differences in lexical bundle types across the six-textbook series. When frequency scores are normalised per 1,000 words, there was no difference in the number of lexical bundles at any size ($X^2 > 0.43$). However, when looking at the actual lexical bundles themselves, there is very little overlap. At 3-words there is only 4.65% overlap in lexical bundles across the series, 2.06% at 4-words, 0.99% for 5-words, and 0.37% for 6-words, yielding a total of 7% overlap in lexical bundles across the series.

Table 1. Number of lexical bundles by type and token for each of the word lengths in the JHSETC and SUBTLEXus.

	JHSETC		SUBTLEXus			
	type	token	type	token	X^2 Type	X^2 token
3-word	1,661	18,739	1,335	10,288	< .001	< .001
4-word	526	4,816	147	996	< .001	< .001
5-word	175	1,430	18	91	< .001	< .001
6-word	73	554	1	4	< .001	< .001

Analysis of frequent lexical bundles

Lexical bundles clearly make up a substantial part of Japanese learners' experience of secondary school English. The previous section demonstrated that lexical bundles may be more common in the textbooks than in native speaker language – but are they qualitatively similar? In their corpus analysis of three separate text genres, Stubbs and Barth (2003) demonstrated that lexical bundles were highly frequent in all of them – but that the actual lexical bundles were qualitatively very different. In this case, because textbooks are supposed to represent authentic, communicative language, we would expect to see very little difference between the JHSETC and the SUBTLEXus.

As a starting point, we considered the top 10 most frequent items for each lexical bundle size. These are displayed in Table 2. Chi-square statistics for lexical bundles discussed in this section can be found in Appendix 1.

Three-word lexical bundles

Fifteen items (30%) appear in the top 50 most common items from the JHSETC that do not appear at all (zero frequency) in the SUBTLEXus and four items (8%) appear in the SUBTLEXus but not in the JHSETC (0 frequency). Further, 22 (44%) items from the JHSETC do not appear in the top 50 for the SUBTLEXus, and 37 (74%) items from the SUBTLEXus do not appear in the top 50 for the JHSETC.

Notably, the bundle *yes I am* – which does not appear at all in SUBTLEXus (0 frequency) – is in the top 10 most frequent items from JHSETC. There are two items

Table 2. The 10 most frequent lexical bundles in the JHSETC and SUBTLEXus corpora at each bundle size.

3-word lexical bundles	
JHSETC	I want to**, do you like†, what do you**, do you have†, a lot of*, I have a*, yes I am, do you know**, I'm going to*, this is my†
SUBTLEXus	I don't know*, what are you*, what do you**, are you doing†, you want to*, you know what†, out of here†, do you think†, I want to**, do you know**
4-word lexical bundles	
JHSETC	la la la la, are you going to†, do you want to*, nice to meet you†, I want to be†, want to be a, what time is it, do you have any, have a lot of, what do you want**
SUBTLEXus	what are you doing*, what do you think*, get out of here†, what do you want**, I don't want to†, what do you mean†, are you talking about†, what are you talking, I don't know what†, how do you know†
5-word lexical bundles	
JHSETC	la la la la la, I want to be a, what do you want to*, are you going to do, what are you going to, how many cds do you, many cds do you have, a junior high school student, I'm going to talk about, what time do you get
SUBTLEXus	what are you talking about, what are you doing here, you know what I mean, what are you gonna do, I don't know what to†, in the middle of the†, what do you want to **, know what you're talking about, what are we gonna do, what do you know about †
6-word lexical bundles	
JHSETC	la la la la la la, how many cds do you have, what are you going to do, what time do you get up, called to say I love you, just called to say I love, I'm a junior high school student, I'm going to talk about my, what do you want to do, I just called to say I
SUBTLEXus	get the hell out of here

*** Items that appear the top 10 of the other corpus; * items that appear in the top 50, † items that do not appear in the top 50; underlined items do not appear at all.*

in the top 10 for JHSETC that do not appear in the top 50 of SUBTLEXus (*do you like*, *this is my*) and three items in the top 10 of the SUBTLEXus that do not appear in the top 50 of JHSETC (*are you doing*, *out of here*, *do you think*). Three items appear in the top 10 items for both the JHSETC and the SUBTLEXus, but have a significant difference in frequency in the two corpora: *what do you, do you know*, and *I want to*.

The top 50 from the JHSETC contains seven yes/no response bundles, four of which, along with the bundle *yes I am*, do not appear in the SUBTLEXus at all. The remaining three do not appear in the top 50 from the SUBTLEXus. In fact, the only such bundle in the SUBTLEXus top 50 is *no no no*, which although qualitatively different to those from the JHSETC in that it does not follow the set 'yes/no + pronoun + verb' pattern, is relatively common in the SUBTLEXus with 33 occurrences, but does not appear at all in the JHSETC. Related to these, *do you~* patterns are also the most common lexico-grammatical patterns in both lists. These are displayed in Table 3

Table 3. Comparison of yes/no fragment bundles, do you~ and what do you~ bundles in the JHSETC and the SUBTLEXus corpus.

Yes/no fragments	
JHSETC	yes I am (100 – 0)*, yes I do (88 – 4), no I don't (65 – 15), no I'm not (65 – 6), yes I have (60 – 0), yes it is (46 – 0), yes I did (42 – 0)
SUBTLEXus	no no no (0 – 33)
Do you~ bundles in the most common 50 items for each corpus	
JHSETC	do you like (171 – 10), do you have (165 – 17), do you know (98 – 45), do you want (87 – 40), do you play (56-0)
SUBTLEXus	do you know (98 – 45), do you think (38 – 48), do you want (87 – 40), do you mean (14 – 25)
What do you~ bundles in the most common 50 items for each corpus	
JHSETC	what do you want (30 – 20), what do you think (26 – 23), what do you have (23 – 0), what do you do (22 – 0)
SUBTLEXus	what do you think (26 – 23), what do you want (30 – 20), what do you mean (10 – 19), what do you know (0 – 8)

** Raw frequency scores are given in brackets for references as (JHSETC – SUBTLEXus).*

along with the 4-word continuations of these *do you~* patterns – *what do you* patterns, which we discuss in more detail in the following section.

Notably in the yes/no response bundles, only three of the items (*yes I do*, *no I don't, no I'm not*) appear at all in the SUBTLEXus corpus, and these are very low-frequency and do not appear in the most frequent 50 3-word bundles. The overall difference between these items is significant – *yes I do*, *no I don't*, and *no I'm not.* This is not, however, to say that question fragments are less frequent in the SUBTLEXus. To elicit a yes or no response of the type seen in the textbooks, a *do you~* or *are you~* question would be required. Both the JHSETC and the SUBTLEXus contain one *are you~* bundle each (*are you going* and *are you doing* respectively), as well as five *do you~* bundles for the JHSETC and four for the SUBTLEXus. This suggests then that although questions which elicit yes/no responses are equally common in both corpora, there is a large difference in the way these are responded to. Unfortunately, because the lines of the SUBTLEXus are randomised, it is not possible to say exactly how *are you~* and *do you~* bundles are responded to, however this is taken up below, briefly as part of the discussion of 4-word bundles, and in more detail for 6-word lexical bundles.

In both corpora, *do you~* bundles are the most common (see Table 3), but the verbs that end the pattern are quite different. The lexical bundle *do you play* appears in all six of the textbook series in the JHSETC, but does not appear at all in the SUBTLEXus, and in fact, this lexical bundle only appears 136 times in the full SUBTLEXus (2.7/per million). Turning to the most common *do you~* bundles from the SUBTLEXus, we see that they are less frequent than in JHSETC. However, the difference between two of the bundles is not significant (*do you think*, *do you mean*), suggesting that they are well represented in JHSETC. The other two lexical bundles occur more frequently in the JHSETC (*do you know*, *do you want*) which, while over-representing their frequency of occurrence, should give students the opportunity to acquire them.

Four-word lexical bundles

Forty (80%) of the 4-word lexical bundles in the top 50 from JHSETC do not appear

at all in the SUBTLEXus. Of the remaining 10 items, six appear in the top 50 of the SUBTLEXus. Sixteen (32%) items in the top 50 of the SUBTLEXus do not appear at all in the JHSETC. One of these contains a word that does not appear in the textbooks – *bitch* from *son of a bitch* – and would be considered inappropriate for junior high school students, but does represent an example of very casual language, something which seems to be missing from textbooks.

Only one lexical bundle, *what do you want*, appears in the top 10 of both lists. Five items from the JHSETC top 10 do not appear in the SUBTLEXus list. Notable in Table 2 is the phrase *what are you talking*, which appears in the top 10 most frequent items in SUBTLEXus but not at all in the JHSETC. In fact, *what are you* was also in the top 10 for the SUBTLEXus for the 3-word bundles, too, but only in the top 50 for the JHSETC. Interestingly *what do you mean,* arguably a polite synonym for *what are you talking about*, does appear in the textbooks 10 times, once again suggesting that the language contained in the textbooks is politer than that of the SUBTLEXus.

In terms of lexico-grammatical constructions, there is a great deal of variation in the types of bundles present in both corpora. However, in both *what do you~* phrases are the most common and are shown in Table 3. These are all extended versions of *do you~* patterns. The difference in frequency is significant for these items overall (X^2 (5) = 42.903, $p < .001$), but insignificant for the individual bundles – *what do you think*, *what do you want*, *what do you mean*. However, the difference is significant for the lexical bundles *what do you know, what do you have* and *what do you do*, which all have zero occurrences in SUBTLEXus.

Five-word lexical bundles

The first notable point with the 5-word lexical bundles is that there are only 18 in the SUBTLEXus. Importantly, only three of these appear in the JHSETC and of these only two are in the top 50. There is only one lexical bundle from the top 10 of either corpus that appears in the other: *what do you want to* (third most common in the JHSETC and seventh in the SUBTLEXus). The only other item that appears in both corpora is *how long have you been*, ranked at number 30 in the JHSETC and number 19 in the SUBTLEXus.

Unlike the smaller lexical bundle sizes, for 5-word sequences, there is very little repetition in terms of lexico-grammatical patterns. Apart from three patterns which occur twice each (*are you going to~*, *do you want to~*, and *do you have any~*), all the items occur only once in the JHSETC. Interestingly, none of the three repeated patterns appear in the SUBTLEXus. Similarly, in the SUBTLEXus only two patterns appear more than once (*what are you ~ing ~* and *what are ~ gonna do*). As with the smaller lexical bundles, these are predominantly versions of the *do you~* pattern, however, unlike previous lexical bundle sizes, we have reached a stage where the items cease to be frequent (or even appear) in the SUBTLEXus list.

Six-word lexical bundles

While there are 73 6-word lexical bundles in the JHSETC, there is only one in the SUBTLEXus (*get the hell out of here)* and it does not correspond to any in the JHSETC, likely because the language is not appropriate for a junior high school textbook.

Three lexical bundles that appear in the top 10 items from the JHSETC are from the longer pattern *I just called to say I love you* which is from the song 'Stand by Me' by Stevie Wonder that appears in four of the textbook series. The most frequent bundle is *la la la la la la*, which again comes from a song ('Sing' by Joe Raposo, appearing in three of the series) and would not normally be considered a lexical bundle. After the song lyrics, the most common 6-word bundles in the corpus are *how many CDs do you have* and *what are you going to do*, both with 22 occurrences. Interestingly neither of these appear in the abridged SUBTLEXus, and *how many CDs do you have* does not appear in the full SUBTLEXus. We also only see one recurring pattern in the top 50 items – *what do you want to~* (*what do you want to do* ranked at number 9 with 13 occurrences on the JHSETC and *what do you want to be* ranked at number 11, with 9 occurrences in the JHSETC), neither of which occur in the sampled SUBTLEXus, however they do appear in the full SUBTLEXus corpus (*what do you want to do* is frequent with 7.47/ million occurrences, and *what do you want to be* is quite uncommon with 0.68/million occurrences).

As mentioned above, the lexical bundle *what are you going to do* appears in the

textbooks 22 times, while *what are you gonna do* does not appear at all. In the full SUBTLEXus, the lexical bundle occurs 801 times (16/million), but significantly less than the 5-word bundle *what are you gonna do* containing the less formal *gonna* (1,438 occurrences, or 28/million – X^2 (1) = 181.23, $p < .001$). This indicates that the less formal, arguably more 'conversational' language is more common in the SUBTLEXus reference corpus than in the JHSETC. If we look at this more closely and assume that contracted forms such as *I'm*, *we're*, *gonna* and *wanna* are indicative of casual language, whereas *I am*, *we are*, *going to*, and *want to* are indicative of more formal language, we would expect casual language to contain a larger number of lexical bundles that include a contracted form. This is indeed the case – the JHSETC contains 176 lexical bundle types (7.7% of the list) including a contraction, whereas the SUBTLEXus contains 259 bundle types (17.24%). The difference is significant X^2 (1) = 15.84, $p < .001$.

There is only one 6-word item in the SUBTLEXus, and this is something most would consider inappropriate for a secondary school audience; therefore, it is fair to say that nothing is *missing* as such from the textbooks in terms of 6-word lexical bundles – although it is arguably another example of the more casual register used in the SUBTLEXus.

Discussion

The overview of lexical bundles in the JHSETC revealed that recurrent strings of words are very frequent in the textbooks – more so than in conversational English. Further, this is true of all bundle sizes. The fact that the JHSETC has more repetitions of lexical bundles could be considered beneficial, as repetition is an important component in sequence acquisition (Ellis, 1996; Webb, Newton, & Chang, 2013 − also discussed at length in Chapter 6). In other words, if certain lexical bundles are an important aspect of conversational competence, repeating them will help them be learned.

An important follow up question is whether the lexical bundles in the JHSETC are representative of conversational English. The analysis demonstrated that to an extent the 3-word lexical bundles found in the JHSETC are like those found in the

SUBTLEXus corpus. As the lexical bundles become longer, they seem to deviate from the patterns seen in actual conversational English: 4-word bundles only show a small amount of overlap in the corpora, and 5- and 6-word bundles are almost completely different. Further, the high-utility items that we see for the 3-word bundles appear to be, for the most part, embedded in considerably lower utility longer lexical bundles. A good illustrative example is the phrase *how many CDs do you have?*, which is likely of low utility. However, it contains high utility bundles such as *do you have*. This example, despite being the most frequent 6-word lexical bundle in the JHSETC, does not appear in the full SUBTLEXus corpus, the COCA, or the BNC. Neither is the pattern *how many ~ do you have?* particularly frequent in the full SUBTLEXus (88 occurrences or 1.72/million), with the most frequent fillers being *children* and *kids* (both with just 10 occurrences each). With *How many children do you have?* being the most frequent bundle in English, the inclusion of *how many ~ do you have?* is hard to justify for a teenage audience. This example points to a limitation of the current study; we did not systematically examine the relationship between lexical bundles embedded into larger lexical bundles. This would be an interesting question to explore in future research.

Although we see promising results at small lexical bundle sizes, we must conclude that the language presented in junior high school English textbooks is not sufficiently representative of language outside the classroom. This finding aligns with previous research looking at other kinds of textbooks for more advanced learners (Chen, 2010; Mcaleese, 2013; Nguyen & Ishitobi, 2012; Wood, 2010b) as well as findings from Koya (2004) and Tsai (2014). Perhaps a part of this is the tendency for textbook designers to script and tailor dialogues to specific goals – generally the illustration of a certain grammar point – without reference to real language (see Gilmore, 2007b for a discussion). Put another way, the lexical bundles in the JHSETC are examples of grammatically correct language – however, they are not examples of high-frequency probable language (Lewis, 1993). A good example of this is the lexical bundle *what time do you get up.* This is one of only three 6-word lexical bundles that appears in all of the six-textbook series, yet it only appears twice in the full SUBTLEXus (0.03/million). Unlike the instances in the JHSETC, both are used for a very specific purpose as shown in Examples (1) and (2):

1. What time do you get up? Why? Because we wanna cook you breakfast.
2. We're civilised people. Thank you. What time do you get up? About 8, I guess. I'll fix you breakfast.

In both instances from the SUBTLEXus, the question is being used to inform the interlocutor that breakfast will be ready, and therefore we would like to know what time you want it. In the textbooks, however, the item seems to be included purely as an example of a question used in relation to talking about time, usually as a simple question–answer pair (*'What time do you get up?'*, *'I get up at ~'*). Ishihara and Cohen (2010) point out that in many cases textbooks fail to take pragmatics into account. This certainly seems to be the case with the illustrative example here, as well as throughout the junior high school textbooks analysed in the current study.

More generally, the question/answer fragments identified in the corpora reveal several patterns. First, there are many formulaic question and answer patterns in the JHSETC which do not appear in the SUBTLEXus (e.g., *yes I am*, *no I'm not*). Second, the way native speakers answer yes/no questions is complex, and often answers do not explicitly state *yes* or *no* (Hockey, Rossen-Knill, Spejewski, Stone, & Isard, 1997). Corpus studies, such as Biber et al. (2004), have found that although yes-no question fragments are very common, the units of analysis are generally a fragment of the questions themselves and not the responses – presumably because no specific responses to these questions are frequent enough to appear in their lists of lexical bundles. Arguably, the inclusion of so many formulaic yes/no response fragments is beneficial for production, as it gives learners a simple way to construct their responses. However, it indicates to learners that this is what they should produce in the real world, and they will be listening for a *yes* or *no* that in many cases will not actually come. Third, there is a difference in the level of formality. High-frequency question fragments in the SUBTLEXus, such as *what are we gonna do*, are not present in the textbooks. Further, the SUBTLEXus bundles contain significantly more contractions. The analysis suggests that the language contained in textbooks is far politer than standard, everyday conversation. An in-depth analysis of register within the two corpora goes beyond the scope of this study but would be an interesting area for further research.

The current findings have implications for 'conversational' approaches and textbooks. There is a large literature in the Japanese context trying to explain the often poor level of achievement in English (see for example Hagerman, 2009). Could it simply be that students do get good at English, but it is in a genre that is misaligned with English that is encountered in the real world? Given some of the research discussed in the Introduction (Koya, 2004; Römer, 2004, 2005; Tsai, 2014) and the findings from the current study, this seems like a possibility. Thus, we can theorise that replacing language in the textbooks with language that is (i) attuned to the needs of junior high school students, (ii) represents the language they are likely to use and encounter and (iii) incorporates frequent lexical bundles, as derived from corpora, should be beneficial. More precisely, there may be frames like *how many ~ do you have?*, which are not frequent in everyday language, but which are deemed to be important for a particular grammar point. A corpus could be used to inform how such language is presented in textbooks so that it most closely matches the real world. The SUBTLEXus shows us that this frame most often comes up in the context of *how many kids/children do you have*. An appropriate adaptation of this for junior high school students might be *how many brothers/sisters do you have*. Additionally, there are instances where the utility of textbook language can be increased simply by making a lexical bundle less formal; for instance, replacing *what are we going to do* with *what are we gonna do*. However, in other cases, there is not a more frequent equivalent. This is the case with *what time do you get up,* simply because this is not something we often say. In fact, the pattern *what time do you~* itself is rather low-frequency with only 131 (2.56/million) occurrences in the full SUBTLEXus.

This does not mean that language examples like *what time do you~* should not be included in the textbooks per se. Rather it could perhaps be better demonstrated with something more open in meaning, allowing for it to be more easily combined with other more frequent language. For example, *what time do you get up?* could be replaced by *what time do you get off?* (27 occurrences in the SUBTLEXus – 0.53/million), although this is potentially unrelated to junior high school students in that they generally do not have jobs. The similar *what time do you finish?* (seven occurrences in the SUBTLEXus, or 0.14/million), or *what time do you want to ~* (six occurrences in the SUBTLEXus, or 1.2/million) are other alternatives. None of these

examples are what would be considered 'high-frequency' language, but are more open in meaning and can more realistically be incorporated into dialogues that comprise a range of topics. For example, one activity that most junior high school students can be expected to engage in is that of after-school club activities, and it is not difficult to create a dialogue that contains *what time do you finish? I want to talk to you* (5.14/million in the full SUBTLEXus). Or we can easily imagine a situation where a character, upon hearing his classmate has basketball/football/piano/etc. practise, asks, '*What time do you get off? Let's go to* (15.80/million) *the mall after.'*

Another very common 5-word bundle in the JHSETC is *do you have any pets* which only occurs 5 times (0.09 /million) in the full SUBTLEXus, and thus cannot be considered high-utility language. Yet by changing the word *pets* to *idea*(*s*), the frequency becomes 631, or 12.57/million. This is a large difference, and perhaps more importantly, it seems unlikely that most junior high school students are interested in whether someone has pets. However, for most of them – especially in their third and final year – topics such as what high school their friends are planning to go to almost certainly will be top of mind. It is not difficult, then, to imagine conversations where students ask each other *'Do you have any idea what high school you want to go to?'* or *'What are you gonna do next year?'*.

These are very simple examples but by considering the things students are likely to say, and combining that with data from corpora, coming up with textbooks full of high-utility language attuned to students' needs should not be difficult. Improving the quality of textbook content requires a top-down process of starting with overall topics that teenagers are likely to talk about, and working from there to engineer realistic dialogues that contain high-frequency language, distributed in proportion to what we would expect in a real conversation. Put simply, textbook designers are recommended to ask *what are you talking about?* and base materials on that.

Conclusion

The current study, working on the framework laid out by Stubbs and Barth (2003), compared lexical bundles in foundational, junior high school 'communicative' EFL textbooks and a corpus of spoken English and has highlighted major differences

between the two. Crucially, to follow the precepts of the communicative approach, such textbooks should: (i) present and model authentic conversational language; and (ii) the relative frequency of the input should mirror that of actual spoken English (i.e., high-frequency language should be encountered more frequently than low-frequency language). Our findings suggest that both of these elements are lacking. While studies of advanced learners in EAP contexts (Biber et al., 2004; Chen, 2010; Wood, 2010b) have identified similar issues, this study gives insights into the kind of input very early learners receive. This is important because, as will be demonstrated in the following chapter, early learners do develop a processing advantage for the lexical bundles in their textbooks (as would be expected form a usage-based approach). We conclude that the textbooks series under investigation are not representative of anything other than themselves; that is, they are a separate genre of English, not the representation of general conversational English that they claim to be. Thus, despite the claims that designers may make (Ogura, 2008), the textbooks fail in their role of providing students with authentic samples of the language they are likely to need in the real world.

Chapter 4. Is What You Put in What You Get Out?

As discussed in Chapter 2, usage-based approaches to second language acquisition put a premium on the linguistic input that learners receive and predict that any sequences of words that learners encounter frequently will experience a processing advantage. The current study explores the processing of 'lexical bundles' in beginner learners of English in Japanese secondary school. To do this we use a phrasal judgment task, testing students on items taken directly from their teaching materials. Students responded to lexical bundles that occurred in their textbooks significantly faster and more accurately compared with non-lexical bundles, and were sensitive to the frequency of occurrence in their textbooks. This study shows, perhaps for the first time, that even very low level, beginner secondary students are sensitive to the frequency of lexical bundles which appear in the input they receive from teaching materials, a finding that has important implications for teaching and material design. The study presented in this chapter was published in *Applied Linguistics,* as 'Is What You Put in What You Get Out? – Textbook-derived Lexical Bundle Processing in Beginner English Learners'. (Northbrook & Conklin, 2018a).

Introduction

It seems obvious that if we try to build something with poor materials rather than the best ones that the end product will be of lesser quality. This sentiment is nicely encapsulated in Computer Science by the saying *garbage in, garbage out*, and refers to the fact that the input computers receive wholly determines the quality of their output. However, is this true in second language learning? More precisely what is the relationship between learners' input and their subsequent processing? The current research will address this question by examining the processing of 'lexical bundles' taken from the textbooks examined in Chapter 3, by Japanese secondary school students learning English – an as of yet largely unexplored age and proficiency group in formulaic language research.

As discussed in previous chapters, loosely defined, lexical bundles are sequences of words that occur frequently in language. They are generally corpus derived and

may constitute a complete utterance (*how's it going*) but often do not (*a lot of the*). They fall under the broad heading of formulaic language, which is referred to by many different names in the literature (see Chapter 2, and also Wray, 2002 for a list of more than 50 such terms). In applied linguistics, there is a growing body of research that suggests formulaic language more broadly, and lexical bundles specifically, are processed more quickly than novel or less frequent sequences of words (Conklin & Schmitt, 2012; see also Chapter 2). Importantly, formulaic language is thought to contribute to fluent and accurate language use (Pawley & Syder, 1983) that facilitates real-time communication for both speaker and listener (Boers et al., 2006; Kuiper, 2004; Pawley & Syder, 1983) and reduces demands on working memory (Conklin & Schmitt, 2008; Kuiper & Haggo, 1984). However, important questions remain about how formulaic language becomes 'formulaic', thereby facilitating processing, as well as how early formulaicity starts to emerge in second language speakers. One view is that language learning, representation and processing are all reliant on exposure, which is the position taken by 'usage-based' approaches.

Usage-based approaches put a premium on the linguistic input that learners receive. More specific to lexical bundles, a usage-based approach would say that any sequence of words that learners encounter frequently will, all else being equal, experience a processing advantage. Thus, if junior high school students repeatedly encounter a phrase like *how many CDs do you have* in their English textbooks (see Chapter 3), this sequence will experience a processing advantage, despite the fact that junior high school students today are unlikely to play music using CDs. In a usage-based approach to first and second language acquisition, experience of and exposure to language results in high frequency, repetitive sequences of words being stored in long-term memory (e.g., Barlow & Kemmer, 2000; Bybee, 1998, 2006; Bybee, 2013; Ellis, 2002a, 2003; Tyler, 2010). Evidence for this from the field of formulaic language in general will be discussed in detail below.

In Chapter 3, we looked at the lexical bundles that Japanese secondary school students encounter in their English textbooks and compared the most frequent 50 lexical bundles at different bundle sizes. We would expect items that are frequent in

the textbooks to be frequent in native language, and items that are low frequency in the textbooks to be low frequency in native language also. We found that this was the case for 3-word lexical bundles, but larger sizes bore little to no resemblance to lexical bundles in the native corpus. Importantly, a usage-based view of language learning would predict that students will extract and store in memory the lexical bundles that they encounter frequently in their textbooks (as well as any frequently occurring formulaic language that they encounter in- and outside of the classroom) and will therefore demonstrate a processing advantage for textbook lexical bundles, regardless of whether they are authentic or not. Thus, if we take frequency of occurrence as a measure of experience, and students are indeed sensitive to the frequency of occurrence of lexical bundles in their textbooks, we would expect them to respond to lexical bundles derived from their textbooks faster and more accurately than novel sequences of words which do not appear in their books, irrespective of whether they represent 'authentic' native-speaker lexical bundles. Before turning to the current study, we will take a closer look at how non-idiomatic, corpus derived formulaic sequences such as lexical bundles, collocations and binomials are thought to be acquired and processed, a topic which has already been discussed in Chapter 2, but which is worth reitterating here in order to contextulise the current research.

Formulaic sequences in non-native speakers

In native speakers, a processing advantage for formulaic language is well attested (see Conklin & Schmitt, 2012; Wray, 2012). However, whether this extends to second language speakers (L2) remains a somewhat open question. In particular this is the case for non-idiomatic formulaic language such as lexical bundles. As will be taken up below, several studies have tried addressing this question, but they have not unequivocally answered it. Further, to our knowledge, no studies have attempted to ascertain whether or not very low proficiency teenage learners have a processing advantage for formulaic language.

Three self-paced reading experiments explored whether lexical bundles would be read faster compared to non-lexical bundles by native speakers (Tremblay, Derwing,

& Libben, 2009; Tremblay et al., 2011) and non-native speakers (Babaei et al., 2015). All of them conducted multiple versions of a self-paced reading task, testing word-by-word reading, chunk-by-chunk reading and full sentence reading. Tremblay et al. (2011) found advantages for sentences containing lexical bundles in all three tasks with native speakers. Tremblay et al. (2009), on the other hand, only found a lexical bundle advantage in the chunk and sentence reading tasks with native speakers, while Babaei et al. (2015) found no advantage for lexical bundles compared with non-lexical bundles in any tasks for moderately proficient non-native speakers in an EFL setting.

Using eye-tracking, Valsecchi et al. (2013) investigated the reading of lexical bundles by native and lower-proficiency non-native speakers. They found that while native speakers had a processing advantage for lexical bundles compared with non-lexical bundles, non-native speakers did not. The authors concluded that this null result was due to lack of proficiency and that a processing advantage would emerge later with higher proficiency. Importantly, in the only work on low-proficiency EFL students, Isobe (2011) conducted a phrasal judgment task where participants were instructed to judge whether the word order of a lexical bundle was correct or not. The author found lexical bundles were responded to significantly faster and with greater accuracy. Further, in a second experiment, involving a timed familiarity judgement task, the author found familiarity and phrasal frequency to be important in predicting response times for lexical bundles. Similarly, Jiang and Nekrasova (2007) conducted a phrasal judgement task and found that both native and non-native speakers judged corpus-derived units faster and more accurately than novel control sequences. Ellis et al. (2008) similarly found that non-native speakers were sensitive to the frequency of corpus-derived sequences. Both of these studies were conducted with high-proficiency students in an ESL setting at American universities.

Although the literature shows a somewhat mixed picture, the majority of evidence suggests that lexical bundles do proffer processing benefits to non-native speakers, particularly at the higher levels of proficiency. Importantly, if exposure to lexical bundles underpins the processing advantage for them, we need to have a good measure of the input learners receive. Generally, studies examining the processing of

lexical bundles identify them in native speaker corpora (e.g., Babaei et al., 2015; Isobe, 2011; Tremblay et al., 2009; Tremblay et al., 2011 all used items from Biber et al., 1999's study of the BNC corpus, and Jiang & Nekrasova 2007 used items from the Bank of English). While this may be a valid method of identifying lexical bundles for high proficiency speakers, especially those living in L1 environments, it is likely that lower proficiency speakers in EFL contexts where input is mainly limited to classroom study will have a different experience of the L2. Put another way, we may find a processing advantage for lexical bundles in lower proficiency speakers if we test them on lexical bundles they encounter in their primary source of input – their textbooks. One goal of this study is to test this assumption by matching experimental stimuli to the population under study using items derived directly from their materials.

How does a processing advantage for lexical bundles emerge?

In Chapter 2, I likened the optimisation of language processing to a computer's internet browsing cache – by storing information pertaining to frequently accessed webpages, we browse the internet faster and more efficiently without having to load large data files such as images every time they need to be accessed. In the same way, formulaic language is said to reduce processing load by 'caching' language into chunks that can be accessed quickly and easily without needing to retrieve and compute their individual components (Goldberg, 2006; Wray, 2002, 2008; Wray & Perkins, 2000). In a usage-based approach, when language users encounter utterances, they store them and look for patterns amongst them (Tomasello, 2003). More specifically, if we encounter the examples of *in front of the* in 1., 2. and 3. below, taken from the SUBTLEXus (Brysbaert & New, 2009), each instance will be stored and the linguistic pattern will be detected on subsequent encounters. Importantly, the first time a sequence of words is encountered, it will be processed in a word-by-word manner, however having done so it leaves an imprint in memory that is strengthened by subsequent exposure (e.g., Logan, 1988). Thus, when reading 1. there would be no difference between *in front of the* and any other sequence of words in the same sentence when reading it for the first time. It is a novel, original string of words.

However, when reading 2. and subsequently 3. there should be a ‘boost’ in processing speed because we already have an imprint of this ‘cached’ in memory.

1. Tomorrow you're going in front of the Commissioner.
2. I didn't want to bring this up in front of the kids...
3. Stand up in front of the windows.

This is important because our online processing resources are quite limited, while our memory capacity is vast (Crick, 1979). A seminal finding by Miller (1956) demonstrated that our working memory can hold seven plus or minus two items, which is often referred to as the ‘magical number seven’ or *Miller’s Law*. Subsequent research has disputed how much information can be held in working memory (e.g., Cowan, 2001; Mathy & Feldman, 2012) and while there is disagreement on the precise amount of information that can be held in working memory, there is complete agreement that there are severe constraints on it. This has serious implications when we consider a task like reading sentences. To comprehend a sentence, readers need to hold the individual words in working memory and integrate the meaning of the sentence with their developing understanding of the text. If the median sentence length in a text is more than 15 words, for instance (cf. Kornai, 2008 discussing journalistic prose), readers will very quickly exceed the capacity of their working memory. In contrast to working memory, long-term memory is a relatively abundant resource. Thus, to compensate for a lack of working memory, the brain may make use of long-term memory by storing frequently used strings of words.

As we have seen, usage-based models of language learning predict that repeated occurrences of linguistic patterns, like lexical bundles, lead to their entrenchment in memory. A detailed description of the differences between the various usage-based approaches has already been discussed in Chapter 2, however as Higham (1997) points out, a common factor among all of them is that chunking performance can be predicted by a metric derived from the frequency at which features occur within a given stimulus. Recall from Chapter 2 that some models (e.g., Goldberg, 2006; Wray,

2002) predict that only items that are frequent enough are stored in memory, while others (e.g., Bybee, 1998, 2006; Bybee & Hopper, 2001; Mclaughlin et al., 1983) predict that every exposure to a phrase influences subsequent entrenchment and processing (see Arnon & Snider, 2010). The frequency with which units of language occur, then, is a driving force behind chunking processes and, all else being equal,[3] each exposure to a sequence of words or sounds will affect its subsequent processing (Bod et al., 2003; Bybee, 1998, 2006; Bybee & Hopper, 2001; Mclaughlin et al., 1983). This has important implications for language learning material design—something which, in the case of EFL contexts, often constitutes students' primary experience of the language—and has an important consequence: *if we expose students to language that is not sufficiently representative of the target language, it is this non-nativelike language that will become entrenched in memory.* Thus, we need to ask whether the adage of 'garbage in, garbage out' is as relevant to our classrooms as it is to computer programming.

In the case of Japanese junior high school students, the teaching materials used in class, in other words what students are exposed to, will (primarily) dictate what they gain a processing advantage for. A good example that illustrates the problem with the language of the textbooks is *do you hear* and *do you play*. In the SUBTLEXus corpus *do you hear* occurs 1,176 times (23 times per million) and in the JHSETC it does not appear at all. In contrast, *do you play* occurs 136 times (2.7 times per million) in the SUBTLEXus and 56 times (395 times per million) in the JHSETC. Thus, in native speaker English *do you hear* is much more frequent and should elicit a processing

[3] It should be noted that while frequency of exposure is a key metric in usage-based approaches, it is not the only property of the input that contributes to learning. While each exposure to a stimuli can be expected to affect subsequent processing, this may occur to a greater or lesser degree depending on other factors such as recency, Zipfian distribution and saliency. See for example Ellis (2006, 2009, 2012) for detailed discussions on this, as well as Ebbinghaus (1885) and more recently Pavlik and Anderson (2008) for a more general discussions of recency effects in memory and learning. Although we do not wish to overly simplify what we know about usage-based language development, when textbooks are the 'source' of language exposure, it would be very difficult, if not impossible, to disentangle these factors. Therefore, in the current study we take properties such as recency, distribution and, although we do also attempt to control for it by only including non-taught lexical bundles, saliency, to be captured by frequency of occurrence and familiarity.

advantage relative to *do you play*, at least for speakers who encounter nativelike input. For Japanese students learning English, the other pattern is more frequent and should therefore experience the processing advantage. According to a usage-based model, this should lead to a non-nativelike processing advantage in Japanese junior high school students.

Because the input has serious implications for processing, exploring the input that students receive is an important endeavour. Study 1, presented in the previous chapter, analysed the language contained in English textbooks for Japanese lower secondary school students. As discussed, these textbooks were intended to be 'communicative', and as such should contain examples of authentic language – a hallmark of the communicative approach. Recall that the 16 textbooks for junior high school English approved by the Japanese Ministry of Education (collectively labelled the 'Junior High School English Textbook Corpus', or the 'JHSETC') consisted of 152,966 words, and contained a total of 25,539 lexical bundle tokens compared with 11,379 tokens in a matched reference corpus. This indicates that the language contained in the books is indeed highly repetitive, and should provide optimal conditions for the acquisition of 'chunked' language. However, the lexical bundles in the textbooks, although abundant, were found to be qualitatively different to and follow very different distributional patterns to the reference corpus – in other words, to authentic native-speaker language. As with similar studies (e.g., Nguyen & Ishitobi, 2012; Ogura, 2008), this research concluded that the language contained in textbooks does not serve the purpose of exposing students to authentic language. In the current study, we will explore whether the frequent lexical bundles in the textbook corpus (e.g., *do you play*) elicit a processing advantage relative to possible sequences of words that do not appear in the textbooks (e.g., *do you hear*).

Methodology

Experimental stimuli

Twenty 3-word lexical bundles covering a range of frequencies were selected from the 'New Horizon' (New Horizon English Course 2013) sub-corpus of the JHSETC, consisting of three textbooks, which we refer to as the New Horizon Corpus (NHC). The whole corpus draws on all of the Ministry approved textbooks for secondary schools in Japan. However, a school will only use one textbook series. Therefore it would be inaccurate to base stimuli on the whole corpus. As the students in the current study used the 'New Horizon' series, the lexical bundles and associated frequency information comes from these books. The language presented in the textbooks is, for the most part, presented as oral language in the form of dialogues and monologues. Thus, as with Study 1 (Chapter 2), the SUBTLEXus was used as a reference corpus, which, as mentioned previously, is a corpus of subtitles taken from US films and television series, and as such usually involves people in social interactions (Brysbaert & New, 2009). Further, while the language of the SUBTLEXus is scripted, it is meant to represent natural language in the same way that the language in textbooks is intended to. The US version of the SUBTLEX was chosen because the language used in the Japanese textbooks is based on standard American English.

In order to retrieve the 3-word bundles, Antconc's (Anthony, 2014) 'Clusters' function was used with a cut off frequency of four times in the NHC. Items were then selected from this list based on several criteria. First, items were chosen from a range of frequencies of occurrence. This was done to ensure the results were representative of the fact that lexical bundles were more or less frequent in the textbooks. Second, because we were interested in the role *teaching materials* play, and not the effects of teaching, we selected lexical bundles that were not explicitly taught in the textbooks. This was ascertained by consulting the word lists in the books, and rejecting any lexical bundles which appeared. The students' (participants') teachers were also asked about how they approached teaching from the textbook, and none of them reported teaching any phrases that were not listed as items for explicit instruction. A third, and

important criterion, was that it had to be possible to create a control sequence by changing only one word in each lexical bundle, using only words that appeared in the textbooks to create a 3-word lexical bundle of zero frequency in the NHC. That is, each individual word in both the lexical bundle and control item appeared in the NHC, but only the original lexical bundle appeared as a string of words. This was done to ensure any differences in reading times found between lexical bundles and control phrases are due to whole string processing and not simply because a student has not been exposed to a given word. For example, *do you play* was selected as a lexical bundle, and had 16 occurrences in the NHC and 136 in the SUBTLEXus. We changed the word *play* to *hear* to create the control item *do you hear*, which had zero occurrences in the NHC but 1,183 in the SUBTLEXus. Crucially, the word *hear* occurred in the NHC. All controls were matched as closely as possible for individual word frequency with their respective bundles. This was calculated as the mean frequency of all three words in the lexical bundle and was confirmed with a *t*-test. There was no significant difference between the word frequency of main items and control items ($t(41.98) = 0.53$, $p = 0.59$), the main items and ungrammatical items ($t(41.98) = -0.52$, $p = 0.60$) or the controls and ungrammatical items ($t(42) = -1.04$, $p = 0.31$). Any differences in response time during the experiment, then, can be assumed to be a result of whole phrase frequency, not of the individual words contained in each lexical bundle.

We selected 20 items that fit these three criteria, and created 20 matched control phrases for them. Because the main task involved a 'yes'/'no' decision about whether a phrase was possible, we also created another matched set of items by scrambling the words to produce a set of ungrammatical phrases (e.g., *you with do)* which had zero frequency in both the textbook corpus and the reference corpus. This resulted in a total of 60 experimental items. Finally, 190 filler phrases were added to mask any similarities between the NHC derived lexical bundles and control phrases. We selected 3-word lexical bundles from the NHC, again from a range of frequency bands, and scrambled the word order of half of them to make 95 grammatical fillers (e.g., *play the piano* – four occurrences in the NHC), and 95 ungrammatical fillers

(e.g., *he yes does* – zero occurrences in the NHC). The ungrammatical items were checked in both the NHC and the SUBTLEXus to ensure that they had a frequency score of zero. In total, the experiment consisted of 250 items.

Participants

Participants for this study consisted of 35 grade nine students from a private secondary school in Japan. A summary of the participants is shown in Table 4. All students spoke Japanese as their first language. No students reported reading, hearing or speaking impairments. Students' English proficiency was assessed by asking them to rate themselves on a seven-point scale across four categories – speaking ability, reading ability, listening ability and ability to use grammar accurately. We took the average of these four ratings to create a 'proficiency' metric. There were some differences in the age at which students were first exposed to English. Eleven students (31%) reported that they currently studied English outside of school at 'juku', or 'cram school'. We conducted a *t*-test to assess if there was a difference in proficiency between the students who did and did not study English outside of school – those who did scored significantly higher for proficiency $t(797.78) = 7.45$, $p < .001$ and therefore this variable should largely be captured by proficiency. We also asked students to report on use of English outside the classroom. Twenty-four students (69%) reported nothing, seven students reported watching films in English and two students had been abroad. This was also significantly related to proficiency $t(3070.9) = 22.81$, $p < .001$ and therefore should fundamentally be captured by this variable in the analysis.

Table 4. Summary of non-native speakers' age, years of studying English, and self-rating of English proficiency for participants in Study 2.

	Age	Age of first contact	Years studying English	Speaking	Listening	Reading	Grammar	Overall proficiency
Mean	14.82	6.91	7.91	3.23	3.66	3.29	3.29	3.36
SD	0.38	3.49	3.50	1.06	1.16	0.99	0.96	0.95

Note: Speaking, Listening, Reading and Grammar are self-rated on a seven-point scale (1 = very weak, 7 = very strong). Overall proficiency is calculated as the mean of these four categories.

Procedure

Students were tested at their school. They completed a phrasal judgement task first, followed by a language background survey which included questions related to the students' use of English outside of school, as well as a familiarity rating task for all of the lexical bundles. Ungrammatical items were assumed to have a familiarity of zero. All instructions were given in Japanese, and the experiment was run using Inquisit (2014). Two hundred and 53-word sequences were displayed for a simple 'yes' or 'no' judgement task. Students were instructed to press a 'yes' key if the phrase was possible in English, and a 'no' key if it was not. For each trial a '*' appeared in the centre of the screen for 700 milliseconds, followed by the target lexical bundle. Students were instructed to respond as quickly as possible. They were given 20 practise items with feedback telling them whether they were correct or not.

Analysis and Results

The data was analysed using mixed effects modelling, with the package lme4 (Bates, Maechler, Bolker, & Walker, 2015). Response time was modelled as the response variable, with textbook frequency, mean word frequency, familiarity, proficiency, phrase length (the number of characters in a given lexical bundle), trial number and age of first contact as fixed effects. Participants and items (the lexical bundles) were included as random-effect factors in the model. Reaction times under 400 milliseconds or over 1999 milliseconds were removed, which accounted for less than 0.7% of the data. Inaccurate responses were removed for the response time analysis (described below). Because there were a large number of errors (23% of the data), we also modelled the data using accuracy as the response variable and used the same fixed and random effects. Table 5 summarises mean response times and accuracy scores.

Some predictors were correlated and so were residualised by fitting linear models and creating a new metric consisting of the residual (i.e. information not accounted for by correlating factors, Baayen, 2008). For example, textbook frequency

Table 5. Mean response times in milliseconds and standard deviations in brackets in each of the three experimental conditions for Study 2.

	Lexical Bundles	Non-LBs	Ungrammatical
Response time	1072.61 (325.59)	1186.65 (362.47)	1285.59 (368.71)
Accuracy	89% (382.54)	67% (153.44)	73% (208.59)

was correlated with phrase length and familiarity, and so textbook frequency was used as the response variable in a model with phrase length and familiarity as predictors. Similar models were made for any correlated factors, including previously residualised response variables. The resulting residuals all correlated significantly with their related variables ($p < .0001$): textbook frequency ($r = .89$) mean word frequency ($r = .80$), proficiency ($r = .92$) and familiarity ($r = .99$). All continuous predictors were centred to avoid having a change in slope that might correlate with a change in intercept (see Baayen, 2008). The accuracy model also included response time as a predictor, which we will label 'response speed' to avoid confusion with the same metric as used in the response time model. It was not correlated with any other predictors.

Response time was modelled using a Liner Mixed Effects model. Because accuracy is a binary variable, accuracy was modelled using a Generalised Linear Mixed Effects model. For both response time and accuracy, a simple model was fit by residual phrase frequency, and with subject and lexical bundle as random effects. Complexity was then built into this model by adding new predictors in a step-by-step manner. Each new predictor was tested as a discrete factor, and for interactions with all other predictors. These were also tested as adjustments for both the subject and lexical bundle random effects where theoretically relevant (as recommended by Barr, Levy, Scheepers, & Tily, 2013). Each iteration of the model was tested against the previous best fit to see if the change was significant (likelihood ratio test, $p \leq .05$). This process was continued until all predictors and interactions had been tested both as fixed and random effects. Table 6 summarises the best fit for both of the models.

Table 6. Summary of fixed and random effects for the best-fit model for Study 2.

	Response time			Accuracy		
	Estimates	CI	p	OR	CI	p
Fixed Parts						
(intercept)	1193.07	1131.41 – 1254.73	<.001	10.72	4.10 – 28.05	<.001
Residual Textbook Frequency	-6.34	-11.49 – -1.19	.018	1.09	1.01 – 1.18	.025
Residual Proficiency	-72.25	-119.60 – -24.89	.005	1.34	1.11 – 1.63	.003
Residual Familiarity				1.25	1.10 – 1.41	<.001
Response Speed				1.00	1.00 – 1.00	.001
Phrase Type [Main Item]	-97.59	-161.73 – -33.45	.004			
Phrase Type [Ungrammatical Control]	121.37	62.27 – 180.46	<.001			
Phrase Length	24.45	11.77 – 37.13	<.001			
Trial Number	-0.36	-0.54 – -0.17	<.001			
Phrase Type [Ungrammatical Control] * Response Speed				1.00	1.00 – 1.00	.005
Random Parts						
σ2	68478.713					
τ00, Lexical Bundle	4973.613			0.705		
τ00, Participant	17412.132			0.575		
ρ01						
N Lexical Bundle	60			60		
N Participant	35			35		
ICC Lexical Bundle	0.055			0.154		
ICC Participant	0.192			0.126		
Observations	1514			1975		
R2 / Ω02	.380 / .377					
Deviance				1549.987		

Response time. The best fit for the model of response times included residual textbook frequency, residual proficiency, phrase type, phrase length and trial number. Predictors and interactions were tested in the random effects structure where appropriate. The model fit was significantly improved by adjusting subject by residual textbook frequency (X^2 (2) = 10.0, p = .007). As predicted, the most prominent effect is phrase type. A post-hoc analysis of the differences of LSMEANS shows that the difference between all phrase types is also significant (lexical bundle vs. control phrase, t(61.4) = 2.98, p = 0.004; lexical bundle vs ungrammatical phrase, t(55.4) = -7.49, p = <.0001; control phrases vs. ungrammatical, t(61.9) = -4.03, p = <.0001).

Residual proficiency and phrase length were significant factors in the model, such that more proficient participants responded more quickly than less proficient ones and shorter items were responded to more quickly than longer ones. Trial number was also a significant factor meaning that participants got faster over the course of the experiment which is a normal phenomenon (Baayen, Davidson, & Bates, 2008). Importantly, there was also a significant effect of residual textbook frequency (X^2 (1) = 5.65, p = .018), which indicates that more frequent sequences were responded to more quickly. There was no significant effect of age of first contact (X^2 (1) = 0.06, p = .80) or residual familiarity (X^2 (1) = 1.92, p = .17). There was also no significant effect of mean word frequency (X^2 (1) = 0.13, p = .72), indicating that students are making use of phrasal frequency but not the frequency of individual words. This is not surprising, as items were controlled for this.

Accuracy. The best fit for the Accuracy response variable included residual textbook frequency, residual proficiency, residual familiarity and response speed. There was also a significant interaction between phrase type in the ungrammatical control condition and response speed. For lexical bundles and control phrases, accuracy decreases as speed of response becomes slower. This is interesting in itself, as we would expect a speed-accuracy trade-off whereby responding faster leads to making more mistakes. For ungrammatical items, however, we see a slight increase in accuracy as response time increases. Although there is no interaction between lexical bundle frequency and response time (X^2 (1) = 0.28, p = .59) or familiarity (X^2 (1) =

3.00, p = .08), given the fact that frequency and familiarity also predict accuracy in the model it seems likely that the lexical bundles responded to faster are the higher frequency and more familiar items. This makes sense as we would not expect any facilitation for frequency or familiarity with ungrammatical items. It is possible that for ungrammatical items, ones that are responded to quickly are more prone to mistakes. This would also account for the slower response time for these items. All other predictors were insignificant – phrase length (X^2 (1) = 0.11, p = .74), residual mean word frequency (X^2 (1) = 0.03, p = .86), trial number (X^2 (1) = 1.31, p = .25), age of first contact (X^2 (1) = 1.66, p = .19) or phrase type (X^2 (2) = 4.86, p = .08), although the later does approach significance. The model was significantly improved by adjusting subject by phrase type (X^2 (5) = 70.93, p = < .0001), suggesting that students did not respond to the different phrase types in the same way.

Discussion

This study shows, for the first time, that even very low proficiency, beginner secondary students are sensitive to recurring lexical bundles from their input – the textbooks. Importantly, they demonstrate an effect of frequency such that lexical bundles that appear more frequently in the books are responded to more quickly and accurately than less frequent ones. This indicates that beginners are sensitive to the frequency of lexical patterns in their input, which highlights the importance of presenting learners with authentic formulaic language in their textbooks.

As predicted, there was a significant difference in response time between lexical bundles and matched control phrases. The controls were possible using only words from the textbooks but did not appear as a whole string in them. Because the students had been exposed to all the words in the lexical bundles and control items to the same degree, if they were processing lexical bundles in a word-by-word fashion we would not expect any difference between the two conditions. Further, including individual word frequency did not improve the fit of the model for either response time or accuracy. This finding is in line with similar studies (Ellis et al., 2008; Isobe, 2011;

Jiang & Nekrasova, 2007) showing that L2 learners responded to lexical bundles faster in a phrasal judgement task, and supports the hypothesis that *even* low proficiency non-native learners are sensitive to the exact sequences of words that they are exposed to – as predicted by a usage-based approach.

Interestingly, phrase type does not significantly predict accuracy. However, we do see a significant interaction between item type and response speed whereby the relationship between speed of response is different for ungrammatical control phrases and that of lexical bundles and grammatical control phrases. In terms of the response time analysis, the students are faster at responding to sequences from their textbooks than what to them are novel sequences, and finally ungrammatical items, in that order. For accuracy, however, correct responses are increased for lexical bundles and control phrases that are responded to quickly, and decreases as speed of response decreases. For ungrammatical items, we see the opposite: accuracy increases as response speed decreases. One explanation may be found in the concepts of 'nativelike fluency' and 'nativelike selection' as put forward by Pawley and Syder (1983), where the difference between a likely utterance and a possible utterance is one of formulaicity (see Chapter 2). In this context, fluency can be thought of as an unconscious process, whereas nativelike selection, or more simply whether a phrase sounds natural or not to the test taker, can be thought of as a conscious, explicit decision-making process that involves collocational knowledge. It may be that students are sensitive to these aspects unequally, and if a given item is not immediately recognised (nativelike fluency) a participants' first reaction may be to reject it – however where conscious processing takes over, participants then judge the item based on form and collocational knowledge, or to put it another way, whether the sequence sounds likely to them (i.e., nativelike selection). Where conscious processing does not take over in time, we are likely to see many errors, which would account for the relationship between speed of response and accuracy. This may also go some way towards explaining the fact that familiarity is only significant in the model for accuracy – it may be that familiarity, at least in this context, is related to conscious decision making processes, such as judging

the likelihood of a phrase being grammatical or not based on collocational knowledge, or put simply, whether it sounds natural or not.

Students were sensitive to the frequency of lexical bundles, which was reflected in both response time and accuracy. Put simply, more frequent items were responded to faster and more accurately. This is important because it indicates that not only are students sensitive to whether or not items appeared in their books, but to the degree that they are exposed to a given item. In the current study, experimental items were selected across the frequency continuum and results suggest a gradual decrease in processing speed in accordance with the increased frequency of a given item. This finding is noteworthy because it provides strong evidence that the input given to students matters. Studies have shown explicit instruction is effective for formulaic sequence learning (e.g., Laufer & Girsai, 2008; Webb & Kagimoto, 2009) however the items studied here were neither explicitly taught, nor necessarily 'highly' frequent. Webb and Kagimoto (2009) as well as others, for example, Szudarski and Carter (2016) and Pellicer-Sánchez (2017), found evidence for collocational knowledge in language learners only at the level of recognition and recall. Further, in the case of the studies by Webb and Kagimoto (2009) and Szudarski and Carter (2016), learning gains resulted from focused, explicit instruction. Pellicer-Sánchez (2017) found evidence of incidental learning similar to our findings, however, no significant differences were found between four and eight exposures, whereas we found clear differences across the frequency continuum. This topic will be taken up in detail in Chapter 6, however one explanation may be that exposure in the current study is over a three year period, whereas in Pellicer-Sánchez's study all repetitions occurred in a short space of time (during the experiment); we know from memory research that memory effects become more robust over time (Ebbinghaus, 1885). Here, in the current study, we see evidence of nativelike processing for lexical bundles learned incidentally on a continuous scale – lower frequency items are responded to slower and with more errors, while higher frequency items are responded to increasingly faster and with fewer mistakes. This means that the actual input—even if it is only intended to provide supporting context in demonstrating a particular word, item or

grammatical construction selected for instruction—is attributing to a student's fluency in the language.

Importantly, proficiency is also a significant predictor for both response time and accuracy, yet appears to be unrelated to frequency. Siyanova-Chanturia, Conklin, and van Heuven (2011) found that sensitivity to frequency of formulaic language, in this case for binomials, only emerged at higher proficiency levels. Such a finding could indicate that a processing advantage only emerges after a learner encounters formulaic language a certain number of times – an argument also put forward by Babaei et al. (2015) and Valsecchi et al. (2013) to explain the null-result in their studies. However, it may be that these studies would have found a processing advantage with a different set of experimental stimuli – i.e., items matched to the input those participants had encountered. The current study addresses this concern by using items taken directly from the students' textbooks, and therefore we know they have encountered them. By taking frequency information from the textbooks, we should have a fairly accurate estimate as to how often the students have encountered the lexical bundles in the study. A problem with Siyanova-Chanturia, Conklin, and van Heuven (2011), Babaei et al. (2015), Valsecchi et al. (2013), and indeed all studies to date with non-native speakers, is that we cannot be sure of what degree participants had encountered all of the items tested, if at all. For high-level non-natives who have spent considerable time in an ESL environment, native speaker norms may be justified – but can we say the same for lower-level non-native speakers? Evidence that frequency information is related to an individual learner's exposure to the language may come from the fact that adjusting the subject random effect by lexical bundle frequency significantly improves the fit of the response time model. This provides an indication that each individual's relative exposure underlies the frequency advantage that he/she demonstrates.

This study also highlights a further important consideration, this time not in terms of methodology, but in terms how English education is being approached in Japan. There is a commonly held belief that students do not spend enough time learning English, which underpins their eventual attainment that is often considered deficient because it is not sufficiently 'nativelike'. This has led to a steady increase in time

allocated to English first by including English in the primary school curriculum and then by increasing the number of hours in English at secondary school from 105 hours to 140 hours (announced by Ministry of Education 2008a; 2008b and implemented from 2011). However, increasing exposure is only one part of the equation. The input that students receive must also be considered. Importantly, the current study demonstrates that beginning learners in an ELF context show a very nativelike formulaic benefit to lexical bundles. Thus, they are showing the onset of a phenomenon that is evident in native speakers and advanced non-natives. Then the key question is what are they showing a formulaic advantage to – is this actually authentic language? If they are simply showing an advantage to genre-specific textbook language that is not reflective of actual native speaker English, this is hugely problematic and points to an important role for textbook developers. This is especially true since, as was demonstrated in Chapter 3, we do not necessarily need to make large changes to materials to make them more authentic.

Conclusion

This study is important because it demonstrates a clear processing advantage for the lexical bundles that Japanese junior high school students encounter in their textbooks. The findings show that students are sensitive to the phrasal frequency of items embedded in their textbooks, and respond to them faster and more accurately. Our results are indicative of a usage-based approach to language learning, in which exposure to language underpins entrenchment in memory and processing speed, *and* is evident in early EFL learners, a population that is underrepresented in the literature. Unfortunately, it also demonstrates the computer adage – *what we put in is what we get out*. The junior high school students receive input—the lexical bundles in their textbooks—and they show a clear processing advantage for them; it is simply that these lexical bundles are not representative of native speaker formulaic language. Crucially, these findings highlight the need to carefully consider the input students are given.

Chapter 5. "That Sounds About Right"

The current study explores the perceived naturalness of lexical bundles. Naturalness ratings of native English speakers and non-native speakers, once again with Japanese as their first language, were compared in relation to corpus derived frequencies from the BNC and the JHSETC (Chapter 3). Both groups' rating scores were significantly predicted by lexical bundle frequency. However, the groups were sensitive to different metrics; the native speakers' performance was best predicted by the BNC, and the Japanese speakers' by the JHSETC. This is taken as evidence that learner intuitions are affected by the input they receive from teaching materials, and that these intuitions may stay with learners long term. Chapter 5 has been submitted as '"That Sounds About Right" - An Examination of Lexical Bundle Naturalness Intuitions in Native and Non-native English Speakers' and is currently under review (Northbrook, Conklin, under review).

Introduction

Anyone who has learned a second language has likely had the experience of being told something they said or wrote is, *'Okay, but not how a native speaker would say it.'* In Hymes (1972) taxonomy of situational competence this relates to the forth parameter, namely that language should be *performed.* In other words, language needs to be 'done' in the real world, not simply be possible, well formed (feasible), or appropriate (the first three parameters). Arguably, this is one of the greatest challenges language learners face: not distinguishing between grammatical and non-grammatical utterances, but distinguishing between *possible* utterances and *likely utterances.* Early work in the area of formulaic language highlighted this fact and proposed a new way at looking at language – not as simply a set of syntactic rules and words, as generative grammar approaches had proposed, but as a memory-based system using lexical elements (Bolinger, 1975). Sinclair (1991) referred to this as a system that considers both an open choice principle, and the *idiom* principle. The 'open choice' principle refers to speakers' ability to construct unique and creative utterances using a system of rules and connected items. However, Sinclair argued that this is not the way in which we primarily produce language. Rather, we default to the *idiom principle* – that is,

where possible we use pre-constructed formulaic—or, 'performed'—language that is entrenched in long-term memory (see for example Conklin & Schmitt, 2012).

Following this, Pawley and Syder (1983) coined the terms 'nativelike fluency' and 'nativelike selection', arguing that simply exercising the creative potential of syntactic rules is not sufficient to sound 'nativelike' when speaking a language. Recall from the discussion in Chapter 2 that although native speakers could say *'Pleasant morning.'* or, *'Could you aid me in this task?'* they tend not to do so. This is not because these phrases are ungrammatical, but rather because there are more common, socially institutionalised phrases available (e.g., *'Good morning.'* and *'Could you help me with this?'*). Most people would likely agree that these more common alternatives sound more natural, are understood more readily and are preferred by native speakers. At the same time, there are utterances such as, *'Thanks very much.'* that are highly frequent expressions and commonly used by native speakers, but which are in fact ungrammatical. We must conclude, then, that communicative competence in a language is not simply the ability to produce grammatically correct utterances. It also encompasses the ability to select frequent and conventional patterns that are *likely* in the language (see Biber et al., 1999; Cowie, 1998; Ellis, 2001; Howarth, 1998; Skehan, 1998 among others for further discussion on this), and therefore whether learners are able to do this or not is an important issue in EFL and second language acquisition. Before we consider studies that have looked at learner intuitions of formulaic sequences, it is first necessary to reiterate some of the points made above in previous chapters and define the concept of formulaicity in the context of the current study.

Formulaic sequences in second language learning

As discussed in Chapter 2, there are many different types of formulaic language: idioms: *kick the bucket*, collocations: *plastic surgery*, binomials: *bride and groom*, phrasal verbs: *rely on*, lexical bundles: *I don't want to*, to name but a few. The current study, as with the other studies presented in this thesis, focuses on lexical bundles, which are non-idiomatic sequences of words that are defined purely based on their frequency of occurrence (Biber, 2009; Biber et al., 1999; Stubbs & Barth, 2003) and

can be thought of as extended collocations (Biber et al., 1999). At the early stages of second language acquisition, learners are thought to use memorised formulaic language, which allows them to express themselves above and beyond their current abilities (Myles et al., 1998; Myles et al., 1999; Nattinger & DeCarrucio, 1992; Wray, 2002). It is not necessary, for example, to understand that the phrase '*That sounds great*' is made up of a pronoun acting as a noun phrase, which is modified by a verb phrase consisting of a third person singular verb in the present tense combined with an adjective – or indeed, to worry that the aforementioned, *'Thanks very much'* is, technically speaking, ungrammatical. Learners can simply use them as is. As proficiency advances, however, learners tend to construct much of their speech using rules rather than lexicalised routines (Foster, 2001; Nesselhauf, 2003; Skehan, 1998), perhaps believing that high-proficiency language requires originality. This is mistaken, however, because studies show that the use of formulaic language, far from indicating that a learner is a 'beginner', is in fact a mark of nativelikeness, and therefore proficiency (e.g., Boers et al., 2006; Dickinson, 2012).

Dickinson (2012) trained students on specific formulaic sequences and found that those who practised and actively used them gave presentations that were more engaging and easier to understand than those who did not. In a similar study, Taguchi (2007) found that instructing and drilling formulaic sequences with learners of Japanese as a second language led to a higher sensitivity to discourse features and was useful for developing learners' oral production ability. Wray and Fitzpatrick (2008) similarly concluded that by acquiring formulaic sequences that are common in conversations, learners not only increase their fluency and naturalness in the language, but also increase their confidence in using the language. In another study, Boers et al. (2006) trained two groups of students for 22 hours each – one using a traditional words-and-rules approach; the other was trained to use formulaic sequences. Subsequently, two judges rated the group that had formulaic training as having higher proficiency.

Despite these findings, very little has been said about how well non-native speakers are able to judge the 'naturalness' of sequences they encounter. Does instruction help learners to differentiate between utterances such as *can you aid me with this task?* and *can you help me with this?*, which in turn would allow them to

better assess the formualicity of their own speech and 'select' nativelike expressions? In one of the few studies looking specifically at learners' ability to judge formulaic sequences, Edmonds (2013) administered an online naturalness judgement task to French natives, and two groups of learners of French living in France (half long stay and half short stay). All groups judged conventional expressions as being more natural and did so faster than control phrases. Importantly, this included the short-stay French learners which, as Edmonds concludes, indicates that even learners who have little relative exposure to the target language are "well on their way to nativelike selection" (p. 95). In two studies that looked at adjective-noun collocation intuitions, Siyanova and Schmitt (2008) and Siyanova-Chanturia and Spina (2015) asked native and non-native speakers to judge the frequency of collocations, which we take here to be related to their naturalness. Similar to Edmonds (2013), Siyanova-Chanturia and Spina (2015) found that both native and non-native speaker intuitions correlated strongly with corpus frequency. This was, however, only for high-frequency items. Neither group were able to judge medium or low frequency pairs, although the native speakers were able to accurately judge very low frequency items. Siyanova and Schmitt (2008) had somewhat similar findings. They showed that native speaker intuitions correlated strongly with BNC frequency scores. However, the non-native group displayed poor collocational intuitions in comparison to the native speakers.

Overall it seems that while learners may be able to accurately judge formulaic sequences, their intuitions may not be entirely accurate or in tune with authentic language use. An explanation for this may be found in usage-based approaches to second language acquisition which, as already discussed in previous chapters, put a premium on the linguistic input that learners receive. In such an approach, experience of and exposure to language results in high frequency, repetitive sequences of words being stored in long-term memory (Barlow & Kemmer, 2000; Bybee, 1998, 2006; Bybee, 2013; Ellis, 2002a, 2002b, 2003), which would influence learners' intuitions regarding formulaic sequences. Thus, the question arises: is it that the second language learners cited above are not sensitive to collocation, or is it that their intuitions are simply *different*? That is, do they treat formulaic language as compositional and not as chunks, or is it simply that they have, through exposure to 'un-nativelike' English (i.e., 'textbook English') developed un-nativelike intuitions?

In order to answer this question, it is necessary to consider the input English learners receive. In the case of English as a Foreign Language (EFL) students, this is, as previously stated, primarily classroom instruction and materials (Meunier, 2012; Nunan, 1991). The studies discussed in Chapter 3 focusing on secondary school textbooks have found the dialogues in them to be inauthentic (Nguyen & Ishitobi, 2012; Römer, 2004; 2005, see also Gilmore, 2004) in that they do not accurately represent the situations students need to be prepared for. Looking at formulaic language more specifically, Study 1 (Chapter 3) examined lexical bundles in English textbooks for Japanese secondary school students and found them to occur frequently, though they were qualitatively different to those found in naturally produced English. Study 2, presented in Chapter 4, found that junior high school students had a processing advantage on a phrasal judgment task for lexical bundles from their textbooks but not for matched controls. It appears that the learners were exposed to un-nativelike formulaic language and had an advantage for it, irrespective of native-speaker frequency distributions. Crucially, the lexical bundles tested in Study 2 (Chapter 3) were, as discussed, selected *because* they occurred in the textbook series that the participants had used over the course of their three years of junior high school and the students were tested while in their final year of junior high school. Subsequent exposure to English might help students overcome their initial exposure and become more nativelike. The current research investigates whether adult Japanese learners' of English intuitions are based on the frequency distributions of the classroom English from their junior high school textbooks or whether they have become more nativelike. To address this, we asked adult native speakers of English and adult non-native, Japanese learners of English to rate the naturalness of sentences containing lexical bundles and examined whether their ratings were related to a native speaker corpus (the British National Corpus, BNC) or the Japanese junior high school English textbook corpus (JHSETC).

Methodology

Participants

Participants for this study consisted of 24 native speakers from the University of

Nottingham and 23 native Japanese speakers. The native speakers participated for course credit. The Japanese speakers were all studying English as a second language, were recruited from an online e-learning platform targeted at intermediate learners, and were a sample of convenience. The Japanese participants were asked to provide background information about themselves, which is summarised in Table 7. In addition to what can be seen in the table, we asked the participants to indicate their profession: five were university students, and the remaining were in full-time employment (ten working in an office, three as engineers, one as a lawyer, one as a medical officer and one as a waitress). All of the Japanese participants reported that they lived in Japan, had never lived in an English-speaking country and had attended a typical Japanese-medium school. Participant-based variables were included in the analysis.

Table 7. Summary of non-native speakers' age and self-rating of English proficiency for participants in Study 3.

	Age	Age of first contact	Speaking	Listening	Reading	Overall proficiency
Mean	31.15	12.55	3.23	3.66	3.29	3.36
SD	6.38	0.91	1.06	1.16	0.99	0.95

Note: Speaking, Listening and, Reading are self-rated on a seven-point scale (1 = very weak, 7 = very strong). Overall proficiency is calculated as the mean of these three categories.

Materials

We created 17 passages consisting of a line of text to provide context, followed by a short dialogue. One portion of each dialogue contained either a high-frequency lexical bundle or a low-frequency control phrase matched for meaning (see below for lexical bundle and control selection criteria). There were two versions of each passage (a total of 34 possible items). Participants saw either version A or B, but not both, so each participant judged 17 items. Lexical bundles are ideal for looking at questions of naturalness, because unlike idioms, they are not easily identifiable when embedded into a sentence. For example, items such as, *'James' aunt kicked the bucket / died last week'*, might make it clear that the focus is on idioms and potentially influence

participants' responses. In contrast items like, '*Really, I don't know why/ I'm not sure why I agreed to this'*, should make the focus on formulaic language less apparent. Each item included a context line and dialogue that served to contextualise and mask the lexical bundle. In the following example, everything other than the underlined portion is supporting context:

> Context: Andy is going on a double date with his friends.
>
> Lexical bundle: 'Really, I don't know why I agreed to this,' he grumbled to his friends Simon and Alice. 'You know this isn't my sort of thing.'
>
> Or:
>
> Control: 'Really, I'm not sure why I agreed to this,' he grumbled to his friends Simon and Alice. 'You know this isn't my sort of thing.'

In lexical bundle studies, controls are normally made by switching a single word for another matched for frequency. To re-use the example given earlier in Chapter 3, *in the middle of the'* (1,513 occurrences in the BNC corpus) might become *'in the front of the'* (47 occurrences in the BNC corpus). This was not possible here, because the lexical bundles and controls needed to have the same meaning so that they could be used in the same contextual sentences. Therefore, we first selected 17 lexical bundles with a minimum frequency of over 3 per million occurrences in the BNC corpus. We then selected synonymous expressions, ensuring that their frequency was as low as possible in comparison to the lexical bundles. For example, the only difference between the two versions in the example above is the lexical bundle, *'I don't know why'* (frequency = 651 in the BNC corpus) and *'I'm not sure why'* (15 occurrences in the BNC corpus). Thus, any difference between the two can be attributed to the presence or absence of a lexical bundle. However, because it was not possible to carefully control the frequency of individual words in the lexical bundles and controls, there was a significant difference in individual word frequencies (t = 4.59, p = <.0001). Thus, word frequency was included as a variable in the analysis.

Procedure

The passages were shown to participants in an offline naturalness judgement task. Participants were shown each passage individually on a computer screen and were asked to rate the underlined portion for how natural they thought it sounded on a 7-point scale (7 = very natural and 1 = very unnatural). The items were counterbalanced so that participants saw an equal number of lexical bundles and control items, but never both versions of the same item. Following this, participants were asked to provide information about their language background.

Analysis and Results

The data was analysed using mixed effects modelling, with the package lme4 (Bates et al., 2015). Naturalness rating was modelled as the response variable with condition, phrasal frequency, mean word frequency (calculated as the BNC frequency of each word divided by the number of words), and sequence length (calculated as the number of characters in a sequence) as fixed effects. Because we are interested in differences between how the native speakers rate items compared with Japanese English learners, we needed two separate phrasal frequency metrics tailored to each of the two groups. As a 'native' metric we used values from the BNC. For the Japanese speakers, selecting a representative frequency metric is difficult as we know little about the relative exposure to English of each participant. What we do know, however, is that each participant learned English at school from the age of 13 using Ministry of Education approved textbooks. Frequency scores from a corpus of these textbooks—the JHSETC (Chapter 3)—proved a robust indicator of online processing in a task with junior high school students (Chapter 4). Therefore, we used the JHSETC as an indication of participants' exposure to lexical bundles (and the controls). We also considered using scores from the JHSETC for mean word frequency, however at the individual-word level, frequency scores between the BNC and the JHSETC were strongly correlated (r =.86, p = <.001) and so we this was deemed unnecessary. For Japanese participants, we also included proficiency, age of first contact with English and age as factors in the model. Participants and items (lexical bundles) were included as random-effect factors in the model. Table 8 summarises mean naturalness ratings.

Table 8. Mean naturalness ratings and standard deviations for lexical bundles and controls for native and non-native participants in Study 3.

	English Speakers		Japanese Speakers	
	Lexical bundle	Control	Lexical Bundle	Control
Mean Ratings	5.47	4.44	4.56	3.89
SD	1.82	2.02	2.00	2.01

Some predictors were correlated and so were residualised by fitting linear models and creating a new metric consisting of the residual (i.e., information not accounted for by correlating factors, Baayen, 2008). For example, BNC phrasal frequency was correlated with JHSETC frequency and mean-word frequency, so BNC phrasal frequency was fitted as the response variable in a model with JHSETC frequency and word-frequency as predictors. Other correlated factors were fitted in a similar manner. The resulting residuals all correlated significantly with their related variables (< .001): JHSETC phrasal frequency (r = .96), BNC phrasal frequency (r =.72), word frequency (r =.76), length (r =.76), and proficiency (r =.89). All continuous predictors were centred to avoid having a change in slope that might correlate with a change in intercept (see Baayen, 2008).

For both native and non-native speakers, a simple model was fit with residual JHSETC frequency, and with subject and lexical bundle as random effects. Complexity was then built into this model by adding new predictors in a step-by-step manner. Each new predictor was tested as a discrete factor, and for interactions with all other predictors. These were also tested as adjustments for both the subject and lexical bundle random effects where relevant (as recommended by Barr et al., 2013). Each iteration of the model was tested against the previous best fit to see if the change was significant (likelihood ratio test, $p \leq .05$). This process was continued until all predictors and interactions had been tested both as fixed and random effects. Table 9 summarises the best fit for both models.

Whether an item was classified as a lexical bundle or a control phrase made no difference for either native (X^2 (1) = 0.89, p = .34) or non-native speakers (X^2 (1) =

Table 9. Summary of fixed and random effects for the best-fit model for Study 3.

	Japanese Speakers			Native English Speakers		
Fixed Parts	Estimates	CI	p	Estimates	CI	p
(Intercept)				4.95	4.45 – 5.45	<0.001
Residual Textbook Frequency	0.39	0.14 – 0.63	0.004			
Age of First Acquisition	-3.88	-6.98 – -0.78	0.014			
Residual Length	-2.53	-4.81 – -0.25	0.038	0.14	0.01 – 0.26	0.037
Age of First Acquisition * Mean Word Frequency	0.31	0.08 – 0.55	0.009			
Mean Word Frequency * Residual Length	0.19	0.02 – 0.37	0.037			
Residual BNC Frequency				0.24	0.07 – 0.41	0.004
Random Effects						
σ2	3.07			2.11		
τ00, Lexical Bundle	0.48			0.76		
τ00, Participant	0.3			0.79		
N Lexical Bundle	34			34		
N Participant	23			24		
ICC Lexical Bundle	0.12			0.21		
ICC Participant	0.08			0.22		
Observations	392			408		
R2 / Ω02	0.096 / 0.278			0.092 / 0.477		

0.02, p = .87) indicating that lexical bundlehood cannot be reduced to a simple binary metric. However, there were significant effects based on lexical bundle frequency. The data was modelled with both native-speaker frequency values and frequency values from the JHSETC. Native speaker ratings were predicted best by the BNC phrasal frequency, and there was no significant effect of the JHSETC values (X^2 (1) = 1.00, p = .32). Japanese speaker ratings, however, are predicted best by the JHSETC phrasal frequency but not the BNC values (X^2 (2) = 0.09, p = .96). This is an important finding and will be taken up in the Discussion.

For both groups, word frequency was not a significant predictor (native X^2 (1) = 0.05, p = .82, Japanese X^2 (1) = 0.25, p = .87) indicating that differences are related to the complete phrase, not simply to the words within them. However, there were significant interactions for the Japanese speakers related to word frequency: first, between word frequency and age of first acquisition whereby participants who started learning English at a younger age tended to rate items with lower mean word frequency higher on the naturalness scale, and items with higher mean word frequency as lower on the naturalness scale. Japanese speakers who started learning later, however, seem to make minimal use of word frequency. There was also an interaction between word frequency and residual phrasal length whereby items shorter in length and with lower word frequency tended to be rated higher, and those shorter in length and with higher word frequency lower on the naturalness scale. For longer items, this pattern was reversed. These interactions will be taken up in the Discussion.

Age of first acquisition was also significant for Japanese speakers as a discrete variable, and residual phrasal length was significant for both native and Japanese speakers. There was no significant effect of residualised proficiency (X^2 (2) = 0.16, p = .92) or age (X^2 (2) = 1.64, p = .44) for Japanese speakers. Finally, there were no significant interactions for native speakers, and neither were the native nor the Japanese speaker models improved by adjusting random effects by any of the fixed effect variables.

Discussion

Although this study did not find a difference between lexical bundles and non-lexical bundles *(I don't know why* vs. *I'm not sure why*), within the category of 'lexical bundle' native and non-native speakers display some interesting differences. The lexical bundles that native speakers and non-native speakers rate as more or less natural are not the same. The native speaker group rate items that are high frequency in the BNC as more natural and there is no JHSETC frequency effect. Native English speakers have not been exposed to Japanese English textbooks, and therefore we would not expect frequency in the JHSETC to be a significant predictor of their behaviour – which was indeed the case. In contrast, the Japanese learners of English are sensitive to frequency of occurrence in their (early) English textbooks (JHSETC); that is, items that are higher frequency in the JHSETC are judged as more natural, and those with a lower frequency, less so. BNC frequency metrics did not predict the Japanese speakers' naturalness-ratings.

As mentioned briefly in the introduction, and at length in Chapter 3, the detailed analysis of the lexical bundles in the JHSETC demonstrated that they are quantitatively and qualitatively different from those in a native-speaker corpus. The current study provides additional evidence that the lexical bundles in the JHSETC do not represent native-speaker English; if they did, the native speakers would rate ones occurring more frequently in the JHSETC as more natural. Notably, the Japanese speakers were exposed to language from the JHSETC on average 15 years before the current study. The fact that we see an effect of the JHSETC, a corpus that reflects the learners' earliest experience with English, demonstrates the importance of teaching materials targeted at beginners. In addition to the findings outlined in Chapter 3, other studies focusing on secondary school textbooks have found the dialogues in them to be inauthentic (Nguyen & Ishitobi, 2012; Römer, 2004, 2005), in that they do not accurately represent the situations students need to be prepared for. Based on the findings of this study, not only is this problematic because we are not preparing our students for English use in the real world, but arguably presenting inauthentic, 'unnatural' English to students is also damaging in the long-term as evidenced by their intuitions about the naturalness of English lexical bundles.

As described above, there was an interaction between word frequency and age of first acquisition whereby Japanese participants who started learning English at an early age rated items with lower word frequency higher on the naturalness scale, and items with higher mean word frequency as lower on the naturalness scale. One explanation for this could be that learners who started learning English earlier have more experience with English, and therefore have had more exposure to lower-frequency words. Further, this pattern is in line with findings from native-speakers where greater frequency effects are evident in low frequency than high frequency words (for a discussion see Conklin, 2020). In contrast, Japanese speakers who started learning later demonstrated no effect of word frequency, indicating that they did not have sufficient exposure to English. Word frequency interacted with phrasal length, such that items that were shorter in length and with lower word frequency were associated with relatively higher ratings, and those shorter in length and with higher word frequency were associated with relatively lower ratings on the naturalness scale. This seems to suggest much the same rating behaviour; there are greater frequency effects in low frequency items, which was more evident in shorter items.

The current study—and indeed, the other studies presented in this thesis—is limited in that it only focuses on one specific group of speakers: Japanese English learners. We chose this population because it continues on from, and adds evidence to findings from Study 1 and 2 (chapters 3 and 4). Simultaneously, corpora such as the JHSETC are essential for generating phrasal frequency metrics tailored to a specific groups' experience of English. In order to test the role of early input in other populations, similar corpora would be needed (this will be discussed in Chapter 7). However, the current research provides an indication of the lasting impact of the language materials that young learners are exposed to. In the intervening time between exposure to the textbook language and the current study, the learners will have been exposed to English in other ways (including their current English study in the e-learning platform from which they were recruited). Crucially, this additional exposure was not reflected by naturalness ratings that were in line with the native speaker ones and related to a native speaker corpus. It may be that frequency metrics from another source, say, for example the COCA (Davies, 2008) or the SUBTLEXus (Brysbaert & New, 2009), may better predict Japanese speakers' rating scores because the Japanese

learners may have had more exposure to American English, or indeed scores taken directly from materials used in the e-learning environment. This warrants further research, particularly to address whether very recent exposure to materials optimised with high-frequency lexical bundles (i.e., ones that could be embedded in their e-learning platform) would 'overwrite' the (un-nativelike) lexical bundles that learners' have entrenched in long term memory.

Overall, as with the previous study (Chapter 3), the results from this study align well with usage-based accounts of language acquisition, which maintain that language develops in reaction to use and exposure (Barlow & Kemmer, 2000; Bybee, 2006; Ellis et al., 2013; Tyler, 2010). In the previous chapter, I demonstrated a clear processing advantage on a phrasal judgement task for the lexical bundles that Japanese junior high school students encounter in their textbooks. The students were sensitive to the phrasal frequency of items embedded in their early textbooks and responded to them faster and more accurately. Such findings indicate that exposure to language underpins entrenchment in memory – even from the very earliest stages of learning. That the frequency from these textbooks predict the naturalness ratings of *adult* Japanese learners of English suggests that the processing advantages they developed early on may be retained over time, and even in some cases—for example where there is a large discrepancy between the frequency distribution of an item in the textbooks and in naturally-occurring language—influence their judgement and prevent them from making (fully) nativelike evaluations of lexical bundles. These findings once again highlight the need to carefully consider the input students are given, even at the earliest stages of learning, as they may have a far-ranging influence.

Conclusion

This study demonstrates that language learners develop intuitions of English based on the actual English they have learned and been exposed to, which supports a usage-based view of language acquisition. If the input learners receive is not nativelike, their *intuitions* will not be nativelike. Crucially, once formed, these intuitions may become entrenched in memory and stay with learners long-term; the current findings demonstrate an influence of English materials that the learners encountered on average

15 years previously. Notably, these findings highlight the need to carefully consider the input students are given and for textbook designers to create contents that reference native speaker English, rather than simply creating examples to demonstrate grammar points without consideration of the utility of the examples in which they are embedded.

Chapter 6 - "Did You See That?"

The current study further explores the development of a processing advantage for lexical bundles, and asks whether learning materials can be optimised via input enhancement in order to facilitate it. To address these questions, Japanese learners of English read a series of three short stories repeating low-frequency, 3-word lexical bundles. Half of these bundles were underlined, that is, enhanced, while the other half were not. After reading, participants performed a phrasal judgement task including the lexical bundles from the story, as well as control items not presented in the story (half of which were repeated in each judgement task and half which were new). Responses were faster to items repeated in the stories, and there was an initial, but short-lived, benefit from textual enhancement. A post-test conducted one week later indicated that students retained some knowledge of the lexical bundles from the stories, as well as for items that only appeared in the phrasal judgement tasks. Chapter 6 has been submitted as '"Did you see that?" - The role of repetition and enhancement on lexical bundle processing in English learning materials' and is currently under review (Northbrook, Conklin, & Allen, under review).

Introduction

As has been discussed already at several points throughout this thesis, native speakers of a language tend to express concepts using the same words each time. As a broad category, these repeated word sequences are referred to as 'formulaic sequences' and are said to aid communication by freeing up processing resources to increase fluency in both production for the speaker (Conklin & Schmitt, 2008; Kuiper & Haggo, 1984) and comprehension for the listener or reader (Boers et al., 2006; Kuiper, 2004; Pawley & Syder, 1983). In Chapter 2, I described and discussed seminal work by Pawley and Syder (1983) who roughly divide the benefits of formulaic sequences into two aspects: nativelike selection and nativelike fluency. In other words, native speakers sound natural because they select nativelike language, and they speak fluently because they default to conventionalised expressions that they store—either as wholes or as routines—and use them without the need to resort to resource-draining computational processes (see also Ellis, 2001). This reduces the processing load and allows longer

stretches of speech to be processed in working memory (Kuiper & Haggo, 1984).

Nativelike selection, on the other hand, can be considered a sensitivity to collocational preferences (Wray, 2000, p. 463). It is not simply a matter of applying syntactic rules; it is also a sensitivity to more familiar or more 'probable' utterances (Ellis, 2001). Indeed, grammaticality alone is neither sufficient nor necessary for an utterance to sound 'natural' to a native speaker: few would say that, *'Would you aid me in this task?'* is more natural than, *'Would you help me with this?'*, although both are grammatical. In the same way, few would argue that, *'Thanks very much'* is an unnatural sounding expression, though technically it is ungrammatical. Recall that this relates closely to Hymes' (1972) fourth parameter of situational competence, which states that language in the real world should be performed—or put more simply, language should be 'done'—not just be possible, well-formed, or appropriate (the first three of Hymes' parameters). As Skehan puts it, "learners who restrict themselves to item-and-rule approaches to language will be forever marked as non-members of the speech community they aspire to" (1998, p. 39). For second-language learners, then, the use of formulaic language is a mark of nativelikeness, and therefore proficiency (Boers et al., 2006; Dickinson, 2012; Kuiper, 2004; Wray & Fitzpatrick, 2008).

One important type of formulaic sequence, used as a unit of analysis throughout this thesis, is the 'lexical bundle', which is loosely defined as a sequence of words that occurs frequently in language (Biber, 2009; Biber et al., 1999) and which can be thought of as an extended collocation. Lexical bundles are generally corpus derived and may constitute a complete utterance (*how's it going*) but often do not (*a lot of the*). In applied linguistics, there is a growing body of research that suggests formulaic language more broadly, and lexical bundles specifically, are processed more quickly than novel or less frequent sequences of words (Conklin, 2020; Conklin & Schmitt, 2012; Siyanova-Chanturia & Van Lancker Sidtis, 2019). This has been referred to in the literature as a 'formulaic advantage' and was discussed at length in Chapter 2. We would expect that learning formulaic sequences and developing a processing advantage for them would be highly beneficial to learners of English as a second or foreign language. However, important questions remain about the emergence of the formulaic advantage and its relationship to the input that learners receive.

Learning formulaic sequences

While formulaic language can benefit second language learners, it also seems to be an area of difficulty. On one hand, at early stages learners appear to use formulaic sequences to boost their skills; prefabricated chunks they have been taught or have memorised allow learners to use structures that they may not have completely mastered (Myles et al., 1998; Myles et al., 1999; Nattinger & DeCarrucio, 1992; Wray, 2002). As proficiency increases, however, learners seem to use less formulaic language and instead prefer to construct much of their speech using rules rather than lexicalised routines (Skehan, 1998), perhaps mistakenly believing that high-proficiency language requires originality and creativeness. Learners also lack the kind of collocational intuitions that mark native speaker speech, even at advanced stages (Siyanova & Schmitt, 2008).

Part of the reason that learners lack nativelike intuitions and knowledge about formulaic language may be due to the materials they encounter, and the studies presented in chapters 3–5 have provided evidence of this. In psychology, this is referred to as the 'mere-exposure effect' (Zajonc, 2001), which states that we develop preferences for the things we are regularly exposed to. An anecdotal, but poignant example of this is the so-called 'linguists-disease', where linguists who regularly create ungrammatical sentences for use in experiments lose their ability to recognise those very sentences as erroneous (see Snyder, 2000). What this tells us is that language users become used to—and even prefer—the language they are exposed to, regardless of its acceptability. Therefore, repeated exposure to non-nativelike input in teaching materials may result in preferences for non-nativelike English. The study presented in Chapter 4 found evidence for this: Japanese secondary school students demonstrated a processing advantage for non-nativelike lexical bundles embedded in their English textbooks. Furthermore, in the naturalness-judgment study with Japanese speakers presented in Chapter 5, we found participants preferred lexical bundles that they had been exposed to in their junior high school textbooks over higher-frequency, more 'nativelike' alternatives (as defined by native-speaker judgments).

These findings can be explained by a usage-based perspective (discussed in Chapter 2, and throughout this thesis). Under such an approach, the formulaic

advantage emerges from repeated exposure to sequences of words, which are stored in long-term memory (e.g., Barlow & Kemmer, 2000; Bybee, 1998, 2006; Bybee, 2013; Tyler, 2010). Earlier, in Chapter 2, also summarised in Chapter 4, I described this optimisation of language in terms of a computer's internet browsing cache. A web-browser stores information pertaining to frequently accessed webpages, which allows us to browse the internet faster and more efficiently without having to load large data files, such as images, every time they need to be accessed. Similarly, a usage-based approach posits a viewpoint whereby formulaic language reduces the processing load for both speakers and listeners by 'caching' language such that it can be accessed quickly and easily without needing to retrieve and compute the individual components (Goldberg, 2006; Wray, 2002, 2008; Wray & Perkins, 2000). Thus, the first time a sequence of words is encountered, it is processed in a word-by-word manner. However having done so, this leaves an imprint in memory that is strengthened by subsequent exposure (e.g., Logan, 1988).

Accordingly, frequency is thought to underpin learning and is a common factor in all approaches that predict the development of a processing advantage for formulaic sequences (Higham, 1997). There is some argument surrounding the exact mechanisms by which this happens, though it is generally explained in terms of memory-based retrieval (e.g., Bolinger, 1975; Logan, 1988) or by a process of algorithmic fine-tuning (DeKeyser, 2015). A key difference between approaches lies in the number of times a sequence of words would need to be encountered for it to be processed as a formulaic unit, rather than computationally. This was taken up in detail in Chapter 2, but in brief, some models, generally referred to as 'threshold-approaches' (e.g., Goldberg, 2006; Wray, 2002), predict that only items that are frequent enough—i.e., surpass some threshold—are stored in memory. Others, posit a 'continuous-approach' that predicts each and every exposure to a phrase influences subsequent entrenchment and processing (e.g., Bybee, 1998, 2006; Bybee & Hopper, 2001; Mclaughlin et al., 1983).

Subsequently, an important question in formulaic language research has been the number of times a sequence needs to be experienced in order for it to be learned. Is one exposure enough? Or does a sequence need to be encountered multiple times for a processing advantage to emerge, as predicted by a threshold-approach, and if so, how

many times is sufficient? This will now be discussed at length in the following section, before turning to a second, but equally important topic in the literature: the benefit of enhancement in learning materials.

Optimising the Input: Repetition and Enhancement

Studies investigating the number of repetitions needed to learn formulaic sequences have come to different conclusions. Indeed, this is not a problem limited to formulaic sequence research; studies exploring the number of encounters to learn single words have also come up with very different results: e.g., six repetitions in Rott (1999), eight repetitions in Horst, Cobb, and Meara (1998), and more than 10 repetitions in Pellicer-Sánchez and Schmitt (2010); Pigada and Schmitt (2006); Webb (2007). This variation, however, is almost certainly due to the differences in other factors across the studies, which are likely to interact with the frequency of exposure effect (Pellicer-Sánchez & Boers, 2019), meaning it is difficult to draw clear conclusions.

Returning to formulaic language, a number of studies have explored the role of frequency of exposure on learning. For instance, Durrant and Schmitt (2010) investigated the effects of repetition on the learning of collocations. The researchers assigned participants to one of three training conditions: single exposure, verbatim repetition and varied repetition. The results suggested that adult L2 learners do retain collocational information, and evidence for this was found after just two repetitions. However, participants only demonstrated knowledge of a small number of the items tested, leading the researchers to conclude that, "substantial exposure to a second language [is required] in order to gradually build up knowledge of a large number of collocational pairs" (p.182).

In another study, Peters (2014) investigated the effects of frequency of exposure (one, three, or five occurrences) during explicit instruction on the form recall of collocations. She found evidence for a large and positive frequency effect whereby increased frequency of exposure in instruction led to higher recall in a post-test. Items that were repeated five times were recalled three times more often than items presented only once. Further, participants retained much of what they had learned, as demonstrated in post-tests.

In a similar study, Pellicer-Sánchez (2017) studied the incidental acquisition of collocations by L2 Learners who read a story containing target collocations that were repeated either four times or eight times. One week later their participants' knowledge of form and meaning of the same items was assessed. The author found evidence of learning; however, no significant differences were found between four and eight exposures. Because only two levels of repetition were used, the study does not indicate how many occurrences are sufficient for learning, nor whether there are differences across initial frequency levels; it is quite possible that (some) learning occurred after a single exposure, although this assertion could be questioned given the results from other studies such as Peters (2014).

The three studies discussed thus far seem to suggest formulaic sequences are learned after a relatively small number of repetitions. Conversely, in their study, Webb et al. (2013) found far more repetitions were required. Following an approach applied by Horst et al. (1998), the researchers studied the learning of verb-noun collocations from reading texts. Taiwanese EFL learners were given one of four versions of a story which they read along with an audio version. The stories contained one, five, 10 or 15 repetitions of target collocations. An immediate post-test showed that a single encounter with collocations was not sufficient for learning to occur. For most learners, even 10 repetitions was insufficient and as many as 15 repetitions were required to see sizable learning gains, which is considerably more than that suggested by other studies.

It is very difficult to draw conclusions about the exact number of times a formulaic sequence may need to be repeated for learning to occur from the currently existing research. Indeed, we may never be able to point to an exact number (Peters, 2014). However, what is clear from the studies cited above is that repetition is beneficial. This was emphasised by Webb and Nation (2017) who suggest that, in terms of vocabulary learning, *repetition and the quality of attention* at each encounter are two key factors underpinning learning. In addition to repetition, the quality of attention is important because learners may not always attend to sequences, such as collocations or lexical bundles, when they consist of familiar words (see Godfroid, Boers, & Housen, 2013). One way that educators can help students attend to, and thus learn, sequences may be by enhancing the input they receive. Such 'input

enhancement' has been an important topic in the formulaic sequence learning literature and can be defined as any pedagogical technique specifically designed to draw learners' attention to features of the language input (Kim, 2006) and which assumes that altering learners' input stimulates their processing and therefore learning of said input (Sharwood Smith, 1981, 2008).

In one input enhancement study, Webb and Kagimoto (2009) investigated the effects of receptive and productive vocabulary tasks on the learning of collocations and their meaning in Japanese English learners. For the receptive task, enhanced sentences were used (achieved by bolding the collocations) and for the productive task, a cloze task was used. They found increased gains in collocational knowledge in both tasks, with the receptive task being particularly effective for participants having lower proficiency levels.

In another study, Peters (2012) explored the learning of words and formulaic sequences in second-language German, asking whether it could be facilitated by drawing attention to items either by directing learners' attention to them in the task instructions, or via text enhancement (bolding and underlining). Learners were told to pay attention to both formulaic sequences and individual words in one group, or to unfamiliar vocabulary in general in another group. In terms of input enhancement, half of the target items were presented bolded and underlined, the other half were not. Directing attention to sequences in instructions yielded no gains. On the other hand, a form recall test showed significant gains with textual enhancement.

Szudarski and Carter (2016) examined the learning of verb-noun and adjective-noun collocations in two conditions (enhancement via underlining vs. no enhancement) while reading six stories read over a span of three weeks. A post-test was conducted two weeks after the treatment and included a battery of tests designed to tap into collocational knowledge. The tests revealed that while the no enhancement condition did not facilitate learning, textual enhancement lead to improved form recall and recognition. Notably, this study also included a measure of repetition: each story included 10 collocations which occurred once, and 10 which occurred twice; making for a total of either six repetitions or 12. The researchers found that 12 occurrences combined with enhancement improved learning at the level of form recall. At the level of recognition, gains were already found at six repetitions and there was no further

improvement from 12 repetitions. Crucially, the unenhanced condition did not seem to promote learning regardless of the number of repetitions. Overall, these studies suggest enhancement is beneficial and can help learners to learn target sequences.

In contrast, a study by Szudarski and Conklin (2014) did not point to a benefit for enhancement. They used a judgement task to measure the effects of rote rehearsal and enhanced rote rehearsal (using underlining) on the learning of four different types of collocation embedded in sentences. They found evidence of learning, but there seemed to be no additional gains via enhancement beyond the effect of rote-rehearsal. Further, in a post-test six weeks later, they found the effects to be short-term, concluding that recycling and reinforcing input over time is essential (see also Durrant & Schmitt, 2010). In sum, this study suggests that the effect of repetition is more important than that of enhancement.

Clearly, the way learners interact with and learn formulaic sequences from textual input is complex, and the role of repetition (frequency) and enhancement cannot easily be teased apart. The current study aims to address these issues using a novel experimental design, intended to explore the relationship between repetition and enhancement of lexical bundles embedded in specially designed learning materials. To this end, the following research questions are addressed:

1. Do phrasal judgements become faster with repeated exposure to lexical bundles embedded in texts—thereby evidencing learning through repetition—and if so, after how many repetitions does this learning occur?
2. Does enhancement of lexical bundles, via underlining, facilitate learning?
3. Are any learning benefits still apparent one week later?

In order to answer these questions, Japanese undergraduate learners of English read a series of short stories containing low-frequency 3-word lexical bundles. Half of the lexical bundles were underlined in the text to draw attention to them (referred to as Story Enhanced), the other half were left unmarked (referred to as Story Unenhanced). The students were tested after each story-reading with a phrasal judgement task (PJT), containing the lexical bundles in the story as well as control items that were not included in the story. These control lexical bundles were included to assess any test

effects that might come about from repeatedly seeing a set of items (i.e., in the PJT itself): Half of these lexical bundles were repeated in each judgment task (referred to as PJT Repeated) and half were new in each judgement task (referred to as PJT Unique).

Methodology

Experimental stimuli

Three different stories of approximately 300 words were created such that each had two sets of ten 3-word lexical bundles (enhanced vs. unenhanced). To ensure the stories were at an appropriate level for the participants, they were checked for their lexical coverage using the Lextutor Classic Vocabulary Profiler (Cobb, n.d.). Over 95% of the words in all three stories were rated as K1-K2, and as such should not be challenging for university-level students who had been learning English for more than seven years. The lexical bundles were selected so that they fit into the stories with the constraint that they were 'low' frequency: less than three occurrences per-million in the COCA (Davies, 2008). The final set of experimental items had a mean frequency of just 38.6 occurrences (SD = 23.9) or 0.8 occurrences per million words in the COCA. Low-frequency lexical bundles were used because they are nativelike, as evidenced by their occurrence in the corpus, but due to their low frequency are unlikely to have been previously encountered by the learners. Frequency is, however, included in the statistical models to account for the possibility of participants having encountered the items at some point.

One set of the lexical bundles was underlined in the stories in order to draw attention to them and participants were instructed to pay attention to these phrases. No mention was made of the unenhanced lexical bundles. In order to ensure that any difference in the items was controlled for, items were counterbalanced by creating versions A and B of the stories, such that the underlined items in version A were not underlined in version B and the underlined items in version B were not underlined in version A. For example, one of the stories contains the sentence: 'I set off home feeling happy, but a little tired and hungry.' In version A, 'set off home' was underlined (I <u>set off home</u> feeling happy, but a little tired and hungry) and in version B

'tired and hungry' was underlined (I set off home feeling happy, but a little tired and hungry). Participants were randomly assigned to one of the versions.

In addition to the lexical bundles that appeared in texts, two sets of items were created that only appeared in the PJTs (see procedure below). One set of items consisted of ten lexical bundles with the same maximum-frequency constraint described above. They all appeared in each iteration of the PJT (referred to as PJT Repeated). The inclusion of PJT Repeated items provides an indication of the impact of repetition within the PJT itself. The second set consisted of four sets of 10 unique items, again following the same constraints, but that only appeared in one PJT (referred to as PJT Unique). The PJT Unique items provide an indication of any practice effects (i.e. getting better at the phrasal judgment task more generally, as opposed to the effect of exposure to specific lexical bundles). All of the sets of items were matched for phrasal frequency and word frequency using the COCA (p's < .05). This was important to ensure that any differences between item types could be attributed to the experimental manipulations and not to one set of lexical bundles being higher-frequency than another (see Bannard & Matthews, 2008).

In total there were 40 experimental items in each phase of the experiment. In addition, there were three sets of 90 'filler' lexical bundles consisting of 25 grammatical items and 65 ungrammatical items. None of the fillers appeared in the stories, and like the PJT Unique items, the fillers were different in each iteration of the PJT. The ungrammatical items had a frequency of zero in the COCA. None of the filler items were included in the analysis. Including fillers, each participant responded to a total of 130 lexical bundles in each of the three PJTs.

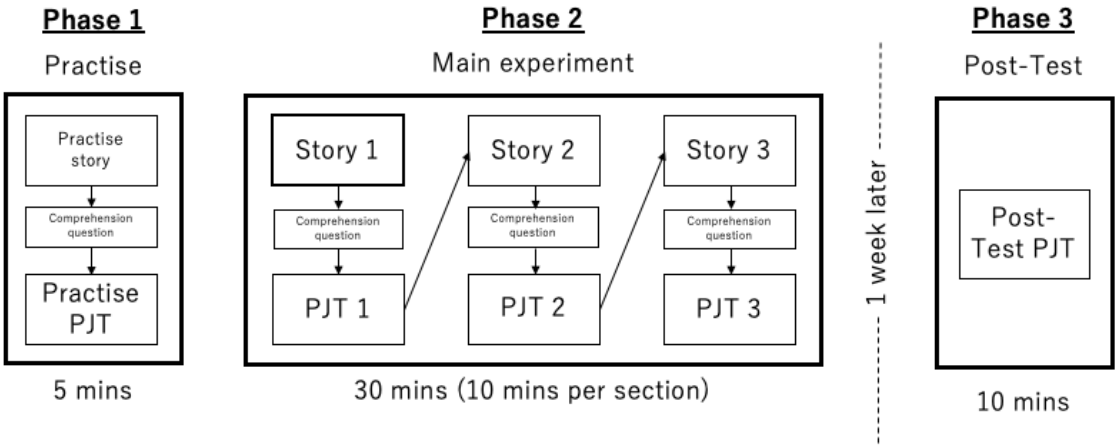
Participants

Thirty non-native undergraduate students at a Japanese university participated in the study. All participants spoke Japanese as their first language and were paid for their participation. The average age of participants was 20.1 (SD=0.9), and the average age of first contact with English was 8.2 (SD=4.0). Participants all had TOEFL ITP scores and were all in the CEFL B1 to B2 range (M=501.1, SD=105.3). These scores were used as a metric of proficiency in the analysis.

Procedure

The participants were tested during a regular English class. The experiment was conducted using Inquisit 4.0.7.0 [Computer software] (2014). The flow of the experiment is displayed in Figure 1. Students were presented with a short story to read, displayed on a single screen. Following this, they were given three 'yes' / 'no' comprehension questions to ensure they had read and understood the story. Following the comprehension questions, the students completed a PJT. The 130 lexical bundles (40 main stimuli and 90 fillers) were displayed one at a time and participants were instructed to press a 'yes' key if the phrase was possible in English, and a 'no' key if it was not. For each trial a '*' appeared in the center of the screen for 700 milliseconds, followed by the target item. Participants were instructed to respond as quickly as possible. The experiment was split into three phases. Phase 1 was a practise phase consisting of a story and 20 practise items with feedback. None of the experimental stimuli were used in this phase, and the data was not included in the subsequent analysis. Phase 2 was the main experiment and consisted of three iterations of stories, comprehension questions and PJT sequence, one after the other in quick succession. A different story was used each time. Finally, Phase 3 consisted of a post-test conducted one week later (again during the participants' regular English class), and only included a fourth repetition of the PJT, once again using the 40 experimental items and 90 fillers.

Figure 1. Experiment Flow for Study 4.



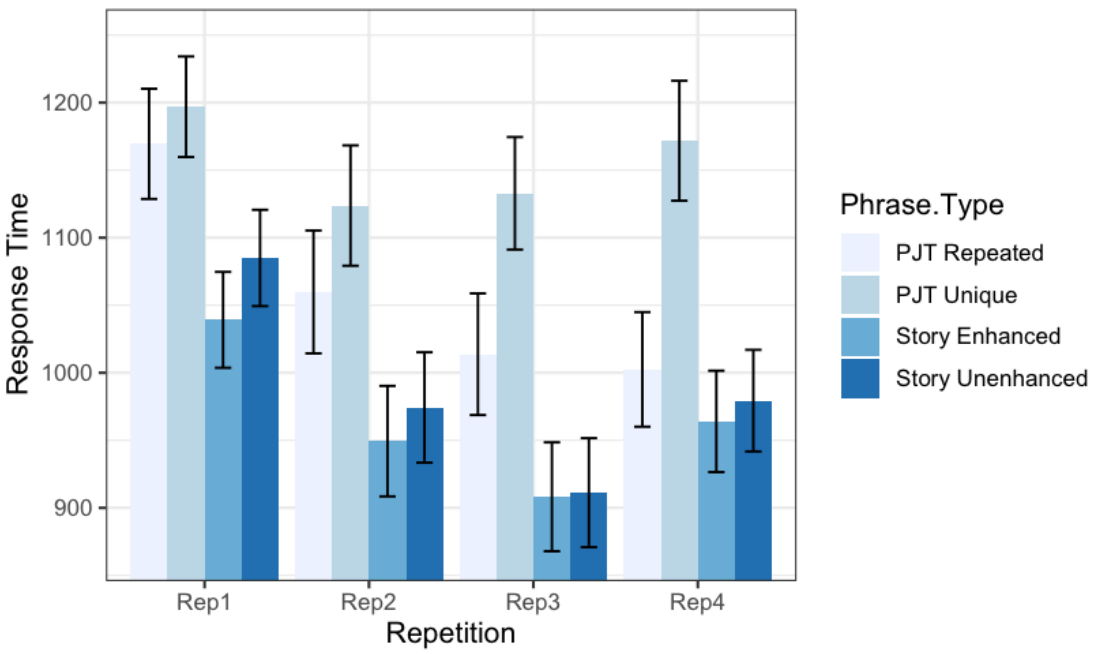
Analysis and Results

Reaction times under 400 milliseconds and inaccurate responses were removed from the data (0.03% and 12.32% of the data, respectively). Further, to ensure that the students had read the stories (and had therefore been exposed to the target lexical bundles), participants with a score of less than 6 out of 9 on the comprehension questions were removed. This resulted in three participants being removed. In addition, one participant who did not complete the post-test was removed from the analysis, leaving 26 participants in the final analysis. Table 10 summarises mean response times, and Figure 2 displays the mean response times across repetitions in all four of the experimental conditions.

Table 10. Mean response times in milliseconds and standard deviations in brackets in each of the four experimental conditions in Study 4.

	Story Enhanced	Story Unenhanced	PJT Repeated	PJT Unique
Reaction Times	971 (272)	990 (294)	1067 (299)	1130 (305)

Figure 2. Mean response times in milliseconds for experimental conditions across repetition for Study 4.



The data was analysed in R version 3.0.2 (R Core Development Team, 2013) using mixed-effects modelling, with the packages *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017). Response time (ms) was used as the dependent (response) variable. The independent predictors of particular interest included the four-level factorial predictor lexical bundle type (Story Enhanced, Story Unenhanced, PJT Repeated, PJT Unique) and the four-level factorial predictor repetition number (Rep1–Rep4). In addition, four control predictors were included: phrasal frequency (defined as the raw frequency of an item in the COCA), length (the number of characters in a given lexical bundle), proficiency (TOEFL IBT score), and trial number *(1–130)*. Finally, subjects and items (the lexical bundles) were included as random-effect factors in the model.

Phrasal frequency and length were correlated and so were residualised by fitting linear models and creating a new metric consisting of the residual (i.e., information not accounted for by correlations among predictors, Baayen, 2008). The resulting residuals correlated significantly with the original predictors: phrasal frequency (r = .99), length (r = .99). All continuous predictors were centered to avoid having a change in slope that might correlate with a change in intercept (see Baayen, 2008).

An initial model was calculated with the above fixed effects and relevant interactions, as well as random intercepts for subjects and items. Automatic backward model selection was performed for fixed effects and interactions using the 'step' function in *lmerTest*, which applies log-likelihood testing by first removing non-significant interactions then main effects. Following this, by-subject and by-item random slopes were added successively for all significant predictors and interactions and were retained if they significantly improved the model. The final model is presented in Table 11.

The final model included repetition, lexical bundle type, phrasal frequency, proficiency, and phrasal length as fixed effects. Trial number was not significant (X^2 (1) = 0.58, p = 0.4463). There were two significant interactions between lexical bundle type and repetition (X^2 (9) = 22.65, p = 0.007), and between phrasal frequency and repetition (X^2 (3) = 16.40, p < 0.001). In addition, random by-subject slopes for repetition significantly improved the model (X^2 (9) = 67.58, p < 0.001), indicating that

Table 11. Summary of Significant Predictors for Study 4.

		Response Time	
Predictors	Estimates	CI	p
(Intercept)	1170.94	1090.93 – 1250.95	<0.001
Repetition [Rep2]	-110.13	-158.23 – -62.04	<0.001
Repetition [Rep3]	-156.52	-209.39 – -103.65	<0.001
Repetition [Rep4]	-167.16	-222.42 – -111.91	<0.001
Phrase.Type [Story Enhanced]	-130.31	-209.38 – -51.23	0.001
Phrase.Type [Story Unenhanced]	-84.51	-163.83 – -5.19	0.037
Residual.Phrasal.Frequency	-51.24	-78.67 – -23.81	<0.001
Residual.Phrasal.Length	63.29	37.16 – 89.43	<0.001
TOEFL	-54.61	-99.21 – -10.02	0.016
Repetition [Rep3] * Phrase.Type [PJT Unique]	91.69	20.76 – 162.61	0.011
Repetition [Rep4] * Phrase.Type [PJT Unique]	141.84	45.13 – 238.56	0.004
Repetition [Rep4] * Phrase.Type [Story Enhanced]	91.82	32.57 – 151.07	0.002
Repetition [Rep4] * Phrase.Type [Story Unenhanced]	61.40	1.52 – 121.28	0.044
Repetition [Rep2] * Residual.Phrasal.Frequency	27.18	3.58 – 50.79	0.024
Repetition [Rep3] * Residual.Phrasal.Frequency	43.15	20.31 – 66.00	<0.001
Random Effects			
σ2	49925.63		
τ00, Lexical Bundle	8452.65		
τ00, Participant	16876.4		
ICC	0.39		
N Lexical Bundle	70		
N Participant	26		
Observations	3562		
R2 / Ω02	0.158 / 0.483		

not all participants responded to repeated exposures to lexical bundles in the same manner.

Both of the key experimental variables, lexical bundle type and repetition, were significant predictors of response times in the PJT. In terms of lexical bundle type, the story conditions (Story Enhanced, Story Unenhanced) were both significant as standalone factor levels, but the two PJT factor levels (PJT Repeated, PJT Unique) were not significant outside of the interaction with repetition (discussed below). Among all fixed effects, repetition was the strongest predictor of response time. Participants' phrasal judgment behaviour changed over the course of the experiment with responses in each repetition becoming faster until the fourth, delayed repetition, when response times slowed. Crucially, these changes differed depending on the type of lexical bundle, making it necessary to consider the interaction between these two factors. Therefore, a post hoc analysis of the differences of LSMEANS was carried out to explore these differences. Contrasts of interest that were considered included those between Story Enhanced and Story Unenhanced items, as well as those between PJT Repeated and PJT Unique items, all at different stages of the experiment (i.e., repetition stages). The difference between Story Enhanced and PJT Unique items was not considered relevant except at Rep4. Statistics for all comparisons are included in Appendix 2.

For all item types, there was a significant speed up in response times between Rep1 and Rep2. Because this was also true of both types of the PJT only items, it is assumed to be in part due to participants improving at the task in general. This assumption was supported by the fact that no difference was observed between PJT Unique items and PJT Repeated items in Rep1, which is expected because both item types were encountered for the first time.

Importantly, however, there was an expected difference between the response times for items that did or did not appear in the story. In the first PJT (Rep1), the story items had already been encountered once during the reading phase, while the two types of PJT items were encountered for the first time. At Rep1, there was a significant difference between the Story Enhanced and Story Unenhanced items, with significantly faster response times for the enhanced items. However, this difference disappeared at Rep2 and Rep3. For the two types of PJT only items, there was no

difference between them at Rep1 and Rep2, though a significant difference appeared at Rep3, with faster responses for PJT Repeated items than PJT Unique items. This indicates that repeated exposure to lexical bundles in the PJT itself was sufficient to facilitate responses, and this emerged on the third exposure (evidenced by faster processing between Rep2 and Rep3 for PJT Repeated items, but not for PJT Unique items).

Turning to the delayed task, recall that Rep4 was carried out one week after the main experiment, and without a story (i.e., it was a single repetition of the PJT). At this stage, there was no significant difference between Story Enhanced and Unenhanced items or between PJT Repeated items and either of the story conditions. This is noteworthy because at Rep1–3, responses in the story conditions were significantly faster than those in the PJT Repeated items. Nevertheless, Rep4 responses in both story conditions and the PJT Repeated condition were significantly faster than those in the PJT Unique condition. This was expected given the additional exposure to these items in the stories and/or PJT: Story items were seen six times (three in the story and three in the PJT) and the PJT Repeated items were seen three times. In contrast, items in the unique condition were seen only once.

Importantly, for the story conditions, as well as the PJT Repeated condition, there were significant differences in response times between Rep1 and Rep4 but no such differences between Rep2 and Rep4. This indicates that participants' performance 'regressed' to that of Rep2 rather than to initial exposure in Rep1. This was not the case in the PJT Unique items, where performance for Rep1 and Rep4 did not differ. Taken together, this indicates that the participants retained knowledge of repeated lexical bundles and did not simply get better at the PJT itself.

There was also a significant effect of residual phrasal frequency in the model, which interacted with repetition. The frequency of the lexical bundles influenced processing at all repetitions whereby higher frequency items were responded to faster, though this effect was less pronounced in Rep2 and Rep3. Length was significant in that shorter items were responded to faster than longer ones. Proficiency was also significant as a main effect but it did not interact with other predictors, indicating that although higher proficiency participants were slightly faster at the PJT overall, there was no difference in their judgment behaviour for the different item types.

Discussion

This study demonstrates that learners are sensitive to recurring sequences of words in their input – in this case, stories designed to give repeated exposure. Participants' response behaviour in the repeated phrasal judgment task (PJT) changed between repetitions, supporting a continuous approach to formulaic sequence learning whereby each and every exposure to a sequence of words impacts processing in an iterative manner. This finding is consistent with those in similar studies which have found that increments of repetition result in incremental benefits for formulaic sequence learning (Durrant & Schmitt, 2010; Pellicer-Sánchez, 2017; Peters, 2014). Furthermore, the results help to explain some of the conflicting findings about the benefit of enhancement found in the literature (Peters, 2012; Szudarski & Carter, 2016; Webb & Kagimoto, 2009).

As predicted, responses to repeated lexical bundles were faster in each iteration of the PJT, and this was above and beyond practice effects (i.e., getting faster at the task itself). This indicates that learners do learn repeated sequences of words in the materials that are presented to them and that this begins at the initial exposure. These findings are consistent with previous research that has found a benefit of repetition; notably, Durrant and Schmitt (2010), who found evidence of learning after just two repetitions (albeit in just a small number of the items tested), and Peters (2014) who found collocations that were repeated five times were recalled three times more often than items presented only once. The current findings are also consistent with Pellicer-Sánchez (2017), who found evidence of learning for new collocations after only four repetitions – of course, learning may have occurred earlier but it was not tested. An important additional insight from the present study is the observation that learning of lexical bundles occurs from a single exposure. Items in both the Story Enhanced and Story Unenhanced conditions were responded to significantly faster than either of the PJT conditions at the very first repetition, indicating that there was a memory trace for these items following a single exposure from Story 1. Thus, when using a highly sensitive measure of visual phrase recognition (i.e., millisecond responses on the PJT), learning can be observed even from an initial exposure.

The present findings differ markedly with those reported by Webb et al. (2013),

who suggested that 10 repetitions was insufficient for learning and that as many as 15 repetitions were required for sizable learning gains. This discrepancy may be explained by an important task difference between the current study and the majority of studies in the literature. With the exception of Szudarski and Conklin (2014), who also used a phrasal judgment task, previous studies in the literature have used measures of form and/or meaning recall and/or recognition to assess learning. It may be that the phrasal judgement task is more sensitive to early stages of learning than recognition and recall tasks because participants do not need to overtly recognise or recall the items. In other words, initial encounters may leave a trace in memory that helps to optimise performance on a task like the PJT, but the memory trace is not strong enough and/or sufficiently developed to be evident on tasks that test more explicit knowledge. This task difference may therefore, in part at least, explain why the current study finds evidence of learning after a single exposure, while others indicate that more exposures are needed.

The current study showed evidence of a change in processing behaviour after the very first exposure. As with the other studies presented in this thesis (Chapters 3–5), these results align well with usage-based accounts of language learning, which maintain that language develops in reaction to use and exposure (Barlow & Kemmer, 2000; Bybee, 2006; Ellis et al., 2013; Tyler, 2010). As mentioned briefly in the introduction, and also discussed in previous chapters, one key question in the literature is whether a processing advantage only develops after a given number of exposures (a 'frequency-threshold' approach) or whether processing improves incrementally with each and every exposure (a 'continuous' approach). Both views are encompassed under a usage-based approach, and stand in contrast to a 'words-and-rules' approach (see Arnon & Snider, 2010, as well as the discussion in Chapter 2), although they make different predictions about performance. The current results are most in line with continuous approaches that predict that each and every exposure to a sequence of words or sounds in the input affect subsequent processing (Bod et al., 2003; Bybee, 1998, 2006; Bybee & Hopper, 2001; Mclaughlin et al., 1983). This is important in terms of designing teaching materials, because in line with arguments made previously in Chapter 3, it means that the quality of supporting materials—and not just items selected for instruction—is important. Learners acquire knowledge of the sequences

contained in teaching materials, even when: a) the sequences are not intended to be taught or learned, and b) when they are not nativelike (which was precisely the case for Japanese secondary school students studied in Chapter 4 who showed a processing benefit for textbook lexical bundles over nativelike ones).

Arguably, in the current study we might be most interested in the judgment behaviour at the fourth, delayed repetition as this indicates durable learning (see Schmitt, 2010). The delayed PJT reveals that participants responded to lexical bundles significantly faster after a gap of one week than they did when *first* encountering them following Story 1. Importantly, this was true of both the Story Enhanced and Unenhanced items as well as the PJT Repeated items. Notably, one week later, not only was there no difference between enhanced and unenhanced items, there was also no difference between these items and the items repeated in the PJT. This indicates that just three exposures to lexical bundles in the judgment task, without the context of the stories, was sufficient to leave a trace in memory. In other words, participants gained some phrasal knowledge of the lexical bundles in three exposures in the PJT. The additional repetitions for the items that occurred in the stories did not seem to result in any additional gains. Thus, it appears that encountering sequences in context had no lasting benefit. A future version of this study might include an additional set of lexical bundles in stories that are not tested in a PJT until the fourth, delayed repetition. This would allow us to better differentiate the benefit conferred by repetition in a story context and a task.

Another key finding is that textual enhancement was beneficial, though only at the first exposure. By the second PJT repetition, there was no additional benefit of underlining. Such a finding is consistent with the wider literature where enhancement has only sometimes been found to be beneficial. Peters (2012), for example, found a benefit for textual enhancement for a form recall test. Webb and Kagimoto (2009) found evidence of learning of collocations in enhanced sentences, presented as a collocation glossed with a translated meaning in the L1, and three example sentences using the collocation which was further highlighted via bolding. However, neither of these studies included a measure of repetition as part of their design or analysis; the full focus was on the effectiveness of the treatments administered, which represents a major difference to the current study. If the current study had included only one

repetition of the story followed by the initial PJT, we would have come to a very similar conclusion regarding the effectiveness of enhancement. In contrast, Szudarski and Conklin (2014) included repetition (in the form of rote-rehearsal) and found no effect of enhancement. Thus, much like in the current study, when there are multiple exposures, the benefit of enhancement is not as apparent. Taking these studies together, it appears that there is an initial benefit of enhancement but that repetition provides a greater benefit.

Seemingly in contrast with the conclusion that repetition confers greater gains than enhancement is the study by Szudarski and Carter (2016) which found that repetition only benefited learners when combined with enhancement; their unenhanced condition yielded no benefit regardless of the number of repetitions. In the current study, repetition seemed to quickly override any effects of enhancement, which, as discussed above, was only significant in the very first repetition of the PJT. An explanation for this may once again be found in differences in experimental design. Szudarski and Carter (2016) presented collocations over six stories, over a period of three weeks. Each story was around 600 words long – double the length of the current study. It might be that the lexical bundles included in the current study were more noticeable and therefore enhancement was not necessary. In a sense, both the enhanced and unenhanced items may have been noticeable due to their density and repetition in stories and PJT, in something that has been termed an 'input flood' and which is considered to be a type of enhancement (Boers et al., 2017). In addition, and to reiterate the point made elsewhere, it may be that the PJT is a more sensitive measure than form recall/ recognition tasks that have generally been used to test learning. This seems likely, given that the current study and Szudarski and Conklin (2014) both used a phrasal judgment task, and both found quite similar results in terms of effects of repetition and enhancement.

It also seems likely that the PJT itself encourages learners to focus their attention on lexical bundles. As stated in the introduction, learners may not always attend to sequences such as collocations or lexical bundles as they consist of familiar words, which attract little attention (see Godfroid et al., 2013). That is, language learners may understand the meaning of a formulaic sequence from its constituent parts, and therefore may fail to allocate it sufficient attentional resources. It may be that the PJT

makes the lexical bundles salient, by presenting them out of context as obviously 'complete' units, thus encouraging learners to pay attention to them in reading (even in the unenhanced condition). In a similar vein, Peters (2012) hypothesised that including glosses with texts may serve as an attention-drawing technique, which was also effective in Webb and Kagimoto (2009). Further, it may be that the act of explicitly testing items makes them more salient in subsequent readings, as in the old adage of *what gets measured, gets managed.* Thus, in the current study, participants attended to enhanced items most in the first repetition in the story, but after that the presence of items in the PJT caused them to notice even the unenhanced items in reading. More precisely, occurrence in the PJT might have overridden any effect of enhancement. While we do not believe this in anyway diminishes findings from the current study, clearly the phrasal judgment task itself is important, and therefore we should be cautious when comparing our findings with those of other studies. Further research will need to tease apart the benefit conferred by the PJT, repetition in context and enhancement.

Notably, and in addition to the findings discussed already, learners were sensitive to phrasal frequency, such that items having a higher frequency in the COCA were responded to faster. This aligns well with what we know about formulaic sequence processing (e.g., Conklin, 2020; Conklin & Schmitt, 2012), such that more frequent items are responded to more quickly. The frequency effect provides evidence that the participants had previous exposure to these low frequency items and that their exposure represented the continuum in natural language: the lowest frequency items had been encountered less than the more frequent ones. Further, there was an interaction between COCA frequency and repetition whereby frequency had the strongest effect in the very first repetition and then again in the delayed, fourth repetition; in the second and third repetitions, it played a comparatively minor role. This finding is expected, because it is well established that along with the relative frequency at which an item occurs in language—here measured by the COCA—recency and repetition also play an important part in determining processing behaviour (e.g., Ellis, 2002a; Ellis & Wulff, 2015; Wood, 2010c). Therefore, as repetitions increase the relative impact of previous experience decreases, particularly in the case of the items in the study which were chosen precisely because they were low

frequency. Notably, we see these effects in the current study with second language learners.

Conclusion

The current study asked whether embedding lexical bundles into learning materials helps university students to develop a processing advantage for those items, and whether textual enhancement can further facilitate processing. We found that Japanese university students responded faster to lexical bundles repeated in a series of (three) stories. There was an initial, but short-lived, benefit for textual enhancement, but overall repetition was seen to be the most powerful driving force in the learning of items, which extended to lexical bundles included repetitively—but out of context—in the phrasal judgement task. Importantly, a delayed post-test indicated that students had faster processing for lexical bundles encountered a week earlier than those that had not been encountered. Thus, (some) knowledge of the previously encountered lexical bundles was retained.

Usage-based approaches to second language acquisition put a premium on the linguistic input learners receive and predict that any sequence of words that they encounter frequently will be entrenched in long-term memory. In the context of EFL learners, 'input' normally, as also mentioned in previous chapters, means the teaching materials used in their classroom (Meunier, 2012; Nunan, 1991). This study shows that including high-frequency lexical bundles in texts is beneficial to learners, as evidenced by faster response behaviour in a phrasal judgment task. Notably, just repeatedly encountering items in the task also benefits processing, clearly demonstrating the importance of repeated exposure in an overt task.

Chapter 7. Materials Matter

This chapter provides a general discussion of the results of the four empirical chapters. Each study has provided data that adds to our understanding of lexical bundle processing in second language learners. In this chapter, I discuss how these findings enrich our understanding of the relationship between input and development of a processing advantage for formulaic sequences, and the implications this has for the design of language learning materials. I also present some general conclusions that highlight the key contributions made by this thesis, as well as some of its limitations and some directions for future research.

Summary of studies

Following an overview of the literature on formulaic language as it pertains to second language learning, and in particular the development of a 'processing advantage' to such language, Study 1, presented in Chapter 3, provided the groundwork for the subsequent empirical studies and asked whether the lexical bundles in communicative Japanese junior high school textbooks are representative of conversational English. I used a corpus-based approach to compare the most frequent lexical bundles in the textbooks (called the *Junior high school English Textbook Corpus* – JHSETC) to those in an English reference corpus (the SUBTLEXus). The study found that although lexical bundles are very frequent in the textbooks and conform relatively well to English patterns at shorter lengths (3-word lexical bundles), they deviate considerably at longer ones (4-, 5- and 6-words). In a sense, the textbooks can be considered beneficial as they do provide repetitive expose to lexical bundles, however to follow the precepts of the communicative approach—as the Japanese *Ministry of Education, Culture, Science, Sports and Technology (MEXT)* state that they should (Ministry of Education Culture Science Sports and Technology (MEXT), 2003)—, they should: (i) present and model authentic conversational language; and (ii) the relative frequency of the input should mirror that of actual spoken English (i.e., high-frequency language should be encountered more frequently than low-frequency language). The findings suggest that both of these elements are lacking, which raises serious concerns as to the quality of Japanese junior high school English textbooks.

Following this study, the next logical question was to ask to what extent these materials impacted students' processing: do Japanese participants display a processing advantage for high-frequency lexical bundles derived from the JHSETC, as would be predicted by a usage-based approach? To answer this, Study 2 (presented in Chapter 4) used a phrasal judgment task in which junior high school students were asked to judge whether a 3-word lexical bundle was possible in English or not by pressing a 'yes' or 'no' key. Response times and accuracy were measured. Students responded to lexical bundles that occurred in their textbooks significantly faster and more accurately compared to control items, and were sensitive to the frequency of occurrence in their textbooks (e.g., textbook bundle = *going to visit* vs. control item = *going to say*). Notably, the target items were not explicity taught to students, but rather were simply present in the input. Based on these findings, it appears that the qualitiy of the learning materials impacts students' processing. The implications for this were discussed in Chapter 4, and will be extended later in the current chapter.

Study 3 (Chapter 5) explored the perceived naturalness of lexical bundles and compared performance by a group of adult Japanese learners of English, who would have been exposed to the textbooks in junior high school, and native English speakers. Naturalness ratings were taken from both groups and were compared in relation to corpus derived frequencies from the BNC and the JHSETC. Both groups' ratings were predicted by lexical bundle frequency. However, the groups were sensitive to different metrics; the native speakers' performance was best predicted by the BNC, and the Japanese speakers' by the JHSETC. This was taken as evidence that learner intuitions are affected by the input they receive in their teaching materials, and that their intuitions are impacted by their early input many years later.

Study 4 (Chapter 6) asked whether we can optimise the materials containing lexical bundles, via repetition and enhancement, in order to facilitate their processing. Japanese learners of English read a series of short stories repeating low-frequency 3-word lexical bundles; half of these were underlined in the text, while the other half were unenhanced. After each story, participants performed a phrasal judgement task on the lexical bundles from the story as well as ones not seen in the story (half of which were repeated in each task and half of which were new in each). Repetition facilitated processing, and there was an initial, but short lived, benefit for textual

enhancement. A post-test conducted one week later indicated that students retained knowledge of the lexical bundles, and also provided evidence for the learning of items that only appeared in the phrasal judgement tasks.

Developing a 'processing advantage' for textbook-language

The three experimental studies discussed in Chapters 4, 5 and 6 have provided evidence that language develops in accordance to exposure and experience with linguistic input (e.g., Barlow & Kemmer, 2000; Bybee, 2006; Ellis et al., 2013; Tyler, 2010). Throughout this thesis I have discussed a rich body of research indicating that both native and non-native speakers have a processing advantage for high-frequency sequences of words, which can be accounted for by usage (see Chapter 2 for an overview). In native speakers, at least, research findings are fairly clear: all else being equal, higher frequency sequences are processed faster than lower frequency ones. Increasingly, this has been evident in studies exploring non-native, second language speakers. However, important questions remain regarding the stage at which, in a learner's development, these processing advantages emerge.

One key question in the literature is whether a processing advantage only develops after a given or particular number of exposures (a 'frequency-threshold' approach) or whether processing improves incrementally with each and every exposure (a 'continuous' approach). As discussed at length in Chapter 2, as well as briefly in Chapter 6, such views stand in contrast to the 'words-and-rules' approach (e.g., Pinker, 1999; Pinker & Ullman, 2002) because they are usage-based, and predict a frequency-driven processing advantage for sequences that can be explained in terms of memory-based retrieval (e.g., Bolinger, 1975; Logan, 1988) or by a process of fine-tuning (DeKeyser, 2015) of algorithmic rules. Our results are most in line with continuous approaches that predict each and every exposure to a sequence of words or sounds in the input will affect its subsequent processing (Bod et al., 2003 2003; Bybee, 1998, 2006; Bybee & Hopper, 2001; Mclaughlin et al., 1983).

In Studies 2 and 3 (Chapters 4 and 5) the frequency measures used to analyse response data were derived from the JHSETC (introduced in Chapter 3). These analyses revealed a relationship with processing speed in early L2 learners, whereby

higher frequency items in the JHSETC were responded to faster and lower frequency ones more slowly across the continuum (Study 2). Frequency in the JHSETC also predicted non-native speaker's naturalness judgements; once again, across the frequency continuum (Study 3). In Chapter 6, the experiment used specially engineered stories, designed to give learners repetitive input to low frequency lexical bundles (that were identified in the COCA corpus). In this study, faster processing was evident for lexical bundles after just one exposure. Notably, such findings could not be accounted for by a threshold-approach, for example, as proposed by Goldberg (2006) and Wray (2002, 2008). Because an advantage is seen after one repetition, the results are also largely incompatible with Logan's (1988) Instance Theory, which predicts that utterances are initially constructed word-by-word and are later replaced with larger constructions over time through use, i.e., 'chunked' (Logan, 2005; Robinson & Ha, 1993; Skehan, 1998).

The current findings also conflict with views holding that a processing advantage does not emerge in non-native speakers until later in the development process at higher proficiency levels (Babaei et al., 2015; Schmitt et al., 2004; Schmitt & Underwood, 2004; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Underwood et al., 2004; Valsecchi et al., 2013). Proficiency was not a significant predictor in Study 2 or Study 3, and only played a minor role in Study 4, where it showed that lower proficiency levels yielded slower responses overall. However, the overall pattern of becoming faster with repeated occurrences and the initial benefit of enhancement was true for both lower and higher proficiency participants (i.e., repetition and enhancement speeded judgments). Overall, this indicates that the appearance of a processing advantage is *not* simply something that emerges later in the learning process (i.e., with increased proficiency). An important difference between prior research and the current studies is that there was a high degree of certainty about the input that learners' received in the classroom, due to the analysis of the JHSETC (although there was some uncertainty about learners' exposure outside of the classroom – discussed below, in the final section of this chapter). Generally, previous studies examining the processing of lexical bundles have identified them in native speaker corpora (e.g., Babaei et al., 2015; Isobe, 2011; Tremblay et al., 2009; Tremblay et al., 2011 all used items from Biber et al., 1999's study of the BNC

corpus, and Jiang & Nekrasova 2007 used items from the Bank of English) and native speakers showed a processing advantage for them (e.g., Arnon & Snider, 2010; Jiang & Nekrasova, 2007; Tremblay et al., 2009; Tremblay et al., 2011 amongst many others). When we test the same formulaic language on non-native speakers, there is an assumption that their input was the same as that of the native speakers (although it will have been less for them). However, the current findings indicate that this may not be an accurate assumption. The lack of a processing advantage for non-native speakers (e.g., Babaei et al., 2015; Schmitt et al., 2004; Underwood et al., 2004) might *not* be attributable to a lack of proficiency, but rather to the non-nativelike input they receive.

As was discussed in Chapter 2, using native speaker corpora to identify lexical bundles may only be valid for very high proficiency L2 speakers, and even then, maybe only for those living in L1 environments. It is likely that lower proficiency speakers and/or those in EFL contexts, where input is mainly limited to classroom study and teaching materials (Meunier, 2012; Nunan, 1991), will have a different experience of English. This means we cannot be certain that the participants in the studies in the wider literature had actually encountered the items tested in the respective studies, or if they had, whether their rate of exposure was the same as that of native speakers (Conklin, 2020). Arguably, proficiency has been used in the literature as a measure of L2 *exposure to the linguistic units studied*. Lower-proficiency leaners are thought to have had less relative exposure to English, and therefore have encountered the formulaic sequences fewer times, while higher-proficiency learners will have had more exposure, and therefore display nativelike processing. This does not mean, however, that lower-proficiency learners process language in a fundamentally different way, as the studies cited above seem to suggest; it only means that they have not developed a processing advantage for the specific items used in those experiments, or if they have, that frequency metrics derived from native-speaker corpora do not bring those advantages out in the statistical analyses. When we account for the input, we see that even low-level, beginner EFL students demonstrate nativelike processing (Study 2, Chapter 4), either for different items or in relation to a different metric. This was clearly demonstrated in Study 3 (Chapter 5), where Japanese speakers showed no difference between phrasal type—lexical bundle or not—based on BNC phrasal frequency. If we had left it at that, as most past studies

have done, we would have concluded that the non-native speakers are not sensitive to lexical bundles, whereas the native speakers are. This would have been inaccurate. The Japanese speakers were sensitive to English lexical bundles: however, they were sensitive to different ones, and a different predictor (JHSETC frequency) was required to demonstrate this in the analysis. Therefore, higher proficiency learners perhaps do not display different processing styles (e.g., formulaic vs. computational) but instead lower-level learners display the same processing style, but for different (hitherto untested) items. While this does not necessarily diminish the findings of previous studies, it is an important distinction to make.

To conclude this section, when a non-native speakers' prior linguistic experience is controlled for, we see a processing advantage even in early L2 learners. Further, and importantly, this advantage emerges for lexical bundles present in the input that may be 'un-nativelike' in their construction and frequency distribution (Study 1, Chapter 3). That is, development seems to occur in a continuous, incremental manner and the mechanisms responsible operate in reaction to input, regardless of whether it is 'nativelike' or not. Thus, it is important to take this prior experience (i.e., early input) into account if we want to accurately account for the processing of formulaic sequences in non-native speakers. This may be achieved by creating corpora from linguistic data such as school English textbooks, as in Study 1. This will be discussed in greater detail below in relation to suggestions for further research. Importantly, it is also essential for practitioners to take into account the quality of supporting materials and 'input' that we give our students. This will now be discussed in detail.

Optimising the input

As educators, there is a need to create teaching materials that promote the acquisition of particular features of language. Typically, this has been: 1. grammatical patterns and 2. words, or language related to some functional feature of language (Gilmore, 2007a, 2015). Often, very little thought seems to go into the surrounding material in which the targeted feature is embedded: which was very much the case in the Japanese English textbooks as demonstrated by the corpus study in Chapter 3, and as was discussed above. Learners are sensitive to the statistical distributions of language, and

'acquire' sequences of words from materials even if they are not explicitly taught. Thus, teachers and material makers should set as a priority the making of quality supporting materials. To put it simply, while the goals of teaching materials are important, development and experience of language are synonymous (de Bot & Larson-Freeman, 2011). We must, therefore, take into account the overall *experience*, not simply the features we wish to teach. Failing to do so risks promoting the development of a very un-nativelike language system.

A question that naturally leads from this is: what can we, as educators, do to optimise learning materials to make them more beneficial, while still fulfilling the two priorities outlined? This will the focus of this section, starting first with optimisation and promotion of 'perceived fluency'—that is, helping students to 'select' more nativelike language in order to sound more natural and fluent—and, second, in relation to improving the overall quality of materials, so that students are exposed to examples in a way that is representative of, and distributed in, a nativelike way.

Perceived Fluency

To reiterate the points made previously, a key factor in the development of a 'processing advantage' for formulaic language is the input individuals receive (Ellis, 2002b; Ellis & Wulff, 2015; Wulff, 2019). Thus, *what we put in is what we get out*, or to use the adage cited in Chapter 3, *'garbage in, garbage out'*. Developing a formulaic advantage is important because formulaic language in general, and lexical bundles specifically, are thought to aid communication because they free up cognitive resources (resulting in nativelike fluency) and are also predictable and expected by listeners (resulting in nativelike selection) (Pawley & Syder, 1983). In the current body of work, this 'input' primarily means the textbooks, or in the case of Study 4, specially designed materials. Many have emphasised that formulaic language should be an integral part of language pedagogy (e.g., Asare, 2006a, 2006b; Lewis, 1993; Okuwaki, 2012) but the norm is still to create materials around grammatical constructs, and often the actual input provided via supporting text is considered secondary (if at all) (Gilmore, 2015). This appears to be the case in the textbooks that made up the JHSETC, which mirrored findings from other studies that have found

textbooks to be inauthentic (e.g., Gilmore, 2004; Nguyen & Ishitobi, 2012; Ogura, 2008; Römer, 2004, 2005).

The corpus study (Chapter 3) demonstrated that many of the most frequent expressions in these books were disconnected from real-language use; for example, *'What time do you get up?'*, or '*how many CDs do you have?'*. The latter example, despite being the most frequent 6-word expression in the JHSETC, was neither highly frequent in a native speaker corpus (the SUBTLEXus, Brysbaert & New, 2009) nor appropriate given the reality that teenagers today do not often buy CDs (BuzzFeed, 2014; Nielsen, 2014). The language in the textbooks was created to demonstrate grammatical features of language but in doing so ignored frequency-distributions of naturally occurring language. On the basis that a processing advantage (and with it nativelike fluency and selection) develops due to usage, as a vast literature suggests it does (see Chapter 2 and above), we must ask ourselves what impact exposure to examples like these has on students' communicative abilities. Studies have shown that teaching formulaic sequences increases perceived oral fluency in second language speakers (e.g., Boers et al., 2006; Dickinson, 2012; Gatbonton & Segalowitz, 2005; Taguchi, 2007). While none of the studies in my thesis looked at this directly, the findings do have important implications for percieved fluency: to what extent does learning non-nativelike intuitions impact an English learners' communicative ability? That is, if training students with *high-frequency* chunks of English helps them to sound more fluent, does training them with the wrong, or low-frequency chunks mean they sound less fluent?

For example, if a learner is primarily taught—via explicit instruction, but also has (implicit) exposure to and practise with textbook language—to expect a formulaic yes/no response such as 'yes, I do' or 'no I can't' in response to yes/no questions, they will not be prepared for the kind of answers many native speakers may actually use (see Hockey et al., 1997). Native speakers would generally say 'yes', or even more likely 'yah', in response to a question like 'Do you want to go and see The Lion King?' To my knowledge, researchers have not looked at the impact of overusing—or expecting—low frequency sequences of words in conversation ('yes I do' vs. 'yah'). However, some studies have looked at *non*-nativelike sequences, and these can give us some insight into the impact of unexpected speech. Millar (2010) explored the

processing of malformed collocations (derived from a corpus of English-learning students) in native-speakers using a self-paced reading experiment, and found "persuasive evidence that learner deviation from target language formulaicity places an increased processing burden on native speaker addressees" (p. 142). It has to be emphasised that Millar's study presented an extreme case: some of the collocations studied were plausible, but low frequency (e.g., *'best partner'* vs. *'ideal partner'*), whereas others were arguably ungrammatical ('*marriage life'* vs. *'married life*'). However, from his findings we can hypothesise that a similar burden may be placed on learners who expect certain formulaic responses (*'yes, I am*', *'I'm fine thank you and you'*, etc.). In other words, the Japanese students' teaching materials have led them to expect formulaic sequences that do not occur in nativelike language. When they use non-nativelike sequences, this will likely impact their perceived fluency, place a burden on any native-speaker interlocutors, as well as resulting in a processing cost when they encounter authentic native formulaic language that they have not encountered.

Relevant to this discussion, a similar point has also often been made by researchers advocating for the teaching of pragmatics alongside language (e.g., Taguchi, 2008). By way of example to illustrate this, a Japanese learner of English, upon arriving in the UK, might ask a newly made acquaintance over a beer, *'What time do you get up?'*. While this lexical bundle will have been encountered numerous times in textbooks and the classroom, which may seem authoritative, learners likely will not understand that people do not ask questions such as this except in very specific circumstances (i.e., in this case when it is essential to know when they get up, so that breakfast is ready on time). As discussed in Chapter 2, Hymes (1972) states that along with being possible, well-formed, and appropriate, utterances should also be 'performed', or actually 'done' in the real world. Clearly, the findings from the corpus study, in Chapter 3, violate this principle: much of the language included in the textbooks is not 'done'. On the contrary, the analysis demonstrated that it is fairly unlikely language (e.g., *'How many CDs do you have?'*) and therefore may be a mark of *non*-nativelikeness and lacking in fluency. Thus, while the findings from Studies 2, 3 and 4 indicate that encountering these sequences yields a processing advantage, arguably learning the language from these books does very little—if anything—to

promote the kind of perceived fluency discussed in the literature (e.g., Boers et al., 2006; Dickinson, 2012; Gatbonton & Segalowitz, 2005; Taguchi, 2007) and, as highlighted in this section, potentially even hinders it. This points to the need for further research.

Optimising teaching materials for exposure

As was discussed in Chapters 4 and 5, the lexical bundles selected for use in the experiments were not explicitly taught to Japanese learners of English, but rather were simply present in their materials; thus, they emerged from texts. In one sense, this repetitiveness can be seen as beneficial to learners, because there are more opportunities for the sequences to be acquired, and indeed, the need to repeat—or 'recycle' language targeted for instruction is important. However, the analysis of the textbook lexical bundles (Chapter 3) demonstrated that the language Japanese secondary school students are exposed to is 'non-nativelike' both in terms of the quality of lexical bundles (as described in the previous section) and their *quantity*.

An important question in the second language learning literature has been how many occurrences are needed to incidentally acquire formulaic language (this was discussed in Chapter 6). Durrant and Schmitt (2010); Webb et al. (2013) both asked whether multiple exposures to collocations in texts (in sentences in Durrant & Schmitt, 2010 and in longer texts in Webb et al. 2013) would lead to learning them. Both studies found that they did, even if, in the case of Webb et al. (2013), one of the conditions required as many as 15 repetitions for it to be taken up. A recent study from Conklin and Carrol (in submission) found a processing advantage emerged for novel binomials (*wires and pipes* compared with *pipes and wires*) in native speakers after five repetitions of an item in a story context in an eye-tracking study. Such findings are in line with results from Study 4 (Chapter 6), as well as those from Study 2 (Chapter 4) and 3 (Chapter 5) and are important because they highlight the importance of repetition of the language we want students to learn, but also show that the emergence of a processing advantage happens automatically (Hasher & Zacks, 1984), regardless of learning goals (or the quality of content).

Pellicer-Sánchez and Boers (2019) point out that, "non-manipulated authentic

texts are extremely unlikely to expose readers to, for example, as many as 15 instances of the same collocation in a short time span" (pp. 156-157), which would allow them to be learned. In the current research, the textbooks cannot be considered to be optimised in the sense that the lexical bundles are not purposely included or targeted for instruction. In addition, they cannot be considered 'authentic', nor is the way students and teachers use them the same as authentic texts are typically used. As a former secondary school English teacher in Japan, I know that students do not simply read textbooks once: rather, each section tends to be read multiple times, studied, listened to via audio recordings and/or the teacher reads it aloud, as well as having elements produced by rote in 'repeat after me' style exercises. Thus, the raw-frequency scores of lexical bundles in the JHESTC can be considered a very conservative indicator of the actual number of times students may encounter the same sequence. This means there is ample opportunity to learn lexical bundles incidentally – as evidenced by the emergence of a processing advantage for them (demonstrated in Chapters 4 and 5). Thus, even if as many as 15 instances is required, the materials themselves do not necessarily need to contain 15 occurrences.

To come back to a concluding statement from Study 1 (Chapter 3), textbook designers and educators do not necessarily have to make large changes to the textbooks. Rather, by consulting corpora—something which textbooks designers do not routinely do (Gilmore, 2015)—much can be done to increase the utility of materials. As I demonstrated in the discussion of Chapter 3, this can be as simple as checking that examples chosen to demonstrate a particular point (i.e., type of construction) are high frequency, and then engineering dialogues and texts to include high-frequency lexical bundles. To give a concrete example, while *'how many CDs do you have'* is not high frequency, if we would like to teach *'how many something do you have'*, higher utility sequences could be created. Looking at the SUBTLEXus, it would suggest inserting a word such as 'children' (*'how many children do you have*', 10 occurrences in the SUBTLEXus). Admittedly this would not yield a phrase that is likely useful to junior high school students. However, something like, 'brothers and sisters' (which, as a complete string, *'how many brothers and sisters do you have?'*) would seem like a reasonable phrase for junior high school students. While it is relatively infrequent with only one hit in the SUBTLEXus, the bundle *brothers and*

*sister*s occurs 177 times.

Through the selection of well-chosen supporting materials, we can also supplement teaching materials to give students additional exposure to high-frequency items we would like them to acquire. Crucially, Gilmore (2011) demonstrates that learners prefer authentic materials and find them more motivating. However, supplementary materials (that need to be developed by educators) are unlikely to provide sufficient exposure for their acquisition because classroom time is limited (Siyanova-Chanturia & Pellicer-Sánchez, 2019) and teachers are often focused on getting through the required textbook material (e.g., Browne & Wada, 1998; Creese, 2008). One alternative, might be to prepare supplementary materials that give students focused input of key items, followed by more informal tasks, such as watching a film or TV show where the target formulaic language occurs, as a way to reinforce learning with larger quantities of input. For example, the chunk *what are you doing* is very frequent in the SUBTLEXus (14,531 - 284.92/per million) and was identified as one of the most frequent 4-word lexical native-speaker bundles in Study 1 (Chapter 3). It is also relatively frequent in the JHSETC (18 occurrences), but not enough so to warrant inclusion in the top-10 4-word lexical bundles from these materials (Chapter 3). Another relatively common lexical bundle in the JHSETC was *what do you mean* (10 occurrences). In order to provide additional exposure to these items, while also reinforcing their usage in natural language, students could be asked to watch the film 'Toy Story' which includes 5 occurrences of *what are you doing* and 4 of *what do you mean*. This is a very simple example, but with planning it should be possible to engineer (explicitly taught) materials in a way that maximises (incidental) learning from more natural exposure. Although only anecdotal at this point, I have been using such a 'two-track' approach successfully with (high-proficiency) learners of English as a second language, and this will be the focus of future research.

Limitations and directions for future work

The studies presented in this thesis have several limitations, and it is important to acknowledge these. Further, the limitations, as well as the findings in this thesis, point to potential avenues for future research.

One limitation, which also points to a greatly under-explored area, is that only one of the studies in this thesis was with young, beginning English learners. This population is important because we have little knowledge about the initial stages of acquisition by young learners. In addition, early L2 knowledge provides the foundation for later acquisition. Having a better understanding of the early stages of learning might help to explain later performance. Indeed, as demonstrated in Study 3 (Chapter 4), intuitions developed during the initial stage of learning may stay with learners long-term. Ideally more studies should be carried out with students at the secondary school level. However, this is often extremely challenging, as became apparent over the course of carrying out Study 2 (Chapter 4). Time to work with students at secondary schools is often very limited; teachers do not want to give up already limited and valuable teaching time and students either do not want, or more likely, are unable to, participate easily outside of school hours. For the thesis, this restricted what research could (and could not) be carried out with this very interesting population. However, many questions have been left unanswered. For example, would secondary school students perform like the adults in the offline-naturalness rating task? This seems very likely, given that JHSETC frequency was the best predictor of rating behaviour in the adults (Chapter 5); however, we cannot conclude this directly from the current data. A different, but related, avenue would be to explore whether adults show the same processing behaviour for textbook derived lexical bundles as the secondary school students did. If so, this would back up the findings presented in Chapter 5 and provide additional evidence that intuitions last long-term.

Another limitation is that it is impossible to be sure what English input the students had outside of the classroom. We can ask about it—and did for each experiment with a language-background questionnaire—but we can never be sure about the accuracy of this self-report. Crucially, recent research has demonstrated the impact of L2 language input outside of the classroom. De Wilde, Brysbaert, and Eyckmans (2019), in a large-scale study exploring the impact of out-of-school exposure on vocabulary learning, found that young learners can and do build knowledge of vocabulary in a foreign language through exposure alone. Puimège and Peters (2019) showed significant variation in the amount of outside-exposure Flemish children have to English between age groups (10 to 12), suggesting that children are

exposed to more English as they get older, and concluded that they pick up words though outside-exposure. Notably, Puimège and Peters (2019) found a positive relationship between gaming and TV viewing, with gaming significantly predicting word knowledge in a meaning recall test. However, these studies were from a very different cultural context. Arguably children in the European context have far more opportunities to encounter English given that much of their media is in English, while this is generally not the case in Japan where most media is either produced in-country, or in the case of foreign films and TV, dubbed. Thus, while Japanese children are much less likely to encounter English outside of the classroom, the recent research (De Wilde et al., 2019) indicates the important role this can have on the development of foreign language knowledge, making it imperative to establish, with some degree of certainty, the contexts in which participants encounter and use English. It also, as already discussed in the previous section, points to a need to explore ways in which teachers can effectively encourage outside-exposure (for example, by setting watching films with particular items in English as homework).

An important contribution of this thesis was showing that young, beginner learners of English, exhibit a strong processing advantage for highly repetitive lexical bundles even when they are not explicitly taught. Demonstrating this was possible because a measure of exposure (textbook frequency) was established for the group under investigation. Thus, exposure has processing implications, which makes it imperative that we gain clearer insight into the varied types and amount of exposure that learners have. In reality, little is known about the kind of exposure to English most learners have, and few corpus studies of English learner materials exist, and none to my knowledge have looked the actual lexical bundles that exist in teaching materials. Further, studies that do explore learner input (e.g., Koya, 2004; Mcaleese, 2013; Tsai, 2014) have tended to take the same approach as psycholinguistic processing studies: they start with pre-defined lists of formulaic sequences from native speaker corpora, and look for occurrences of them in textbooks. They have not, however, considered the naturally occurring formulaic sequences that *are* in them, which sadly might not match the nativelike formulaic language that one would purposefully attempt to identify. Future research should aim to compile corpora of 'input' matched to the learners tested in behavioural studies. Although, as has been

stated elsewhere in this thesis, ideally the input that learners receive would be better matched to native-speaker language and appropriate for the needs of the learners.

Finally, the current study focused only on learners from a single context: that of Japan. This was a particularly appropriate context in which to situate the current study, not the least because it is one with which I am intimately familiar, but also because the Japanese Ministry of Education (MEXT) dictates the content of textbooks to publishing companies (Ministry of Education Culture Science Sports and Technology (MEXT), 2011) and requires all schools to use one of the six alternatives. This makes it relatively easy to address questions about language exposure, because it is possible to gain a clear picture of exactly what the vast majority of Japanese students have been exposed to. In some contexts (such as many European secondary schools) no such official textbooks exist, while in other countries they do (e.g., Taiwan, which has a system very similar to that of Japan and is the topic of an in-progress research project that fell outside of the scope of this thesis). An interesting avenue for future research would be to replicate the current studies with learners from other countries. Further, as mentioned briefly above, more research is necessary to ascertain the nature of the input learners receive, not just in their early years, but as higher-level, continuing ESL and EFL learners. With some effort, it should be possible to create corpora informed by the input identified by studies such as those conducted by De Wilde et al. (2019) and Puimège and Peters (2019) which can be combined with measures of formal learning (such as the JHSETC, Chapter 3). In doing so, we may better understand and explore the usage-based development of L2 English.

Final Conclusions

The studies presented in this thesis are important because they clearly suggest careful consideration needs to be given to the quality of the materials we present students if they are to develop nativelike language intuitions. Japanese Junior high school students' materials expose them to language that is distributed in a non-nativelike way and is therefore presenting learners with an experience of English that is fundamentally different to what they may encounter in the real world (Chapter 3). Early language learners acquire the formulaic sequences from this kind input, even if

they are not explicitly taught (Chapter 4, also Chapter 6), and the intuitions they develop from this appear to stay with them and affect processing long-term (Chapter 5). Further, the research demonstrated that materials can be optimised with lexical bundles to facilitate their learning via repetition and enhancement (Chapter 4), which would allow learners to benefit more from them, but not actually require a major shift in the kind of materials teachers use.

The foundation of usage-based approaches is that development happens in reaction to input. Therefore, the quality of an individual's language is defined by the input he/she receives. Measures should, then, be taken to ensure the input given to learners is as high quality as possible. In a nutshell, the findings from this study can be summarised thus: *materials matter*.

References

Anderson, F. (1993). The enigma of the college classroom: Nails that don't stick up. In P. Wadden (Ed.), *A handbook for teaching English at Japanese colleges and universities* (pp. 101-110). Hong Kong: Oxford University Press.

Anderson, J. R. (1992). Automaticity and the ACT* Theory. *The American Journal of Psychology, 105*(2), 165-180.

Andrade, M., & Williams, K. (2009). Foreign Language Learning Anxiety in Japanese EFL University Classes: Physical, Emotional, Expressive, and Verbal Reactions. *Sophia Junior College Faculty Journal, 29*, 1-24.

Anthony, L. (Producer). (2014). AntConc (Version 3.4.3) [Computer Software]. Retrieved from http://www.laurenceanthony.net/

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language, 92*, 265–280.

Arnon, I., & Priva, U. C. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech, 56*(3), 349-371.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*(1), 67-82.

Asare, K. (2006a). The Treatment of Lexis in the Japanese University Classroom: Classroom Analysis. *Fukuoka kokusai daigaku kiyo, 38*(3), 855-886.

Asare, K. (2006b). The treatment of lexis in the Japanese university classroom: vocabulary issues. *Fukuoka kokusai daigaku kiyo, 37*(4), 1171-1197.

Aspinall, R. W. (2006). Using the paradigm of ‘small cultures’ to explain policy failure in the case of foreign language education in Japan. *Japan Forum, 18*(2), 255-274.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon, 5*(3), 436-461.

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.

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial Explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech, 56*(3), 329-347.

Babaei, S., Taleb Najafabadi, F., & Fotovatnia, Z. (2015). Processing of lexical bundles by Persian speaking learners of English. *The Journal of Teaching Language Skills (JTLS), 6*(4), 1-18.

Bailey, K. M. (1999). *Washback in language testing (TOEFL Monograph Series Report No. 15)*. Princeton, NJ: Educational Testing Service.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*(3), 241-248.

Barlow, M., & Kemmer, S. (2000). *Usage Based Models of Language*. United States: CSLI.

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. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bednarek, M. (2012). “Get us the hell out of here”: Key words and trigrams in fictional television series. *International Journal of Corpus Linguistics, 17*(1), 35-63.

Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing, 8*(4), 243-257.

Biber, D. (2006). *University Language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins Publishing Company.

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

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., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Essex: Pearson Education Limited.

Bobrow, S. A., & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition, 1*(3), 343-346.

Bod, R., Hay, J., & Jannedy, S. (2003). Probabilistic Linguistics. Cambridge, Massachusetts: The MIT Press.

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

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: putting a lexical approach to the test. *Language Teaching Research, 10*(3), 245–261.

Bolinger, D. (1975). Meaning and Memory. *Forum Linguisticum, 1*, 2-14.

Brants, T., & Franz, A. (2006). *Web 1T 5-gram Version 1 LDC2006T13 DVD.* Philadelphia: Linguistic Data Consortium.

Breeze, R. (2013). Lexical bundles across four legal genres. *International Journal of Corpus Linguistics, 18*(2), 229-253.

Brown, J. D. (2000). University entrance examinations: Strategies for creating positive washback on English language teaching in Japan. *Shiken: JALT Testing & Evaluation SIG Newsletter, 3*(2), 2-7.

Browne, C., & Wada, M. (1998). Current Issues in High School English Teaching in Japan: An Exploratory Survey. *Language, Culture and Curriculum, 11*(1), 97-112.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990.

Butler, Y. G., & Iino, M. (2005). Current Japanese reforms in English language education: The 2003 "Action Plan". *Language Policy, 4*(1), 25-45.

BuzzFeed. (2014). Who Buys Music Anymore? A Statistical Inquiry. Retrieved from https://www.buzzfeed.com/reggieugwu/who-buys-music-anymore-a-statistical-inquiry?utm_term=.wyjzEAY42#.dvvxA96o4

Bybee, J. (1998). The Emergent Lexicon. *Proceedings of the Chicago Linguistic Society 34*, 421-455.

Bybee, J. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language, 82*(4), 711-733.

Bybee, J. (2013). Usage-based Theory and Exemplar Representations of Constructions. In T. Hoffman & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 49-69). Oxford: Oxford University Press.

Bybee, J., & Hopper, P. (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLoS One, 5*(6), 1-8.

Chen, L. (2010). An Investigation of Lexical Bundles in ESP Textbooks and Electrical Engineering Introductory Textbooks. In D. Wood (Ed.), *Perspectives on formulaic language* (pp. 107-125). London: Continuum.

Cobb, T. (n.d.). Web Vocabprofile, an adaptation of Heatley, Nation & Coxhead's (2002) Range. Retrieved from https://www.lextutor.ca/vp/eng/

Conklin, K. (2020). Processing single- and multi-word items. In S. Webb (Ed.), *Routledge Handbook of Vocabulary Studies* (pp. 174-188). New York: Routledge.

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.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423.

Cowan, N. (2001). The Magical Number 4 in Short-term Memory: A Reconsideration of Mental Storage Capacity. *Behavioral and Brain Sciences, 24*(1), 87-114.

Cowie, A. P. (1998). Phraselogical dictionaries: some east-west comparions. In A. P. Cowie (Ed.), *Phrasology: theory, analysis and applications* (pp. 209-228). Oxford: Oxford University Press.

Creese, A. (2008). Supporting Talk? Partnership Teachers in Classroom Interaction. *International Journal of Bilingual Education and Bilingualism, 9*(4), 434-453.

Crick, F. H. C. (1979). Thinking about the brain. *Scientific American, 241*(2), 219-232.

Cuetos, F., Glez-Nosti, M., Barbón, A., & Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica, 32*(2), 133-143.

Cutrone, P. (2009). Overcoming Japanese EFL Learners' fear of speaking. *University of Reading language studies working papers, 1*, 55-63.

Cutrone, P. (2010). Helping Japanese ESL/EFL Learners Overcome Difficulties in Intercultural Communication. *Journal of the Faculty of Global Communication, University of Nagasaki, 11*, 11-22.

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

de Bot, K., & Larson-Freeman, D. (2011). Researching SLD from a DST perspective. In M. H. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to Second Language Development: Methods and techniques* (pp. 5-24). Amsterdam: John Benjamins.

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2019). Learning English Through Out-of‐School Exposure: How Do Word‐Related Variables and Proficiency Influence Receptive Vocabulary Learning? *Language Learning*, 1-33.

DeKeyser, R. (2015). Skill Acquisition Theory. In B. Van Patten & J. Williams (Eds.), *Theories in Second Language Acquisition* (Second Edition ed., pp. 94-112): Taylor & Francis.

Dickinson, P. (2012). Improving second language academic presentations with formulaic sequences. *Bulletin of Niigata University of International and Information Studies Department of Information Culture, 15*, 25-36.

Dimitropoulou, M., Dunabeitia, J. A., Aviles, A., Corral, J., & Carreiras, M. (2010). Subtitle-based word frequencies as the best estimate of reading behavior: the case of greek. *Frontiers in Psychology, 1*, 1-218.

Doyon, P. (2000). Shyness in the Japanese EFl class: Why it is a problem, what it is, what causes it, and what to do about it. *The Language Teacher, 24*(1), 11-16.

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

Ebbinghaus, H. (1885). *Memory: A Contribution to Experimental Psychology. Translated by Henry A. Ruger and Clara E. Bussenius. (1913)*. New York: Teachers College Columbia.

Edmonds, A. (2013). Conventional Expressions. *Studies in Second Language Acquisition, 36*(1), 69-99.

Ellis, N. C. (1996). Sequencing in SLA: Phonological Memory, Chunking, and Points of Order. *Studies in Second Language Acquisition, 18*, 91-126.

Ellis, N. C. (2001). Memory for Language. In P. J. Robinson (Ed.), *Cognition and Second Language Acquisition* (pp. 33-68). New York: Cambridge University Press.

Ellis, N. C. (2002a). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143-188.

Ellis, N. C. (2002b). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition, 24*(2), 297-339.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In Doughty & Long (Eds.), *Handbook of Second Language Acquisition* (pp. 63-103). Oxford: Blackwell.

Ellis, N. C. (2006). Language Acquisition as Rational Contingency Learning. *Applied Linguistics, 27*(1), 1-24.

Ellis, N. C. (2009). Optimizing the Input: Frequency and Sampling in Usage-based and Form-focussed Learning. In M. Long & C. Dought (Eds.), *The Handbook of Second and Foreign Language Teaching* (pp. 139-158). Oxford: Blackwell Publishing.

Ellis, N. C. (2012). Formulaic Language and Second Language Acquisition: Zipf and the Phrasal Teddy Bear. *Annual Review of Applied Linguistics, 32*, 17-44.

Ellis, N. C., Frey, E., & Jalkanen, I. (2009). The Psycholinguistic reality of collocation and semantic prosody (1). In U. Römer & R. Schulze (Eds.), *Exploring the Lexis-grammar interface* (pp. 89-114). Amsterdam: John Benjamins.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2013). Usage-Based Language: Investigating the Latent Structures That Underpin Acquisition. *Language Learning, 63*(1), 25-51.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *Tesol Quarterly, 42*(3), 375-396.

Ellis, N. C., & Wulff, S. (2015). Usage-Based Approaches to SLA. In *Theories in Second Language Acquisition* (Second Edition ed., pp. 75-93): Taylor & Francis.

Elman, J. L. (2009). On the Meaning of Words and Dinosaur Bones: Lexical Knowledge Without a Lexicon. *Cognitive Science, 33*(4), 547–582.

Elman, J. L. (2011). Lexical knowledge without a lexicon? *The Mental Lexicon, 6*(1), 1-33.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29-62.

Eskildsen, S. W., & Cadierno, T. (2007). Are recurring multi-word expressions really syntactic freezes? Second Language Acquisition from the perspective of Usage-Based Linguistics. In M. Nenonen & S. Niemi (Eds.), *Collocations and idioms 1: Papers from the First Nordic Conference on Syntactic Freezes*. Joensuu, Finland: Joensuu University Press.

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, UK: Longman.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking Communicative Language Teaching: A Focus on Access to Fluency. *The Canadian Modern Language Review / La revue canadienne des langues vivantes, 61*(3), 325-353.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*(2), 256-281.

Gewirtz, P. (1996). On "I Know It When I See It". *The Yale Law Journal, 105*(4), 1023-1047.

Gilmore, A. (2004). A comparison of textbook and authentic interactions. *English Language Teaching Journal, 59*(4), 363-374.

Gilmore, A. (2007a). Authentic materials and authenticity in foreign language learning. *Language Teaching, 40*(02), 97-118.

Gilmore, A. (2007b). *Getting real in the language classroom: Developing Japanese students' communicative competence with authentic materials.* University of Nottingham, Retrieved from http://etheses.nottingham.ac.uk/1928/1/478936.pdf

Gilmore, A. (2011). "I Prefer Not Text": Developing Japanese Learners' Communicative Competence with Authentic Materials. *Language Learning, 61*(3), 786-819.

Gilmore, A. (2015). Research into practice: The influence of discourse studies on language descriptions and task design in published ELT materials. *Language Teaching, 48*(4), 506-530.

Godfroid, A., Boers, F., & Housen, A. (2013). AN EYE FOR WORDS: Gauging the Role of Attention in Incidental L2 Vocabulary Acquisition by Means of Eye-Tracking. *Studies in Second Language Acquisition, 35*(3), 483-517.

Goldberg, A. E. (2006). *Constructions at Work: the nature of generalization in language*. Oxford: Oxford University Press.

Gorsuch, G. J. (2001). Japanese EFL Teachers' Perceptions of Communicative, Audiolingual and Yakudoku Activities: The Plan Versus the Reality. *Education Policy Analysis Archives, 9*(10), 1-27. Retrieved from http://epaa.asu.edu/ojs/article/view/339/465

Grant, L. E. (2005). Frequency of 'core idioms' in the British National Corpus (BNC). *International Journal of Corpus Linguistics, 10*(4), 429–451.

Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics, 18*(1), 109-136.

Guest, M. (2008). A Comparative Analysis of the Japanese University Entrance Senta Shiken Based on a 25-Year Gap. *The Japan Association for Language Teaching Journal, 30*(1), 85-103.

Hagerman, C. (2009). English language policy and practice in Japan. *Osaka Jogakuin University, Departmental Bulletin Paper, 6*, 47-64.

Hakuta, K. (1974). Prefabricated Structure in Patterns and the Emergence Structure of Second Language Acquisition. *Language Learning, 24*(2), 287-297.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: the case of frequency of occurrence. *American Psychologist, 39*(12), 1372-1388.

Hashimoto, K. (2009). Cultivating "Japanese Who Can Use English": Problems and Contradictions in Government Policy. *Asian Studies Review, 33*(1), 21-42.

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

Higham, P. A. (1997). Chunks are not enough: The insufficiency of feature frequency-based explanations of artificial grammar learning. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 51*(2), 126-138.

Hinenoya, K., & Gatbonton, E. (2000). Ethnocentrism, Cultural Traits, Beliefs, and English Proficiency: A Japanese Sample. *The Modern Language Journal, 84*(2), 225-240.

Hockey, B. A., Rossen-Knill, D., Spejewski, B., Stone, M., & Isard, S. (1997). Can you predict answers to Y/N questions? Yes, No and Stuff. In *Proceedings of Eurospeech 1997* (pp. 2267–2270). Rhodes, Greece.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring Second Language Vocabulary through Reading. *Reading in a Foreign Language, 11*(2), 207-223.

Howarth, P. (1998). The phraseology of learners' academic writng. In A. P. Cowie (Ed.), *Phrasology: theory, analysis and applications* (pp. 161-186). Oxford: Oxford University Press.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth: Penguin.

Inquisit 4.0.7.0 [Computer software]. (2014). Seattle, WA: Millisecond Software.

Ishihara, N., & Cohen, A. D. (2010). *Teaching and Learning Pragmatics: where language and culture meet*. Edignburgh: Pearson.

Isobe, Y. (2011). Representation and processing of formulaic sequences in L2 mental lexicon: How do Japanese EFL learners process multi-word expressions. *JACET Kansai Journal, 13*, 38-49.

Jiang, N., & Nekrasova, T. M. (2007). The Processing of Formulaic Sequences by Second Language Speakers. *The Modern Language Journal, 91*(3), 433-445.

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. Meaning Overrides Frequency in Idiomatic and Compositional Multiword Chunks. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society.* (pp. 692-697). Austin: Cognitive Science Society.

Jurafsky, D. (2003). Probabilistic modeling in psycholingustics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 39-95). Cambridge, Massachusetts: MIT Press.

Kapatsinski, V., & Radicke, J. (2009). Frequency and the Emergence of Prefabs: Evidence from Monitoring. In R. Corrigan, E. Moravcsik, H. Ouali, & K. Wheatley (Eds.), *Formulaic Language. Vol. II: Acquisition, loss, psychological reality, functional explanations* (pp. 499-520). Amsterdam: John Benjamins.

Kikuchi, K. (2006). Revisiting English Entrance Examinations at Japanese Universities after a Decade. *The Japan Association for Language Teaching Journal, 28*(1), 77-96.

Kikuchi, K., & Browne, C. (2009). English Educational Policy for High Schools in Japan : Ideals vs. Reality. *Regional Language Centre Journal, 40*(2), 172-191.

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.

Kim, Y. (2006). Effects of Input Elaboration on Vocabulary Acquisition Through Reading by Korean Learners of English as a Foreign Language. *Tesol Quarterly, 40*(2), 341-373.

Kimura, Y., Nakata, Y., & Okumura, T. (2001). Language Learning Motivation of EFL Learners in Japan - A Cross-Sectional Analysis of Various Learning Milieus. *The Japan Association for Language Teaching Journal, 23*(1), 47-68.

King, J. (2012). Silence in the Second Language Classrooms of Japanese Universities. *Applied Linguistics, 34*(3), 325-343.

Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *English Language Teaching Journal, 59*(4), 322-332.

Kornai, A. (2008). *Mathematical Linguistics*. New York: Springer Verlag.

Koya, T. (2004). Collocation Research Based on Corpora Collected from Secondary School Textbooks in Japan and in the UK. *Dialogue, 3*, 7-18.

Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, processing and use* (pp. 37-54). Amsterdam: John Benjamins.

Kuiper, K., & Haggo, D. (1984). Livestock auctions, oral poetry, and ordinary language. *Language in Society, 13*(2), 205-234.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1-26.

Kwan, C. H. (2002). Why the Japanese are Poor English Speakers - A Proposal to Reform English Language Instruction in Japan. Retrieved from http://www.rieti.go.jp/en/columns/a01_0053.html

Kwon, Y.-E., & Lee, E.-J. (2014). Lexical Bundles in the Korean EFL Teacher Talk Corpus: A Comparison Between Non-native and Native English Teachers. *The Journal of Asia TEFL, 11*(3), 73-103.

Kyriacou, M., Conklin, K., & Thompson, D. (2019). Passivizability of Idioms: Has the Wrong Tree Been Barked Up? *Language and Speech, Online First*, 1-32.

Lancker, D. R. V., & Kempler, D. (1987). Comprehension of Familiar Phrases by Left- but Not by Right-Hemisphere Damaged Patients. *Brain and Language, 32*(2), 265-277.

Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford: Oxford University Press.

Laufer, B., & Girsai, N. (2008). Form-focused Instruction in Second Language Vocabulary Learning: A Case for Contrastive Analysis and Translation. *Applied Linguistics, 29*(4), 694-716.

Leedham, M. (2012). Combining intuition with corpus linguistic analysis: A study of marked lexical chunks in four Chinese students' undergraduate assignments. *Nordic Journal of English Studies, 11*(3), 155-187.

Lewis, M. (1993). *The lexical approach*. Hove: Language Teaching Publications.

Logan, G. D. (1988). Toward an Instance Theory of Automatization. *Psychological Review, 95*(4), 492-527.

Logan, G. D. (2005). Attention, Automaticity and Executive Control. In A. F. Healy (Ed.), *Experimental Cognitive Psychology and Its Applications* (pp. 129-153). Washington: American Psychological Assotiation.

Logan, G. D., & Stadler, M. A. (1991). Mechanisms of performance improvement in consistent mapping memory search: Automaticity or strategy shift? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(3), 478-496.

Loveday, L. J. (1986). Japanese Sociolinguistics: An Introductory Survey. *Journal of Pragmatics, 10*(3), 287-326.

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition, 122*(3), 346-362.

Mcaleese, P. (2013). *Investigating Multi-Word Items in a Contemporary ELT Course Book.* Paper presented at the Japan Association for Language Teaching 2012 Conference Proceedings, Tokyo.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book*. Oxon: Routledge.

Mclaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: an Information-processing perspective. *Language Learning, 33*(2), 135-158.

Meunier, F. (2012). Formulaic Language and Language Teaching. *Annual Review of Applied Linguistics, 32*(2), 111-129.

Millar, N. (2010). The Processing of Malformed Formulaic Language. *Applied Linguistics, 32*(2), 129-148.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review, 63*(2), 81-97.

Ministry of Education Culture Science Sports and Technology (MEXT). (2003). Action Plan to Cultivate "Japanese with English abilities". Retrieved from http://warp.ndl.go.jp/info:ndljp/pid/286794/www.mext.go.jp/b_menu/houdou/15/03/03033101/001.pdf

Ministry of Education Culture Science Sports and Technology (MEXT). (2011). Course of Study Section 9 Foreign Languages. Retrieved from http://www.mext.go.jp/english/elsec/1303755.htm

Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or Rule? Exploring the Role of Formulaic Language in Classroom Foreign Language Learning. *Language Learning, 48*(3), 323-363.

Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French L2: A Basis for Creative Construction? *Studies in Second Language Acquisition, 21*(1), 40-80.

Nattinger, J. R., & DeCarrucio, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Nelson, R. (2018). How 'chunky' is language? Some estimates based on Sinclair's Idiom Principle. *Corpora, 13*(3), 431-460.

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

Nguyen, H. t., & Ishitobi, N. (2012). Ordering Fast Food: Service Encounters in Real-Life Interaction and in Textbook Dialogs. *Japan Association for Language Teaching Journal, 34*(2), 151-186.

Nielsen. (2014). 2014 Nielsen Music Report. Retrieved from http://s0.thejazzline.com/tjl/uploads/2015/03/nielsen-2014-year-end-music-report-us.pdf

Nishino, T. (2008). Japanese secondary school teachers' beliefs and practices regarding Communicative Language Teaching: An exploratory Survey. *Japan Association for Language Teaching Journal, 30*(1), 27-50.

Nishino, T. (2011). Japanese High School Teachers' Beliefs and Practices Regading Communicative Language Teaching. *JALT Journal, 33*(2), 131-156.

Nishino, T., & Watanabe, M. (2008). Communication-oriented policies versus classroom realities in Japan. *Tesol Quarterly, 42*(1), 133-138.

Northbrook, J., & Conklin, K. (2018a). Is What You Put in What You Get Out? - Textbook-derived Lexical Bundle Processing in Beginner English Learners. *Applied Linguistics, 40*(5), 816-833.

Northbrook, J., & Conklin, K. (2018b). "What are you talking about?" An analysis of lexical bundles in Japanese junior high school textbooks. *International Journal of Corpus Linguistics, 23*(3), 311-334.

Nozaki, K. N. (1993). The Japanese student and the foreign teacher. In P. Wadden (Ed.), *A Handbook for Teaching English at Japanese Colleges and Universities* (pp. 27-33). Hong Kong: Oxford University press.

Nunan, D. (1991). *Language teaching methodology: A textbook for teachers*. Hertfordshire: Prentice Hall.

O'Donnell, K. (2005). Japanese Secondary English Teachers: Negotiation of Educational Roles in the Face of Curricular Reform. *Language, Culture and Curriculum, 18*(3), 300-315.

Ogura, F. (2008). Communicative competence and senior high school oral communication textbooks in Japan. *The Language Teacher, 32*(12), 3-8.

Okuwaki, N. (2012). Formulaicity of Language: Its Pervasiveness and the Processing Advantage in Language Use. *The Tsuru University Review, 75*, 1-11.

Okuwaki, N. (2014). The Development of Formulaic Language in L2. *The Tsuru University Review, 79*, 43-57.

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied, 14*(2), 101-117.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards (Ed.), *Language and Communication* (pp. 191-225): Longman.

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

Pellicer-Sánchez, A., & Boers, F. (2019). Teaching and Learning of Formulaic Language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.),

Understanding Formulaic Language: a second Language Acquisition Perspective (pp. 153-173). New York: Routledge.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language, 22*(1), 31–55.

Peters, E. (2012). Learning German formulaic sequences: the effect of two attention-drawing techniques. *The Language Learning Journal, 40*(1), 65-79.

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.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1-28.

Pinker, S. (1995). *The Language Instinct*. New York: Harper Collins.

Pinker, S. (1999). *Words and Rules: the ingrediants of langaguage*. New York: Basic Books.

Pinker, S., & Ullman, M. T. (2002). The Past-Tense Debate: The past and future of the past tense. *Trends in Cognitive Sciences, 6*(11), 456-463.

Puimège, E., & Peters, E. (2019). Learners' English Vocabulary Knowledge Prior to Formal Instruction: The Role of Learner-Related and Word-Related Variables. *Language Learning, 69*(4), 943-977.

Rammell, C. S., Sidtis, D. V. L., & Pisoni, D. B. (2017). Perception of formulaic and novel expressions under acoustic degradation. *The Mental Lexicon, 12*(2), 234-262.

Rapley, D. J. (2010). Learning to speak English: Japanese junior high school student views. *The Language Teacher, 34*(6), 33-40.

Read, J., & Nation, P. (2004). Measurement of formulaic sequences. In B. Harley & J. H. Hulstijn (Eds.), *Formulaic Sequences Acquisition, processing and use* (Vol. 9, pp. 23-36). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Robinson, P. J., & Ha, M. A. (1993). Instance Theory and Second Language Rule Learning Under Explicit Conditions. *Studies in Second Language Acquisition, 15*(4), 413-438.

Römer, U. (2004). Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and Language Learners* (pp. 151-168). Amsterdam: John Benjamins.

Römer, U. (2005). *Progressives, Patterns, Pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics*. Amsterdam: John Benjamins Publishing Company.

Rott, S. (1999). The Effect of Exposure Frequency on Intermediate Language Learners' Incidental Vocabulary Acquisition and Retention through Reading. *SSLA, 21*(2), 589–619.

Sato, K., & Kleinsasser, R. C. (2004). Beliefs, practices, and interactions of teachers in a Japanese high school English department. *Teaching and Teacher Education, 20*(8), 797-816.

Schmidt, R. (1992). Psychological Mechanisms Underlying Second Language Fluency. *Studies in Second Language Acquisition, 14*(4), 357-385.

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Basingstoke: Palgrave.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In B. Harley & J. H. Hulstijn (Eds.), *Formulaic Sequences Acquisition, processing and use* (Vol. 9, pp. 127-152). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In B. Harley & J. H. Hulstijn (Eds.), *Formulaic Sequences Acquisition, processing and use* (Vol. 9, pp. 173-190). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Seargeant, P. (2009). *The Idea of English in Japan: Ideology and the Evolution of a Global Language*. Bristol: Multilingual Matters.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning Artificial Grammars With Competitive Chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 592-608.

Shannon, C. E. (1949). *The Mathematical Theory of Communication*. Urbana and Chicago: University of Illinois Press.

Shantz, K. (2017). Phrase frequency, proficiency and grammaticality interact in non-native processing:

Implications for theories of SLA. *Second Language Research, 33*(1), 91-118.

Sharwood Smith, M. (1981). Consciousness-Raising and the Second Language Learner1. *Applied Linguistics, 2*(2), 159-168.

Sharwood Smith, M. (2008). Input Enhancement in Instructed SLA. *Studies in Second Language Acquisition, 15*(2), 165-179.

Sidtis, D., Canterucci, G., & Katsnelson, D. (2009). Effects of neurological damage on production of formulaic language. *Clin Linguist Phon, 23*(4), 270-284.

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 and Collocation*. Oxford: Oxford University Press.

Siyanova, A. (2010). *On-line Processing of Multi-word Sequences in a First and Second Language: Evidence from Eye-tracking and ERP.* University of Nottingham,

Siyanova, A., & Schmitt, N. (2008). L2 Learner Production and Processing of Collocation: A Multi-study Perspective. *The Canadian Modern Language Review, 64*(3), 429-458.

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, 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., & Janssen, N. (2018). Production of familiar phrases: Frequency effects in native speakers and second language learners. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(12), 2009-2018.

Siyanova-Chanturia, A., & Martinez, R. (2014). The Idiom Principle Revisited. *Applied Linguistics, 35*(5), 549-556.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2019). *Understanding Formulaic Language: a second Language Acquisition Perspective*. New York: Routledge.

Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of Native Speaker and Second Language Learner Intuition of Collocation Frequency. *Language Learning, 65*(3), 533-562.

Siyanova-Chanturia, A., & Van Lancker Sidtis, D. (2019). What online processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language: a second Language Acquisition Perspective* (pp. 38-61). New York: Routledge.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skinner, B. F. (1957). *Verbal Behavior*. Vermont: Echo Point Books & Publishing.

Snyder, W. (2000). An Experimental Investigation of Syntactic Satiation Effects. *Linguistic Inquiry, 31*(3), 575-582.

Sonbul, S. (2014). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition, 18*(3), 419-437.

Sonbul, S., & Schmitt, N. (2013). Explicit and Implicit Lexical Knowledge: Acquisition of Collocations Under Different Input Conditions. *Language Learning, 63*(1), 121-159.

Sorhus, H. B. (1977). To Hear Ourselves—Implications for Teaching English as a Second Language. *English language Teaching Journal, 31*(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.

Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language, 54*(2), 161-184.

Stout, M. (2003). Not guilty as charged: Do the university entrance exams in Japan affect what is taught? *ETJ Journal, 4*(1), 1-7.

Stubbs, M., & Barth, I. (2003). Using recurrent phrases as text-type discriminators: A quantitative method and some finding. *Functions of Language, 10*(1), 61-104.

Swinney, D. A., & Cutler, A. (1979). The Access and Processing of Idiomatic Expressions. *Journal of Verbal Learning and Verbal Behavior, 18*, 523-534.

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics, 26*(2), 245-265.

Szudarski, P., & Conklin, K. (2014). Short- and Long-Term Effects of Rote Rehearsal on ESL Learners' Processing of L2 Collocations. *Tesol Quarterly, 48*(4), 833-842.

Taguchi, N. (2007). Chunk learning and the development of spoken discourse in a Japanese as a foreign language classroom. *Language Teaching Research, 11*(4), 433-457.

Taguchi, N. (2008). The role of learning environment in the development of pragmatic comprehension: A comparison of gains between efl and esl learners. *Studies in Second Language Acquisition, 30*(4), 423-452.

Tode, T. (2008). *Effects of freqency in classroom second language learning: Quasi-experiment and stimulated-recall*. Bern, Switzerland: Peter Lang.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*: Harvard University Press.

Tremblay, A., & Baayen, R. H. (2010). Holistic Processing of Regular Four-word Sequences: A Behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151-173). London: Continuum.

Tremblay, A., Derwing, B., & Libben, G. (2009). Are Lexical Bundles Stored and Processed as Single Units? *Working Papers of the Linguistics Circle, 19*(1), 258-279.

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.

Tremblay, A., & Tucker, B. V. (2010). The Effects of N-gram Probabilistic Measures on the Recognition and Production of Four-word Sequences. *The Mental Lexicon, 6*(2), 302-324.

Tsai, K.-J. (2014). Profiling the collocation use in ELT textbooks and learner writing. *Language Teaching Research, 19*(6), 723-740.

Tukahara, N. (2002). The Sociolinguistic Situation of English in Japan. *Revista de Sociolingüística*, 1-10. Retrieved from http://www6.gencat.net/llengcat/noves/hm02tardor/internacional/a_nobuyuki.pdf

Tyler, A. (2010). Usage-Based Approaches to Language and Their Applications to Second Language Learning. *Annual Review of Applied Linguistics, 30*, 270-291.

Ullman, M. T., & Walenski, M. (2005). Reply: Moving past the past tense. *Brain and Language, 93*(2), 248-252.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In B. Harley & J. H. Hulstijn (Eds.), *Formulaic Sequences Acquisition, processing and use* (Vol. 9, pp. 153-172). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Valsecchi, M., Künstler, V., Saage, S., White, B. J., Mukherjee, J., & Gegenfurtner, K. R. (2013). Advantage in Reading Lexical Bundles is Reduced in Non-Native Speakers. *Journal of Eye Movement Research, 6*(5), 1-15.

van Heuven, W. J., 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*: Blackwell Publishing Ltd.

Van Lancker-Sidtis, D. (2004). When novel sentences spoken or heard for the first time in the history of the universe are not enough: toward a dual-process

model of language. *International Journal of Language & Communication Disorders, 39*(1), 1-44.

Van Lancker-Sidtis, D., & Rallon, G. (2004). Tracking the incidence of formulaic expressions in everyday speech: methods for classification and verification. *Language & Communication, 24*(3), 207-240.

Webb, S. (2007). The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics, 28*(1), 46–65.

Webb, S., & Kagimoto, E. (2009). The Effects of Vocabulary Learning on Collocation and Meaning. *Tesol Quarterly, 43*(1), 55-77.

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford, UK: Oxford University Press.

Webb, S., Newton, J., & Chang, A. (2013). Incidental Learning of Collocation. *Language Learning, 63*(1), 91-120.

Weinert, R. (1995). The Role of Formulaic Language in Second Language Acquisition: A Review. *Applied Linguistics, 16*(2), 180-205.

Wolter, B., & Gyllstad, H. (2013). Frequency of Input and L2 Collocational Processing. *Studies in Second Language Acquisition, 35*(3), 451-482.

Wolter, B., & Yamashita, J. (2014). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, 1-29.

Wood, D. (2010a). *Formulaic Language and Second Language Speech Fluency: Background, evidence and classroom applications*. London: Continuum.

Wood, D. (2010b). Lexical Clusters in an EAP Textbook Corpus. In D. Wood (Ed.), *Perspectives on formulaic language* (pp. 88-106). London: Continuum.

Wood, D. (2010c). *Perspectives on Formulaic Language*. London: Continuum.

Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching, 32*(4), 213-231.

Wray, A. (2000). Formulaic Sequences in Second Language Teaching: Principle and Practice. *Applied Linguistics, 21*(4), 463-489.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.

Wray, A. (2012). What Do We (Think We) Know About Formulaic Language? An Evaluation of the Current State of Play. *Annual Review of Applied Linguistics, 32*, 231-254.

Wray, A., & Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorized language as a gauge of nativelike competence. In F. Meunier & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 123-147). Amsterdam: John Benjamins.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language & Communication, 20*(1), 1-28.

Wulff, S. (2019). Acquisition of Formulaic Language from a Usage-Based Perspective. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language: a second Language Acquisition Perspective* (pp. 19-37). New York: Routledge.

Yaoyu, W., & Lei, L. (2011). Lexical Bundles in the Academic Writing of Advanced Chinese EFL Learners. *RELC Journal, 42*(2), 155-166.

Yoshida, K. (2003). Language Education Policy in Japan: the problem of espoused objectives versus practice. *The Modern Language Journal, 87*(2), 290-292.

Zajonc, R. B. (2001). Mere Exposure: A Gateway to the Subliminal. *Current Directions in Psychological Science, 10*(6), 224-228.

Appendices

Appendix 1. Chi-square scores testing the difference between JHSETC frequency and SUBTLEXus frequency (ranked in the order they are discussed in the paper).

Lexical bundle	df	$\chi 2$	p
what do you	1	17.966	< .001
do you know	1	19.643	< .001
I want to	1	80.161	< .001
yes I do	1	76.696	< .001
no I don't	1	31.25	< .001
No I'm not	1	49.028	< .001
do you like	1	143.21	< .001
do you have	1	120.35	< .001
do you want	1	76.696	< .001
do you play	1	56	< .001
do you think	1	1.1628	= 0.28
do you mean	1	3.1026	= 0.08
what do you think	1	0.18367	= 0.67
what do you want	1	2	= 0.16
what do you mean	1	2.7931	= 0.09
what do you know	1	8	= 0.004
what do you have	1	23	< .001
what do you do	1	22	< .001

Appendix 2a. LSMEANS statistics for Response Time by Repetition.

Rep	Contrast	Estimate	Std. Error	df	t value	lower	upper	Pr(>|t|)	
Rep1	Rep2	95.81986	17.15783	34.5	5.5846	60.96948	130.67024	2.852e-06	***
Rep1	Rep3	131.15014	20.16308	27.2	6.5045	89.79071	172.50956	5.501e-07	***
Rep1	Rep4	93.24734	22.50680	35.3	4.1431	47.56862	138.92606	0.0002038	***
Rep2	Rep3	35.33028	15.82674	41.7	2.2323	3.38431	67.27625	0.0310201	*
Rep2	Rep4	-2.57252	19.01025	51.2	-0.1353	-40.73328	35.58825	0.8928873	
Rep3	Rep4	-37.90280	22.21226	35.8	-1.7064	-82.95826	7.15266	0.0965873	.

Appendix 2b. LSMEANS statistics for Response Time by Lexical Bundle Type.

LB Type	Contrast	Estimate	Std. Error	df	t value	lower	upper	Pr(>|t|)	
PJT Repeated	PJT Unique	-94.96360	34.03652	52.3	-2.7900	-163.25471	-26.67248	0.0073385	**
PJT Repeated	Story Enhanced	96.20623	35.91565	49.1	2.6787	24.03301	168.37946	0.0100303	*
PJT Repeated	Story Unenhanced	73.91644	35.94019	49.1	2.0567	1.69594	146.13694	0.0450527	*
PJT Unique	Story Enhanced	191.16983	27.89128	58.1	6.8541	135.34237	246.99729	5.093e-09	***
PJT Unique	Story Unenhanced	168.88004	27.90712	58.3	6.0515	113.02307	224.73701	1.105e-07	***
Story Enhanced	Story Unenhanced	-22.28979	10.32408	3429.0	-2.1590	-42.53176	-2.04782	0.0309186	*

Appendix 2c. LSMEANS statistics for Repetition: Lexical Bundle Type interactions by lexical bundle type.

	Rep	Contrast	Estimate	Std. Error	df	t value	lower	upper	Pr(>|t|)	
Story Enhanced	Rep1	Rep2	89.80131	23.56946	121.3	3.8101	43.14032	136.46229	0.0002196	***
	Rep1	Rep3	131.01466	26.06379	75.6	5.0267	79.09994	182.92939	3.255e-06	***
	Rep1	Rep4	75.19331	26.84228	72.8	2.8013	21.69373	128.69288	0.0065163	**
	Rep2	Rep3	41.21336	22.39296	164.7	1.8405	-3.00100	85.42771	0.0675006	.
	Rep2	Rep4	-14.60800	23.72960	131.4	-0.6156	-61.54947	32.33347	0.5392217	
	Rep3	Rep4	-55.82136	26.56770	75.0	-2.1011	-108.74638	-2.89634	0.0389900	*
Story Hidden	Rep1	Rep2	110.67302	23.70356	123.9	4.6690	63.75654	157.58951	7.741e-06	***
	Rep1	Rep3	173.76148	26.33249	78.7	6.5987	121.34525	226.17772	4.376e-09	***
	Rep1	Rep4	105.61224	27.16943	76.3	3.8872	51.50329	159.72120	0.0002147	***
	Rep2	Rep3	63.08846	22.45572	166.4	2.8095	18.75353	107.42339	0.0055574	**
	Rep2	Rep4	-5.06078	23.86580	134.4	-0.2121	-52.26176	42.14019	0.8323879	
	Rep3	Rep4	-68.14924	26.83733	78.2	-2.5393	-121.57653	-14.72195	0.0130907	*
PJT Repeated	Rep1	Rep2	109.65298	24.53671	141.6	4.4689	61.14735	158.15862	1.597e-05	***
	Rep1	Rep3	155.75475	26.97380	86.5	5.7743	102.13696	209.37254	1.186e-07	***
	Rep1	Rep4	167.01379	28.18693	88.1	5.9252	110.99930	223.02828	5.925e-08	***
	Rep2	Rep3	46.10176	23.32257	191.8	1.9767	0.10013	92.10340	0.0495083	*
	Rep2	Rep4	57.36081	25.11717	163.2	2.2837	7.76423	106.95738	0.0236770	*
	Rep3	Rep4	11.25904	27.83269	90.0	0.4045	-44.03540	66.55349	0.6867858	
PJT Unique	Rep1	Rep2	73.15211	35.07378	412.3	2.0857	4.20637	142.09785	0.0376235	*
	Rep1	Rep3	64.06965	33.60201	191.5	1.9067	-2.20796	130.34726	0.0580540	.
	Rep1	Rep4	25.17002	47.31253	105.5	0.5320	-68.63642	118.97646	0.5958477	
	Rep2	Rep3	-9.08246	36.14710	394.8	-0.2513	-80.14736	61.98244	0.8017409	
	Rep2	Rep4	-47.98209	47.77649	108.3	-1.0043	-142.68014	46.71596	0.3174705	
	Rep3	Rep4	-38.89963	47.09000	100.0	-0.8261	-132.32534	54.52608	0.4107318	

Appendix 2d. LSMEANS statistics for Repetition: Lexical Bundle Type interactions by repetition.

	LB Type	Contrast	Estimate	Std. Error	df	t value	lower	upper	Pr(>|t|)	
Rep1	PJT Repeated	PJT Unique	-27.45616	41.98380	93.7	-0.6540	-110.81974	55.90742	0.5147332	
	PJT Repeated	Story Enhanced	130.30929	40.34512	77.9	3.2299	49.98632	210.63227	0.0018148	**
	PJT Repeated	Story Unenhanced	84.51014	40.47227	78.7	2.0881	3.94760	165.07267	0.0400223	*
	PJT Unique	Story Enhanced	157.76545	36.77559	119.9	4.2900	84.95190	230.57901	3.639e-05	***
	PJT Unique	Story Unenhanced	111.96630	36.91580	121.7	3.0330	38.88601	185.04658	0.0029600	**
	Story Enhanced	Story Unenhanced	-45.79916	20.52394	3406.1	-2.2315	-86.03965	-5.55867	0.0257129	*
Rep2	PJT Repeated	PJT Unique	-63.95704	44.13053	99.5	-1.4493	-151.51654	23.60247	0.1504080	
	PJT Repeated	Story Enhanced	110.45762	40.25017	77.1	2.7443	30.31173	190.60350	0.0075395	**
	PJT Repeated	Story Unenhanced	85.53017	40.23572	76.9	2.1257	5.40888	165.65147	0.0367360	*
	PJT Unique	Story Enhanced	174.41465	39.27872	127.3	4.4404	96.69078	252.13852	1.927e-05	***
	PJT Unique	Story Unenhanced	149.48721	39.27428	127.0	3.8062	71.77054	227.20388	0.0002184	***
	Story Enhanced	Story Unenhanced	-24.92744	20.21659	3404.1	-1.2330	-64.56533	14.71045	0.2176539	
Rep3	PJT Repeated	PJT Unique	-119.14126	41.71991	86.3	-2.8557	-202.07336	-36.20916	0.0053772	**
	PJT Repeated	Story Enhanced	105.56921	40.21406	76.9	2.6252	25.49065	185.64776	0.0104444	*
	PJT Repeated	Story Unenhanced	102.51687	40.30023	77.4	2.5438	22.27525	182.75849	0.0129572	*
	PJT Unique	Story Enhanced	224.71047	36.53136	107.8	6.1512	152.29771	297.12323	1.331e-08	***
	PJT Unique	Story Unenhanced	221.65813	36.56521	108.4	6.0620	149.18283	294.13343	1.990e-08	***
	Story Enhanced	Story Unenhanced	-3.05234	20.29437	3408.4	-0.1504	-42.84270	36.73802	0.8804554	
Rep4	PJT Repeated	PJT Unique	-169.29993	46.68240	71.3	-3.6266	-262.37459	-76.22527	0.0005355	***
	PJT Repeated	Story Enhanced	38.48881	40.24947	77.5	0.9563	-41.64970	118.62731	0.3419158	
	PJT Repeated	Story Unenhanced	23.10859	40.34583	78.1	0.5728	-57.21238	103.42955	0.5684510	
	PJT Unique	Story Enhanced	207.78874	42.31124	79.6	4.9110	123.57962	291.99786	4.748e-06	***
	PJT Unique	Story Unenhanced	192.40852	42.39387	80.1	4.5386	108.04314	276.77390	1.969e-05	***
	Story Enhanced	Story Unenhanced	-15.38022	20.70944	3410.0	-0.7427	-55.98439	25.22395	0.4577344	