University of Alberta

An analysis of the processing of multiword units in sentence reading and unit presentation using eye movement data: Implications for theories of MWUs

by

Georgina Columbus

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor in Philosophy

Department of Linguistics

Spring 2012
Edmonton, Alberta

Abstract

Multiword units (MWUs) have been the subject of much research in psycholinguistics, due to their syntactic and semantic idiosyncrasies. While studies traditionally focused on idioms (*a piece of cake*), more recent work has focused on another type: lexical bundles (*in the middle of*). How MWUs are stored and retrieved remains a central question in the literature, the answer to which will add to our understanding of language processing. To date though, there have been few investigations comparing the processing of different types of MWUs.

This dissertation aims to fill that gap through analyses of eye movement data during normal sentence reading and trigram reading. The sentence reading results suggest that the familiarity rating for the MWU types analysed here is a relevant predictor of MWU processing. Surprisingly however, individual word frequency has more predictive capacity for MWU reading times than does MWU frequency. Much of the variance is explained by individual word frequency instead. Overall, the three MWU types investigated here are distinguished from one another in fixation durations on words, particularly for idioms and lexical bundles. For sentence reading times, in contrast, the effects of the MWU types are cancelled out, suggesting that processing difficulties may have been resolved at the sentence level.

The second study investigates MWU type effects while reading them without context. Each MWU in this study is a trigram taken from the Google Web1T n-gram corpus (Brants & Franz, 2006) using stratified sampling across n-

gram frequencies. The trigrams were coded for MWU type based on the categories used in Chapter 1. The results show that MWU effects are visible at the trigram level even without context. Somewhat surprisingly, however, there is also evidence of MWU types affecting processing of the first word in the first fixation duration, and of the first bigram in the subgaze duration. The findings suggest the semantic composition of MWUs is apparent to the reader very early. Our results support a usage-based model of language access and storage, such as those put forward by Bybee (e.g., 2006), Pierrehumbert (2001) and Bod (1998), where individual and unit frequency both affect reading times.

Acknowledgments

I wish to thank the many people who have helped me complete this dissertation. Firstly, thank you for all the guidance, mentoring and support to my supervisors Patrick Bolger and Harald Baayen, my former supervisors at the University of Alberta, John Newman and Gary Libben, and my Masters and Honours supervisor, without whose influence I wouldn't be studying MWUs at all, Kon Kuiper at University of Canterbury. Thanks, too, to the fantastic staff at the Department of Linguistics office who helped ensure that I could find jobs and get paid and processed varia for me: Elizabeth French, Jana Tomasovic, Diane McKen & Oliver Rossier. Special thanks to the team of linguists in the graduate offices, both former and current, with their helpful comments on my work, particularly Antoine Tremblay and Laura Teddiman, and Cyrus Shaoul from Psychology. Other people whose discussions and feedback have aided this dissertation are Norbert Schmitt, Kathy Conklin, Walter van Heuven, Jen Hay, Anna Siyanova-Chanturia, Kon Kuiper, Alison Wray, Janet van Hell, Chris Westbury, Debra Titone, and Octavian Ion; many thanks to you all as well. All errors, of course, remain my own.

Thanks are also due to the many lab assistants at the Centre for Comparative Psycholinguistics who ran experimental sessions: Kyla Coole, Reagan Cyr, Janelle Dickout, Christen Flottemesch, Simon Fung, Allison Husband, Mikyung Lee, Danielle Markewic, Caelan Marr, Vince Porretta, Laura Teddiman and Antoine Tremblay; to the CCP lab managers over the years: Anamaria Popescu, Chelsi Mitchell, Lindsay Griener, Karin Nault, Laura

Teddiman, and Alexandra Mateu Martin; as well as to the excellent technical support staff at the Arts Resource Centre and the Department of Linguistics: Lee Ramsdell, Tom Welz, Al Dennington and Clare Peters.

Funding for this dissertation was provided by the University of Alberta's Provost Doctoral Scholarship, the GRA Rice Communications Scholarship, the GSA Graduate Student Research Assistant Award, the GSA Doctoral Travel Fund Award, numerous Travel Awards from the Department of Linguistics at the University of Alberta, and funding from the Mental Lexicon research group through their CFI grant for early experimental costs (SSHRC MCRI 412-2001-1009; Gary Libben PI).

On a more personal note, huge gratitude is owed to the people who helped me get through the dissertation: to my family, especially Lisa Dooling for her hospitality on many conference trips, and to my parents for their support, especially during their own times of stress, and to my other siblings for their long phone calls and cards and treats from home; to Tamara Sorenson Duncan and her husband James for the hospitality and moral support; to Laura Teddiman for lab-related and relaxation-related support, and finally, to my hugely supportive and fellow academic husband Andrew D.F. Ross, who 'gets it', and helped me to focus and write and keep on working when the light at the end of the tunnel seemed too far away. *Arohanui.*

Table of Contents

List of Tables

List of Figures

Introduction

This dissertation investigates the effects of reading and processing multiword units (MWUs) in eye movement paradigms. The psycholinguistics of reading is an area which remains under-examined in terms of the details of figurative versus literal language, and of holistic versus decompositional access to meanings of words in MWUs. These experiments attempt to add to the current knowledge by examining the role of several variables in MWU processing. Additionally, it aims to determine whether MWUs are processed differently dependent on their types. That is, while all MWUs may have high probabilities for one word leading to the next in a string, not all MWUs have the same level of overall frequency (e.g., *a pig in a poke* is considerably less frequent than *Would you like to*) in everyday language. Moreover, not all MWUs have the same degree of semantic unity. That is, there may be differences among MWUs in the degree to which each word contributes to the overall unit meaning, e.g., *kick the bucket* means 'to die', while *in the middle of* means simply 'in the middle of'.

MWUs are combinations of words which frequently co-occur, such as idioms (*too many cooks spoil the broth, between a rock and a hard place*), restricted collocations (*closely guarded (secret), to rain heavily* but not **rain strongly*), and lexical bundles (*in the middle of the, would you like to*). Each of these types represents a different level of semantic and/or syntactic unity, and has different frequency relationships between the component parts and the whole. The taxonomies on which the categories are based – by Alexander (1984), Mel'čuk (1995, 1996, 1998), Gläser (1983), Moon (1998a, 1998b), Howarth (1996), Fontenelle (1998), Pawley and Syder (1983), Benson, Benson and Ilson (1986), Kuiper (2001, 2009), Wray (2002, 2008) *inter alia* in the European field of phraseology – include semantic as well as syntactic qualities of the MWU.

In our definition, idioms are semantically united, with none or few of the component words retaining their basic meaning. For example, in *to let the cat out of the bag*, a secret (*the cat*) is *let out*, while *to blow your top* is to have an angry outburst, where *top* possibly refers to 'head', and *a piece of cake* (a noun) is something that is simple and/or easy to do (adjective). These examples contrast with non-idiomatic items which have no figurative connotation, e.g., *She ate a piece of cake*. Idioms have strong mutual attraction between the component words in the string, but in general are infrequent MWUs in most constructed corpora (see e.g., Moon, 1998b), and are not common in formal writing with the exception of (often punning)[1] news headlines.

A category of MWUs that is less semantically unified than idioms is restricted collocations. As the name implies, restricted collocations have restrictions on the choice of the lexical items which may co-occur with them. For instance, an area may have a *severe drought* but not a **hard drought*, while *rain* is *heavy* or *hard* but not **harsh, *severe* or **strong*. Additionally, the two- or three-word combinations may have a specialized usage, such as *best by/before* and *use by/*before* to indicate when a consumable is considered fresh until or safe to eat, respectively (Mel'čuk, 1995, 1996, 1998). Restricted collocation items frequently co-occur, though they may not be extremely frequent in language use. Different

[1] For example, from the CTV News website on September 19th, 2011: "Canadian Tire *rolling out* online tire business".

genres have different restricted collocations, such as *Yours faithfully* in a business letter and *a heavy loss* in news reportage.

In contrast, a less cohesive category is that of the lexical bundle. Lexical bundles are strictly statistical – they are strings of frequent co-occurring words with no semantic bond at the supralexical level. For example, the word string *in the middle of* is a lexical bundle as the words are very frequently used together in this order, but there is no competing semantic understanding between the unit and the component words. Similarly, *I would like to* is a syntactic structure which may carry some pragmatic meaning (i.e., higher politeness over the alternate form *I want to/ I wanna*). However, *I would like to* carries no semantic meaning other than that of the component words. Syntactically and semantically, then, lexical bundles follow normal compositional rules of a language. They are also highly frequent in language, regardless of genre.

Phraseology encompasses the linguistic subfields of morphology, phonology, semantics, syntax, and more recently, psycholinguistics. At the heart of the research are questions surrounding MWU storage and production, the relation of MWU components to one another, and their semantic transparency and syntactic flexibility. However, MWUs are not homogenous. Indeed, much research has concentrated on one type: idioms (e.g., *to let the cat out of the bag*). Idioms have been investigated for their literal versus non-literal qualities, as well as examining why some are syntatically fixed (e.g., Gibbs, 1986; Cacciari & Tabossi, 1988; Nunberg, Sag & Wasow, 1994). Meanwhile the fixed co-occurrence, often without semantic motivation, of restricted collocations (e.g., *to* *pay attention to, to take care to, *to take attention to, *to pay care of*) has been the focus of other studies (e.g., Mel'čuk, 1998; Howarth, 1998). From psycholinguistic and pedagogical perspectives though, the question of how MWUs are processed or learned is unanswered. We do not yet know, for example, whether all MWUs are processed or learned in the same way, nor whether they are processed or learned differently than compositional structures. We also do not know how L1 (native) and L2 (non-native) speakers differ with respect to MWU processing, though we do know that L2 speakers have difficulty in mastering accurate and appropriate use of MWUs (Wray, 2002; Lewis, 1993). These questions must be answered before models of language processing and storage (e.g., Bod's [1998] or Pierrehumbert's [2001] exemplar-based theories, or Sprenger, Levelt & Kempen's [2006] Superlemma theory), and language pedagogy (e.g., the Lexical Approach of Lewis, 1993, *inter alia*) can fully take MWUs into account. Thus, the studies in this dissertation aim to describe the processing of different types of MWUs in native and non-native speakers of English.

Background

Much research has been conducted on the learning and storage of phrases. For the past few decades, the focus on MWUs has primarily come from the fields of applied linguistics and psycholinguistics. Many studies have attempted to ascertain, for example, what types of MWUs are used across all genres and to what extent, (e.g., Altenberg, 1998; Biber, Conrad & Reppen, 1998), and how much formulaic and figurative language is available to aphasic patients (Van Lancker & Kempler, 1987; Papagno & Caporali, 2007). For instance, both Biber, Conrad and Reppen (1998) and Altenberg (1998) investigated various corpora to determine the most common strings of co-occurring words in English. Conversely, Van Lancker and Kempler (1987) found that formulaic and familiar language (e.g., sayings and nursery rhymes) had distinct access and storage characteristics in the novel language production of aphasics. Their results showed that left-hemisphere damaged patients had significantly better recognition for familiar phrases compared to novel language than patients with right-hemisphere damage. That is, Van Lancker and Kempler suggested that the right hemisphere may hold the key to MWU processing.

Other studies have addressed how idiomatic and figurative language is accessed and processed (e.g., Swinney & Cutler, 1979; Gibbs, 1980, 1985, 1986; Schweigert, 1986; Titone & Connine, 1994, 1999; Rataj & Jáskowski, 2008; Conner et al., 2011; Siyanova-Chanturia, Conklin, Kaan & van Heuven, 2011). For example, Bobrow and Bell (1973) proposed that figurative MWUs are stored in a separate lexicon, allowing figurative processing when literal processing fails. In contrast, Swinney and Cutler (1979), after conducting an experiment where participants were required to judge whether a string of words was meaningful, concluded that MWUs are stored like long words, and that figurative and literal meanings undergo not serial but simultaneous processing.

Several paradigms have been used to investigate MWUs. In the past few years, research into MWUs has ventured into cloze testing. Researchers have recently begun using this paradigm to test the associative links between the individual items that make up the MWU (e.g., Conklin & Schmitt, 2008; Kuiper, Columbus & Schmitt, 2009). Schmitt (2005) used an informal cloze questionnaire on lexical bundles, variable expressions, and idioms. He deduced that participants responded to lexical bundles more slowly and less accurately than from the other types, potentially because their meanings are the sum of their parts. His observations supported previous findings (particularly in special population research, e.g., Conner et al., 2011) that revealing the first word(s) of the idioms facilitated recognition of the entire idiom. This could mean that the initial part of an idiom, the 'idiom key' (Cacciari & Tabossi, 1988), acts in a similar manner to the uniqueness point in spoken word recognition (Marslen-Wilson & Tyler, 1980). Evidence for such a ‘key’ suggests that idioms may be processed in a manner different from other MWU types. This theory will be investigated further in the two online studies in this dissertation.

Moreover, a great deal of research has been undertaken to establish the MWU ‘processing advantage’: whether MWUs are read or processed faster than compositional strings (e.g., Cacciari & Tabossi, 1988; Cacciari & Glucksberg, 1991; Titone & Connine, 1994, 1999; Schmitt & Underwood, 2004; Schmitt, Grandage & Adolphs, 2004; Tremblay, Derwing, Libben & Westbury, 2011). Other MWU investigations have shown that MWUs are processed holistically, rather than by the items that compose them (Underwood, Schmitt & Galpin, 2004; Tremblay, Derwing, et al., 2011; Tremblay, Baayen, et al., 2011; Siyanova-Chanturia, Conklin & Schmitt, 2011). Some have even looked at pedagogy (e.g., Siyanova & Schmitt, 2007). Studies which relate to processing are described in more detail below.

Researchers have recently begun to use eye-movement data to investigate MWU questions. Underwood, Schmitt and Galpin (2004) investigated reading times of the final word in an MWU sequence (presented phrase-finally) compared to terminal words in control (non-MWU) sentences. They looked at the first fixation durations (the duration of time the participants fixed their pupils on a word for the first time when reading the stimuli) and found significantly shorter fixation durations for MWU words over control words. The MWUs themselves, however, were not separated into structural or semantic groups but rather collapsed across such categories as transparent metaphors, lexical phrases and proverbs. Siyanova-Chanturia, Conklin and Schmitt's (2011) eye-movement study on the other hand, investigated only one type of MWU – idioms. This study found that non-native speakers of English processed idioms at the same speed as novel constructions, while the native speakers showed relatively faster reading times for idioms; they found both groups subconsciously accessed MWU frequency, as shown in the fixation durations. Both these findings suggest that eye-tracking is a promising paradigm for further MWU research, particularly in light of the differences found between L1 and L2 readers. In this tradition, chapters 1 and 2 of this dissertation are eye-movement studies: Chapter 1 measuring MWU reading in sentence context, and Chapter 2 measuring context-free MWU trigram reading.

The key issues in all the research outlined above are as follows: semantic transparency, syntactic features, and co-occurrence probabilities. Namely, some

MWUs are not decomposable from their parts (e.g., *act the goat*), while others are syntactically inflexible, or 'fixed' (e.g., **the bucket has been kicked by her* from *she has kicked the bucket*), and other are classed as MWUs purely because of their frequent co-occurence patterns (e.g., *in the middle of* or *I'm gonna*). Thus there have been many studies into the storage, processing and make-up of MWUs and metaphoric language. Yet, with the exception of Molinaro and Carreiras' (2010) investigation of predictability effects, there have been no comparative investigations of the processing differences in MWU types, and fewer still between L1 and L2 English speakers' processing. The results from studies involving more than one type of MWU may have been negatively affected by the differences between the items in semantic relationships, syntactic composition, etc. If we do find differences in processing between various MWU types, then such findings will encourage reanalyses of some results to determine if any prior insignificance is now significant (and vice versa). Another point of interest is that there have not yet been studies using the classifications for any of the MWU types defined in the European tradition over the last forty years or so. As Vanlancker-Sidtis (2004) has suggested, the field is in need of empirical evidence of the classifications of different types of MWUs, and of neurological evidence of how such types are processed, in order to further our models of language processing and pathology. That is, the study of phraseological units would benefit from having a defined set of variables associated with broad classes of MWUs, which could then be considered in future MWU investigations and potentially clarify the source of some effects. Furthermore, identifying which variables may aid or cause delays in reading MWUs will add to the pedagogical debate on how MWUs should be taught to promote L2 proficiency. This set of studies aims to fill these gaps by conducting eye-movement experiments on processing of a subset of three MWU types, by both L1 and higher-intermediate/proficient L2 English speaking participants.

Chapters 1 and 2 report on eye-tracking experiments that yielded reading times. The results of both the experiments are analysed through regression modelling. The predictors in the models are chosen through a stepwise selection procedure. The predictor variables of interest are various, continuous frequency measures (e.g., word frequencies, whole-MWU frequencies, and familiarity ratings) and MWU type, a categorical variable. Since MWUs represent a wide range of multiword strings of varying structure and semantic cohesion, we focus on three broadly-defined MWU types using feature-based classifications from traditional phraseology.[2] The types are not intended to capture the entire range of MWUs, but rather are groups with similar syntactic and semantic features to each other, with no absolute delineation between one group and the next. The regression analyses will help determine which variables best predict MWU processing as measured by single and summed fixation durations, and if and when MWUs are advantageous to processing time. Overall, the results from each study will inform us about any differences in processing between three MWU feature-

[2] The historical difference between traditional phraseology studies and idiom/metaphor studies is beyond the scope of this paper. Suffice it to say that until relatively recently the North American formal syntactic and semantic approach to idioms and metaphors operated quite independently from the European tradition of typological and cultural phraseology (predominantly focused on routine formulae, restricted collocations and proverbs).

based types. In Chapter 1, we also address L1 and L2 groups. As noted above, not much is yet known about the processing of MWUs with different inherent features, or of MWUs in varied positions within sentences. For this reason, the two studies are exploratory in part, investigating how MWUs are processed. The underlying questions for the combined studies are as follows:

1. Are the three different types of MWUs (chosen for this study) that have been posited by some scholars reflected in the eye-movement data? And if so, how?
2. Do individual MWU type predictors such as frequency and semantic transparency play a role in processing? And if so, which predictors and what roles?
3. Do L1 and L2 speakers of English process MWUs differently with respect to predictors of reading times, rather than just reading times themselves?

On methodologies employed

Materials and Design

The working definitions for the stimuli and selection process for this dissertation are given in each paper. But overall, the stimuli have been randomly selected from corpora. As such, the stimuli are exemplars of language usage, rather than experimenter creations. This is intended in part to avoid experimenter-bias effects (Forster, 2000). Additionally, this allows for analysis of MWU reading in naturalistic contexts, since the stimuli for Chapter 1 are taken straight from a corpus with the whole sentence context. The stimuli for Chapter 2 are taken from a different corpus, the Google 1T Web corpus (Brants & Franz, 2006), as naturally occurring strings of three words. Reading times for MWUs presented both in 'natural' sentence context and alone out of context will add to our understanding of MWU effects on reading, particularly with regard to the long-held claim of MWU processing advantages.

There are 250 stimuli for the sentence reading experiments with 50 sentences per condition, plus 100 additional fillers, taken from the British National Corpus (BNC; BNC Consortium, 1994/2001). The three MWU conditions illustrate a range of the MWU spectrum: from the most neutral and compositional extreme of lexical bundles to the most unexpected (in a compositional sense) extreme of idioms, with the restricted collocation group representing MWUs with both compositional and figurative properties. Both practice and experimental trials are randomly ordered. The normal sentence reading involves the participants reading sentences while wearing a head-mounted eye-tracking system, with regular rest and recalibration breaks. Participants also answered comprehension questions intermittently to ensure they were paying attention to the reading task rather than scanning.

The trigram study uses a separate set of 1000 trigrams taken from the Google Web 1T corpus (Brants & Franz, 2006). The trigrams are categorised into the three broad MWU types used in the sentence reading study. Again, both practice and experimental trials are randomly ordered. The task involves trigrams presented alone and participants created sentences using the trigrams and probe words for approximately 20% of the trials. There were regular rest and recalibration breaks.

Analysis

The main predictor used across the three papers is MWU type (we focus on restricted collocations, lexical bundles, and figurative idioms). The reading times and eye-movement data were calculated using the Data Viewer® programme (S-R Research™). These were analyzed with a mixed-effects regression analysis (lmer), using the lme4 package (Bates, Maechler & Bolker, 2011) in R. The predictors for each study are explained in more detail within the relevant papers. The variables employed are listed below.

The dependent variables for the eye-movement sentence study are First Fixation Duration (i.e., the amount of time the reader fixates on a word for the first time, before any reading ahead or coming back to the word), Sentence Reading Time (i.e., the summed fixation durations for all words in the sentence) and total Word Reading Time (i.e., the summed fixations for each word). The first fixation measure allows us to account for the effect of the prior context on the words within the MWU during the first pass, as well as the effect of the MWU on the first pass of the rest of the sentence. That is, in sentences first fixation durations provide data on the processing of a word based on initial parsing of the previous words and no parsing of the upcoming words. The word reading time measure instead offers more information on how words are processed *in toto* during the sentence reading, including whether extra fixation time is necessary to integrate the MWU into the sentence. On the other hand, we include sentence reading time to evaluate the common claim of MWU advantages to processing. Predictors other than MWU Type in this study are: frequency of individual words, frequency of the MWU, MWU familiarity, semantic transparency of the MWU, position of the MWU relative to the word being read, words within versus outside the MWU string, position of the first word of the MWU, word status (i.e., function vs. content word), complexity of sentence (using Flesch-Kincaid readability scores to account for, rather than control, the differences in sentence difficulty), number of words in the sentence (to account for different sentence lengths between stimuli), font size, and sentential position (i.e., which word the reader's eye is fixating on in comparison to the rest of the sentence).[3] The fixations are analysed using linear multiple regression modelling with mixed effects to determine the relevant predictors for MWU processing.

For the trigram study, the dependent variables are First Fixation Duration on the first word (i.e., the duration of only the first fixation on the first word of the trigram), Subgaze Duration on the first bigram (i.e., the summed fixations on the first two words before any regression is made), and Total Fixation Durations (i.e., summed over all three words) for the entire trigram. Some variables in this study are the same as for the sentence study, namely MWU Type, word frequency, word status, MWU frequency; others are unique to this study: constituency rating for the trigram, the length of each word in the trigram, the frequency (i.e., transitional probability) of the first bigram and of the second bigram, and the frequency of the

[3] An analysis including co-occurrence probabilities and cloze probabilities is not yet possible due to the size and construction of this dataset, and so is set aside for a future investigation.

split bigram of the first and third words. Again, analysis is with multiple regression mixed effect modelling.

As mentioned above, reading studies using eye-tracking are a particularly useful method for assessing language processing, as the fixations and movements of the pupils can be registered at an extremely high frequency. Typical reading in English is left-to-right, but not word-by-word necessarily. That is, readers often skip unimportant words (*a, the, to*) when they are visible in parafoveal or peripheral vision, and move to the next meaningful word in a progressive saccade (forward jump). When a reader (mostly subconsciously) feels that the meaning of the words is not clear from the order they have read them in, or the time they spent reading each word in the first pass, they will often make a regressive saccade (backward jump) to the word that was thought to be the cause of the problem. For example, someone may read Bever's (1970) notorious garden-path sentence, *The horse raced over the barn fell*, and show difficulty on the verb *fell*, having incorrectly interpreted the earlier adjective *raced* as the sentence's main verb. This would trigger a regressive saccade back to the area surrounding *raced*, and likely lead to slower fixations on the words during this reading pass. Additionally, the reader is more likely to start reading at the word *horse* than *the*, because of the following: a) *the* occurs at the start of the sentence and is only 3 letters long, meaning it is less likely to be 'landed' on by the first fixation on the sentence but is within the normal span of foveal vision; and b) *the* is a short function word which is more often skipped because of the ability to process it through parafoveal or peripheral vision. Given the structure of idioms, restricted collocations and lexical bundles as described above, it seems clear then that an eye movement analysis of MWU reading could identify subtle differences in processing of these three MWU types. Certainly, the current main reading models may predict different outcomes dependent on the factors involved in reading different types of MWUs.

There are two major reading models in the eye-movement literature: SWIFT (Engbert, Nuthmann, Richter & Kliegl, 2005) and E-Z Reader (Reichle et al., 1998; Reichle, Rayner & Pollatsek, 1999). Broadly speaking, SWIFT posits parallel processing of several words at once through parafoveal vision, as well as focused attention on the word in the foveal view. In contrast, E-Z Reader advances a predominantly serial processing approach, where words outside of foveal view are not accessed beyond more basic information such as frequency and length. These models pose certain predictions regarding the reading of MWUs. For example, in the E-Z Reader model we would expect the frequency of the individual words to be important in word processing measures, but semantic access to the majority of the MWU would not be possible, resulting in a lack of semantic-based MWU effects at the first fixation duration on words within MWUs. On the other hand, the SWIFT model would allow for some semantic access to MWUs in the parafoveal view, but has the same advantage from frequency and length information of the words immediately surrounding the word in foveal view.

In terms of the methodology of these studies, the variables in each analysis will add to our understanding of both reading models and language models. By investigating first fixation durations, total word reading times and total sentence reading times for the sentence study, we can address the processing of words within both a MWU and a MWU sentence at incremental stages. Additionally, by looking at MWUs within a sentence and at MWUs presented unaccompanied and out of context, we have an opportunity to evaluate effects of predictability and integration, using the non-MWU words in each sentence as a form of auto-control, while being able to focus on MWU-only effects in the trigram presentation. The inclusion of both is important given the different predictions a parallel reading theory and a serial reading theory would make for when the MWU is accessed. With respect to language acquisition, storage and retrieval, by including MWUs within sentences as well as including isolated MWUs, we will contribute to evaluations of the holistic versus exemplar-based accounts of language.

The remainder of this dissertation is structured as follows. Chapter 1 is a study of native and non-native English speakers' processing of MWUs. This is an eye movement study with 250 sentences containing MWUs presented on single lines.

Chapter 2 expands on the findings from the sentence eye movement study by investigating frequent three-word strings (trigrams) of differing frequency relationships and differing MWU types. This study is also an eye movement study, but with 1000 MWUs presented as trigrams, without context, to native speakers only. The analysis here also uses mixed effects modelling, and considers the evidence with respect to various reading models.

Finally, the General Discussion and Conclusions summarises the findings on MWU processing with a focus on the implications of the results to language storage and retrieval, and to reading models.

References

Alexander, R. J. (1984). Fixed expressions in English: Reference books and the teacher. *ELT-Journal, 38*, 127-134.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications*. (pp. 101-122). Oxford: Clarendon Press.

Bates, D., Maechler, M. & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 Classes [Computer software manual]. Available online at: http://CRAN.R-project.org/package=lme4 (R package version 0.999375-39).

Benson, M., Benson, E., & Ilson, R. (Eds.). (1986). *The BBI combinatory dictionary of English: A guide to word combinations.* (First ed.). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language.* (pp. 279-362). New York: Wiley.

Biber, D., Conrad, S. & Reppen, R. (1998). Lexico-grammar. In D. Biber, S. Conrad, & R. Reppen, *Corpus Linguistics: Investigating language structure and use*. (pp. 84-105). Cambridge: Cambridge University Press.

BNC Consortium. (1994/2001). *The British National Corpus*, *version 2 (BNC World).* Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/

Bobrow, S. & Bell, S. (1973). On catching on to idiomatic expression. *Memory and Cognition, 1*, 343-346.

Bod, R. (1998). *Beyond grammar: An experience-based theory of language.* Stanford, CA: CSLI Publications/ Cambridge University Press.

Brants, T. & Franz, A. (2006). *Google 1T N-gram Web Corpus*. Google Inc. Available at Linguistics Data Consortium, Pennsylvania.

Cacciari, C. & Glucksberg, S. (1991). Understanding idiomatic expressions: The contribution of word meanings. In G. B. Simpson (Ed.), *Understanding word and sentence.* (pp. 217-240). North Holland: Elsevier.

Cacciari, C. & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language, 27,* 668-683.

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.

Conner, P. S., Hyun J., O'Connor Wells, B., Anema, I., Goral, M., Monéreau-Merry, M-M., Rubino, D., Kuckuk, R. & Obler, L. K. (2011). Age-related differences in idiom production in adulthood. *Clinical Linguistics and Phonetics*, *25*(10), 899-912.

Engbert, R., Nuthmann, A., Richter, E. & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777-813.

Fontenelle, T. (1998). Discovering significant lexical functions in dictionary entries. In A. P. Cowie (Ed.), *Phraseology. Theory, analysis, and applications*. (pp. 189-207). Clarendon Press-OUP: Oxford/New York.

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory and Cognition, 28*(7), 1109-1115.

Gibbs, R. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory and Cognition, 8,* 449-456.

Gibbs, R. (1985). On the process of understanding idioms. *Journal of Psycholinguistic Research, 14*(5), 465-472.

Gibbs, R. (1986). Skating on thin ice: Literal meaning and understanding idioms in conversation. *Discourse Processes, 9,* 17-30.

Gläser, R. (1983). *Phraseologie der englischen Sprache.* Niemeyer: Tübingen.

Howarth, P. (1996). *Phraseology in English academic writing.* Niemeyer: Tübingen.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications*. (pp. 161-188). Oxford: Oxford University Press.

Kuiper, K. (2001). Linguistic registers and formulaic performance. *New Zealand Journal of Sociology,16*(1), 151-164.

Kuiper, K. (2009). *Formulaic genres.* (1st ed.) Basingstoke: Palgrave Macmillan.

Kuiper, K., Columbus, G. & Schmitt, N. (2009). The acquisition of phrasal vocabulary. In S. Foster-Cohen (Ed.), *Advances in language acquisition.* (pp. 215-240). Basingstoke: Palgrave Macmillan.

Lewis, M. (1993). *The lexical approach*. Hove: LTP.

Marslen-Wilson, W. D. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8,* 1-71.

Mel'čuk, I. (1995). Phrasemes in language and phraseology in linguistics. In

Everaert, M., van der Linden, E-J., Schenk, A. & Schreuder R. (Eds,), *Idioms. Structural and psychological perspectives*. (pp. 167-232). Hillsdale, N.J./Hove: Lawrence Erlbaum Associates.

Mel'čuk, I. (1996). Lexical functions: A tool for the description of lexical relations in the lexicon. In L. Wanner (Ed.), *Lexical Functions in lexicography and natural language processing.* (pp. 37-102). Amsterdam/Philadelphia: Benjamins.

Mel'čuk, I. (1998). Collocations and lexical functions. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications.* (pp. 23-53). Oxford: Clarendon Press.

Molinaro, N. & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology*, *83*(3), 176-190.

Moon, R. (1998a). *Fixed expressions and idioms in English.* Oxford: Clarendon.

Moon, R. (1998b). Frequencies and forms of phrasal lexemes in English. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications.* (pp. 79-100). Oxford: Clarendon.

Nunberg, G., Sag, I. & Wasow, T. (1994). Idioms. *Language, 70*(3), 491-538.

Papagno, C. & Caporali, A. (2007). Testing idiom comprehension in aphasic patients: The effects of task and idiom type. *Brain and Language, 100,* 208-220.

Pawley, A. & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication.* (pp. 191-226). London: Longman.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure*. (pp. 137-157). Amsterdam: John Benjamins.

Rataj, K. & Jáskowski, P. (2008). ERPs of novel metaphoric, literal and anomalous utterances. Poster presented at the Sixth International Conference on the Mental Lexicon, University of Alberta, Banff, Canada, October 8th-10th 2008.

Reichle, E. D., Pollatsek, A., Fisher, D. L. & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125-157.

Reichle, E. D., Rayner, K. & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research, 39*, 4403-4411.

Schmitt, N. (2005). Formulaic language: Fixed and varied. *Estudios de lingüística inglesa aplicada*, *6*, 13-39.

Schmitt, N., Grandage, S. & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences*. (pp. 127-151). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Schmitt, N. & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Formulaic sequences.* (pp. 173-189). Amsterdam/Philadelphia: Benjamins.

Schweigert, W. (1986). The comprehension of familiar and less familiar idioms.

Journal of Psycholinguistic Research, 15(1), 33-45.

Siyanova-Chanturia, A., Conklin, K., Kaan, E. & van Heuven, W. (2011). Representation and processing of frequent phrases in the brain. Paper presented at ESCOP 2011, 17th Meeting of the European Society for Cognitive Psychology. Basque Centre on Cognition, Brain and Language, San Sebastian, Spain, September 29th-October 2nd, 2011.

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, A. & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics in Language Teaching, 45*(2), 119-139.

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

Swinney, D. & Cutler, A., (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior 18*, 523–534.

Titone, D. & Connine, C. (1994). Comprehension of idiomatic expressions: Effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*(3), 1126-1138.

Titone, D. & Connine, C. (1999). The compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics, 31*, 1655-1674.

Tremblay, A., Baayen, R. H., Derwing, B. L., Libben, G., Tucker, B. & Westbury,

C. (2011). Empirical evidence for an inflationist lexicon. Paper presented at the Organized session on empirically examining parsimony and redundancy in usage-based models at 2011 Linguistic Society of America Annual Meeting. Pittsburgh, PA.

Tremblay, A., Derwing, B. L., Libben, G. & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading experiments, word and sentence recall tasks, and off-line semantic ratings. *Language Learning, 61*, 569-613.

Underwood, G., Schmitt, N. & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences.* (pp. 153-172). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Van Lancker, D. & Kempler, D. (1987). Comprehension of familiar phrases by left- but not by right-hemisphere damaged patients. *Brain and Language, 32,* 265-277.

Vanlancker-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 Communication Disorders, 39*(1), 1-44.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.

Chapter 1. Processing Multiword Units: Degrees of idiomaticity seen through eye movements.

MWUs are defined as commonly co-occurring n-grams. That is, regardless of the overall frequency of the complete word string, there is a high probability of each consecutive component occurring given the components before it. Recent research into MWUs has ranged from studies of semantic and syntactic relations, on the one hand, to co-occurrence phenomena, driven by large corpus studies into 'n-grams' (frequent word strings of *n*-length) on the other. Taxonomic, syntactic and semantic theory on MWUs has had a longer history than the computationally-driven n-gram studies. Researchers in MWU theory include Mel'čuk (1996, 1998) on restricted collocations, Nunberg, Wasow and Sag (1994) and Wasow, Sag and Nunberg (1980) on idioms, Pawley and Syder (1983) on 'routinized' formulae, and Wray (2002) on formulaic language. The three specific types of MWUs under investigation in the current study are idioms, restricted collocations, and lexical bundles, i.e., the categories described in the Introduction. Recall that these three subtypes stem from the set of categories described in European phraseological tradition. To our knowledge, the definitions given by the above authors, which distinguish between kinds of MWUs, have not been tested experimentally.

The most commonly investigated MWU is the idiom. Idioms are word strings of varying lengths which have a figurative or metaphorical interpretation. For example, in *to let the cat out of the bag*, the 'cat' is a secret that is not supposed to be revealed. Psycholinguistic studies on idioms often aim to determine differences in storage and processing between figurative and literal word sequences, since there is potential for the figurative meaning to activate the literal meaning and vice versa.

Another, less semantically cohesive, MWU type is the lexical bundle. Lexical bundles are strings of words which co-occur frequently in a language. From a processing perspective, lexical bundles have been studied as a specific type of MWU only by Tremblay (Tremblay, 2009; Tremblay & Baayen, 2010; Tremblay Derwing, Libben & Westbury, 2011), though lexical bundles have often been included in studies as part of n-gram, MWU or 'formulaic sequence' stimuli (e.g., Schmitt & Underwood, 2004; Jiang & Nekrasova, 2007; Bannard & Matthews, 2008; Arnon & Snider, 2010). The results of Tremblay and his colleagues not only confirmed processing advantages for lexical bundles over matched controls, but also revealed a further advantage for syntactically-complete units (e.g., *I don't think so*) over syntactically incomplete frequent four-word strings (e.g., *in the middle of*). Arnon and Snider's results also showed advantages for lexical bundles over compositional low-frequency n-grams. Furthermore, they found that the frequency effects were graded so that higher frequency lexical bundles were more advantaged than medium-frequency lexical bundles. Lexical bundles differ predominantly from idioms in that lexical bundles have the following characteristics: Each string is frequent; the transitions from one word to the next are frequent; and they are both syntactically predictable and semantically regular.

The third type of MWU to be investigated here is the restricted collocation. Restricted collocations are short word strings (typically 2-3 words) which co-occur frequently, such as *pay attention to, have a drink,* or *take leave*. However, unlike lexical bundles, they are not entirely semantically transparent. Restricted collocation meanings, like idioms, may not equal the semantic sum of their parts. In restricted collocations, the meanings of one or more ‘parts’ are typically held constant, while another word derives its meaning from the combination of words, rather than its sense in isolation. This is particularly true for the many restricted collocations which contain a semantically ‘light’ verb, such as *take, make,* or *do*. Previous studies of restricted collocations have often focused on (corpus) frequency, first and second language acquisition, and semantics. Yet there have been few empirical studies on the processing of restricted collocation. Though some studies target ‘collocations’ as the construct of interest, the associated stimuli classify as idioms or lexical bundles in our scheme. One study on the processing of restricted collocations was Siyanova and Schmitt (2008; Study 3), where they investigated the reaction times of native and non-native speakers while rating restricted collocations for ‘commonness’ (viz. subjective frequency). They found strong facilitation for high frequency restricted collocations in both groups, although only the native speakers showed facilitation differences between the medium and high frequency responses, and non-native speakers’ responses were significantly less accurate compared to the objective frequency measures.

Theoretically, restricted collocations, idioms and lexical bundles fill the 'gap' between phrases and atomic lexical units in the mental lexicon. That is, idioms, with their varying syntactic fixedness and semantic opacity, are more word-like than restricted collocations. Restricted collocations are syntactically flexible but have a higher degree of semantic transparency, and the category includes a high number of semantically 'light' verbs (e.g., *take* as in *take part in* is not equal to *take* in *take care*). Restricted collocations are then less word-like than idioms but less phrasally-structured than lexical bundles, which are closest to the (syntactically) phrasal end of the mental lexicon/grammar continuum. Recall that lexical bundles are always transparent, though they may be either complete (e.g., *I don't think so*) or incomplete (e.g., *in the middle of*) syntactic constituents. Lexical bundles differ from 'normal' compositional structures in their joint frequency: Lexical bundles are compositional word strings which have a high n-gram frequency, typically over 10 per million words (Biber et al., 1999). We consider Bybee's view then that there is no absolute boundary between idioms and other MWUs, they are part of a continuum (e.g., Bybee & McClelland, 2005). The three MWU types investigated here are seen as representing gradients, and not discrete categories, along the continuum of single word lexical items to compositional phrases. Figure 1-1 shows a possible visualisation of this continuum. Reading processing on compositional phrases and single words fall beyond the scope of the current project, and will not be discussed further.[1] The classification of items into these three MWU types is based on the definitions and

---

[1] See however Columbus (2010) for a comparison of the reading of the MWU

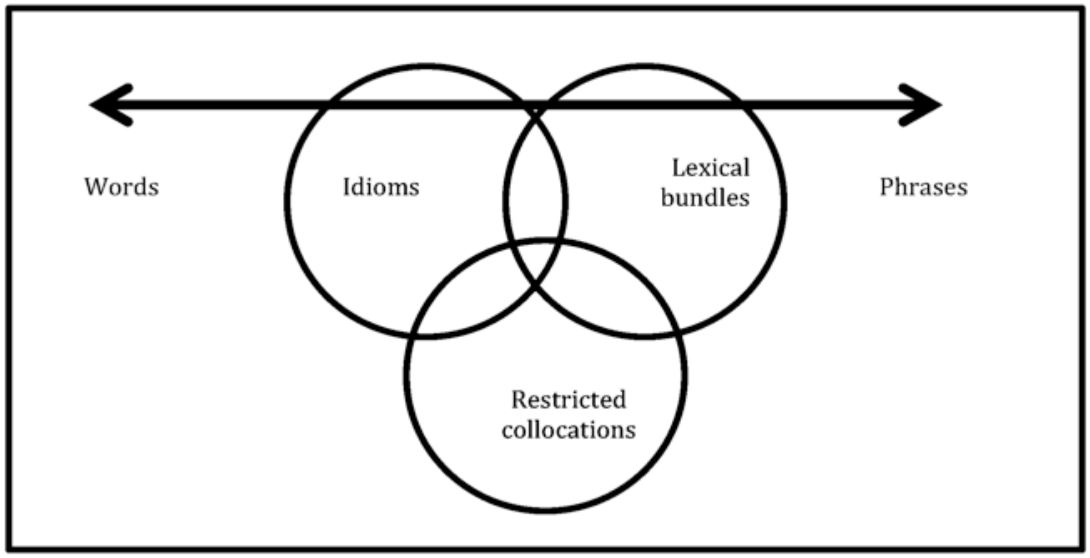

Figure 1-1. Visualisation of MWUs' positions in the continuum between words and phrases.

prior research outlined in the Introduction. (That is, the classification of restricted collocation is based on the definitions in e.g., Mel'čuk, 1996; Howarth, 1996; the idiom definitions are based on e.g., Moon, 1998a; Kuiper, 2000; etc.).

The majority of previous studies on MWUs have focused on lexical access and retrieval of idioms. Such studies have suggested various processing and access theories based on their experimental results. For example, Swinney and Cutler's (1979) Lexical Representation hypothesis adapted Bobrow and Bell's (1973) Idiom List hypothesis, whereby an idiom lexicon is accessed only after failed literal parsing. This led to Di Sciullo and Williams' (1987) description of idioms as 'listemes', which are entered in the lexicon only when they have idiosyncratic properties.

types in this study to control, compositional sentences.

Later models of idiom access included Gibbs' (1980) Direct Access hypothesis, in which the idiom was retrieved whole *before* literal processing was achieved. However, Gibbs later moved to the Decomposition Model, based on results (Gibbs & Nayak, 1989; Nayak & Gibbs, 1990) suggesting that idioms are decomposed through conceptual metaphors linked to the component words. In this theory, the level of semantic transparency is directly related to the level of syntactic (in)flexibility of the idiom, and so ease of access. All propositions are read literally, and idioms which are easily decomposable (e.g., *spill the beans*) are read more slowly than idioms which are non-decomposable (e.g., *kick the bucket*): Indeed, the inability to access the figurative meaning through the literal meaning leads to holistic retrieval of the non-decomposable idiom (Gibbs & Nayak, 1991; Gibbs, Nayak & Cutting, 1989). Cacciari and Tabossi's (1988) Configuration Hypothesis claimed that the whole of the idiom is accessed once the 'idiom key' is reached – that is, once the uniqueness point in the idiom is reached, much like the uniqueness point in a word (Marslen-Wilson, 1975, 1984; Marslen-Wilson & Tyler, 1980) – with complex links between the individual idiom words. Following from this, Titone and Connine (1999) altered the Configuration Hypothesis to their Hybrid Theory, where they suggested that meanings are simultaneously activated for each idiom *and* each word in the idiom. A more recent theory, the Superlemma Approach (Sprenger, Levelt & Kempen, 2006), shares many traits with the Configuration and Hybrid hypotheses, including access to each word individually as well as to the MWU as a whole. This adds explanatory power to their theory, as it can also account for slips of the tongue (or pen) in idioms such as #*scratch the bottom of the barrel* rather than *scrape* (see e.g., Cutting & Bock, 1997; Kuiper, van Egmond, Kempen & Sprenger, 2007, for more on idiom speech errors). Sprenger et al.'s production results show support for a lexico-grammatical model which stores and links the lemmas, the idioms ('superlemmas') and the associated information at both the conceptual and lexical levels. Many recent studies have supported hypotheses like the Configuration model, due to cumulative results which show literal and figurative meanings are comprehended both concurrently and completely (e.g., Tabossi & Zardon, 1993; Titone & Connine, 1999; Vespignani et al., 2010). All these studies regarding access focus on literal/figurative and predictability/probabilistic measures in idioms. Typically, the central question in studies on idioms has remained the same, whether idioms are decomposed word-by-word, and therefore accessed via regular semantic and syntactic links, or whether they are decomposed (semi-) holistically as 'big word' units, and so accessed lexically. With holistic access it follows that idiom-internal content – such as phonology, semantic relatedness of one component to other words, or word frequency – could not be accessed at the same time. Word-by-word decomposition, on the other hand, requires that the idiom's meaning is accessed after an initial parse, such that the idiom cannot affect early processing. However, as will be highlighted below, this question regarding holistic or regular decomposition also applies to less figurative MWUs like lexical bundles and restricted collocations.

Leaving aside for the moment the question of figurative and metaphoric processing, research into other MWUs (as 'formulaic sequences' or similar) has shown processing advantages for frequent word combinations. Again, however, the results cannot in themselves resolve the debate between holistic versus segmental processing. Experimental results from eye movement (Underwood, Schmitt & Galpin, 2004), self-paced reading (Schmitt & Underwood, 2004; Tremblay, 2009), and MWU decision or judgment tasks (Arnon & Snider, 2010; Ellis & Simpson-Vlach, 2009; Jiang & Nekrasova, 2007) all reveal faster reading or reaction times for frequent word strings compared to 'compositional' controls. Bannard and Matthews (2008) found that children are also faster and more accurate at recalling multiword chunks than controls. In Tremblay's studies (Tremblay & Baayen, 2010; Tremblay, 2009; Tremblay et al., 2011), faster processing for MWUs seemed to suggest holistic storage. However, his finding that syntactically-complete lexical bundles are facilitated more than 'incomplete' lexical bundles (e.g., *in the middle of*) suggest that something other than storage of frequent word strings is at work. Tremblay suggests that a usage-based model of access and retrieval also fits his results (Tremblay & Tucker, 2011; Tremblay et al., 2011; Tremblay & Baayen, 2010**)**. Further, his results hint at significant differences between MWU types even at the word level. These results seem to weaken theories of holistic MWU storage. So far as we are aware, to date only Tremblay's (Tremblay & Baayen, 2010; Tremblay, 2009; Tremblay et al., 2011), Bannard and Matthews' (2008), and Arnon and Snider's (2010) investigations into frequent four-word strings have detailed the variables that modulate MWU processing advantages. Unravelling such variables may be key to determining how MWUs are stored and accessed.

A final area of MWU studies is that of L2 competence. While many studies have focused on non-native production of MWUs (e.g., Siyanova & Schmitt, 2007, on rate of MWU usage, and Schmitt, Dörnyei, Adolphs & Durow, 2004, on fluency of production), fewer studies have investigated processing of different MWUs by non-native speakers. Several MWU processing studies have been carried out by Siyanova and colleagues (Siyanova, 2010; Siyanova-Chanturia, Conklin & Schmitt, 2011; Siyanova & Schmitt, 2008; Siyanova-Chanturia, van Heuven & Conklin, 2011). In investigating the processing of restricted collocations, idioms, and irreversible binomials (*salt and pepper, *pepper and salt*) they found that non-native speakers tended to process idioms in the same manner as they would for novel phrases (Siyanova-Chanturia, Conklin & Schmitt, 2011), and that they are better at determining the frequency of collocations when the collocations are highly frequent than not (Siyanova & Schmitt, 2008). Additionally, Siyanova-Chanturia, Conklin and van Heuven (2011) found that the frequency of the MWU was a crucial predictor in reading times and fixation durations on irreversible binomials. However, there remain gaps in our knowledge, including how non-native speakers process different MWU types, and research which considers varying measures of frequency as predictors.

We have seen above that while there is strong evidence for MWU processing to be faster than controls, the root of this advantage is yet to be determined. In particular, we cannot identify all the variables which facilitate and/or inhibit MWU processing. Nor can we yet determine the model of access/retrieval which best explains prior results. Additionally, while idioms have been frequently studied, other investigations have used MWU stimuli with a range of different syntactic and semantic structures, rather than using stimuli of certain MWU categories. This makes it difficult to partial out which predictor has led to the advantages in these studies' results. With respect to the kinds of effects different variables contribute, to our knowledge only the semantically transparent and syntactically 'normal' lexical bundles have been analysed in terms of predictors beyond lexical controls and frequency. Overall, there has not yet been an empirical study into effects of MWU features through MWU types, or other variables which differentiate the processing of MWUs, compared either to controls or to other MWU types. This study aims to add to the understanding of MWU processing through eye movement data. Specifically, our first goal is to determine if there are differences between the three MWU types we have defined here with respect to processing. If there are such differences, our second goal is to determine how individual word frequency, MWU frequency, MWU familiarity, and semantic transparency or idiomaticity variables contribute to the MWU processing. That is, we aim to determine how the semantic, syntactic and occurrence features of different MWUs contribute to their processing. We investigate these questions with regard to native speakers of English (L1), as well as non-native (L2) English speakers, in order to add to the current knowledge and theories of L2 language processing.

As these previous studies show, the majority of MWU studies have been factorial designs with controlled stimuli, and often have analysed only one or two words in the MWU, rather than each word of the unit. In this study, we aim to investigate how MWUs are read in natural sentences taken from an English-language corpus, the British National Corpus. Because these sentences are taken whole from exemplars of spoken and written language, they are allowed to vary randomly along several dimensions including word and sentence length, in word frequency, and in MWU frequency. The stimuli used here, then, are closer to 'natural' sentences. They are not wholly naturalistic, in that the sentences are not related to one another and are not presented in wider contexts. But they are free to vary internally in terms of having fairly uncontrolled relationships between the MWU and the remaining sentence context. We are able, for example, to look specifically at processing effects on the words inside versus outside of the MWU. As such, the results of this study offer a wider scope for generalizability to everyday reading. That is, the widely-claimed advantages of MWUs in processing can be assessed in reading contexts which more closely match spontaneous exposure to MWUs outside the laboratory. Besides the question of processing advantages, the 'natural' sentence stimuli will allow us to determine if the position of the MWU within the sentence affects overall reading time and/or strategy. The analysis of random variation such as we have used here requires linear regression modelling with mixed effects (i.e., fixed variables as well as random, crossed independent variables). Mixed effects regression modelling is an attractive alternative to the limitations imposed by controlled factorial design. This is because the approach allows us to account for the variation we inherited by selecting less restricted stimuli. For example, we can account for the range of word frequencies and so avoid having to create stimuli of matched frequencies.

We can also account for the random sentence structure (e.g., passives, adverbial clause beginnings etc.) through sentence complexity measures, giving us the ability to avoid unintentional structural priming in the task. Thus, the current study offers a new perspective on MWU processing beyond the investigation into effects of MWU Types.

The descriptions above allow us to make some predictions regarding MWUs and models of language and reading. Firstly, if MWUs are stored holistically, then we would expect idiom, lexical bundles and restricted collocations to have similar reading times. But if only semantically irregular MWUs are stored holistically, then only restricted collocations and idioms would have the shortest reading times. If MWUs are instead stored and accessed in an exemplar-based account, then we would expect effects such as word frequency and word status to contribute to the reading times as well as effects from the unit-based predictors. As such, we would also expect graded differences to the reading times for each MWU type. With respect to reading theory, recall the description of the two major reading models, SWIFT and E-Z Reader, in the Introduction. We posit that a parallel reading model such as SWIFT would allow more content from the MWUs themselves in early fixations whereas a serial reading model would not.

The remainder of this chapter is as follows: we begin with definitions of the stimulus categorisation and potential predictors, and then move to the results of the reading experiment for our three dependent measures: first fixation duration, total word reading time, and total sentence reading time.

Method

Participants

Twenty-four native speakers of North American/Canadian English and 29 non-native speakers took part in the eye movement experiment. Participants were recruited either from the Centre of Comparative Psycholinguistics volunteer database (native speakers only) or from first year linguistics courses in the Participation for Credit option. Participants contacted via the database received $15 as compensation while linguistic student participants received course credit. Data from one non-native and three native participants were excluded from the eye movement analysis due to technical difficulties. All participants had normal or corrected-to-normal vision. L2 participants were recruited from the general student body to allow a closer replication of an L2 class in an English environment. Namely, they were multicultural, multilingual (with varied L1s), and had various ages of acquisition and lengths of exposure. All L2 participants met the minimum university entrance requirement for English proficiency, and could be described as ranging from High Intermediate, to Advanced, to (in some cases) Proficient levels in their English speaking skills.

Materials

The stimuli were taken directly from a corpus, to maximise the naturalness of the English MWU sentences. Fifty sentences each from the three stimulus types (restricted collocations, lexical bundles and idioms) were collected from the British National Corpus (BNC; BNC Consortium, 1994/2001). The idioms and restricted collocations used in this experiment were pseudo-randomly selected from the BNC using key words as search terms. They were selected using the classifications described above for each type in the analysis. That is, the MWUs chosen for searches met the graded combinations of qualities including text, genre, and register preferences/features, as well as linguistic attributes such as frequencies, syntactic form and so on. The search items were taken from idiom and restricted collocation dictionaries and studies (e.g., the Benson, Benson & Ilson *Combinatory Dictionary of English*, 1986; Flavell & Flavell's *Dictionary of Idioms and their Origins*, 2006; Moon, 1998b; Kuiper et al.'s *Syntactically Annotated Idiom Dictionary* [SAID], 2003). The lexical bundles were four-word sequences which were pseudo-randomly selected from Tremblay's (2009) lexical bundle list. Tremblay follows Biber et al. (1999) in defining lexical bundles of four words as being four words occurring as a whole unit ten times per million words (Tremblay, 2009, p.8). The lexical bundle sentences were then selected from the BNC, in the same manner that the idiom and restricted collocation sentences were collected.

The initial results for each MWU returned thousands of lines of data. A random sample of either 100 or 200 of all results was conducted using the BNCWeb thinning tool (Lehmann, Hoffmann & Schneider, 2002).[2] Of any given set of hits that contained the target structure, the first or last sentence[3] in the set was selected as the representative stimulus for the study, provided it was both logical and of a feasible length for an eye tracking experiment. That is, sentences

---

[2] The method of using different numbers for the random sample set was to add further random variation to the selection.

[3] The selection was made from the first sentence of one set, the last sentence of the next set, and so on.

which contained nonsensical restarts (e.g., *Then they – Then we went…*), which were too long for single-screen presentation, or which were short responses to questions (e.g., *I don't know!*) were excluded. This quasi-random process minimized the opportunity for experimenter bias (Forster, 2000) while maximizing the potential for generalization.

In addition to the 150 sentences containing MWUs, we also included 100 additional filler sentences. Fifty of these sentences were neutral control sentences to distract from the MWUs. The control sentences were randomly sampled from the BNC by searching for a very common word (*the* or *is*) and selecting 50 pseudo-randomly from the search results in the same manner as the three MWU sentence types above. Sentences that contained idioms, restricted collocations, or frequent lexical bundles of over 2 words were excluded from the sampling process. A further 50 sentences were semantically anomalous sentences, which served to hold participants' attention during the experiment through invoking a surprise element. These sentences were sampled from the BNC in the same method as for the neutral control sentences. Once the set of 50 had been selected, each sentence had one content word replaced in order to make the stimulus nonsensical. These neutral and anomalous sentences are not included in our analysis. (For more on the semantically anomalous and control sentences' reading processing effects, see Columbus, 2010.)[4] All the experimental stimuli are found in Appendix A.

---

[4] The neutral and semantically anomalous control sentences were part of a study which ensured that MWU reading processing was different to both 'normal' reading (i.e., neutral controls) and to disturbed reading (i.e., semantically

Experimental procedure

In this experiment, we look to reading times as a means of describing MWU-reading behaviour. The three time variables we are interested in are First Fixation Duration (or FFD), Total Word Reading Time (or WRT), and Total Sentence Reading Time (or SRT). First Fixation Duration refers to the length of the first time the participant fixates on a word; it is measured for each word in each sentence. The fixation duration measurement is particularly focused on the reading times for words within the MWU region compared to fixations on the words outside of the MWU region. We include this within and outside distinction in order to capture effects of the MWUs on the words before and after they appear, as well as to provide non-MWU words as a type of internal control measure. Word Reading Time is the total gaze duration of all fixations, i.e., including regressive fixations, for each word in the sentence. Like the FFD variable, WRT is analysed for all words in the sentence, not only the target words within the MWU region. Again, we place particular emphasis on comparing reading times for words within versus outside of the MWU region. The Sentence Reading Time measure is the trial dwell time, i.e., the sum of all durations over each word in the entire sentence. This is labelled distinctly so as to emphasise that the areas of interest in this experiment are all the words in the sentences, rather than the MWU region alone. SRT is included to allow for comparison of predictor effects on the whole sentence reading compared to the reading of individual words, particularly in light of any MWU processing advantages.

---

anomalous controls). This study looked at predictors relevant to reading measures of the control sentences versus the MWU sentences (Columbus, 2010).

Procedure

The stimuli were presented on a desktop PC monitor, using Experiment Builder™ software running on Windows XP. The data were collected using an Eyelink II, video-based head-mounted binocular eye tracking system (SR Research™). Eye movements were collected for each word in each sentence at a sampling rate of 250Hz or 500Hz depending on the conditions of the experiment sessions.

Each experiment session was preceded by 12 practice trials, with rest and recalibration breaks occurring after each block (approximately every 4-6 minutes). Both the practice and 250 experimental trials were pseudo-randomly ordered for each participant. Participants were seated at a comfortable distance from the screen (approximately 70cm) and asked to read the sentences for meaning, both silently and as quickly as possible. The participant then cued the next sentence. To ensure they were reading for comprehension, participants also answered yes/no comprehension questions at a rate of approximately one for every five to six questions. The error rate for exclusion was set at >25%. Responses were logged on a Cedrus RB-530 button box. A generic USB foot pedal was used for cueing the stimuli.

The comprehension questions and instructions were presented in white font on a black screen. In order to highlight the difference between stimulus sentences and comprehension questions, the stimulus sentences were presented in yellow, following a fixation cross. To ensure that the stimuli appeared on one line despite random lengths, sentences shorter than 15 words were presented in size 18 Times New Roman font, whereas longer sentences were presented in size 14. (This size difference is accounted for in the analysis.) All stimuli were presented from the centre left of the screen, while instructions and comprehension questions were presented centrally.

Results

Regions of interest (ROIs) were set manually for each word in the experimental stimuli and data matrices based on these ROIs were generated by Dataviewer, a software programme designed for the analysis of eye-movement data (SR Research™). The matrices were then loaded into the statistical programme R (R Development Core Team, 2011). These were then merged with the item statistics. The three dependent variables (FFD, WRT and SRT) were all (natural) log transformed to remove most of the rightward skew, and minimise the possibility of overly influential outliers in the statistical model.

Recall that there are three time measures to be analysed, from small to large: first fixation duration, word reading time, and sentence reading time. The remainder of this results section discusses each measure in the same small-to-large sequence. For each time measure, we first give results and a discussion of first the native speakers' model, which includes only the MWU Type factor and variables which are significant for these data. We then present the non-native speakers' best-fit model, followed finally by the combined group results. The combined group model for each dependent variable consists of only the significant predictors from the L1 and L2 models, with the addition of by-group interactions.

Analysis methodology

Planned analysis uses mixed effects multiple regression modelling using the lmer function in the lme4 package for R (Bates, Maechler & Bolker, 2011). Subject and item were included as crossed random-effect factors. Predictors included in the modelling are only those which are possible for the particular dependent variable at hand (e.g., the frequency of a word in word reading time and first fixation durations, but the mean frequency of the sentences' words in the sentence reading time measure). Variables to be investigated for their predictive capacity are defined below.

Predictors in the analysis

Recall from the introduction that previous experimental evidence for theories of MWU processing have taken varying word and unit occurrence measures into account. Our study follows suit, investigating the predictive capacity of a combination of frequency and familiarity variables as well as semantic, syntactic and positional effects. Table 1-1 lists and defines the variables to be considered for their effects in MWU reading. Table 1-2 lists the numeric predictors' value ranges and means. Note that the MWU types do not differ significantly with regard to individual word frequency, but are significantly different in their MWU frequency. Both these frequency measures are inherent features which are included in the type classification. A complete downloadable list of these predictors' values for each stimulus can be found on at: http://www.ualberta.ca/~columbus/Site/Welcome.html.

Before we turn to results themselves, it is necessary to describe the familiarity and semantic transparency ratings in more detail. Each of these item variables is a residualised mean rating taken from pen and paper or online surveys. These surveys were completed by two separate groups of native and non-native speakers who received course credit for their participation. The familiarity variable accounts for the subjective frequency of the MWUs in the stimuli. Participants gave ratings of between 1 and 4 (least known to best known and used). For the L1 and the L2 datasets, the ratings were averaged for each item over subjects. We then merged these mean item ratings with the full MWU variable data, and tested the ratings values for collinearity with other predictors for the MWUs. The native speaker ratings were shown to be collinear with Word frequency. For each group a linear regression model was made to decorrelate the familiarity rating from any collinear variables. The residuals from the L1 decorrelation model form the Familiarity variable in the L1 MWU dataset, and the L2 decorrelation model residuals form the Familiarity variable for the L2 MWU dataset. Importantly, note that a lower Familiarity score indicates higher familiarity.

Table 1-1. *Predictors in the analyses*

Predictor name	**Predictor definition**	**Models relevant to**		
<u>*Key predictors*</u>		*FFD*	*WRT*	*SRT*
MWU Type	Idiom, restricted collocation or lexical bundle MWU within the sentence	x	x	x
MWU Frequency	Logged frequency of the MWU (from the BNC)[5]	x	x	x
Word Frequency	Logged surface form frequency of individual words (from the BNC)	x	x	–
Mean Word Frequency	Logged mean frequency of all words in the sentence (from the BNC)	–	–	x
MWU Familiarity	Averaged residualised familiarity rating per MWU taken from a familiarity survey (see below)	x	x	x
MWU Semantic Transparency	Averaged residualised semantic transparency rating per MWU taken from a semantic transparency survey (see below)	x	x	x
Sentence Complexity	Logged rating of the difficulty of reading the sentence as gauged by the Flesch Ease of Reading Formula (Flesch, 1948)	x	x	x
Start of MWU	Location of the first word in the MWU within the sentence	x	x	–
Relative Position	Location of the word being read compared to the MWU (= *before, in, after*)	x	x	–
In MWU Region	Whether the word being read is within or outside the MWU itself	x	x	–
Word Status	Whether the word being read is a function word (= *yes*) or a content word (= *no*)	x	x	–

[5] MWU frequency using the Google search engine was also tested in the preliminary modelling. However, the Google measure was not predictive and is thus not discussed.

Combined model

Group	Whether the readers are *L1* or *L2*	x	x	x

Control predictors

Word Length	Length in characters per word	x	x	–
Mean Word Length	Mean (in characters) of words per sentence	–	–	x
Font size	Whether size 14 or size 18 font was used (sentences over 14 words used smaller font to fit in one line)	x	x	x
Sentential Position	Where the word being read is located within the sentence	x	x	–
Sentence Length	How many words are in the sentence	–	–	x

Table 1-2. *Numerical values of predictors' a) value ranges and b) means*

Predictor name	**a) Value range**
Key predictors	
Word Frequency *(logged)*	0.01-62,014.49 occurrences per million words
Word Length	1 – 16 characters
MWU Frequency *(logged)*	0.00995 – 4.69583 per million words
Mean Word Frequency *(logged)*	7.605 – 11.171 per million words
Mean Word Length	1.251 – 1.941 characters
MWU Familiarity	-0.718 – 0.595
MWU Semantic Transparency	-19.182 – 57.519
Sentence Complexity *(logged)*	3.411 – 4.982
Sentence Word Count	3 – 22
	b) Mean *(from logged value)*
Lexical bundle – Word frequency	6.84 per million words
Restricted collocation – Word freq.	6.31 per million words
Idiom – Word freq.	6.43 per million words
Lexical bundle – MWU frequency	0.36 per million words
Restricted collocation – MWU freq.	-1.75 per million words
Idiom – MWU freq.	-2.96 per million words

Like the familiarity variables, the semantic transparency variable in each dataset is a residualised mean rating per item. The semantic transparency predictor is included to account for the level of difficulty or ease speakers have in accessing MWU meanings from the meanings of the individual words in the multiword units. These ratings were formed from magnitude estimations of transparency for each MWU item, which the participants rated based on a common reference. (For a full description of the magnitude estimation method for evaluating MWUs, and the questionnaire on which this was based, see Wulff, 2009. A paper-version of this survey is available in Appendix B.)

Once the semantic transparency ratings were collected into a dataset, they too were averaged per item over subjects before being merged with the full MWU variable set. We then tested the semantic ratings for collinearity with other MWU predictors. Collinearity was found in the L1 semantic transparency data for MWU familiarity and MWU frequency. In the non-native ratings, Word frequency, text Complexity and MWU Familiarity were found to be collinear with semantic transparency. A linear regression model was then run on the averaged items for each speaker group and the collinear variables. The residuals of the L1 and L2 decorrelated models constitute the semantic transparency variable in the respective L1 and L2 datasets for current analysis. Importantly, note that a lower Semantic Transparency score indicates higher transparency.

We turn now to a discussion of the predictors involved in each analysis. Recall that we predicted that we predicted that restricted collocations and idioms would have the shortest reading times if only semantically irregular MWUs were stored holistically. As we shall see below, this was not borne out. However, we did find graded differences in reading times caused by MWUs dependent on their components' word features, such as frequency and word status. We stated earlier that this would be predicted in an exemplar-based account. In addition to these results, we found very early effects in one MWU type which may indicate a parallel processing model of reading. We begin with these, and other, effects found in the first fixation duration results.

Note that as our focus is on *how* MWU sentence types are processed, and how processing differs *between* types, we report the control variables only in the model tables.

Results and discussion: First Fixation Duration

We first investigated the first fixation duration times for each word in each sentence. This dependent variable offers an insight into immediate processing issues on any word. Separate models were first fitted to the log transformed First Fixation Duration for the L1 and L2 data. Predictors in the models were MWU Type, Word Frequency, Relative Position, Familiarity, Complexity, Start of MWU, Word Status and two control variables, Font Size and Sentential Position. Each model was subjected to model criticism, and potentially harmful outliers were removed (SD +/- 2.5; L1 *n*=97%, L2 *n*=97.4%). The models were then fitted to the remaining datapoints (L1 *n*= 27961; L2 *n*= 38542).

L1

For the L1 data, by-subject random slopes for Word Frequency (estimated standard deviation 0.0035) were supported by Likelihood ratio tests. Random intercepts for the Sentence, Word and Subject random effect factors were also supported (estimated standard deviation parameters 0.0276, 0.0465 and 0.1297 respectively; the estimated standard deviation of the residual error was 0.3336). There were interactions for Familiarity by Word Status and for MWU Type by Word Status by Word Frequency. Table 3 lists the log likelihood ratio test values for each predictor added to the model. Following Baayen, Davidson and Bates (2008), absolute values for *t* exceeding 2 were taken to be significant. Predictors that did not reach significance are not listed in this table, which summarises a minimally adequate model.

MWU Types are robustly differentiated in first fixation times of the L1 speakers. Figure 1-2 illustrates the MWU type by word frequency by word status effects when all other variables are held constant. As can be seen in panel 1, first fixations on words in sentences with idioms are not shorter for words of higher frequency, which contrasts with those on words in sentences with restricted collocations or lexical bundles. In fact, high frequency words in restricted collocation and lexical bundle sentences are processed faster, as measured by first fixation durations, than those in idiom sentences. The reverse holds for low frequency words. This interaction was not affected by whether the word was located in the MWU region of interest. We shall see below that this frequency effect is restricted to content words. (The two-way interaction of MWU type by word frequency reveals the same results as seen in Figure 1-2, panel 1.) Note that the frequency effect in panel 1 is not significant for idioms. It unusual that there should be no frequency effect for words in these sentences even when the words occur well before the idiom. Post-hoc analysis revealed no frequency effect for words in idioms, regardless of the position of the first word of the MWU, number of words in the sentence, or the number of function words in comparison to the lexical bundle and restricted collocation sentences. To be certain of ruling out any variables that may have been omitted from the model, we also carefully analysed

Table 1-3. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the First Fixation Duration L1 model*

Predictor	**AIC**	**Df**	***p***
MWU type	32612	8	–
Word frequency	32611	9	0.0653
MWU type * Word frequency	32609	11	0.0396
Word status	32609	12	–
MWU type * Word status	32604	14	0.0171
MWU type * Word status * Word frequency	32595	17	0.0018
Sentential position	32475	18	<0.0001
Familiarity	32476	19	–
Familiarity * Word status	32470	20	0.0104
Font size	32452	21	<0.0001
Relative position	32446	23	0.0053
Complexity	32443	24	0.0364

Note: These are model comparisons, and as such the table has no t-values for the predictors. Each predictor's t-value was significant, however, to have been kept in the model. Predictors which are not significant as main effects (though are significant in interactions) are italicised.

the structure of the idiom stimuli by close inspection of each sentence. This was to determine if the individual subjects or the subjects' positions may have increased the predictability of the idiom. However, there is no difference between type of subjects in the idiom sentences compared to the restricted collocations or lexical bundles, and so we cannot postulate a marker signalling an upcoming idiom to the reader. Nor is there any structural difference in sentence beginnings for the idioms as compared to the restricted collocations. Thus we have no cause for the lack of word frequency effect for first fixations on words in the idiom sentences, even before the idiom unit begins. We concede that the statistical technique may not be sensitive enough to differentiate between the presence of a normal frequency effect before the idiom is reached, and the absence of a frequency effect after the idiom is reached. This was not found in the sub-analyses of the word frequencies by sentence position and relative position, however. It may be that a semantically and syntactically tagged version of the stimuli may allow for a more fine-grained statistical analysis of the structure of the idiom sentence beginnings versus those of the restricted collocations and lexical bundles. Future investigation with such parameters may resolve this potential cause. Nevertheless, there is a possibility that the idiom may have been accessed in parafoveal preview of the initial word(s), as an extreme form of the SWIFT model would predict (Engbert et al., 2005; Hohenstein, Laubrock & Kliegl, 2010). In this model, we would expect some semantic information to be accessed through parafoveal uptake while attention was placed on the word(s) in foveal view. We are unable to either prove or disprove such a preview, however, as this question requires a more structured study than the paradigm used here.

A difference between content word and function word reading also emerged, as seen in Figure 1-2. It is well-known that function words are often skipped (Morrison, 1984). Hence, when they are fixated on, one would also expect shorter first fixation durations on function words (such as reliably shorter durations on function words after a skipped content word; see Kliegl, 2007). However, we do not see such an effect for highly frequent function words in sentences of any MWU type (panel 2). The longer fixation durations for function words in idiom sentences is possibly due to the predictability of words neighbouring the 'key' content words in idioms (see Cacciari & Tabossi, 1988, for more on the idiom key, though note also Tremblay & Tucker, 2011). Additionally, for restricted collocation sentences the effect was greatly reduced, if not significantly inhibitory. There were, however, clearly shorter first fixation durations for the lower frequency function words in lexical bundles. This may be due to their high MWU frequency, which in our data proves significantly greater than that of the restricted collocations (Welch Two-sample t-test: $t = -8.13$, df = 95.202, $p < 0.0001$), and of the idioms ($t = -14.59$, df = 98.748, $p < 0.0001$), as shown in Figure 1-3. As Brysbaert, Drieghe and Vitu (2005) note, skipping (and so shorter fixations) is more common on 'easy' words (i.e., more frequent and more predictable). By definition, lexical bundles consist of highly predictable word strings, with increasing predictability for completing the lexical bundle within the sentence. It is not surprising, then, that lexical bundles have faster first

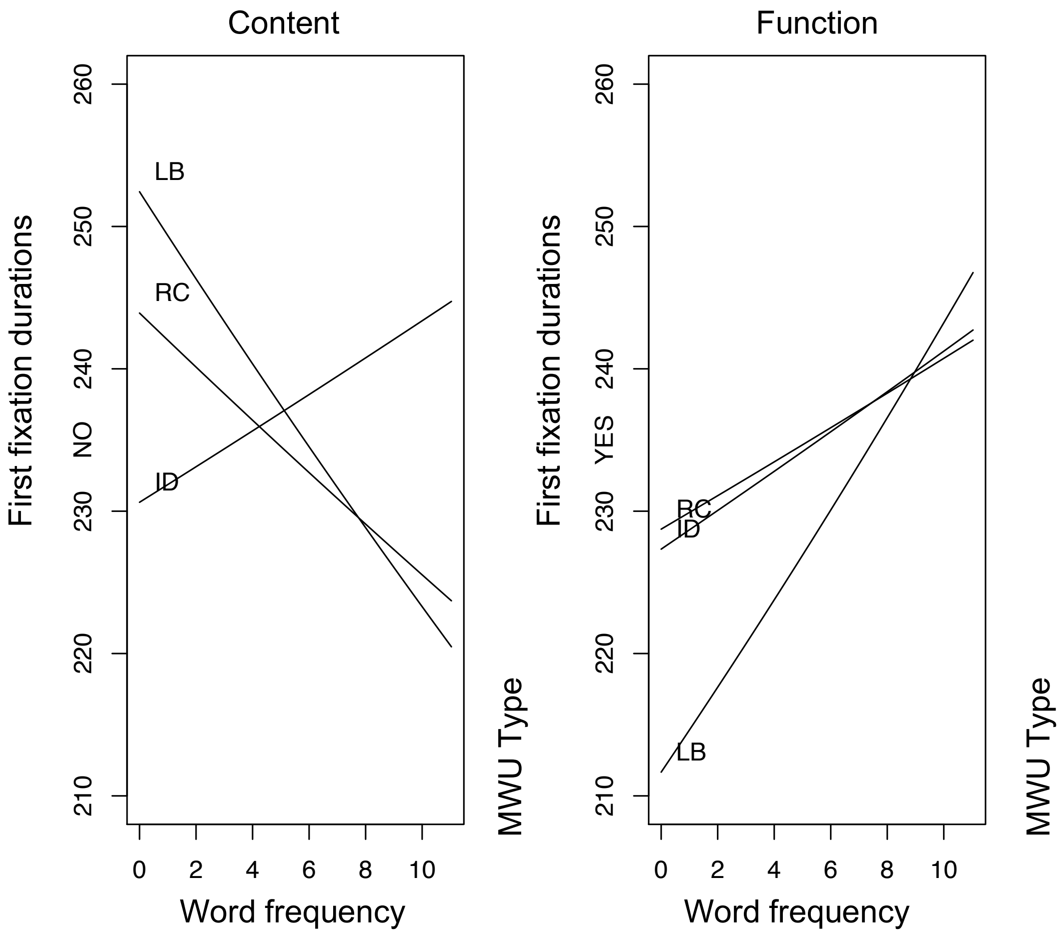

Figure 1-2. Word Frequency interaction with Word Status (Content = panel 1, Function = panel 2) for words in sentences with Idioms, Restricted Collocations, and Lexical Bundles. First fixation durations for L1 group.

fixations for less frequent function words (Figure 1-2, panel 2). However, fixation durations on high frequency function words in lexical bundles were relatively inhibited. This may be a reflection of the strings of function words which make up lexical bundles. That is, if the word *n*+1 in parafoveal view is a function word, the reader will likely pause on the current word for parafoveal information uptake.

Multiple function words in a MWU, as is the case for lexical bundles, may lead to more frequent pauses to access the *n*+1 word. In contrast to lexical bundles, consider the effect of function words in restricted collocations such as *down* in *look down on, cried her eyes out, make light of, get the message over, do his bidding, come across*. Note that function words in restricted collocations, specifically prepositions, carry unique modification of the content word. That is, function words in restricted collocations are often semantically much richer than in normal sentences. For example, *out* in *cry her eyes out* is not directional or deictic but figurative; likewise *across* in *it came across as* does not refer to literal directional opposition. Non-literal prepositions and/or particles are highly frequent in restricted collocations, and occur in nearly half of these MWUs; this could cause significant perturbation of the traditional function word effect. Conversely, many of the head verbs in the restricted collocations are semantically impoverished, so-called 'light verbs' (Jespersen, 1954; Kearns, 1988/2002), for instance *give* in *give rise to*, and *come* in *come across*. The semantic load of these verbs is more similar to that of function words in normal sentences. Additionally, articles used with light verbs cause semantic change in terms of foregrounding and backgrounding (Kearns, 1988/2002). As a consequence, the balance of semantic richness is reversed for the restricted collocations, leading to no significant difference in first fixation durations between the (semantically richer) function words and content words.

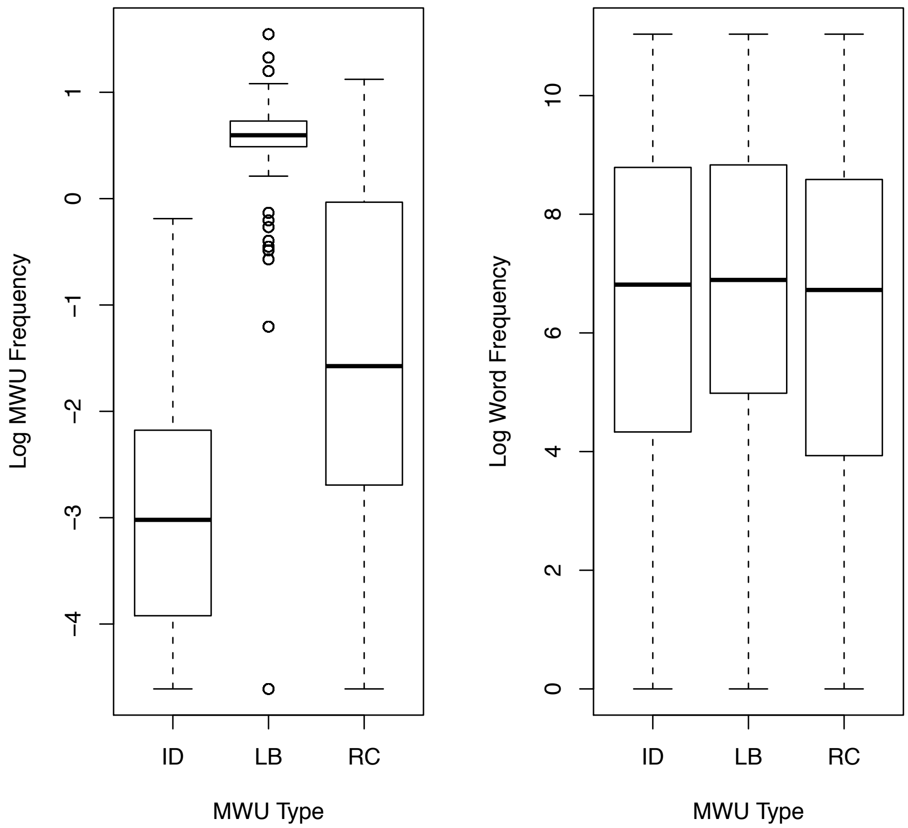

Figure 1-3. MWU type differences in logged MWU frequency (left; significant) and in logged individual word frequency (right; non-significant).

As a final note on word status effects in first fixation durations, function words have longer first fixation durations (Figure 1-2, panel 2) than content words (panel 1) when they are more frequent. This content word interaction is also illustrated in the interaction of MWU type and word frequency in panel 1. There, we see that a greater word frequency affords shorter fixation durations for words in lexical bundles and restricted collocations. The interaction of word frequency by word status indicates that this advantage of roughly 25 ms for high frequency words, as opposed to low frequency words, is restricted to the content words. For function words, this 25 ms advantage is offset by a 35 ms disadvantage, resulting in an overall null effect. This supports findings in other paradigms of ceiling effects on function words (e.g., Bell et al.'s reduction study, 2009). That is, higher word frequency for function words is not a reliable predictor of faster lexical access as it is for content words (see e.g., Gordon & Caramazza, 1985; Segalowitz & Lane, 2000).

The panels in Figure 1-4 illustrate three further effects. Again, note that the figure depicts the effects when all other variables are held constant. As MWU familiarity increases (i.e., more negative scores, leftward in panel 1), fixation durations increase for function words, but decrease for content words. As the MWU becomes more unfamiliar, readers spend more time on content words and less time on function words. This fits with models of reading where parafoveal uptake of function words is more likely the longer fixations on neighbouring content words are (Kliegl, 2007; Rayner et al., 2007). Alternatively, the lack of familiarity may lead to longer fixation on the content words for meaning. Less complex structures, as indexed by the Flesch Reading Ease Formula, also lead to shorter fixations (panel 2). Another consideration for first fixation duration times is the position of the word being read relative to the MWU (panel 3). Words occurring within the MWU have significantly faster first fixations than words occurring both after the MWU and before the MWU.

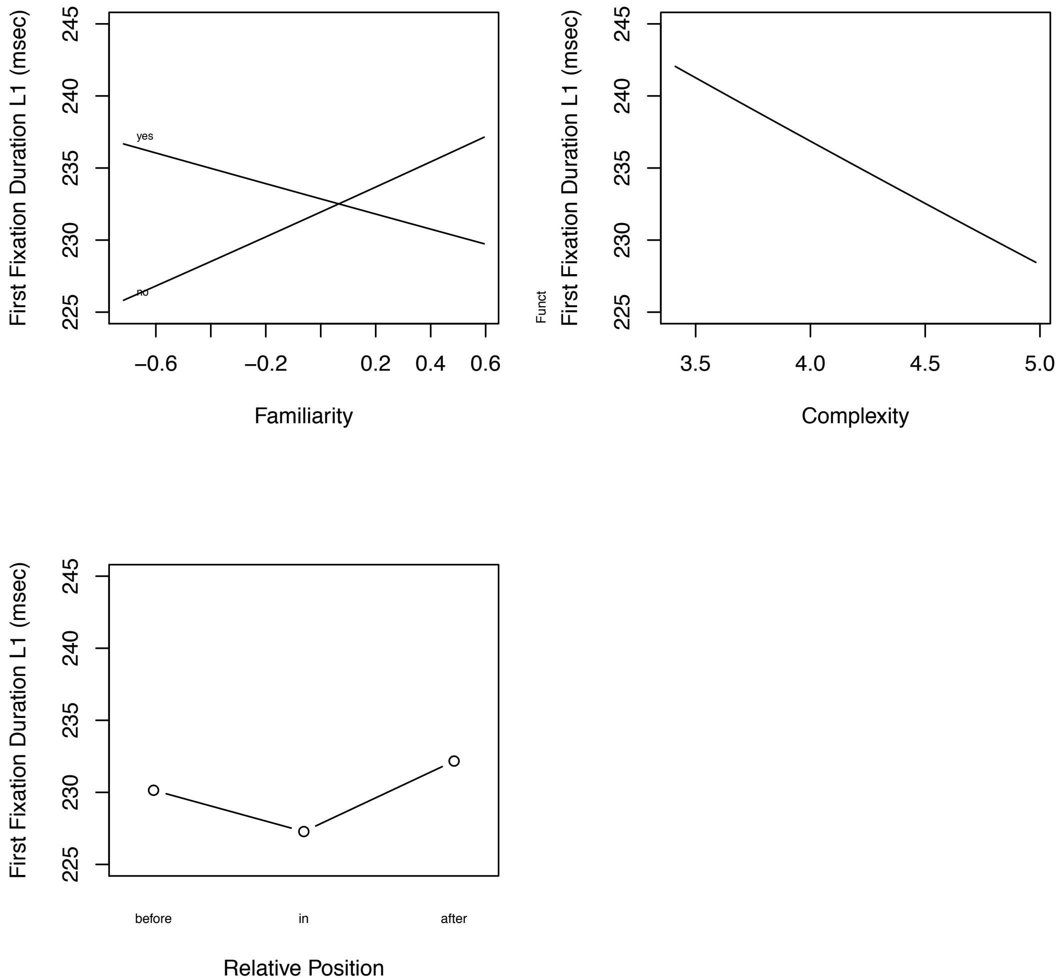

Figure 1-4. Top row panels 1-2: (L) Familiarity interaction with Word Status, (R) Complexity; bottom row panel 3: Relative Position. First fixation durations for L1 group.

Overall, even on the first fixation of a word some recognition of MWU as joint units must occur, as the as the idiom results in particular show this effect is independent of word frequency effects. We turn our attention now to the analysis of the non-native group's first fixation durations.

L2

The model fitted to the first fixation duration of the L2 data required random intercepts for Word, Sentence and Subject (estimated standard deviations 0.0487, 0.0172 and 0.1197 respectively), with the standard deviation of the residual error at 0.3448, and by-subject random contrasts for Word Status (estimated standard deviation 0.0821). Predictors considered were MWU Type, Word Frequency, Word Status and Relative Position, with control variables Font Size and Sentential Position. Table1-4 lists the predictors and the significance of their addition to the model in accounting for the variance.

Table 1-4. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the First Fixation Duration L2 model*

Predictor	**AIC**	**Df**	***p***
MWU type	43704	10	–
Word frequency	43681	11	<0.0001
Word status	43677	12	0.0113
Sentential position	43605	13	<0.0001
Font size	43553	14	<0.0001
Relative position	43549	16	0.0136

Firstly, as we see in Figure 1-5 (panel 1), words in sentences with restricted collocations elicited significantly longer first fixation durations than words in sentences with lexical bundles. This effect is completely independent of whether the word is in the MWU region or not. The words in sentences with idioms elicited average first fixation durations that were intermediate between those of lexical bundles and the words in restricted collocation sentences.

Pairwise comparisons did not reveal any significant differences for idioms, at least not for first fixation durations.

Secondly, words in the MWU area generally have significantly shorter first fixation times among L2 readers than words occurring before the MWU (Figure 1-5, panel 2). First fixation durations on words before the MWU are also significantly shorter than first fixations on words occurring after the MWU. It seems that the MWU introduces some interference which then disrupts/delays the reading times for the remainder of words in the sentence. Alternatively, if the MWU is visible in parafoveal view, the longer first fixation durations could indicate recognition. It is possible that the readers are fixating on the word for longer while extracting more information through parafoveal uptake. Relative Position did not interact with MWU type.

Thirdly, recall that for the L1 first fixation durations, there was an interaction for word frequency by word status (i.e., content versus function words). In contrast, the L2 fixation durations have simple main effect of word frequency across all word types (panel 3). Finally, as we can see in Figure 1-5, panel 4, there was a significant main effect for Word Status: as expected, function words had faster first fixation durations than content words. However, as mentioned earlier, participants showed significant differential sensitivity for this predictor, as indicated by inclusion of significant by-subject random contrasts for Word Status ($\chi^2 = 33.078$, $p < 0.0001$). Some non-native speakers had shorter reading times on the function words within a sentence, while others took longer to

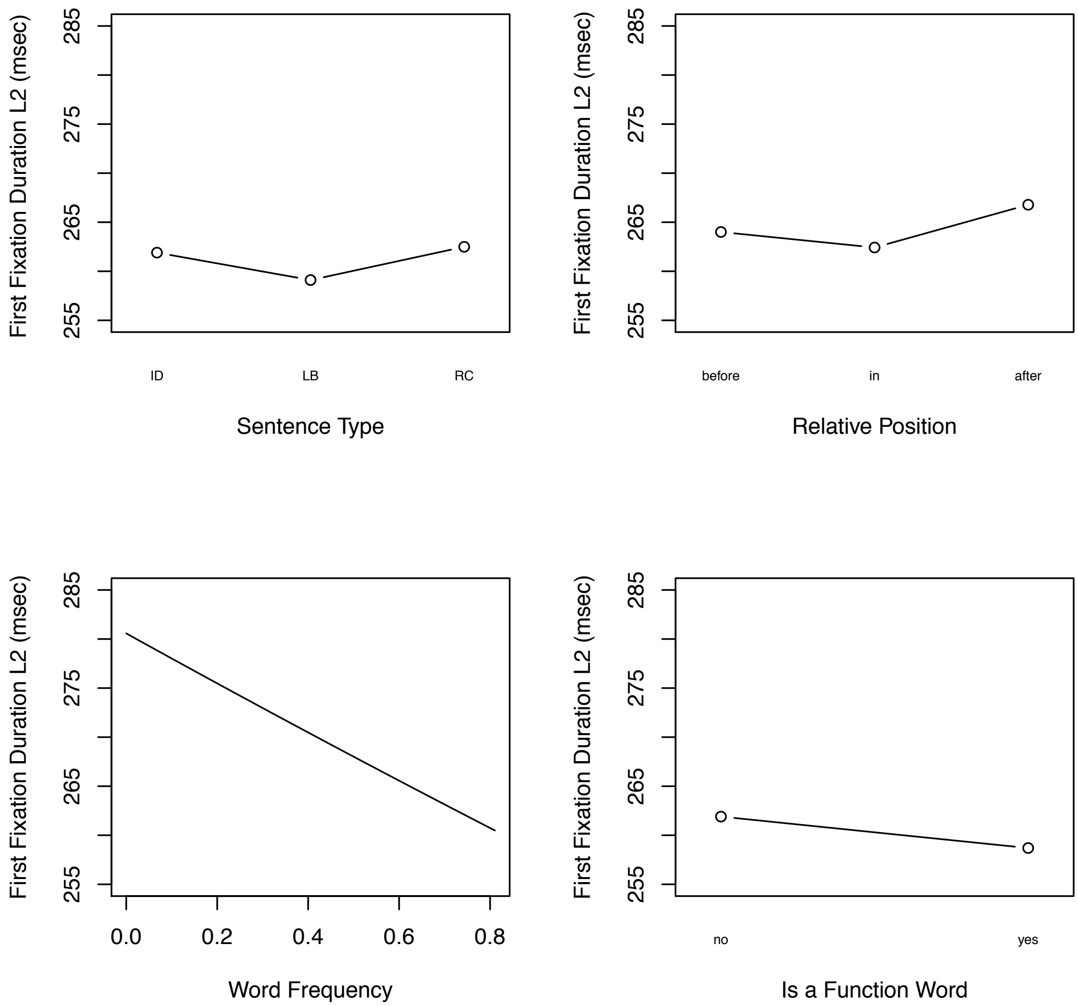

Figure 1-5. Top row panels 1-2: MWU Type, Relative position; bottom row panels 3-4: Word frequency and Word status. First fixation duration for L2 group.

read function words. This varied behaviour in the first fixation durations is a characteristic of the L2 population that was not found for the L1 readers. It is to a more systematic comparison of the two groups of readers that we now turn.

Joint analysis of L1 and L2 readers

The joint data for both groups of speakers were analysed with Group (L1 versus L2) as a between-subjects factor. An initial model was subjected to model criticism and potentially harmful outliers (SD +/- 2.5) were removed (2.8% of datapoints). The model fitted to the remaining datapoints (Both Groups *n*= 64542) required random intercepts for Word, Sentence and Subject (estimated standard deviations 0.0518, 0.0223 and 0.1427 respectively) with the standard deviation of the residual error at 0.34. As in the preceding analyses, predictors considered were MWU Type, Word Frequency, Word Status, Start of MWU, Relative Position, with Font Size and Sentential Position as control variables. The values for the log likelihood ratio test supporting the predictors in this minimally-adequate model are listed in Table 1-5. Of interest to the comparison of L1 and L2 readers' first

Table 1-5. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the First Fixation Duration Joint L1 and L2 model*

Predictor	**AIC**	**Df**	***p***
Group	75667	6	–
MWU type	75668	8	–
Group * MWU type	75663	10	0.016
Word frequency	75657	11	0.0043
Group * Word frequency	75629	12	<0.0001
Word status	75623	13	0.0053
Group * Word status	75618	14	0.0092
Relative position	75507	16	<0.0001
Sentential position	75464	17	<0.0001
Font size	75413	18	<0.0001
MWU start	75410	19	0.0171

fixation durations are the interactions involving Group that reached significance: Group by MWU Type, Group by Word Frequency, and Group by Word Status.

Firstly, we see in Figure 1-6, panel 1, that, compared to native speakers, non-native speakers have small but significant (approximately) 4 ms shorter first fixations on words in lexical bundle sentences relative to those on words in idiom and restricted collocation sentences. The lexical bundles are probably relatively easier to process for L2 speakers due to their semantic transparency. Furthermore, only the non-native speaker group benefits from word frequency in that first fixations are faster with increasing word frequency (Figure 1-6, panel 2). This effect is in accordance with previous studies (Gollan, Montoya, Cera, & Sandoval, 2008; Duyck, Vanderelst, Desmet & Hartsuiker, 2008), where stronger frequency effects have been found for non-native speaker groups over native speaker groups. There was no evidence in this joint analysis supporting a three-way interaction between word status and word frequency by group. In the analysis of native speakers, we saw that a frequency effect was present for content words. Apparently this restricted frequency effect does not survive pooling with the data of more variable non-native speakers. Instead, L2s have significantly faster first fixation durations on function words than the L1 group does (Figure 1-6, panel 3). This fits with the non-native speakers' not having competing meanings for the function words in idioms and restricted collocations.

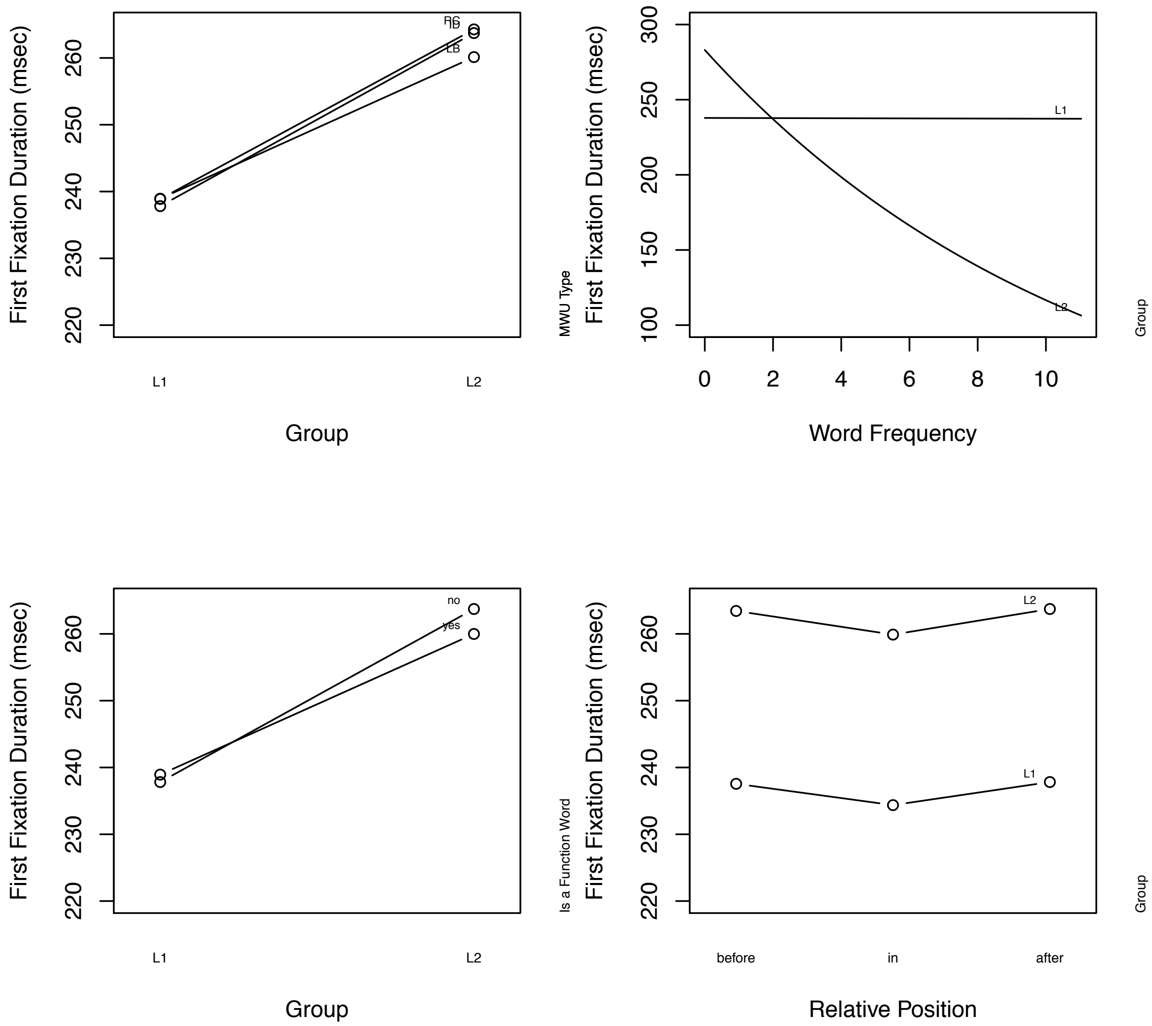

Figure 1-6. Top row panels 1-2: MWU Type by Group, Word frequency by Group (note the different scale on the y-axis); bottom row panels 3-4:, Group by Word status; Relative position of word to MWU in each Group (interaction non-significant). First fixation durations for the combined L1 and L2 speaker group.

As shown in Table 1-5, we find that the further into the sentence the MWU begins, the shorter the first fixation durations are throughout the sentence, for both speaker groups. Additionally, there was no significant difference between the non-native and native speaker first fixation durations on words within the MWU: Both groups had shorter fixations on words within the MWU compared to words before or after the MWU (Figure 1-6, panel 4). Overall, then, non-native speakers are disadvantaged when reading semantically irregular MWUs whereas native speakers are advantaged, at least in the first fixations. The most useful predictors overall for first fixation duration times are discussed in the summary below.

In sum, the first fixation duration data shows that the MWUs significantly affect processing times. That is, durations within the MWU region are significantly shorter than on words following the MWUs. The inherent features of MWUs, then, aid L1 reading across all MWU types, and impede L2 reading of idioms as measured by first fixation durations. In addition, higher frequency words elicited shorter fixations, and function words had shorter initial fixations than content words. These effects were stronger for L2s than L1s. When a MWU appears later in a sentence, this also leads to shorter first fixation durations on average.

The analysis of the native speaker data provides additional information on this subset of speakers that does not emerge in the analysis of the combined groups, as the data introduce relatively more variance. For the L1 speakers, we see that individual word frequency and word status interact with the MWU type. That is, the frequency and word status of the words in the MWU affect reading times. Very frequent content words in lexical bundle and restricted collocation sentences have significantly shorter first fixation durations than frequent words in idiom sentences. Compared to function words in idiom sentences, function words in lexical bundle sentences also have significantly shorter first fixation durations, except for very high frequency function words. It is interesting in this data that the function words do not have the expected general advantage for native speakers of being read faster than content words in the first fixation. With respect to content words, however, the longer first fixations on frequent content words in idiom MWUs suggests radical semantic competition. That is, the idiomatic meaning and the meaning of the individual frequent word may or may not be related through metaphoric extension (e.g., *he'll have to* <u>*pay for*</u> *that* [decision/action etc.] versus *he kicked the bucket*) or decomposition (e.g., *she let the cat* (secret) *out of the bag*). In either case, the reader has to suppress the semantics of a frequent word in an idiom in order to activate the semantics of the entire idiom, causing a delay in processing.

On the other hand, we note that MWU frequency is not a significant predictor in any of the first fixation duration analyses. This is somewhat surprising given previous results for advantages with higher MWU frequency, including Tremblay's (2009) self-paced reading experiment and Arnon and Snider's (2010) phrasal decision task. There are several reasons that MWU frequency may not have contributed to the first fixation durations in this study. The first is that first fixation durations may be too early consider for MWU effects (although we have found MWU-type differences in the analyses above). The second potential reason is that the MWUs in this study are presented within sentences without requiring each word to be looked at, unlike both Tremblay's and Arnon and Snider's designs. There is the possibility that the varying sentence context has obscured subtle MWU frequency effects by including both words within and without of the MWU in this analysis. We ran a post hoc analysis of the L1 first fixation duration model using only the reading times for words occurring within the MWU (and excluding sentence-based predictors). In the model which included only the words within MWUs, we again did not find a significant effect for MWU frequency. Therefore the finding that MWU frequency is not predictive of first fixation durations seems solid. We shall see whether MWU frequency and those predictors which were found for first fixation durations have similar effects on total word reading times in the next section.

Results and discussion: Word Reading Time

Separate models were first fitted to the log transformed Word Reading Times for the L1 and L2 data. Predictors in the models were MWU Type, Word Frequency, MWU Frequency, Relative Position, MWU Region, Start of MWU, Familiarity, Semantic Transparency, Word Status, and two control variables, Font Size and Sentential Position. Each model was subjected to model criticism, and potentially harmful outliers were removed (SD +/- 2.5; L1 n=98.5%, L2 n=98.3%). The models were then fitted to the remaining datapoints (L1 n= 28393; L2 n= 36644).

L1

For the L1 data, likelihood ratio tests supported random intercepts for the Sentence, Word and Subject random effect factors (estimated standard deviation parameters 0.0692, 0.1376 and 0.0733 respectively; the estimated standard deviation of the residual error was 0.4485). By-subject random slopes were supported for Word Frequency (estimated standard deviation 0.0073), as well as random contrasts for Word Status (0.051), and Relative Position (0.0551). Interactions were found for Sentential Position by Font Size, Word Status by Word Frequency, and MWU Type by Relative Position by MWU Start. Table 1-6 lists the log likelihood ratio test values for each predictor added to the model.

Firstly, similar to what we reported for first fixations durations, faster total word reading times are modulated by the frequency of the content words, and not

Table 1-6. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the total Word Reading Time L1 model*

Predictor	**AIC**	**Df**	***p***
MWU type	42519	17	–
Relative position	42498	19	<0.0001
Sentential position (polynomial)	42426	22	<0.0001
Word Status	42420	23	0.0048
Font size	42419	23	<0.0001
Sentential position * Font size	42369	27	<0.0001
Word frequency	42324	28	<0.0001
Word status * Word frequency	42304	29	<0.0001
MWU Start	42292	30	0.0002
MWU Type * MWU Start	42291	32	0.1252
MWU type * relative position * MWU Start	42288	42	0.0103

by function words. Here, word frequency interacts with whether a word is a function word or not. Content words are read considerably faster when they increase in frequency, and have significantly slower word reading times when they are low in frequency. For function words, frequency was not predictive. That is, participants do not read frequent function words like *the* or *is* quickly but they do read frequent words like *happy* more quickly.

Secondly, there is a small but significant effect of where the MWU begins in a sentence (Table 1-6). The further into the sentence a MWU begins, the shorter the total word reading time durations for all words in the sentence. Apparently, the later the MWU occurs in the sentence, the less disturbance there is to reading overall. Furthermore, native speakers clearly differentiate between MWU types in interaction with both the MWU Start and the Relative Position to the MWU**.** In Figure 1-7, panel 1 we see that in restricted collocation sentences words occurring in the MWU have significantly shorter reading times than after the MWU when the MWU begins earlier in the sentence. If the MWU begins later in the sentence, then words in the MWU have significantly longer word reading times, and the words after the MWU have significantly shorter reading durations. In contrast, words occurring before the lexical bundle have significantly slower word reading times than words occurring after or within the lexical bundle, but significantly faster times than words within a restricted collocation MWU. Furthermore, words occurring after the MWU in idiom sentences have significantly faster word reading times than those in lexical bundle sentences. Yet words in idiom sentences are read significantly slower than words in lexical

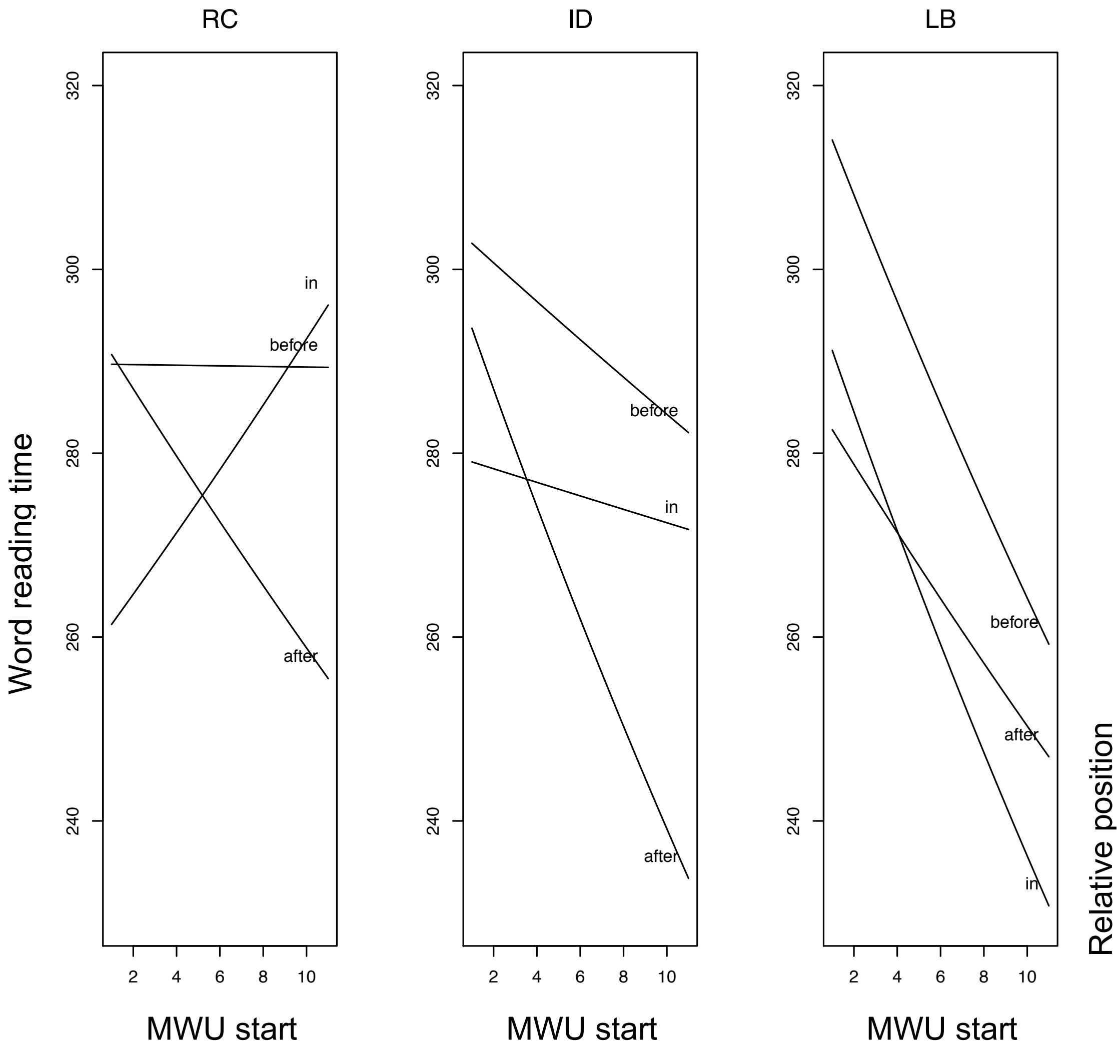

Figure 1-7. Relative position (position of the word relative to the MWU) interaction with MWU Type and MWU Start (the position of the MWU's first word) for restricted collocations (panel 1), idioms (panel 2) and lexical bundles (panel 3). Word reading times for L1 group.

bundles when they occur before or within the MWU region, unless the MWU beings very early in the sentence. This suggests an overall effect of meaning and predictability from the idiom structures. That is, while lexical bundles are highly predictable in terms of word transitions within the MWU, they do not impart any particular sense of resolution. The processing of the idiom creates an information bottleneck, where the meaning is at first hard to resolve, but once processed provides a good frame of reference from which information can be relatively quickly integrated. Such an interpretation is compatible with the findings in the first fixation durations, in that longer word reading times within idioms likely reflect refixations on words prior to and during the idiom unit. This is in line with idioms' typical usage. That is, speakers use an idiom – a commonly-known and understood metaphor and/or analogy – to provide a reference for relating to a situation which is very quickly accessed by the hearer (see e.g., Wray, 2002; Kuiper, 2000; Biber et al., 1999). We turn now to the analysis of word reading time predictors for the L2 data.

L2

The model for the L2 data, like the L1 data discussed above, required random intercepts for Word, Sentence and Subject (estimated standard deviations 0.1878, 0.0491, and 0.1355 respectively), with the standard deviation of the residual error at 0.4646. By-subject random slopes were supported for Word Frequency (estimated standard deviation 0.1827), MWU Frequency (0.0082), and random contrasts for Word Status (0.2507), and MWU Region (0.0315). Predictors were MWU Type, Word Frequency, MWU Region, Word Status, Complexity, and the control variable Sentential Position. (Relative Position to the MWU of the word being read and Font Size were not predictive for the second-language speakers.) Interactions were found for MWU Type by MWU Frequency by Start of MWU, Word Status by MWU Region by Word Frequency, and MWU Type by Semantic Transparency by MWU Region (see Table 1-7).

Table 1-7. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the total Word Reading Time L2 model*

Predictor	**AIC**	**Df**	***p***
MWU type	58740	15	–
Sentential position (polynomial)	58704	18	<0.0001
MWU start	57741	19	<0.0001
MWU frequency	57740	20	–
*MWU start * MWU frequency*	57742	21	–
MWU start * MWU frequency * MWU type	57736	27	0.0074
Word status	57701	28	<0.0001
In MWU region	57700	29	–
Word status * In MWU region	57691	30	0.0011
Word frequency	57672	31	<0.0001
Word status * In MWU region * Word frequency	57622	34	<0.0001
Semantic transparency	57624	35	–
*Semantic transparency * MWU type*	57625	37	–
(Semantic transparency + MWU type + In MWU region)^3	57610	42	0.0003

Figure 8 illustrates a three-way interaction that emerged for the non-native speakers between MWU type, MWU frequency and the start position of the

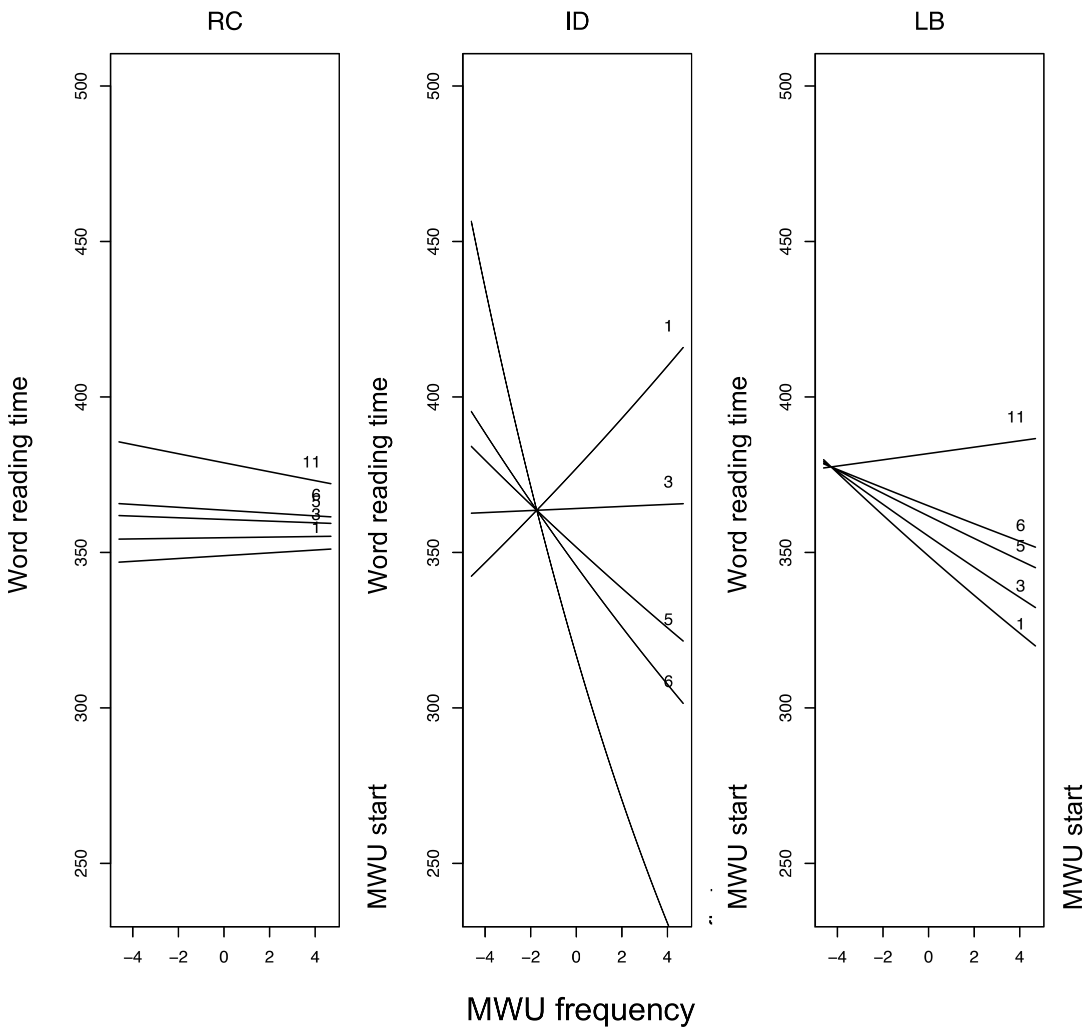

Figure 1-8. Left to right: MWU frequency interaction with the Start of MWU position (viz. number of words into the sentence) by Restricted collocation (panel 1), by Idiom (panel 2), and by Lexical bundle (panel 3) sentences. Word reading times for L2 group.

MWU. It is clear that words in sentences with restricted collocations are significantly less affected by MWU frequency (panel 1) than words in idiom sentences (panel 2) according to when the MWU begins in the sentence. For idiom sentences, the further into the sentence the MWU begins leads to the most advantage for word reading times as MWU frequency increases. Words in lexical bundle sentences (panel 3) appear to follow a similar trend to those in idioms, but this effect is non-significant. On the other hand, words in restricted collocation sentences have almost no decrease in word reading times as the MWUs become more frequent. This decrease for word reading times in idiom sentences is likely to be a manifestation of increasing predictability. Note that this effect for frequent idioms was not found in the first fixation duration data.

Secondly, Figure 1-9, panel 2 shows that function words for non-native readers are not significantly facilitated by word frequency when within the MWU area. The L2 group does, however, have an effect of frequent content word reading (panel 1). While function words are read quickly, there is no frequency effect. But content words have a clear frequency effect, and this is much stronger when occurring outside of the MWU. Content words inside the MWU area have significantly faster word reading times than content words outside of the MWU region. As was the case for the first fixation duration data above, this is clearly semantic competition: the content words' frequencies are presumably trying to connect the words with their canonical meanings, rather than their meanings as part of a wider unit, which may or may not be related to a word unit's meaning. That is, for some idioms and restricted collocations, meaning can be obtained by metaphoric extension of each individual word's meaning while other restricted collocations' or idioms' components could be semantically empty, and take on meaning only in terms of the MWU unit semantics. For example, the idiom *blow*

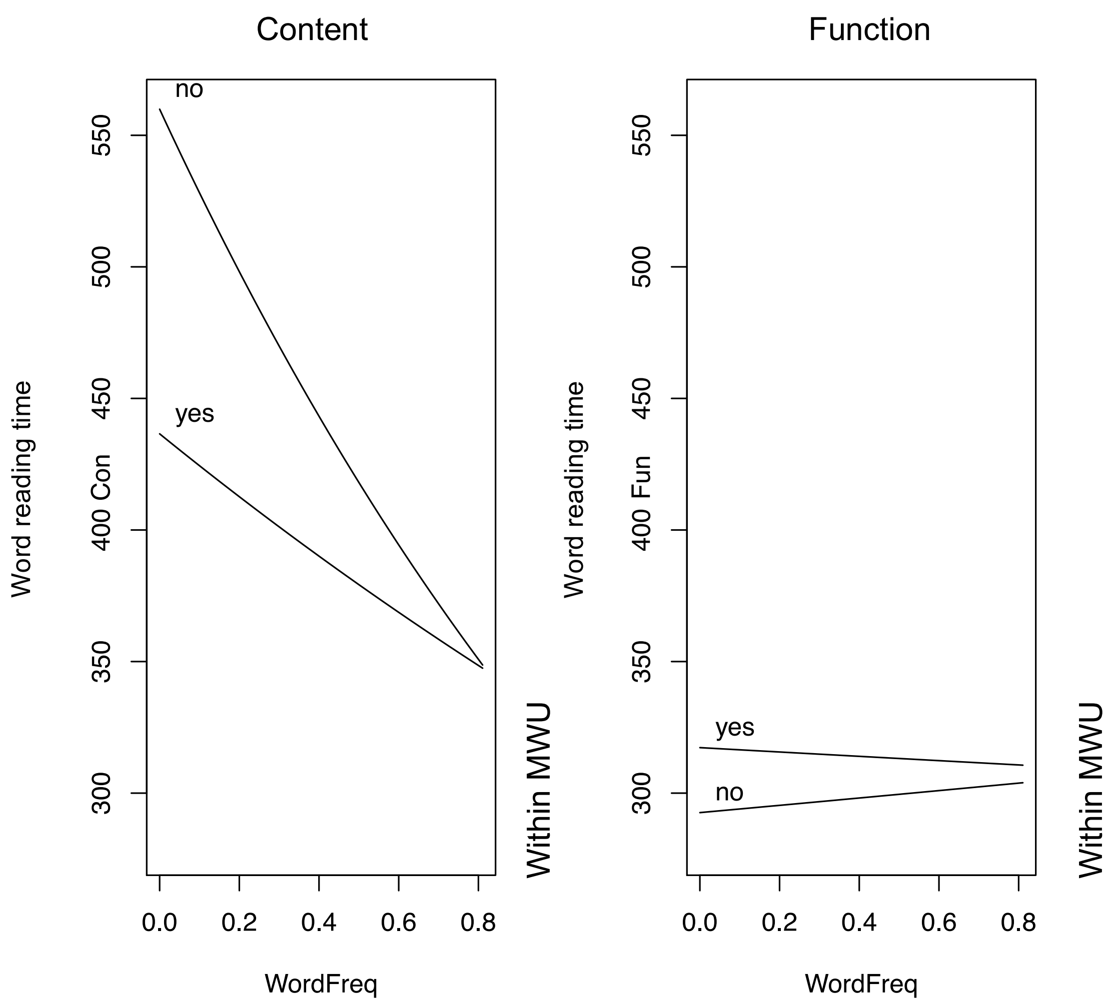

Figure 1-9. Left to right: Word frequency interaction with word position within ('yes') and outside of ('no') of MWU region by Content words (panel 1), and by Function words (panel 2). Word reading times for L2 group.

your top means 'get angry' while *spill the beans* can be broken down into *spill* = 'tell', *the beans* = 'a secret'. Restricted collocation examples include *button up* ('be quiet') vs *pay attention to* (*pay* = 'give', *attention* = 'attention'). L2 word reading times, then, seem to derive from a focus on meaning of content words in sentence processing. There are no effects for function words.

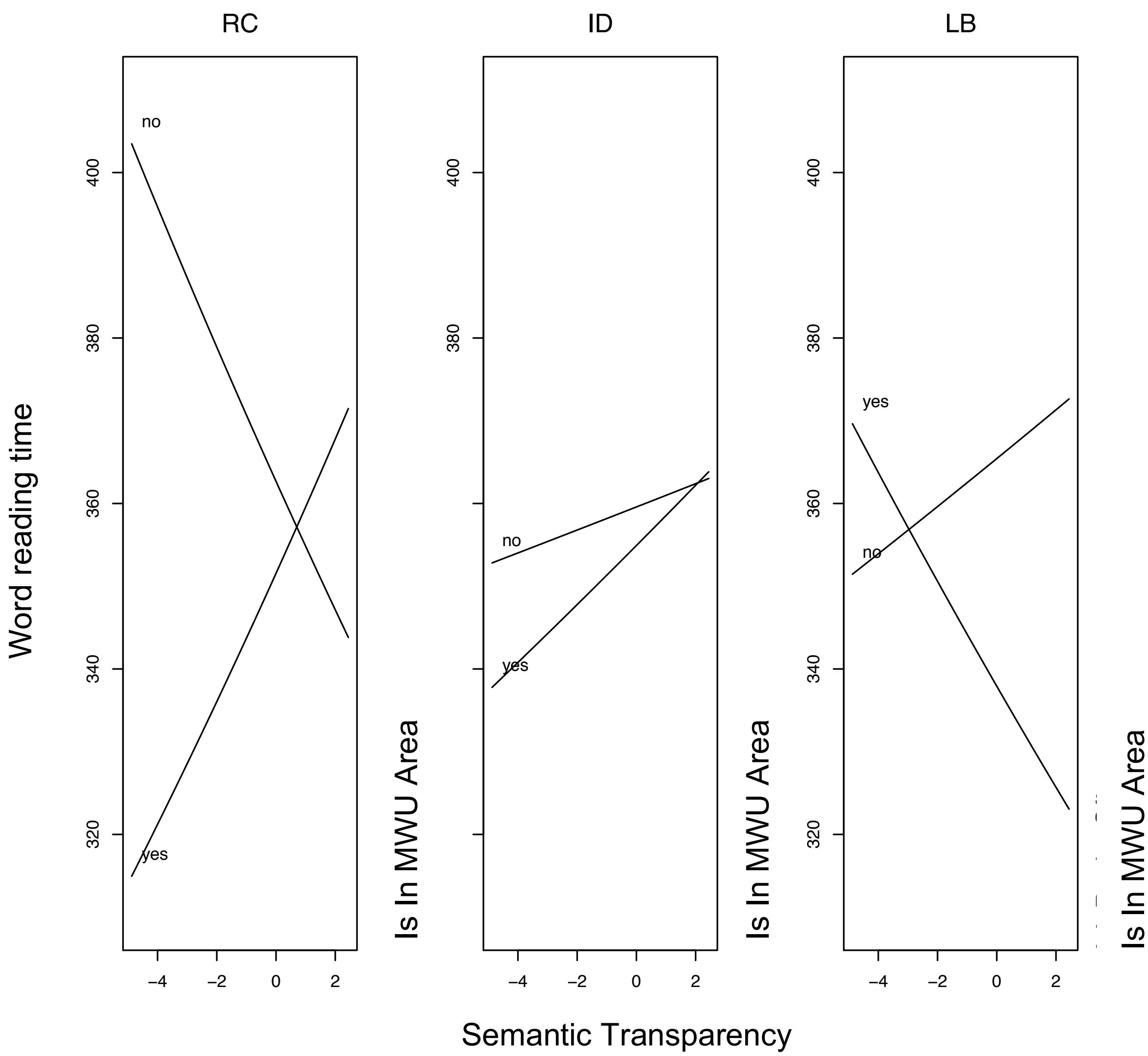

Figure 1-10. Left to right: Semantic transparency interaction with word position in (‘yes’) and out (‘no’) of MWU region by Restricted collocation (panel 1), by Idiom (panel 2), and by Lexical bundle (panel 3) sentences. The more transparent a MWU is rated, the lower the number. Word reading times for L2 group.

Finally, in addition to both frequency effects, word reading times for non-native speakers are also affected by the semantic transparency of the MWU in the sentence, and this also varies by MWU Type. Figure 1-10, panel 1 shows words in restricted collocation sentences are read slower with increasing MWU opacity, yet words in lexical bundle sentences (panel 3) have faster word reading times than both restricted collocations and idioms (panel 2) with higher MWU opacity. This is likely a reflection of the higher transparency of lexical bundles compared to restricted collocations and idioms. When the words are outside of the MWU area, the opposite is found. Words in restricted collocation sentences are facilitated by the transparency of the MWU, while words in lexical bundle and idiomatic sentences are inhibited**.** We turn now to the comparative word reading time analysis between the two speaker groups.

Joint analysis of L1 and L2 readers

A joint analysis model was fitted to data for both groups with the L1 versus L2 Group predictor and interactions added. This model was then trimmed to remove potentially harmful outliers, (SD +/- 2.5) and the model fitted to the remaining datapoints (Both Groups *n*= 65312, 98.4%). The model required random intercepts for Word, Sentence and Subject (estimated standard deviations 0.1897, 0.0547 and 0.2058 respectively) with the standard deviation of the residual error at 0.4638. Predictors for the Both Groups model were MWU Type, Word Frequency, Word Status, Start of MWU, Relative Position, Group, with the control variables Font Size and Sentential Position. Interactions considered were Group by MWU Type, MWU Type by Relative Position, and Group by Word

Status by Word Frequency, Group by Start of MWU, Word Status by Start of MWU, as well as Group by Sentential Position and Sentential Position by Font Size.

Table 1-8. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the total Word Reading Time joint L1 and L2 model*

Predictor	**AIC**	**Df**	***p***
MWU type	101105	7	–
Group	101102	8	0.0211
MWU type * Group	101075	10	<0.0001
Sentential position	101056	11	<0.0001
Sentential position * Group	101031	12	<0.0001
Word status	100965	13	<0.0001
MWU start	100963	14	0.0638
MWU start * Word status	100955	15	0.0011
Group * MWU start	100945	16	0.0007
Word frequency	100831	17	<0.0001
Word frequency * Word status	100825	18	0.0051
Word frequency * Word status * Group	100767	21	<0.0001
Relative position	100724	23	<0.0001
Relative position * MWU type	100705	27	<0.0001
Font size	100707	28	0.5405
Sentential position * Font size	100700	29	0.004

Firstly, in Table 1-8 we note that the L2 word reading times are longer than the L1 readers': this is a significant difference. However, this is to be expected and will not be discussed further. Figure 1-11 illustrates the between-group differences for word frequency by word status (panels 1 and 2), and Figure

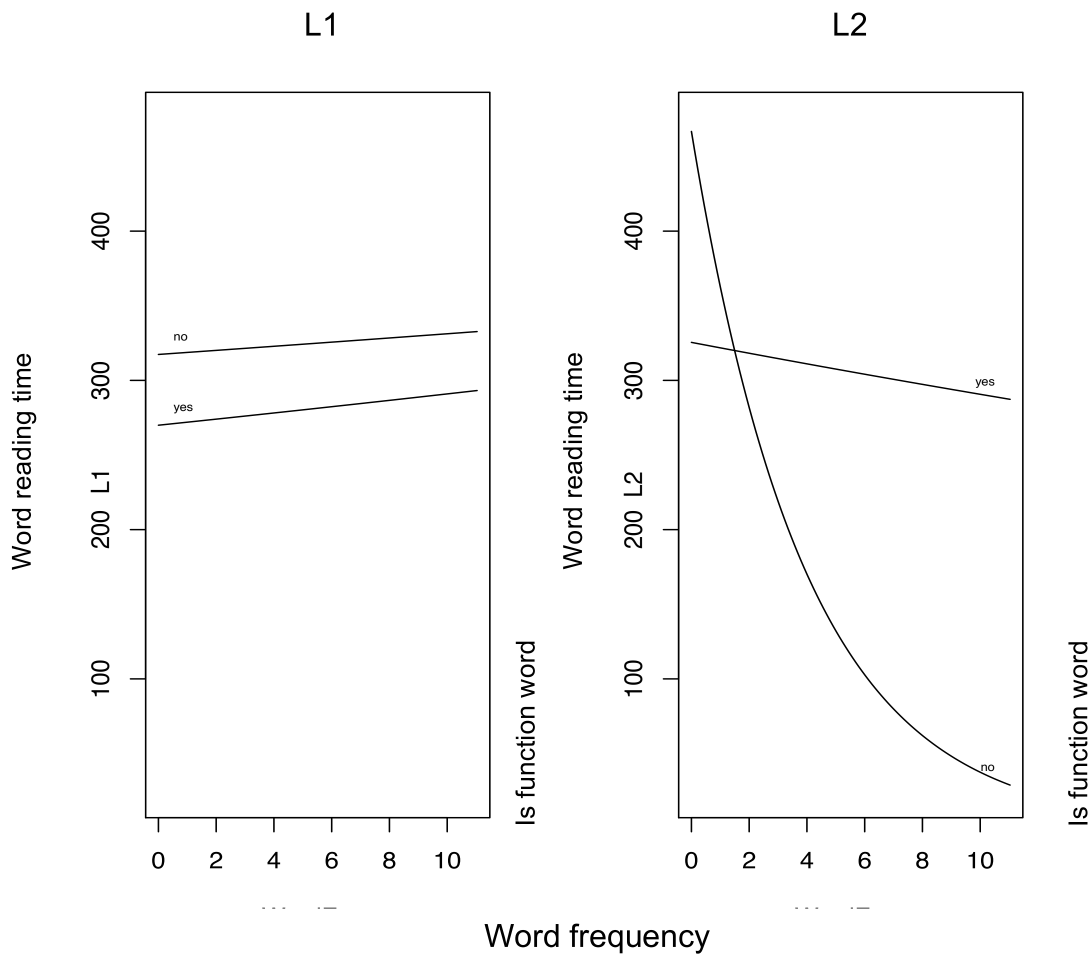

Figure 1-11. Word frequency by Word status for L1 (panel 1) and L2 (panel 2).

Word reading times for L1 and L2 combined groups.

1-12 illustrates the non-significant group differences in relative position by MWU type (panels 1 and 2). We include this figure as it is of note that both speaker groups process the MWUs in the same way within the sentence. Where the word being read occurs in the sentence makes a difference to MWU type for both the native and non-native speakers. In Figure 1-12, we see that words in lexical bundle sentences are the fastest MWU type to be read within the MWU region, and have the largest increase in reading times after the MWU. Idioms are the slowest to be read within the MWU region. However, both groups read words in idiom sentences significantly faster after the MWU region, suggesting that they are able to predict the remainder of the sentence from the content and context of the idiom. This suggests that idioms afford faster processing of words following the MWU. This is in line with theories that claim that MWUs offer advantaged processing in order to allow extra time for parsing or preparation of the following utterance/phrase (e.g., Wray, 2002; Kuiper, 1996). It is fitting, then, that this is not an effect found for the restricted collocation data. Surprisingly though, given Tremblay and colleagues' strong processing advantages mentioned earlier, we have the opposite effect for lexical bundles.

On the other hand, word frequency and word status reading time trends are dissimilar for the L1 and L2 groups (Figure 1-11). Only the native speaker group has an advantage with faster word reading times for function words, and this is relatively stable regardless of the function word's frequency. However, while native speakers are significantly faster at reading function words over content words, non-native speakers read frequent content words significantly faster than frequent function words. This is expected given, as we noted above, non-native speakers tend to have stronger word frequency effects than native speakers, and this is true of function words as well. In contrast, the L1 group displays a ceiling effect for frequent function words, as has been noted by Segalowitz and Lane (2000) and Bell et al. (2009).

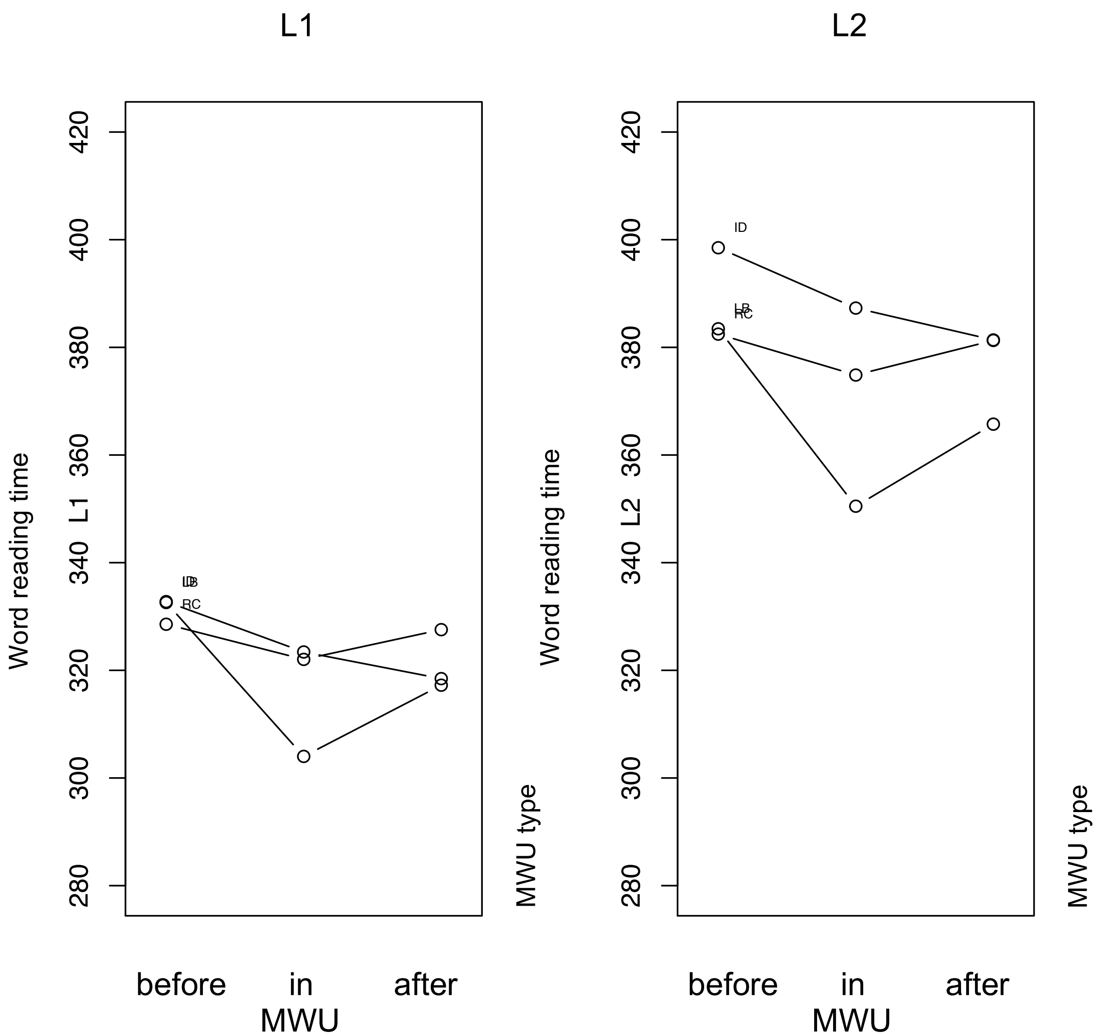

Figure 1-12. Left to right: Relative Position by MWU Type for L1s (panel 1), and L2s (panel 2) (non-significant). Word reading times for L1 and L2 combined group.

A final word-based effect is seen in Figure 1-13. While both the native and non-native speakers read words slower on average when a MWU begins later in the sentence, only non-native speakers have a large inhibitory effect. That is, for

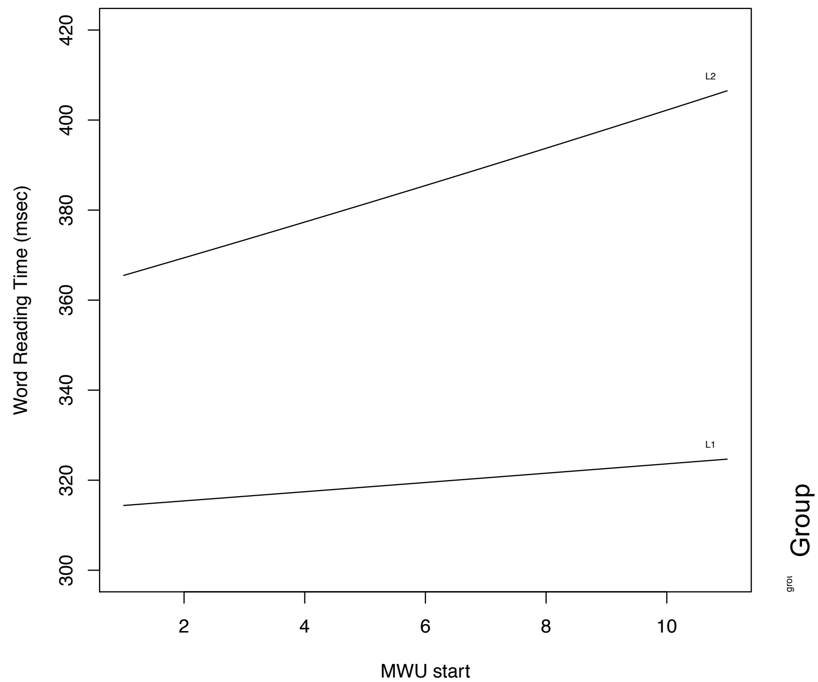

Figure 1-13. Group interaction with MWU Start (panel 2). Word reading times for L1 and L2 combined groups.

the non-native group, significantly slower word reading times are elicited when the MWU begins late in the sentence. Interestingly, the data suggest that the native speakers are not advantaged by increasing predictability as they read further into the sentence.

Overall, the total word reading time data has further revealed MWU type differences, and different (subconscious) processing strategies between the native and non-native reader groups. For example, native speakers read function words faster than content words with lower and mid-frequency, but, unlike the L2 speakers, the L1 group does not read frequent content words faster than frequent function words. In terms of MWU type differences, we find that L1 and L2 speakers both process words in MWUs in the same order, though the difference between idiom and restricted collocations is stronger for non-native speakers.

Note that we do not find a significant difference between groups for MWU frequency. Unlike the L1 and L2 first fixation duration models, there is evidence of a MWU frequency effect for non-natives in the L2 word reading time model. This is of interest as it points to L2 participants having attended to the frequencies of lexical bundles, at least when the MWU occurs early in the sentence. Recall, in the first fixation duration discussion we contemplated that the first fixations may be too early to find a MWU frequency effect (despite other MWU-type effects being found). The interaction of MWU frequency with MWU Start in the L2 data hints at a possible effect being masked by the other words in the sentence. With this in mind, a second post hoc regression model was made for the within-MWU only words, this time for L1 total word reading time (once again removing all sentence-based predictors). Indeed, we do find a MWU frequency effect in the word reading times for L1s. The effect is in an interaction with MWU type, with frequent idioms having significantly shorter word reading times than frequent lexical bundles (t=-2.60) and also shorter than frequent restricted collocations (t=-2.24). Interestingly, the order of effects for the MWU types in the best fit word reading time model is the opposite to the order of effects in the MWU-word only model. That is, when the words surrounding the idiom are not accounted for in the regression model, then words within idioms are the fastest to be read, and lexical bundles are the slowest. This is the order of MWU advantages that we predicted if MWUs were stored holistically. In contrast, when we consider the word reading time model for words in idioms with the full sentence stimulus taken into account, the order is reversed; words in idiom sentences are now the slowest to be read, <u>even when they occur within the MWU</u> (as seen in Figure 1-12, panel 1). It seems that the surrounding context of the sentence does play a role in the processing of the MWU. We turn now to the models which predict the Sentence Reading Times.

Results and discussion: Sentence Reading Time

Separate models were first fitted to the log transformed Sentence Reading Times for the L1 and L2 data. Predictors in the models were Mean Word Frequency, Familiarity, Complexity, and two control variables, Font Size and Sentence Length. Each model was subjected to model criticism, and potentially harmful outliers were removed (SD +/- 2.5; L1 n=98.6%; L2 n=97.7%). The models were then fitted to the remaining datapoints (L1 n= 3,129; L2 n= 4,106).

L1

For the L1 data, likelihood ratio tests supported random intercepts for Sentence and Subject (estimated standard deviation parameters 0.1008 and 0.1741 respectively; the estimated standard deviation of the residual error was 0.1771). An interaction was found for Font Size by Sentence Length. Table 1-9 lists the log likelihood ratio values for the addition of each predictor to the model. Predictors that did not reach significance, among them MWU Type, are not listed in this table, which summarises a minimally adequate model.

Table 1-9. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the total Sentence Reading Time L1 model*

Predictor	**AIC**	**Df**	***p***
Mean word frequency	-454.91	5	–
Font size	-564.73	6	<0.0001
Sentence length	-654.58	7	<0.0001
Font size * Sentence length	-659.68	8	0.0077
Complexity	-672.65	9	0.0001

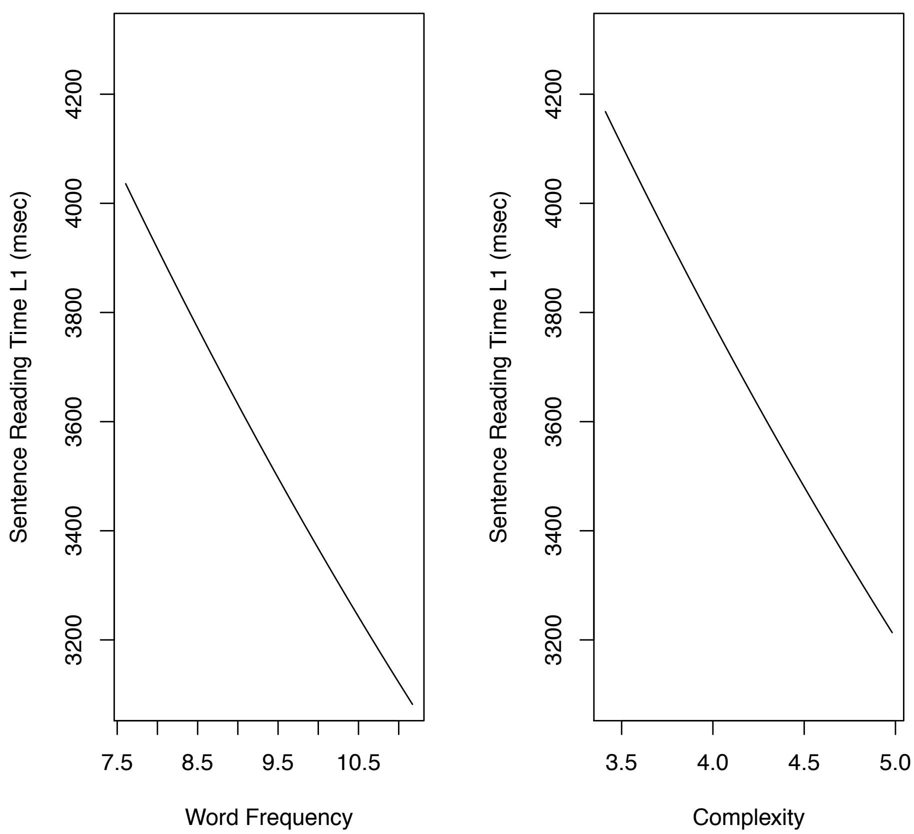

Figure 1-14. Left to right: Word Frequency (panel 1), Complexity (panel 2). Sentence reading times for L1 group.

Sentences with higher mean word frequency were read faster (Figure 1-14, panel 1). The complexity of the sentence structure as gauged by the Flesch Reading Ease Formula was significant, with the expected greater reading times for more complex sentences (Figure 1-14, panel 2).

L2

The model for the L2 data also required random intercepts for Sentence and Subject (estimated standard deviations 0.0708 and 0.3782 respectively), with the standard deviation of the residual error at 0.2124. Predictors included Mean Word Frequency, Familiarity and Complexity, with control variables Font Size 18 and Sentence Length. (Again MWU Type failed to reach significance). An interaction was found for Font Size by Sentence Length (Table 1-10).

Table 1-10. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the total Sentence Reading Time L2 model*

Predictor	**AIC**	**Df**	***p***
Mean Word frequency	1047	5	–
Font size	952	6	<0.0001
Sentence word count	820	7	<0.0001
Font size * Sentence word count	812	8	0.0017
Familiarity	804	9	0.0012
Complexity	798	10	0.005

As would be expected, we find slower reading of more complex sentences as gauged by the Flesch Reading Ease score (Figure 1-15, panel 2). Additionally, the non-native speakers had faster reading times for sentences with higher mean frequency (Figure 1-15, panel 1), supporting Gollan et al.'s (2008) and Duyck et

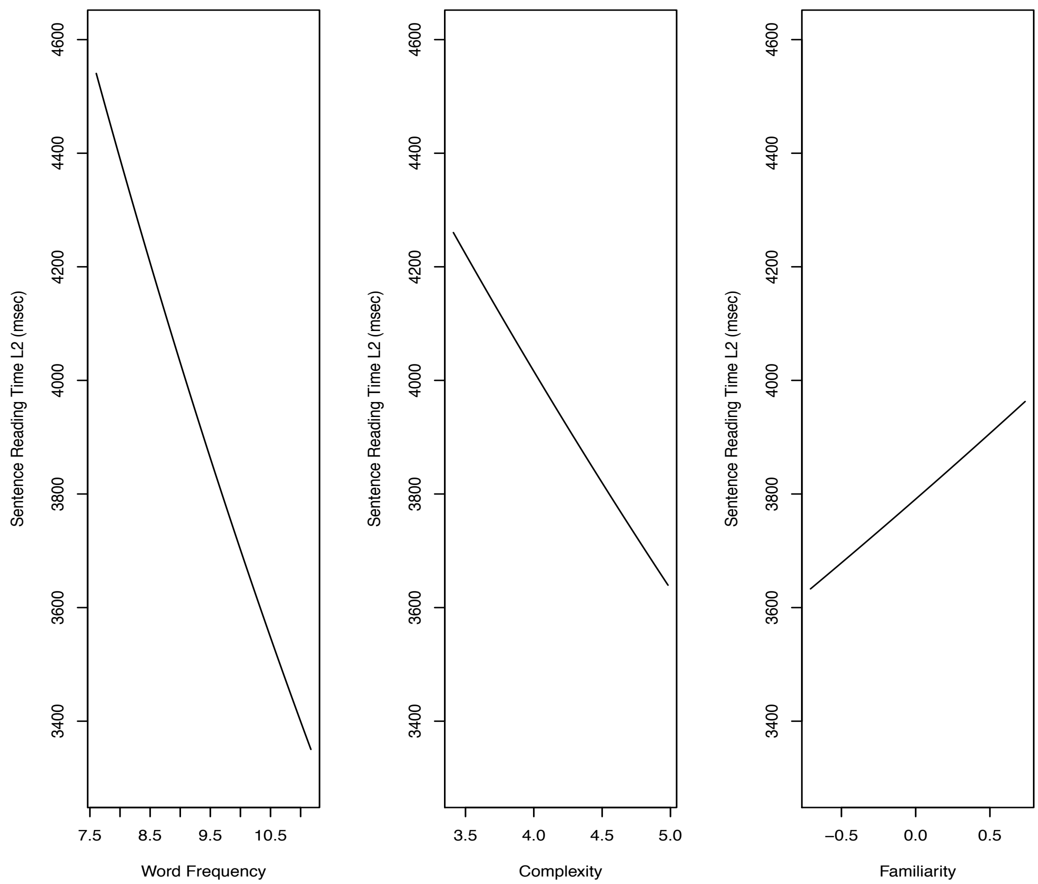

Figure 1-15. Left to right: Word Frequency (panel 1), Complexity (panel 2), and Familiarity (panel 3; higher familiarity is marked with lower numbers). NB lower familiarity scores indicate higher familiarity. Sentence reading times for L2 group.

al.'s (2008) findings of stronger frequency effects in second-language speakers of English. With respect to the familiarity effect, non-native speakers were significantly slower in reading less familiar MWU sentences, but were significantly faster at reading the more familiar MWU sentences (panel 3). This is the only predictor in the L2 data which hints at the MWUs themselves, rather than their inherent qualities.

Joint analysis of L1 and L2 readers

Data for both groups was fitted to a model with significant L1 and L2 effects, with added Group predictors and interactions. This model was then trimmed to remove outliers, (SD +/- 2.5) and the model fitted to the remaining datapoints (Both Groups *n*= 7,222, 97.9%). Sentence and Subject random effects were 0.0804 and 0.3035 respectively, with residuals of 0.1968. Predictors considered in the Both Groups model were MWU Type, Mean Word Frequency, Group, Complexity and Familiarity. Group interactions were investigated for Group by Font Size by Sentence Length, as well as Group by MWU Type by MWU Frequency. MWU Frequency was found to be a predictor in the between-group comparisons, despite failing to reach significance in the individual L1 and L2 models.

Table 1-11 shows significant facilitation for L2 speakers when reading lexical bundles, though more frequent lexical bundles are inhibitory. However, in Figure 1-16, panel 2, we see that in fact idioms are significantly slower for both groups as compared to lexical bundles, and that there is little advantage of MWU Frequency by MWU Type for L2 readers. Instead, we find the L2 readers have significantly less facilitation from frequent lexical bundles than the native speakers do. The three-way interaction also shows the conflict with idiom reading: sentence reading times for native speakers are not faster when reading more frequent idiom sentences, but are faster for reading more frequent restricted collocation sentences.

Table 1-11. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the total Sentence Reading Time joint L1 and L2 model*

Predictor	**AIC**	**Df**	***p***
MWU Type	498.55	6	–
Group	498.55	7	–
MWU Type * Group	494.19	9	0.0153
MWU frequency	496.07	10	–
MWU frequency *MWU type* Group	490.23	15	0.0137
Font size	399.00	16	<0.0001
Mean word frequency	388.11	17	0.0003
Sentence length	265.70	18	<0.0001
Font size * Sentence length	257.22	19	0.0012
Font size * Sentence length * Group	239.56	22	<0.0001
Familiarity	235.83	23	0.0167
Complexity	224.16	24	0.0002
Complexity * Group	217.66	25	0.0036

Between group differences are clear for Familiarity (t= -2.48) and Complexity (t= 2.73) (Figure 1-16, panels 5 and 6). In contrast to the main effects for familiarity and complexity, we find that L2 Group sentence reading times are significantly aided by more familiar MWUs and less complex sentences. Finally, we see in Figure 1-16 (panels 1 and 2) that there is no robust difference for MWU Types by L2 readers, but that L2 readers are significantly slower at MWU reading

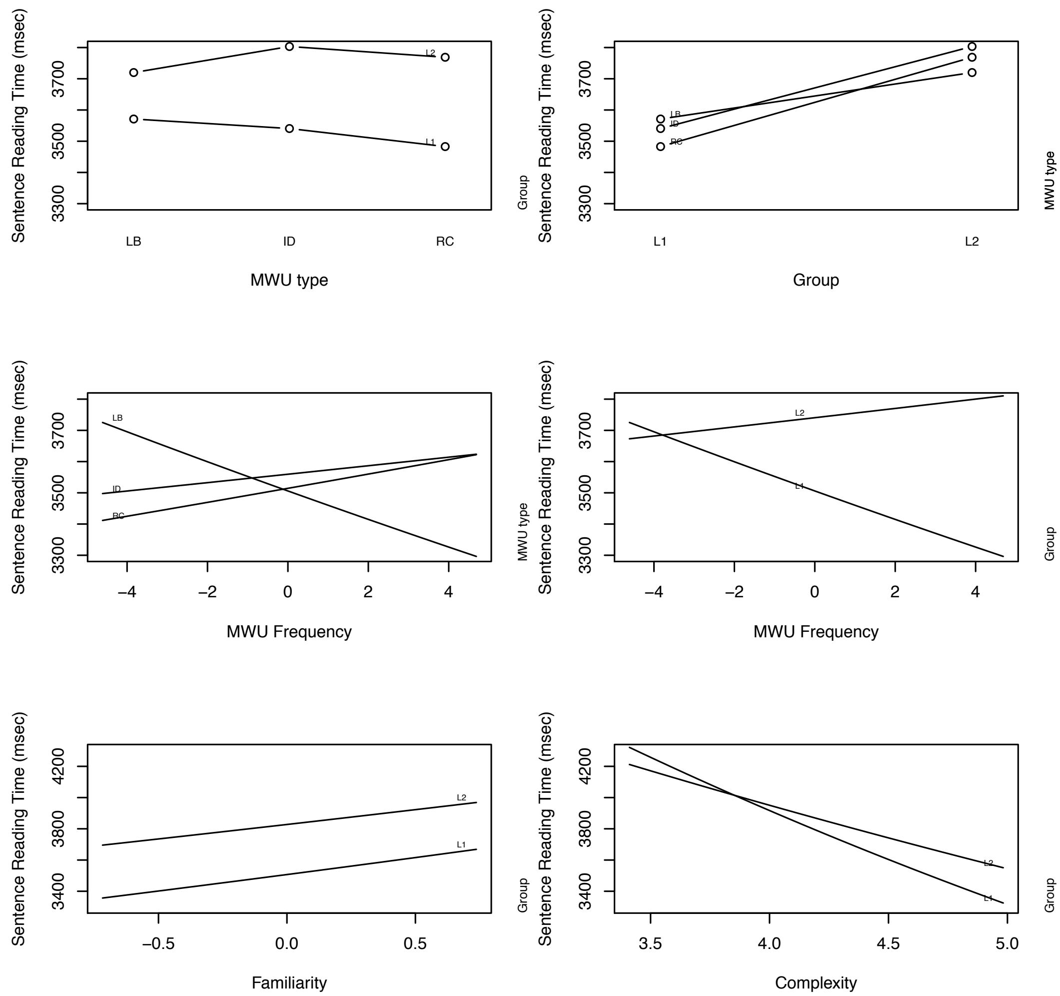

Figure 1-16. Top row left to right: Group by MWU Type (lexical bundle – idiom – restricted collocation) interaction (panel 1), MWU Type by Group interaction (panel 2); middle row: MWU Frequency by MWU Type interaction (panel 3) and MWU Frequency by Group (panel 4); bottom row: Familiarity by Group panel 5; lower scores indicate higher familiarity), Complexity by Group (panel 6). Sentence reading times for L1 and L2 combined group.

than their L1 counterparts. There is, however, a significant difference between L1 and L2 sentence reading times for idioms and lexical bundles.

The sentence reading time results for L1 and L2 English speakers show a stable effect of word frequency, as well as control measures such as Font Size and Sentence Length. More relevant to the MWU processing are the between-group differences in certain frequency- and context-based predictors. That is, the L2 group is significantly facilitated by MWU frequency in sentence reading times, whereas this is not a significant main effect for L1 speakers. Instead, competition from semantics over frequency is apparent in the L1 reading times, but not in the L2 reading times, suggesting that non-natives' reading of multiword units is more reliant on the (in)complexity of the MWU structure than the semantic anomalies. Further, while lower familiarity and higher complexity are significantly inhibitory for both groups, they are both more significantly inhibitory for the L2s compared to the L1 group. MWU frequency is more important for L2 reading times, with weaker slopes for facilitation from more familiar MWUs and more transparent sentences.

General discussion and conclusion

This study set out to determine how three MWU types (idioms, restricted collocations and lexical bundles) embedded in natural sentences might be distinguishable through fixation durations and reading times. Additionally, we asked which predictors were responsible for any differences found. The results above show clearly that words in each MWU type are read in a different manner by native and non-native speakers, as seen in the first fixation durations and word reading times. However, all disturbances and advantages to MWU reading from frequency, probability, position and transparency predictors are resolved at the level of sentence reading time. That is, any delays in fixation durations caused by e.g., semantic opacity of or the individual word frequency within the MWU are only seen at the single word level. This seems to confirm an additive approach to processing during reading, as both the E-Z Reader and SWIFT reading models put forward.

We also note the lack of predictive capacity for MWU types on sentence reading times since this study is, to our knowledge, the first to test the hypothesis that MWUs offer a processing advantage within natural sentences. Certainly there are effects seen at the word level within the MWU, but the combined features of each of the three MWU types investigated here do not affect reading times over the whole sentence. In contrast, we find that there are subtle effects on processing of the words within the MWU when the remaining words from the stimuli are taken into account. In the MWU-words only analysis of word reading times, there was a significant interaction between MWU type and MWU frequency. Additionally, this effect showed the opposite order of advantages to the order from the model with all words in the sentence included. This finding may point to a cancelling-out of the word-level effects in the sentence reading times. Both these results merit further investigation in more tightly controlled sentence studies as well as in other paradigms, such as self-paced listening.

In this study, we did not control for the stimuli as we endeavoured to record 'naturalistic' MWU reading. That is, we used sentences from the British

National Corpus which contained MWUs. These sentences were allowed to vary randomly in terms of sentence length, sentence structure and complexity, word frequency, word length, MWU frequency and MWU length, as well as in the location of the MWUs within the sentences. A consequent drawback to the study design is that the sentence-by-sentence presentation of the stimuli did not allow ‘natural reading’ as would be preferable for eye movement analysis (see e.g., Kliegl, 2007; Rayner et al., 2007). However, while the presentation could be considered less ‘natural’ by excluding paragraph reading, the sentence-by-sentence presentation did avoid accumulative effects of context on the MWU reading. (For more on context effects in reading see McDonald & Shillcock, 2001.) That is, the reader was not able to predict an upcoming MWU on the basis of the context of the accumulated stimuli, an effect which is clear in story-based paradigms and traditional cloze tasks (see e.g., Kuiper, Columbus & Schmitt, 2009). Instead, the unrelated sentences provided a neutral background for the reading of each MWU. Any context effect on prediction was particularly important to avoid for lexical bundles, since they are completely transparent. We have shown here that lexical bundles have less semantic cohesiveness and lower familiarity rates. Finally, as mentioned above, the random variation lessened the potential for either experimenter bias (Forster, 2000) or participants becoming aware of repeated sentence structure.

An additional goal of this study was to describe differences between native and non-native English speakers in processing MWUs while reading. Our results show that MWU type processing effects are different for native versus non-native speakers during word reading and sentence reading. In word reading, there were qualitative differences between L1s and L2s for total word reading time and first fixation durations. Function words had significantly shorter word reading times and first fixation durations for non-native speakers' than content words did. Additionally, non-native speakers gained from word frequency in terms of shorter first fixation durations, and this effect was significantly higher for non-native speakers than native speakers. For both groups, total word fixation durations were shortest for words in lexical bundle sentences, but the effect is stronger for L2s. Only the L2 group data showed an effect of word frequency within versus outside of the MWU region, and the interaction included differences between function words' and content words' frequencies. Reading of words relative to the MWU region is also similar for each group, as both non-native speakers and native speakers have the shortest first fixations on words within the MWU. This indicates that the MWUs' potential to speed processing for native speakers extends to non-native speakers' reading. The similarity of MWU processing is also evident in the sentence reading times, though here the MWU frequency is facilitatory for non-native speakers, unless the sentence has a frequent lexical bundle.

The non-native speaker results have some implications for acquisition models and pedagogy. The most important is that the frequency effects found both for individual words and for the MWU itself support a language learning model where the frequency of not only words but also their associations, such as co-occurrence patterns, are stored and accessed. In an exemplar-based account of acquisition, for example, we could postulate that information such as frequency of use, context of use, sociolinguistic context of use (particularly for idioms, which tend to be less formal compared to academic restricted collocations) is stored in exemplar 'clouds' (Pierrehumbert, 2001) associated with the words within the MWU. In this way, information regarding the semantics of the MWU is stored alongside the individual words, their frequency and appropriate use. Storage such as this accounts for the effects found in this study for both MWU-level and word-level variables such as word frequency, word status, MWU familiarity, MWU transparency and MWU frequency. As the participants in this study were all immersed in an English-speaking environment in Canada, there is the possibility that the frequency effects are based on exposure rather than teaching. Whether these frequency effects are found for non-native speakers who live in a non-English speaking environment is something worth consideration in future studies, particularly with respect to comparisons with formal language education.

Generally, these findings regarding MWU types suggest firstly that the only real disadvantage for high intermediate/ proficient L2 English speakers in encountering MWUs is slower reading times. But secondly, the slower reading of words within an idiom by native speakers compared to within lexical bundles (and to a lesser extent, restricted collocations) runs counter to the theory of MWUs providing processing advantages to the listener/reader. Recall that phraseological theory has claimed a processing advantage for MWUs. The advantage is explained as the holistic storage of MWUs enabling faster retrieval and processing of the unit as a whole, and therefore use of a MWU allows more time to the listener or reader for processing the (de)compositional information in the rest of the proposition. If we put aside the MWU frequency effects in the MWU-words only word reading time analysis, the findings here would seem to support a model of access where the literal processing of an idiom failed before a figurative parse took place. But when we add both the effect of individual word frequency and the inhibition for word reading times after a lexical bundle or restricted collocation unit, it seems more likely that the real processing advantage argument for idioms is a function of predictability. That is, reading idioms requires suppression of the canonical meaning of the component words, and this is more difficult when reading more frequent words. But the cost of reading during the MWU is attenuated by easier sentence completion based on the context of the MWU. In contrast, lexical bundles do not require more effort to read in themselves, but compared to words following idioms, words following lexical bundles are read slower overall. Nonetheless, total sentence fixation times show that there is no discernable difference between the MWU types once the entire sentence has been read fully. Combined with the post hoc results for reading times for MWU-only words, it is clear that context makes a difference to MWU reading.

The predictability advantages argument for faster MWU reading relates to language models, and reading models in particular. We have seen that, for these 150 MWU sentences at least, word frequency and MWU frequency have varying effects on the MWU reading. The fact that the first fixation durations for words in idiom sentences did not show word frequency effects, regardless of the position of the words and the idiom MWUs, might suggest that a form of ‘extreme’ parafoveal preview benefit could be at play. That is, as idioms are shown here to be harder/more costly to read than lexical bundles and restricted collocations, the lack of frequency effect on words coming before the idiom is not only surprising but may illustrate a semantic parafoveal benefit, that is, semantic preview of the coming MWU (see e.g., Kliegl, Risse & Laubrock, 2007; Hohenstein, Laubrock & Kliegl, 2010, on this extension of the SWIFT model). If this were the case, then an idiom in parafoveal view may cause longer first fixation durations on all words which occur prior to the idiom in the sentence, regardless of frequency, and presumably in order to allow maximum uptake of each word in order to resolve the MWU meaning. However, this would require a relatively larger area in the parafoveal view than has previously been considered possible. As our sentences vary in length and MWU placement, we cannot look into this question further in the current study, but future investigations will consider manipulation of this variable. One other possibility is that the structure of the individual idioms may vary so that any effect of frequency is muted by the idioms' fixedness, flexibility, and decompositionality (Titone, p.c.) For example, the idioms may differ from one another in terms of whether they can be passivized, or whether their meanings are metaphorically linked to any of the individual words, as the idiom type covers a broad range. Again, the variation in our idiom stimuli makes this difficult to test without further experimentation.

Overall, our findings indicate that, while there may be areas of overlap, MWUs do not seem to form a strict continuum from most to least figurative and/or most to least syntactically fixed. Instead, it appears that there are feature-based levels of separation which group our 150 MWUs into some of the classifications (i.e., restricted collocations and idioms, as well as lexical bundles) that have previously been posited by phraseologists such as Kuiper (2006), Wray (2002), Nattinger and DeCarrico (1992) and Pawley and Syder (1983). MWU types in this study have distinguished processing times through their individual words' frequencies, the MWUs' joint probabilities, and the semantic relatedness of the words' meanings within the MWU relative to their base meanings. Bybee's contention of a continuum of frequently used combinations then seems less true for MWU *types*. That is, the continuum between individual words and phrasal structure may be filled with MWUs of different types, but MWUs themselves can be grouped into at least the three types investigated here. Additionally, how the MWUs are processed is also divisible in terms of variables with predictive capacity: MWUs are affected most by the frequency of individual words within the MWU and the sentence when they are encountered in a full sentence context.

In conclusion, in this study we have shown that three MWU categories defined by various phraseologists can be grouped through reading times and ratings of MWU familiarity and MWU semantic transparency. Finally, with the exception of word reading time in idioms, we have shown that individual word frequency is still the most consistent predictor for MWU reading when in a sentence, despite the inclusion of the MWU frequencies. This points toward a theory of MWU storage and retrieval that is based on exemplars and probability, in that only a frequency- and exemplar-based theory of language acquisition, storage and access can account for the complex interactions between individual word frequencies, function words, and semantic opacity that has been presented here.

References

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

Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

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

Bates, D., Maechler, M. & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes [Computer software manual]. Available online at: http://CRAN.R-project.org/package=lme4 (R package version 0.999375-39).

Bell, A., Brenier, J. M., Gregory, M., Girand, C. & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*, 92–111.

Benson, M., Benson, E. & Ilson, R. (Eds.). (1986). *The BBI combinatory dictionary of English: A guide to word combinations.* (First ed.). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken and written English.* Harlow: Longman.

BNC Consortium. (1994/2001). *The British National Corpus*, *version 2 (BNC World)*. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/

Bobrow, S. & Bell, S. (1973). On catching on to idiomatic expression. *Memory and Cognition, 1*, 343-346.

Brysbaert, M., Drieghe, D. & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance*. (pp. 53-77). Oxford: Oxford University Press.

Bybee, J. & McClelland, J. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review, 22,* 381-410.

Cacciari, C. & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language, 27,* 668-683.

Columbus, G. (2010). Processing MWUs: Are different types of MWUs psycholinguistically valid? An eye-tracking study. In D. Wood (Ed.), *Perspectives on formulaic language in communication and acquisition*. (pp. 194-210). London and New York: Continuum.

Cutting, J. & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition, 25*(1), 57-71.

Di Sciullo, A. M. & Williams, E. (1987). *On the definition of word*. Cambridge, Mass.: MIT Press.

Duyck, W., Vanderelst, D., Desmet, T. & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin and Review, 15,* 850-855.

Ellis, N. C. & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory, 5*, 61-78.

Engbert, R., Nuthmann, A., Richter, E. & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777-813.

Flavell, R. & Flavell, L. (2006). *Dictionary of idioms.* (Rev. ed.). London: Kyle Cathie Limited.

Flesch, R. (1948). A New Readability Yardstick. *Journal of Applied Psychology*, *32*(3), 221-233.

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory and Cognition, 28*(7), 1109-1115.

Gibbs, R. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory and Cognition, 8,* 449-456.

Gibbs, R. & Nayak, N. (1989). Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology, 21*, 100-138.

Gibbs, R. & Nayak, N. (1991). Why idioms mean what they do. *Journal of Experimental Psychology: General, 120,* 93-95.

Gibbs, R., Nayak, N. & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language, 28*, 576-593.

Gollan, T. H., Montoya, R. I., Cera, C. M. & Sandoval, T. C. (2008). More use almost always means smaller a frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*, 787-814.

Gordon, B. & Caramazza, A. (1985). Lexical access and frequency sensitivity: Frequency saturation and open/closed class equivalence. *Cognition, 21*, 95-115.

Hohenstein, S., Laubrock, J. & Kliegl, R. (2010). Semantic preview benefit during eye movements in reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1150-1170.

Howarth, P. (1996). *Phraseology in English academic writing.* Niemeyer: Tübingen.

Jespersen, O. (1954). *A modern English grammar on historical principles*. London: George Allen and Unwin/Copenhagen: Ejnar Munksgaard.

Jiang, N. & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal, 91*(iii), 433-445.

Kearns, K. (1988/2002). Light verbs in English. Unpublished manuscript retrieved October 27th, 2010, from: http://www.ling.canterbury.ac.nz/people/kearns.shtml.

Kliegl, R. (2007). Towards a perceptual-span theory of distributed processing in reading: A reply to Rayner, Pollatsek, Drieghe, Slattery, & Reichle (2007). *Journal of Experimental Psychology: General, 138*, 530-537.

Kliegl, R., Risse, S. & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word n+2. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1250-1255.

Kuiper, K. (1996). *Smooth talkers: The linguistic performance of auctioneers and sportscasters*. Mahwah, NJ: Lawrence Erlbaum.

Kuiper, K. (2000). On the linguistic properties of formulaic speech. *Oral Tradition, 15*(2), 279-305.

Kuiper, K. (2006). Formulaic speech. In K. Brown (Ed.), *Encyclopedia of language and linguistics.* (2nd ed.). (pp. 597-602). Oxford: Elsevier.

Kuiper, K., Columbus, G. & Schmitt, N. (2009). The acquisition of phrasal vocabulary. In S. Foster-Cohen (Ed.), *Advances in language acquisition*. (pp. 215-240). Basingtoke: Palgrave Macmillan.

Kuiper, K., McCann, H., Quinn, H., Aitchison, T. & van der Veer, K. (2003). *A syntactically annotated idiom database*. Philadelphia: Linguistics Data Consortium, University of Pennsylvania.

Kuiper, K., van Egmond, M-E., Kempen, G. M. & Sprenger, S. (2007). Slipping on superlemmas: Multiword lexical items in speech production. *The Mental Lexicon, 2*(3), 313-357.

Lehmann, H-M., Hoffmann, S. & Schneider, P. (2002). *BNCweb interface, v. 2.0*. http://homepage.mac.com/bncweb/manual/bncwebman-home.htm.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science, 189*, 226-228.

Marslen-Wilson, W. D. (1984). Function and processing in spoken word recognition: A tutorial review. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processing*. (pp. 125-150). Hillsdale NJ: Erlbaum.

Marslen-Wilson, W. D. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8,* 1-71.

McDonald, S. A. & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech, 44,* 295-323.

Mel'čuk, I. (1996). Lexical functions: A tool for the description of lexical relations in the lexicon. In L. Wanner (Ed.), *Lexical Functions in lexicography and natural language processing.* (pp. 37-102). Amsterdam/Philadelphia: Benjamins.

Mel'čuk, I. (1998). Collocations and lexical functions. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications.* (pp. 23-53). Oxford: Clarendon Press.

Moon, R. (1998a). *Fixed expressions and idioms in English.* Oxford: Clarendon Press.

Moon, R. (1998b). Frequencies and forms of phrasal lexemes in English. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications.* (pp. 79-100). Oxford: Clarendon Press.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 667 – 68.

Nattinger, J. R. & DeCarrico J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Nayak, N. & Gibbs, R. (1990). Conceptual knowledge in the interpretation of idioms. *Journal of Experimental Psychology: General, 119*, 315-330.

Nunberg, G., Sag, I. & Wasow, T. (1994). Idioms. *Language, 70*(3), 491-538.

Pawley, A. & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication.* (pp. 191-226). London: Longman.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure*. (pp. 137-157). Amsterdam: John Benjamins.

R Development Core Team. (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Available online at: http://www.R-project.org.

Rayner, K., Pollatsek, A., Drieghe, D., Slattery, T. J. & Reichle, E. D. (2007). Tracking the mind during reading via eye movements: Comments on Kliegl, Nuthmann, & Engbert (2006). *Journal of Experimental Psychology: General, 136*, 520-529.

Schmitt, N., Dörnyei, Z., Adolphs, S. & Durow, V. (2004). Knowledge and acquisition of formulaic sequences: A longitudinal study. In N. Schmitt (Ed.), *Formulaic sequences*. (pp. 55-86). Amsterdam/Philadelphia: Benjamins.

Schmitt, N. & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Formulaic sequences.* (pp. 173-189). Amsterdam/Philadelphia: Benjamins.

Segalowitz, S. J. & Lane, K. (2000). Lexical access of function versus content words. *Brain and Language, 75*, 376-389.

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., van Heuven, W. & Conklin, K. (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, A. & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics in Language Teaching, 45*(2), 119-139.

Siyanova, A. & Schmitt, N. (2008). L2 learner production and processing of collocation: a multi-study perspective. *The Canadian Modern Language Review/ La Revue canadienne des langues vivantes, 64*(3), 429-458.

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

Swinney, D. & Cutler, A., (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior 18*, 523–534.

Tabossi, P. & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory and Language, 32,* 359-372.

Titone, D. & Connine, C. (1999). The compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics, 31*, 1655-1674.

Tremblay, A. (2009). Processing advantages of lexical bundles: Evidence from self-paced reading, word and sentence recall, and free recall with event-related brain potential recordings. Unpublished Ph.D. Dissertation. University of Alberta, Canada.

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 and New York: Continuum.

Tremblay, A., Derwing, B. L., Libben, G. & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading experiments, word and sentence recall tasks, and off-line semantic ratings. *Language Learning, 61*, 569-613.

Tremblay, A. & Tucker, B. V. (2011). The effects of n-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon, 6*(2): 302-324.

Underwood, G., Schmitt, N. & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences.* (pp. 153-172). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S. & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience, 22*(8), 1682–1700.

Wasow, T., Sag, I. & Nunberg, G. (1980). Idioms: An interim report. In *Preprints of the XIIIth International Congress of Linguists*. Nippon Toshi Center, Tokyo, Japan.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Wulff, S. (2009). Converging evidence from corpus and experimental data to capture idiomaticity. *Corpus Linguistics and Linguistics Theory, 5*(1), 131-159.

Chapter 2. When three become one: eye movements and fixations in trigram reading

As we have seen in the Introduction and Chapter 1, multiple word units have long been the topic of typological and theoretical discussion. In addition to the semantic and syntactic issues described in the previous sections, the properties of the words within MWUs present further opportunities for psycholinguistic research. For example, current models of sentence reading take into account the effects of the words surrounding the word being read. Whether a word is short or long, frequent or uncommon, a function or content word, can lead to the word or its neighbours being skipped, read faster, read longer or regressed to. On the other hand, compared to normal 'compositional' strings of words, (e.g., *I bought a sandwich*), the joint probabilities and conditional (i.e., transitional) probabilities of the component words within MWUs are generally high. Because of the closer relationship between words in MWUs, they offer a particularly rich source of stimuli for eye movement studies.

Certainly, previous studies have shown MWU effects in reading, where MWU sentences are read faster than controls. Results in self-paced reading studies, such as those of Tremblay (2009; Tremblay & Baayen, 2010) on lexical bundles, and that of Schmitt and Underwood (2004) on formulaic sequences, support theories that MWUs offer processing advantages. In a previous study on idioms, lexical bundles and restricted collocations against neutral compositional sentences (Columbus, 2010), Columbus found that MWUs presented in context were read faster than controls. Additionally, in Chapter 1 we found differences in reading times between different types of MWUs, even after accounting for semantic and structural differences. Results such as these raise questions regarding not only reading models and MWUs, but also the concept of decompositional versus holistic storage of and access to MWUs (see e.g., Wray, 2002) within the mental lexicon.

What all the previous studies on idioms and formulaic sequences have in common is that, while many have investigated comprehension during online tasks, the experimental stimuli and/or designs have been relatively controlled. We believe this may mean that there is yet more to learn about MWU processing in reading by designing experiments with more random elements. For example, in the majority of idiom studies the stimuli were carefully selected for structure, as in Libben and Titone's study (2008) with SVO four-word idiom sentences. This control limits the context surrounding the idiom. Similarly, Vespignani et al.'s (2010) ERP study measured responses to only the first and second word of the idiom within a sentence. In another ERP study, Laurent et al. (2006) restricted their analysis to the final words of infinitival idioms. Other studies restricted the target of analysis, such as in Schmitt and Underwood's (2004) self-paced reading study, where each word in the idiom was necessarily measured, and words were not able to be skipped naturally. The self-paced paradigm does not allow parafoveal view of surrounding words, and so cannot inform theory on reading processing in situ. Indeed, we have yet to encounter an MWU study where the researchers deliberately maximized the random selection of stimuli.

Another difficulty in eye movement studies on MWUs is that the single-line presentation cannot guarantee that the first word of the MWU itself will be fixated on. That is, the analysis of fixations on the MWU cannot be guaranteed to include the first word in the MWU, since the participants can focus on any part of the sentence or phrase which has a MWU embedded in it. As such, we cannot be sure of where the readers enter the MWU, meaning that we potentially miss important fixation patterns based on parafoveal information from the first MWU component. This is an important consideration in terms of the reading theories (E-Z Reader and SWIFT) discussed in Chapter 1. A study presenting only the minimal unit of a MWU may encourage a first fixation on a first word, particularly if the fixation cross is placed within the first letters of the first word, and so may provide crucial information in terms of reading processing. Certainly, if the first fixations on the first word of the trigram reveal MWU effects, then this would support an 'extreme' version of semantic parafoveal preview and parallel processing, as discussed in Chapter 1.

In what follows, we aim to determine if natural trigrams deliberately taken from a corpus with maximum attention to random selection can provide insight into MWU processing, with the benefit of the first word having to be read. Our study combines three-word MWUs sampled from the Google Web 1T N-gram corpus (Brants & Franz, 2006) with the placement of the reader's first fixation fixed to the first word of the trigram. The stimuli are all MWUs, all three words, and each presented as a single-line trigram. Building on the focus of prior MWU research on holistic access to the MWU, this experiment seeks to extend single-word reading studies to 'big words', as MWUs have often been termed. Further, presenting the trigrams without context lessens the predictive memory retrieval aspect in reading processing (Kliegl, Nuthmann & Engbert, 2006) and so allows measurement of the trigram access only. Our stimuli are thus presented without context, without controlling a priori for individual word or trigram frequencies. Additionally, reading the string is not forced and 'forward-only' as it must be in self-paced reading paradigms, allowing natural skipping but ensuring registration of at least the first word of the trigram.

In the current study, our goal is to provide evidence for independence of the three MWU categories by determining relationships between MWU variables and their eye movement correlates in three-word strings (trigrams). In contrast to Chapter 1, however, we are searching for when such effects take place in simple, context-free trigrams. Put simply: When do MWU type effects occur within trigrams, and how do other MWU-related predictors affect trigram reading time? We investigate these questions by focusing on three specific dependent variables: the first fixation durations on the first words of the trigrams, the subgaze durations on the first bigrams, and lastly the total fixation duration times for the trigram (i.e., the summed durations from each of the words in each trigram).

These particular time variables are chosen because of the relationship between the words in a trigram. In the first fixation duration measure we can investigate whether there are effects from the MWU on the reading of the first word. We believe that this dependent variable will illustrate the expected parafoveal preview benefits for the length and frequency of the second word.

What is less clear is whether there will be any advantage for the first word from the MWU, which includes all words in the trigram. In a parallel processing model of reading, we would expect an indication of the processing of the third word in the trigram on reading of the first word. A holistic or even the Hybrid view of MWU storage and retrieval (Titone & Connine, 1999) would predict that the MWU would be retrieved from the first word and its parafoveal preview, and so the reading of the first word would be faster for MWUs of higher trigram frequency.

On the other hand, the subgaze duration time measure gives insight to the early processing of two-thirds of the trigram, as it includes all fixations on the first two words before any regressions. In this measure, we expect parafoveal preview effects (based on either reading theory) on the first bigram from the MWU's type and MWU's frequency. In a holistic storage model, however, we should expect no individual word frequency effects from the first two words, having instead bigram and trigram frequency advantages from holistically stored MWUs.

Finally, in the total trial fixation duration measure, we are able to see the complete timeframe of the fixations per word and for the MWU as a whole. In this measure, we expect to see differences in reading times between the MWU types. Besides the expected advantage from higher trigram frequency, there is another result to be expected. This is that if idiom MWUs are indeed the most semantically cohesive, then they should be read the fastest, with lexical bundles, being completely compositional, being the slowest trigrams to read.

Method

Participants

Nineteen native speakers of North American/Canadian English took part in the eye movement experiment. All participants were recruited from introductory Linguistics courses offered by the Department of Linguistics at the University of Alberta. Participants received course credit as compensation. All participants had normal or corrected-to-normal vision. Data from two participants were excluded due to difficulties with calibration, leaving 17 participants in the analysis.

Materials

We used the Google Web1T n-gram corpus (Brants & Franz, 2006) as our source of the internal frequencies for all 258,599,481 lower-case trigrams in the corpus. Each trigram has seven frequencies associated with it: 1) first word; 2) second word; 3) third word; 4) first bigram; 5) second bigram; 6) split gram of Word 1 and Word 3; and 7) the whole trigram. We then selectively sampled from this set. We did this by searching for trigrams which had a particular (base 10) log frequency band, say 0.1, for Word 1, Word 2, Word 3, Bigram 1, Bigram 2 and Bigram 3. In the next sample, we searched for trigrams with the same values for all these variables except for a single change in value for, say, Bigram 3. From these two searches the results would show a set of trigrams with the same first word, second word, and third word frequency bands, and first bigram, second bigram, third bigram frequency bands, and a second set where all band values were identical until the third bigram frequency (Table 2-1 below). The third search would return a list with the same band values for all variables except Bigram 2, and so on until sets of trigrams with all combinations of frequencies had been retrieved.

The total number of unique combinations of frequency bands that contained one or more trigrams was approximately 27,000. To narrow the stimulus set we sampled at random 1000 frequency bands from this larger set, choosing one trigram for each frequency band. Potential stimuli were excluded if they contained offensive language, repeated words (e.g., *sale, sale, sale!*), or with the potential to evoke emotional responses from the participants. After being removed, they were replaced with a new sample from the same frequency band group. Using the MWU definitions described in the Introduction and used in the sentence reading experiment in Chapter 2, the 1000 trigram items were classified as idiom, lexical bundle, or restricted collocation trigrams. The items were later blindly reclassified to ensure consistency. A list of all the stimuli and their frequency values are available at:

http://www.ualberta.ca/~columbus/Site/Welcome.html

Table 2-1. *Examples of trigrams in the stimulus set*

Trigram	*w1f*	*w2f*	*w3f*	*b1f*	*b2f*	*b3f*	*Freq. band*
glutton for punishment	0.08	5357.09	5.68	0.03	0.09	0	-1:-1:3:0:-1:-1:-4
electric mixer until	16.93	2.36	103.64	0.11	0.03	0	-1: 1:0:2: 0:-1:-3

Procedure

The stimuli were presented on a desktop PC monitor, using Experiment Builder™ software running on Windows XP. The data were collected using an Eyelink II, video-based head-mounted eye tracking system (SR Research™). Eye movements were collected using pupil-only sampling at a rate of 500Hz.

Each experiment session was preceded by 10 practice trials, and rest and recalibration breaks occurred after each block (approximately every 4-6 minutes). Both the practice trials and 1000 experimental trials were randomly ordered for each participant. The experimental task was reading for meaning. Participants were seated at a comfortable distance from the screen (approximately 70cm) and asked to silently read the phrases for meaning as quickly as possible for normal comprehension. The participant then cued the next sentence by moving their gaze to an invisible boundary (100 pixels wide) on the right side of the monitor. The purpose of the gaze-contingent cue was to prevent participants moving their eyes down to a keyboard, footpedal or mouse before the trial ended, given that only three words were presented at a time. Additionally, we hoped this technique would lessen re-reading the trigram after the normal read-through was completed. To ensure participants were reading for comprehension, participants also created sentences using the most recent phrase and a probe word at a rate of approximately one for every twenty phrases. Responses were contemporaneously graded offline by the experimenter as either grammatically plausible (0), or partially plausible (-0.5), or implausible (-1). Participants with plausibility scores below 75% were to be excluded from the data analysis, but as no participant scored less than 90%, none were excluded in the end.

The stimuli were presented in white, fixed-pitch font (Courier New) on a black screen, following a fixation cross presented between two and three character spaces into the first word of the trigram. All stimuli were presented from the centre left of the screen, while instructions were presented centrally from the top, and key words for the sentence creation task were presented in the top left region of the screen.

Results

Analysis methodology

The statistical analysis to be used is linear mixed effects modelling using the lme4 statistical package (Bates, Maechler & Bolker, 2011) in the statistical programming language R (R Development Core Team, 2011). The approach we use treats subject, word, and trigram as random effects. The main time measures considered as dependent measures were First Fixation Duration (FFD; the duration of only the first fixation on a word during the trial), Subgaze Duration on the first bigram (SGD; the sum of fixations on the first two words before any regressions), and Total Fixation Duration Time (TFD; the sum of all fixations on the words in trigram during the trial). The First Run Fixation Count (*Word Fixation Count >1*), which is technically a dependent measure, is appended to the FFD model. It ensures that each word being analysed has been fixated on at least once. As such, it is included in the FFD model solely as a control measure. It is not a necessary control in the first bigram or whole trigram models. Variables to be investigated for their predictive capacity are described below.

Predictors in the analysis

Control variables considered for analysis included the following: a) the length of each word (*Word Length* for *Word 1, Word 2, Word 3)*; b) whether the word was a function word or a content word (*Word Status* for *Word 1, Word 2, Word 3)*; c) the frequency of each word in the trigram *Word Frequency (Word 1, Word 2, Word 3)*; and d) the order of the stimuli presented to each participant *(Trial Order*). Variables of theoretical interest were as follows: a) *Bigram Frequency* (for words [1+2], [2+3], and [1+3]); b) the frequency of the trigram *(Trigram Frequency)*; and c) the MWU category for each trigram (*MWU Type*). Additionally, we added a variable gleaned from judgements of whether the trigram could be considered syntactically and semantically complete (*Constituency*). This judgement task is described in the next sub-section.

Constituency judgements for trigrams

Lemke, Tremblay and Tucker (2009) found, in their 4-gram lexical bundle phrase production experiments, that n-grams were produced faster if they were full constituents (e.g., *I don't want to*) over those that were not (e.g., *in the middle of*). To account for variance in our trigram reading, we had participants judge the completeness of each trigram.

Twenty-nine graduate students from the departments of Psychology and Linguistics were asked to read a list of trigrams and judge whether the trigram seemed syntactically and semantically complete. All participants were native

English speakers with no prior experimental exposure to this particular set of trigrams. Participants received candy bars as compensation. We included only graduate-level participants with backgrounds in the psychology of language or linguistics in order to avoid prescriptive judgments. The 1000 stimuli were divided into four questionnaires of 250 items each to reduce trial fatigue. Eight participants completed the first set of 250, nine judged the second set, seven judged the third set, and five judged the remaining set of 250 items. The questionnaire explicitly stated that a 'yes' response indicated that an utterance would be understood as 'complete in a conversation'. Finally, the judgements were compiled and averaged over subjects, resulting in a scale that ranged from 0-1 for *Constituency* of each item. The binary choice averaging over multiple judges allowed for a scalar result that would also represent the *semantic* completeness of the three words in the stimuli, which a purely objective constituency factor would not allow. That is, the judgements for the 'completeness' of an item could incorporate both semantic and syntactic considerations, meaning that the responses were not the same across all participants. This led to the scalar rating, which allows the grey areas of 'phrasiness' to be taken into account. The constituency values for each trigram were then associated with the relevant stimuli.

Model results

Frequency counts and the three time measures (first fixation duration, subgaze duration and total fixation duration) were (natural) log transformed to remove most of the rightward skew, and minimise the possibility of overly influential outliers in the statistical model. Trial Order, Word length and frequency variables were additionally scaled and centred. Regions of interest (ROIs) were set for each word in the experimental stimuli. Data matrices ('reports') based on these ROIs were generated by Dataviewer (SR Research™). The reports were then loaded into the statistical programme R (R Development Core Team, 2011). These were then merged with the item statistics.

Results: First Fixation Duration times on the first word in the trigram

For the best-fit model of the first fixation duration variable, the random effect factors were Trigram and Subject (estimated standard deviation parameters 0.0178 and 0.0626 respectively; the estimated standard deviation of the residual error was 0.1173). Predictors included First Run Fixation Count and Trigram frequency, and interactions were supported for MWU Type by Bigram frequency for Words [1+2], Function word (for Word 1) by Word 1 length, and Word 1 frequency by Word 2 length. Following Baayen, Davidson and Bates (2008), absolute values for t exceeding 2 were taken to be significant. Table 2-2 lists the log likelihood ratio test values for the addition of each predictor to the model. Predictors that did not reach significance are not listed in this table, which summarises a minimally adequate model.

In Figure 2-1, panel 1, we see a significant interaction between MWU Type and the frequency of the first bigram. Specifically, when the first word in a restricted collocation trigram has high bigram frequency, it has shorter first fixations than the first word in idiom trigrams as well as the first word of lexical bundles trigrams. (The slope for idioms in panel 1 is not significant, despite appearances.) This interaction is unexpected, as restricted collocations have a supra-lexical level of meaning that lexical bundles do not have, meaning that the restricted collocations should require more time to access the semantics of the unit. As such, the lexical bundles ought to have had the shorter first fixations on the first word, or at least have had the same fixation times as words in restricted collocations. We manually inspected the restricted collocations and a random sample of lexical bundles, matched for number, to determine if it were possible to predict the trigram from the first word of the restricted collocation MWU, and from the lexical bundle MWUs. Post-hoc log likelihood ratio tests on the trigrams' semantic relationships confirmed that, compared to lexical bundles, the semantic content of Word 1 in restricted collocation trigrams is more closely tied to the first bigram and the whole trigram's meaning. The participants seemingly found it easier to recognise the full restricted collocation, but not the full lexical bundle or idiom, when the first bigram was highly frequent. This is even more significant when we include in the test any restricted collocations with an article as the first or the third word, as articles are more likely to be skipped or read parafoveally. Idiom trigrams were not significantly different from either the restricted collocations or lexical bundles in the post-hoc analysis of trigram semantics; since the proportion of idioms in the stimulus set is lower than either restricted collocations or lexical bundles, this is possibly an issue of power. We suggest, however, that the significant difference between restricted collocations and idioms is also based on the ability to view the following word and so access the meaning of the entire trigram. This then causes the first fixation durations on restricted collocations to be significantly shorter than those on idioms when the Bigram1 frequency is high. If this is true, then it supports a theory of semantic parafoveal preview (e.g., Hohenstein, Laubrock & Kliegl, 2010), at least of the following word (i.e., *n*+1).

Table 2-2. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the First Fixation Duration model for Word 1*

Predictor	**AIC**	**Df**	***p***
MWU type	-20248	6	–
Bigram 1 frequency	-20249	7	0.0834
MWU type * Bigram 1 frequency	-20258	9	0.0011
Trigram frequency	-20264	10	0.0076
Word 1 length	-20428	11	<0.0001
Word fixation count >1	-20470	12	<0.0001
Word 1 is Function word	-20532	13	<0.0001
Word 1 length * Word 1 Function	-20772	14	<0.0001
Word 2 length	-20793	15	<0.0001
Word 1 frequency	-20873	16	<0.0001
Word 2 length * Word 1 frequency	-20897	17	<0.0001

Note: These are model comparisons, and as such the table has no t-values for the predictors. Each predictor's t-value was significant, however, to have been kept in the model. Predictors which are not significant as main effects (though are significant in interactions) are italicised.

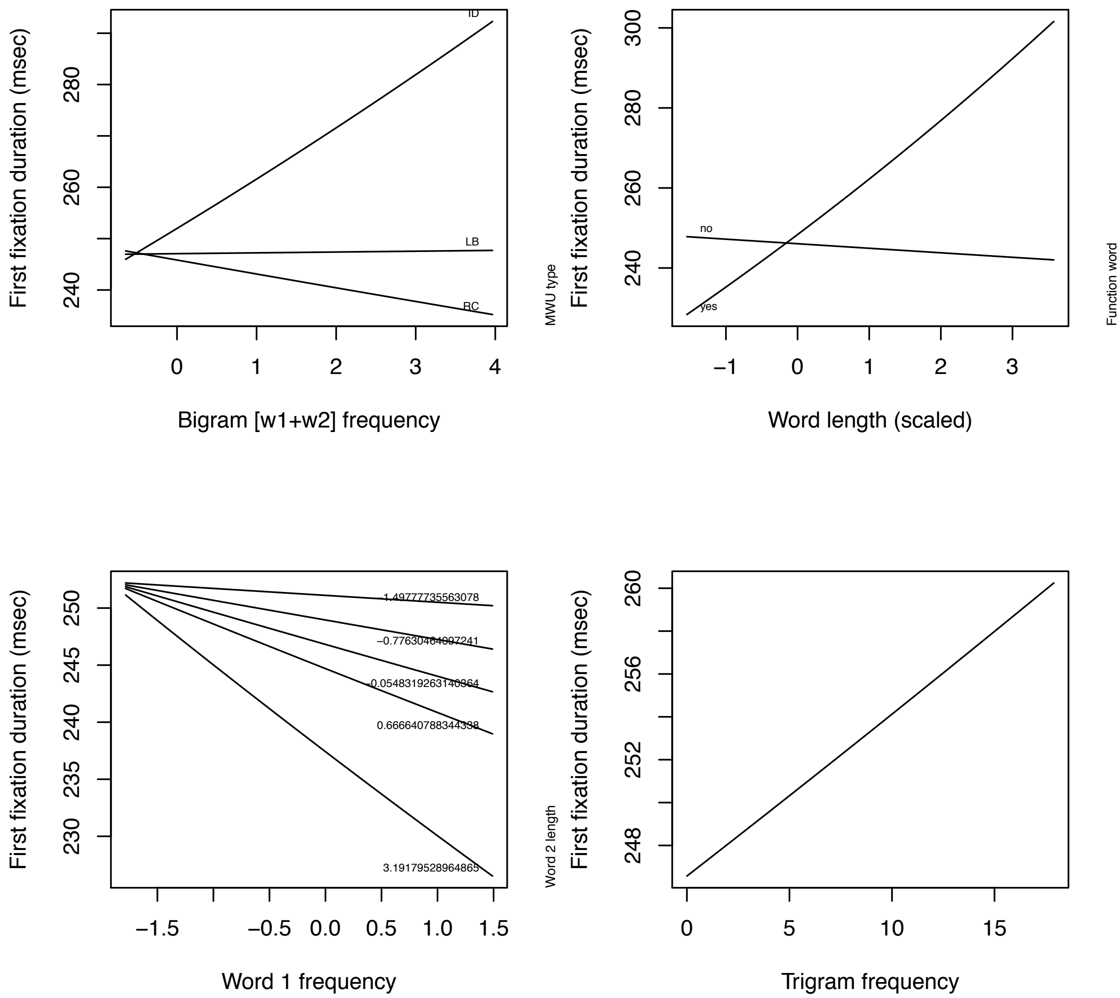

Figure 2-1. First fixation durations on the first word in the trigram. Top panels 1-2: (L) Bigram Frequency interaction with MWU Type (Lexical bundles [non-sig.] – Idioms [non-sig.] – Restricted collocations) for first two words, (R) Scaled Word Length by Word Status; bottom row panels 3-4: (L) Interaction of Word 1 Frequency with Word Length (in quartiles) for Word 2, (R) Trigram frequency. Each figure illustrates effects when all other variables are held constant.

Panel 2 of Figure 2-1 shows the word length interaction with Word status. First fixation durations on content words are relatively stable across Word length. However, there were much shorter first fixation durations for short function words than short content words. Longer function words elicit longer first fixations. Another length effect is found in Figure 2-1, panel 3, where the longer Word 2 is, and the more frequent Word 1 is, the shorter the Word 1 first fixation duration, as predicted by both serial and parallel processing models of reading.[1]

Finally, in panel 4 of Figure 2-1 we see that the frequency of the trigram affects first fixation durations. Somewhat surprisingly, we find that Word 1 has significantly longer first fixation durations as the trigram increases in frequency. However, this is consistent with a parafoveal preview effect extending to Words 2 and 3 while reading Word 1, as put forward by Kliegl and colleagues (Kliegl, Nuthmann & Engbert, 2006; Kliegl, 2007). That is, the more predictable the *n*+1 and *n*+2 words are while reading word *n*, the longer the reader will focus on word *n*, allowing more retrieval time for words *n*+1 and *n*+2. We will return to this reversed frequency effect in FFDs on Word 1 in the General Discussion.

In summary then, the first fixation duration data on the first word of the trigram show clear differentiation between the restricted collocations and both

[1] Both theories have shown that the reader spends less time on the word *n* when *n* is frequent and the word *n*+1 long, presumably since the speaker sees a long word and decides they must move their attention to it as soon as possible. Here, the added frequency of word *n* also contributes to the shorter fixation time.

idioms and lexical bundles when the frequency of the first bigram is high. Additionally, the higher the trigram frequency, the longer the first fixations on the first word, suggesting an effect of the MWU 'whole' on the word 'part' via parallel processing. We turn now to the effects found in the subgaze durations on Bigram 1.

Results: Subgaze Fixation Duration times on the first bigram

Subgaze duration may be a better measure of word properties in that it includes multiple forward-moving fixations on both the first and second words. Recall that subgaze duration is defined here as the sum of all fixations on the first bigram before any regressions are made. For the subgaze duration analysis, the random effect factors were Word, Trigram and Subject (estimated standard deviation parameters 0.0221, 0.1436 and 0.1353 respectively; the estimated standard deviation of the residual error was 0.64). Predictors included the following: a) MWU category, b) Word 1 length, c) Word frequency for Words 1, 2 and 3, and d) Bigram frequency for Words [1+2] and [2+3]. We also included Trial order as a control variable. The log likelihood ratio value of each predictor added to the model is listed in Table 2-3.

Figure 2-2 illustrates the predictors which explain variance in the summed fixations on the first two words of the bigram before any regressions. In panel 1, we see a significant difference among MWU types, where readers have longer

Table 2-3. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the Subgaze Duration model for first bigram*

Predictor	AIC	Df	*p*
MWU type	72930	7	–
Word 1 length	72871	8	<0.0001
Word 1 frequency	72645	9	<0.0001
Word 2 frequency	72381	10	<0.0001
Word 3 frequency	72223	11	<0.0001
Bigram 1 frequency	72176	12	<0.0001
Bigram 2 frequency	72173	13	0.0178
Trial order	72145	14	<0.0001

subgaze durations on the first bigram for lexical bundles than for the first bigrams of restricted collocation and idiom trigrams. This may be related to the interaction of restricted collocation with first bigram frequency in the first fixation durations. That is, reading times for restricted collocations were differentiated from lexical bundles and idioms based on the predictability of the third word of the trigram given the first two words. The longer subgaze durations on Bigram 1 in lexical bundles than those in restricted collocation trigrams are likely due to the unhelpfulness, in semantic terms, of the first bigrams of lexical bundles, as discussed in the First Fixation Duration results. Because the meaning of the lexical bundle is not predictable from the first two words, the reader takes longer subgaze fixations on the words for maximal uptake.

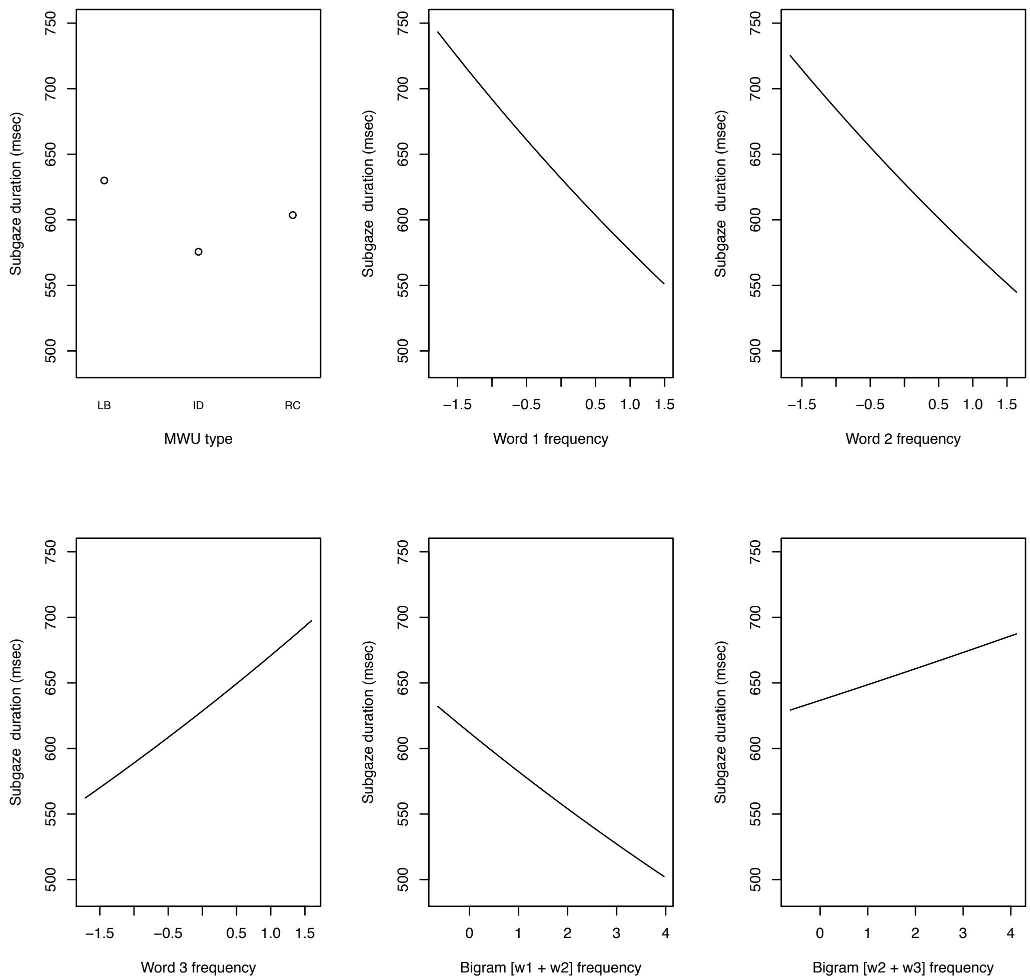

Figure 2-2. Subgaze durations on the first two words in the trigram. Top row panels 1-3: (L) MWU Type (Lexical bundles – Idioms – Restricted collocations), Word 1 Frequency; (R) Word 2 Frequency; bottom row panels 4-6: (L) Word 3 Frequency, Bigram Frequency of Words 1+2, (R) Bigram Frequency for Words 2+3.

Besides these effects from the frequency of Bigram 1, frequencies for Word 1 and Word 2 (panels 2 and 3) also show shorter subgaze durations for more frequent words. However, the slopes for both Word 1 frequency and Word 2 frequency are significantly steeper than that of Bigram 1 frequency. That is, while fixating on the first bigram in the first pass, readers gain more advantage from the frequency of the individual words than of the bigram unit. In contrast with the frequencies of Words 1 and 2, Word 3 frequency (panel 4) elicits significantly longer subgaze durations as frequency increases. This is consistent with a parafoveal preview effect (see e.g., Hohenstein, Laubrock & Kliegl (2010) *inter alia* for more on semantic previews), where the higher frequency of the third word causes longer reading of the Words [1+2] bigram, since the reader is able to focus on the current words, in the knowledge that the next word is frequent and so faster to process. A similar pattern is found for Bigram 2 (Words 2 and 3) frequency (panel 6): the higher the Bigram 2 frequency, the longer the subgaze duration. As with Bigram 1 frequency, the Bigram 2 frequency slope is weaker than the slope for the Word 3 frequency effect. This is likely due to the fact that the current bigram includes one of the two words in Bigram 2, and so the current word is already being processed.

As a final point, note that, unlike in the first fixation data on Word 1, there is no significant effect for trigram frequency in subgaze durations. Recall that on Word 1, trigram frequency increased first fixation durations, whereas now, we find that Bigram 1 frequency (panel 5) actually decreases subgaze durations substantially. It is possible that this strong facilitation at Bigram 1 (i.e., Words [1+2]) is cancelling out the increased first fixation durations on Word 1 of the trigram, leading to a *t*-value of around 0. The MWU type effect, however, is clearer in the subgaze duration data than in the first fixation data, with durations being affected differently based on the MWU type. We shall see in the following section whether these MWU and frequency effects are borne out in the total fixation duration data.

Results: Total Fixation Duration times on words in the trigram

For the total fixation duration variable, random effect factors were included for the Trigram, Word and Subject (estimated standard deviation parameters 0.1081, 0.0081 and 0.1656 respectively; the estimated standard deviation of the residual error was 0.2468). Predictors included 1) the MWU type; b) MWU constituency rating; c) Trigram frequency and d) Word frequency (for Word 1, Word 2 and Word 3), as well as the control variables of e) Word length (for each of the three words) and d) Trial order. The log likelihood ratio test values for each predictor when added to the model are listed in Table 2-4.

Firstly, we see in Figure 2-3, panels 4 and 5 that in contrast to the subgaze duration results, readers' total fixation durations are shorter when Word 3 has higher frequency. That is, the typical facilitatory word frequency effect is found for Word 3 in total fixation durations, instead of the inhibitory effect found in the subgaze durations. This is in line with a parafoveal-on-foveal effect as we have discussed earlier, where the subgaze duration on the first bigram is longer due to the higher frequency of the word in the parafoveal view – Word 3.

Table 2-4. *Increase in goodness of fit (as indicated by AIC) with addition of each predictor in the Total Fixation Duration model for the trigram*

Predictor	**AIC**	**Df**	***p***
MWU type	18029	7	–
Trigram frequency	17940	8	<0.0001
Word 1 length	17902	9	<0.0001
Word 2 length	17706	10	<0.0001
Word 3 length	17505	11	<0.0001
Constituency	17467	12	<0.0001
Word 2 frequency	17446	13	<0.0001
Word 3 frequency	17430	14	<0.0001
Trial order	6648	15	<0.0001

processing of Word 3 while still completing the fixations on the first bigram. Like the Word 3 frequency effect, total fixation durations are significantly shorter across the trigram when the Word 2 (panel 4) frequency is higher. Surprisingly, however, there is no frequency effect for Word 1 in the total fixation durations. Instead, we find strong Word 2 and trigram frequency effects. This is perhaps unexpected in the total trigram reading times, but may be indicative of the irrelevance of the first word to the trigram as a unit. That is, the frequency of Word 2 and the trigram frequency matter for total fixation duration but not for first fixation duration, because much of what was happening earlier in the FFD was simply parafoveal uptake. Note that the internal bigram frequencies are not relevant in the whole trigram fixation durations; this is perhaps due to their effects having been reconciled after the first fixation and subgaze durations. Furthermore, in panel 1 we see a clear picture of the distinctions between the three MWU types.

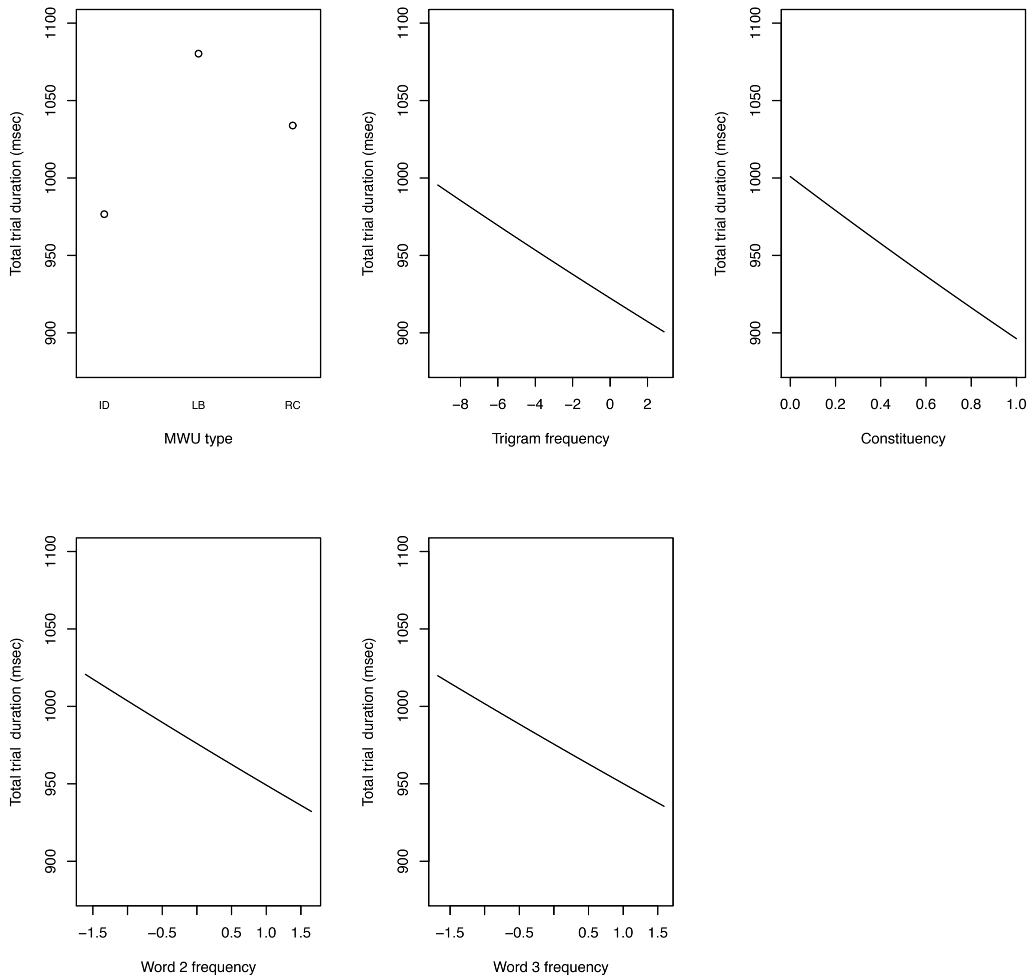

Figure 2-3. Total fixation durations on the trigram. Top row panels 1-3: (L) MWU Type, (Idioms – Lexical bundles – Restricted collocations) Trigram Frequency, (R) Constituency; bottom row panels 4-5: (L) Word 2 Frequency, (R) Word 3 Frequency.

Idiomatic trigrams have the shortest trial fixation durations, followed later by restricted collocation trigrams, and later still by trigrams which are lexical bundles. The order of this effect is the same as for subgaze durations, with the semantically-united trigrams being advantaged over the entirely literal lexical bundles. There are clear differences between the MWU types, then, since they are able to explain the variance in the total trigram fixation durations.

It is perhaps not surprising to find an effect in total fixation durations for greater constituency ratings on a trigram (panel 3). That is, the more highly a trigram was ranked as 'sounding complete', the shorter the total trial fixations across the trigram. This variable was not predictive in the first fixation duration or subgaze duration data, suggesting that Constituency of the trigram is not accessed in the earlier stages of processing. This may appear to go against a semantic preview benefit for the whole phrase during reading of Word 1 and Bigram 1, but is consistent with the trigram frequency inhibition in the first fixation durations, and potentially with the lack of bigram frequency effect here. That is, the attention paid to the trigram parafoveally does capture the trigram frequency in FFD; perhaps the strong inhibition here obscures any advantage from the more 'complete' trigrams. Additionally, in the total fixation duration times, effects from integration of the three trigram words into one unit are included, since it is the sum of all combined fixations on the words in the trigram, which could explain the effect. Future analysis of the second bigram's subgaze durations, and the first fixations on Word 3 may clarify any effect of constituency leading up to the total trial fixation durations.

Finally, one last frequency effect is seen in panel 2 of Figure 2-3. Higher trigram frequency leads to shorter fixation durations on the trigrams. This facilitation contrasts with the fact that the trigram frequency inhibited first fixation durations. Taking this and the lack of trigram frequency effect in the subgaze durations into account, it supports our previous contention that any effect at the subgaze duration level of processing has been cancelled out by the shift from inhibition on the first word fixations to facilitation on the trigram fixation durations. Overall, then, trigram frequency and MWU types are both important in predicting total trigram fixation durations. We shall return to this issue of trigram frequency later in the General Discussion, to which we now turn.

General Discussion

We set out in this paper to determine whether the MWU types used in Chapter 1 can account for some of the variation in word reading times while reading trigrams without context, & to see how word reading times for MWUs presented alone can add to current reading processing theory. We find that MWUs of different semantic/syntactic classifications are indeed read in varying ways, depending on factors such as constituency ratings, and frequencies of the MWUs as well as of their component words and bigrams.

Firstly, only total word fixation durations across the trigram are affected by how complete the trigrams were rated. This measure implicitly incorporated subjects' syntactic & semantic judgments on the phrases. We stated above that the Constituency effect was not found in the earlier time measures, suggesting that the third word had not been incorporated into the trigram in parafoveal preview while processing the first word or first bigram. This would go against the concept of extreme parafoveal preview as posited by Kliegl, Risse & Laubrock (2007).

However, a lack of Consituency effect in the first fixation duration data & subgaze duration data may instead be due to inability to 'wrap up' phrases during first pass reading. (Support for wrap-up effects from the first fixation is found in Warren, White & Reichler, 2009.) This would certainly be consistent with the traditional position of wrap-up effects at clause boundaries. Furthermore, it could mean that any semantic preview benefit of Word 3 during first fixation durations might be obscured by attempts to wrap up the phrase. As we did not account for trigram complexity or the critical areas for wrap-up effects, we cannot provide evidence either for or against these theories; we leave this to a follow-up study.

Critically, the most important finding from this study from a holistic versus decompositional processing viewpoint is that MWU variables are predictive for MWU reading for trigrams, even when presented alone and without context. Firstly, MWU types are differentiated in subgaze durations and in trigram fixation durations, and in the restricted collocation type (only) in interaction with Bigram 1 frequency in first fixation durations. In the subgaze durations and total trial fixations, the expected order of MWU reading times was found. That is, longest reading times were for lexical bundle trigrams, and shortest durations were for idiom trigrams. This goes against a model where decomposition of utterances prefers a literal interpretation (e.g., Bobrow & Bell, 1973; Swinney & Cutler, 1979), as there is no semantic competition or confusion in a lexical bundle. Additionally, this result is possibly supportive of a Configuration Hypothesis model (Cacciari & Tabossi, 1988; Cacciari & Glucksberg, 1991) or Hybrid model (Titone & Connine, 1999), as the idiom MWUs appear to be recognised by the second word. We note that this order is in opposition to the results found in Chapter 1. This suggests that the sentence context has strong effects in MWU recognition and processing, and may lean toward a new interpretation of some prior idiom results. Nevertheless, the order of MWU advantages does support a usage-based model of language, where idioms and restricted collocations have advantaged access from the recognition of the unique combination, such as in Sprenger, Levelt and Kempen's (2006) Superlemma theory of storage and retrieval. The fact that the two semantically-cohesive types, restricted collocations and idioms, can be distinguished from the lexical bundles in each time measure suggests that the semantic access to the whole MWU meaning occurs at even the subgaze duration on the first bigram. For restricted collocations, this occurs at the very first fixation on the first word when the restricted collocation is a part of highly frequent first bigram (i.e., Words [1+2]). Furthermore, the interactions of MWU Type by Bigram 1 frequency, with shorter fixations for restricted collocations of high frequency, shows some processing of the second word is occurring while the fixation is still on the first word, that is, there is the predicted parafoveal preview benefit. In all, the MWU type effects found here are further evidence that MWUs are processed differently in reading depending on their type's inherent features.

As noted in the results section above, the restricted collocation category has significantly shorter subgaze durations than the lexical bundles. Subgaze durations on the first bigrams of lexical bundles are almost significantly longer than on idioms' first bigrams. Further, there are reliably longer total fixation durations on the whole trigram for lexical bundles than for idioms. Meanwhile, there is still the issue of the varying trigram frequency effect. There are significantly longer first fixation durations on the first word of the unit for frequent trigrams, but significantly shorter total fixation durations on each word in more frequent trigrams. On the first bigrams, however, subgaze durations were not significantly different based on trigram frequency. It seems possible that this trigram effect is related to the restricted collocation effects in the first fixation on the first word. For example, in the subgaze duration of the first bigram, two things are altered from the first fixation duration: the first is that the lexical bundles are now shown to have significantly longer subgaze durations than the idioms, the second is that the trigram frequency effect, which had been inhibitory for the FFD, is now non-significant. It is interesting to note that the restricted collocation MWU category is now only significant against lexical bundles (and without the bigram frequency interaction).

It is this difference between restricted collocations and lexical bundles which presents one possible explanation for the lack of trigram frequency effect at the first bigram subgaze duration. Namely, that there is no trigram frequency effect because as a category the lexical bundles have the most frequent MWUs, and the restricted collocations have the next most frequent MWUs, but strong semantic relationships visible from the first word. The difference between the lexical bundles and the restricted collocations at this point is still related to the advantage restricted collocations show in the first fixation durations; thus the differing trigram frequency levels for restricted collocations and lexical bundles in this dataset may lead to the non-significant effect. It was positive at the first fixation duration – i.e., longer for more frequent items – and close to zero at the first bigram subgaze duration. Once the entire trigram has been fixated, however, the trigram frequency is facilitatory. The trigram frequency here has moved from positive, to zero, to negative over the subsequent addition of fixations on further words in the trigram.

We have stated earlier that an important finding in this study is that the frequency of the trigram as a unit is a significant predictor of first fixation durations on Word 1, and total fixation durations on the trigram. This effect is not consistent facilitation, as might be expected. Instead, we find that the first fixation durations on Word 1 are longer when the frequency of the trigram is high, and that there is no trigram frequency effect in the subgaze durations on the first bigram. There are two potential explanations for this change in trigram frequency effect over the course of the trial. The first is a holistic representation theory, where the trigram information is accessed through 'extreme' parafoveal-on-foveal preview benefit. That is, perhaps there is access to the whole trigram when attention is on the first word. This would explain the facilitation for the Bigram 1 interaction with restricted collocations in the first fixation duration, based on the restricted collocation qualities described above. This theory, however, requires a discrete representation of *each* trigram in the Mental Lexicon, in order to have facilitation of all trigrams modulated by frequency in the total trial fixation durations. Such representation has been posited in the past to be highly redundant and costly in terms of memory requirements. Is it truly necessary that each lexical bundle, for example, be memorised even when they are completely compositional (such as *in the middle of*)? Additionally, a holistic representation theory such as the Configuration Hypothesis cannot explain the lack of constituency effects at the recognition of the frequent MWU in FFD, nor the inhibition or null effect for trigram frequency. A second explanation for the trigram frequency effect was mentioned in the FFD results section, and is predicted by parallel reading models such as SWIFT (Engbert et al., 2005; Kliegl, Nuthmann & Engbert, 2006). In this view, the inhibition of the high frequency trigram is caused by the parafoveal preview of the entire trigram from the first word. This requires parafoveal-on-foveal effects for not only n+1, but also n+2 (Kliegl, Risse & Laubrock, 2007). Additionally, these effects are per word, rather than occurring over a holistic memorised unit. Therefore we propose a usage-based model of language as providing the best explanation for our results. This is due to the inability of holistic models to account for how individual word frequency and bigram frequency effects occur when reading MWUs: If the MWU is accessed as a whole, then the frequency of word n+1 or the second bigram should not affect the durations of fixation for the first word or second bigram. These effects are instead present until the total trigram fixation durations. Conversely, a usage-based model such as an exemplar theory can account for both MWU-level and word-level effects, where all associated information with each word would be stored in an exemplar cloud with strong links to the whole MWU.

Crucially for our results, the trigram frequency effects in our data fit an explanation of language storage and retrieval that can accommodate frequency effects at both the word and MWU level, and access to the whole MWU from the first word of the trigram. That is, in the first fixation duration data on Word 1, trigram frequency leads to significantly longer first fixation durations. This requires parafoveal access to Word 3, for the trigram frequency to be of note in the first fixation duration on the first word. In the subgaze duration data for the first bigram, the trigram frequency variable is not predictive, having effectively been cancelled out with current facilitation from the earlier inhibition. Since the longer reading times of Word 1 and Word 2 are included in this reading measure, then we find an overall null effect. Once we reach the total fixation durations on the whole trigram, though, the reader has (fully) processed each of the three words, leading to facilitation for total fixation durations.

On the other hand, the lack of effects found for Bigram 3 frequency (i.e., of Words [1+3]) is of theoretical interest for those MWUs which have empty 'slots'. For example, in our stimulus set we had stimuli such as *talk negatively about* and *so damn far*, where the middle word is completed compositionally by the speaker, and is not fixed as Words 1 and 3 are. This is particularly important for idioms, which can often be considered as templates rather than completely fixed expressions, such as *As happy as a N in N, Up the/a/shit creek without a paddle,* or *N got under N's skin*. That bigram 3 frequency was not significant in any reading time measurement is a problem for idioms of this type. This is because in terms of parafoveal preview, we should be able to access the entire trigram, and therefore find a bigram frequency advantage. However, as idioms and restricted collocations were the minority in this stimulus set, then we cannot be sure that this effect has not been cancelled out by the non-effects for lexical bundle trigrams. Future investigation manipulating the Bigram 3 frequency and idiom category is needed to examine this more thoroughly.

In conclusion, in this study our findings point to MWU type effects beginning at the processing of the first word on the first pass, and continuing over the duration of the trigram reading. While word frequency has proven a significant predictor, we have also found that, in the total fixation duration time for the trigram, effects of whether the trigram is considered a semantic and/or syntactic constituent have shown predictive ability. Additionally, we have seen that MWU types are processed differently in the three dependent time variables investigated here. However, the results also showed that the effects of MWU type and trigram frequency are not consistent throughout the trigram processing. The frequency of the trigram leads to significantly different effects in the different time measurements. With respect to the trigram frequency effects, we have considered two potential accounts, and found that a usage-based reading model has more explanatory value than the theory of holistic representation of MWUs. Further, the evidence supports parafoveal-on-foveal frequency effects as predicted by both serial and parallel models of reading. Yet with evidence for parafoveal preview of more than one word ahead of the eye’s focus, our results point more to a parallel processing theory such as SWIFT. In sum, this study has given further evidence in support of categorical differences between MWU types, and of MWU processing as requiring access to the entire unit from reading of the first part of the MWU. That is, MWUs are treated very similarly to single words in reading experiments. Finally, the results suggest that usage-based models and a parallel reading processor may best explain the reading times for the three types of trigrams investigated here.

References

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*, 390-412.

Bates, D., Maechler, M. & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 Classes [Computer software manual]. Available online at: http://CRAN.R-project.org/package=lme4 (R package version 0.999375-39).

Bobrow, S. & Bell, S. (1973). On catching on to idiomatic expression. *Memory and Cognition, 1*, 343-346.

Brants, T. & Franz, A. (2006). *Google 1T N-gram Web Corpus*. Google Inc. Available at Linguistics Data Consortium, Pennsylvania.

Cacciari, C. & Glucksberg, S. (1991). Understanding idiomatic expressions: The contribution of word meanings. In G. B. Simpson (Ed.), *Understanding word and sentence*. (pp. 217-240). North Holland: Elsevier.

Cacciari, C. & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language, 27,* 668-683.

Columbus, G. (2010). Processing MWUs: Are different types of MWUs psycholinguistically valid? An eye-tracking study. In D. Wood (Ed.), *Perspectives on formulaic language in communication and acquisition*. (pp. 194-210). London/ New York: Continuum.

Engbert, R., Nuthmann, A., Richter, E. & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777-813.

Hohenstein, S., Laubrock, J. & Kliegl, R. (2010). Semantic preview benefit during eye movements in reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1150-1170.

Kliegl, R. (2007). Towards a perceptual-span theory of distributed processing in reading: A reply to Rayner, Pollatsek, Drieghe, Slattery, & Reichle (2007). *Journal of Experimental Psychology: General, 138*, 530-537.

Kliegl, R., Nuthmann, A. & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135,* 12-35.

Kliegl, R., Risse, S. & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word n+2. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1250-1255.

Laurent, J-P., Denhières, G., Passerieux, C. Iakimova, G. & Hardy-Bayle, M-C. (2006). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research, 1068*(1), 151-160.

Lemke, S. F., Tremblay, A. & Tucker, B. V. (2009). Function words of lexical bundles: The relation of frequency and reduction. *Proceedings of Meetings on Acoustics, 6*(1), 060009-9.

Libben, M. & Titone, D. (2008). The multidetermined nature of idiomatic expressions. *Memory and Cognition, 36*, 1103-1131.

R Development Core Team. (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Available online at: http://www.R-project.org.

Schmitt, N. & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Formulaic sequences.* (pp. 173-189). Amsterdam/Philadelphia: Benjamins.

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

Swinney, D. & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior, 18*, 523–534.

Titone, D. & Connine, C. (1999). The compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics, 31*, 1655-1674.

Tremblay, A. (2009). Processing advantages of lexical bundles: Evidence from self-paced reading, word and sentence recall, and free recall with event-related brain potential recordings. Unpublished Ph.D. Dissertation. University of Alberta, Canada.

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 and New York: Continuum.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S. & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience, 22*(8), 1682-700.

Warren, T., White, S. J. & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition*, *111*(1), 132-137.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

General Discussion and Conclusions

In this dissertation I have investigated various aspects of MWU processing through eye movement studies. The results from each study support the concept that MWUs incorporate several types, of which three – idioms, restricted collocations and lexical bundles – have been investigated here. Overall, MWU type differences have been found in sentence reading and trigram reading through fixation data. Additionally, the combined results here support an exemplar-based model of language, and a parallel processing approach in reading.

Recall that in the introduction, we set out several issues in MWU studies and phraseology that were yet to be resolved. One such issue was whether some types of MWUs could be distinguished by objective measures such as reading time results, rather than relying on seemingly subjective classifications, as Vanlancker-Sidtis (2004), among others, has observed. I have shown in each of the studies here that MWUs of idiom, restricted collocation, and lexical bundle types can be and are differentiated in both feature-based and experimental time-based measures. In Chapter 1, the first fixation durations and total word fixation durations were predicted by the type of MWU being read in the sentence. Additionally, the familiarity, semantic transparency, and MWU frequency values were significantly different between the three types of MWUs for the 150 trials in this study. Interestingly, the model results show that these MWU-level variables are not entirely responsible for the differences between MWU types, since there were several models which required both the MWU Type variable and Semantic Transparency *or* MWU Frequency *or* Familiarity to account for significant variance in the reading times. For the trigram data in Chapter 2, there was clear evidence of MWU type predicting subgaze durations and total fixation durations of the trigrams, and for restricted collocations (in interaction with bigram frequency) predicting first fixation durations. Overall, I believe the studies offer strong support for the classifications of these three MWU types in the traditional, non-empirical literature. This is important in terms of MWU experiment design: Since we have found significant differences in MWU processing times depending on the MWU type, then studies which encompass more than one type of MWU ought to include the type as a factor. Additionally, reanalysis of some previous studies' results with MWU type as a factor may return clearer findings.

Secondly, this dissertation aimed to determine the variables relevant in predicting MWU processing. This has been achieved in the sentence reading study. The eye-movement analysis of 150 MWU sentences showed that semantic transparency, familiarity, and MWU frequency were important as control-type variables which accounted for variance within the first fixation duration, word reading time, and sentence reading time data. However, none of these variables was key in predicting the MWU reading times. Instead, individual word frequency, and, in the word reading and first fixation durations, the location of the word with respect to the MWU (before, in, after, outside) was the most predictive. Yet when the analysis was restricted to only the words within the MWUs, we did find a significant interaction between MWU Type and MWU Frequency in predicting word reading times. It seems that when the surrounding context is not visible to the model, there is a clear MWU frequency effect. The fact that this does not occur when all words in the sentence are accounted for in the model shows a clear link between the sentential context and ease of reading MWUs. Future studies should manipulate this context factor to determine if it is a reflection of co-occurrence frequencies and predictability in the surrounding text.

On the other hand, in the trigram reading study (Chapter 2), we see that MWU frequency is both inhibitory (at the first fixation on the first word) and facilitatory (for the combined durations for the whole trigram). In contrast, individual word frequency was a significant variable across reading times, but, in accordance with theories of parafoveal preview, the higher the frequency of the next word(s), the slower the reading of the current word.

Indeed, this issue of parafoveal-on-foveal reading effects is a third concern of these studies. Recall that the sentence reading experiment and the trigram reading experiment had somewhat unexpected results with respect to frequencies and MWU types in the first fixation duration data. In Chapter 1, we found that idioms had no word frequency effects at the first fixation, across all sentence positions. All post hoc analyses to determine the cause of this effect – including the model for only the words within the MWUs – returned nonsignificant results. This led us to propose an extreme parafoveal preview – an extension of the SWIFT model's semantic parafoveal effect (see e.g., Hohenstein, Laubrock & Kliegl, 2010). In this theory, the reader parafoveally gains some semantic information from the MWU, and so lengthens the first fixations on all words in the sentence. This cannot be easily accounted for in a more serial reading theory such as E-Z Reader, since there is no semantic uptake from parafoveal view in that theory. When a similar issue surfaced in Chapter 2, with first fixation durations being inhibited by high MWU frequency, again it seemed that an extreme parafoveal preview was occurring, with enough attention on all three words for the reading of the first word to take significantly longer. We note that the trigram result does not rule out an E-Z Reader model of reading, since it is a frequency-based effect, albeit in a longer word span than is typically discussed in the serial reading literature. Another explanation may be that as the trigram stimuli were presented without a vehicle sentence, the lack of context could explain the inhibition on the stimuli, instead of finding no effect as we found with the idiom sentences from Chapter 1. This is not consistent, though, with the post hoc regression models on the dataset which included only the word within MWUs: there was still no word frequency effect found for idioms when the surrounding words were excluded from the analysis. However, as idioms are not typically frequent MWUs, this question remains open to further investigation.

As I indicated in the introduction, the major theoretical question regarding MWUs is still whether MWUs are processed holistically as one unit, or as individual word units. From the viewpoint of our results, this remains unclear, since there is evidence for both. For instance, in the trigram study we found weak support for holistic access to MWUs based on the inhibition from MWU frequency at the first fixation on the first word. However, the lack of effect from the split bigram frequency and the additional effects from the frequency of each word and each bigram within the trigrams support access to each word in a sequential way, with processing of each word dependent on the features of the words surrounding it. Similar evidence is found in the sentence reading study (Chapter 1), in that the effect of the MWU from the first fixation durations indicates that access to the MWU affects reading of the remainder of the sentence. That is, there were near-instantaneous effects from MWU recognition, which seems consistent with holistic access to the MWU. Taken together, the evidence points to a theory of MWU storage and access which encompasses both individual word access and MWU access. The Hybrid theory (Titone & Connine, 1999) or the similar Superlemma theory (Sprenger, Levelt & Kempen, 2006) discussed earlier would fit well with the data we have, since they both posit links between the individual words and full idioms. Yet such theories may be best understood in an exemplar-based context, such as those described by, for example, Bod (1998) and Bybee (2006). That is, the information retained with each MWU should include not just the frequency and meaning of the unit, but the frequency and meaning of each of the component words, bigram frequencies, appropriate usage in context and genre, and so forth. This way, we can account for differences between e.g., lexical bundles and restricted collocations as well as between e.g., lexical bundles and idioms in terms of processing times and neural components. Furthermore, an account such as this could explain the differences found between prior psycholinguistics studies of MWU. In particular, an exemplar-based approach could also explain the differences between findings in both context-full and context-free experiments. That is, the co-occurrence frequencies for all words occurring in an utterance with a MWU would be stored as weak links with the relevant MWU, complete with those words' own semantic, phonological and frequency information. Thus any activation of the MWU would also access the entire set of contexts, and any matches would result in faster processing of the MWU and the remainder of the sentence. This was found in Chapter 1 for idiom sentences, but not lexical bundles, which is consistent with the idea that a MWU with no semantic unity would not have experience, i.e., exemplars, of highly similar contexts, leading to slower post-MWU reading of lexical bundle sentences when compared to idioms. Since the MWUs occurring without a vehicle sentence have no need to refer to the properties of the words in the context, then the reading of the MWU item is more reliant on access to the semantic content and the frequencies of co-occurrence and the whole unit. Hence, the processing of a MWU is made easier by the access to its combination of properties, and so those with semantic unity such as idioms are more likely to be processed faster when they are presented without a context. We found this advantage for idioms in both the trigram analysis and in the post hoc analysis of the within-MWU subset data in the sentence reading paper. Conversely, the sentence reading study results showed a processing advantage for lexical bundles over idioms, which points to both higher co-occurrence frequencies leading into the lexical bundle MWU, and to integration being unnecessary in pairing the prior context to the lexical bundle itself. In contrast, the idiom must deal with competing semantics for the individual words versus the whole of the MWU and then also integrate the meaning with the prior context.

In terms of non-native speaker MWU processing, an exemplar-based model can best explain the reading times and predictors found here also. In fact, the L2 results highlight the importance of frequency for MWU processing, both for words and for MWUs. Recall that in the sentence reading experiment in Chapter 1, L2s had stronger facilitation from word frequency than the L1s did. The non-native group were also the only readers to have MWU Frequency and Semantic Transparency effects, both in the word reading times. Though the latter occurred in an interaction with MWU Type and In MWU Region, it shows that the reading of the MWU is affected by both MWU-level and word-level variables. This suggests that the L2 readers have a usage-based model of storage and access which, when taken with the general findings of otherwise similar predictors and MWU type processing order, is highly similar to the L1 English representation. To determine how similar the processing framework is, future investigation into exemplar-based representation of MWUs in non-native speakers requires L2 groups with extensive information on or close control of the readers' L1s, length and type of exposure to the L2, and comprehension testing for each of the MWUs.

Further to future experimental design, we have shown here that we can still determine differences in processing, and the variables key to that processing, while using randomly-selected and randomly-constructed sentences. Each MWU stimulus in the sentence reading study came directly from a structured corpus of English (the BNC), with no restrictions on word length, sentence length, MWU length, or sentence structure, let alone word or MWU frequency. Instead, we successfully accounted for the variance in the mixed-effect regression analysis for the eye-movement data, and in the generalized additive models in the ERP data. The results here, then, are generalisable to the majority of naturally-occurring sentences which contain any of these three MWU types.

To sum up, by performing a set of eye-movement tests on three different MWU types, it has been possible to establish that the structural differences between MWUs have led to distinct processing and storage. Overall, the eye-movement experiments discussed here offer support for an exemplar-based model of language storage and access, with conceptual levels for MWUs which are accessed along with all other variables from the lexeme level. From a linguistic perspective, the results from these two studies build on the MWU debate, particularly with respect to clarification of how three MWU subtypes are processed in not only native speakers but also non-native speakers. These results will benefit various fields of linguistic research, particularly regarding language modelling and lexical theory, as well as offer clinical and pedagogical applications.

References

Bod, R. (1998). *Beyond grammar: An experience-based theory of language.* Stanford, CA: CSLI Publications/ Cambridge University Press.

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

Hohenstein, S., Laubrock, J., & Kliegl, R. (2010). Semantic preview benefit during eye movements in reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1150-1170.

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

Titone, D. & Connine, C. (1999). The compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics, 31*, 1655-1674.

Vanlancker-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 Communication Disorders, 39*(1), 1-44.

Appendices

Appendix A. Full stimulus set for Chapter 1. Note that some stimulus numbers are not contiguous as some stimuli were used in the practice trials.

Key: CN – control; ID – idiom; LB – lexical bundle; RC – restricted collocation; SA: semantically anomalous.

1	Only use it to protect the contents of the tree while erasing.	SA
2	The earth is rapidly dried away by the water.	SA
3	This could mean an extra $6 cents.	SA
4	Peregrine's attention was paid by this.	SA
5	In an interview with the Guardian, he attacked the Prime Minister for failing to have a pleasant strategy.	SA
6	So what exactly is gossip -- this cat that is so reviled.	SA
7	The Kiwis know that British is best and have said Klondyke's quality is better than they don't obtain in the far East.	SA
8	If the future is anything to go by, the position will change but little.	SA
9	Just don't include it in the flower, that's all.	SA
10	Using handholds in the wood to steady herself, she spoke back along the beam.	SA
11	This has been gilded in conversations with Mr Keith Turner.	SA
12	And the town was said to be "wet in piety, gifts and wealth".	SA
13	Her frustration reached a pitch in which she screamed and threw birds round the house.	SA
14	Our first serious engine secret was in Amman.	SA
15	A down-stuffed duvet in a printed-cotton onion lay over them.	SA
16	In 1992, everything (apparently) heats with a D.	SA
17	Eyes sparkling, Jill gazes at a relaxed John in the garden of their hideaway wallet.	SA
18	Please sweep in the fun, buy the T-shirt.	SA
19	Ten years later a state of emergency is still in wood, renewed in May 1991.	SA
20	The pictures we form of the places eaten in our study are no exception.	SA
21	In response the Kremlin tightened its reach in the East.	SA
22	JustText, the program in question, is a very high etiquette tool.	SA
23	She lived in a house down the pipe at Brampton.	SA
24	She's changing; in the apple, that is.	SA
25	On warm summer days ponds in South Georgia are occasionally jogged with a grey film.	SA
26	Melt the remaining carrots in a wide-bottomed pan, add the bamboo shoots.	SA
27	She was alone in the concrete.	SA
28	He was mayor of the wall in 1552.	SA
29	Mrs Howard had a ripple.	SA
30	Literary texts are not some static crystalline structure in which we may glimpse a captured swimming past.	SA
31	We were going round and round in a weird dance and nearly read over in the pee.	SA
32	I could see George in her, I was sleepy of it.	SA
33	It remained as in Roman earth.	SA
34	I'm actually in the carpet in Tower Street.	SA
35	Will there be paper employed to actually fill in?	SA
36	Much later they lay in one another's arms, at peace amidst a frenzied information of bedclothes.	SA
37	One in five will graze from a hip fracture.	SA
38	It can mean greenish death.	SA
39	But Andrew Barker spotted the advert in the sea.	SA

40	And banana used to go in his car.	SA
41	I mean ...some time ... in this next term would be the metallic thing really.	SA
42	Twenty readers plus their partners had ringside seats at the Daily Mirror sponsored leaf.	SA
43	Alice sat reticent, watching the air develop.	SA
44	Then at 5 a.m. one morning, the security eggs kidnapped her.	SA
45	He had been gargled at Westminster Abbey Choir School.	SA
46	Jane Pargeter sniffed and prodded with her hair at an imaginary stain in the carpet.	SA
47	The wind licked at the white cloth and nailed up one corner.	SA
48	Scouts were put through their gazelles at the weekend.	SA
49	We could stop off at ... at ... yeah... at road.	SA
50	Girls have only the choice of continuing at school or attending this coffee.	SA
51	I know Freiburg.	CN
52	He's marvellous and he was the curate.	CN
53	A degree of spitting is beneficial.	CN
54	My husband would be a nature poet if he didn't live now, here, in England.	CN
55	How good is the course?	CN
56	Journalists have a professional obligation to protect confidential sources of information.	CN
57	After the gig, we mooch around the band's dressing room.	CN
58	"Very ingenious," he said.	CN
59	Beyond the Chindwin is a restrained account of his experiences as a Column Commander.	CN
60	Freeze the mixture in a shallow metal pie pan.	CN
61	Fortunately this was a case where I had felt very confident of the remedy.	CN
62	"Go to Paris", he advised, "there is a painter there who is a beautiful man".	CN
63	He alleged that a West Ham player assaulted him, in front of his wife.	CN
64	He was an Australian and it sounded nasty the way he said it.	CN
65	It would be much easier if I could do that informally.	CN
66	His artistic fame rested primarily on the series of paintings he did of Queen Victoria.	CN
67	I borrowed a veil from my aunt.	CN
68	Don't take birds' eggs or nests.	CN
69	You don't need this.	CN
70	It is held when the baby is seven days old.	CN
71	They opened the door.	CN
72	More, he was almost the last representative of Tory paternalism.	CN
73	Not surprisingly, Ken was apprehensive about the position I might inherit when I touched down.	CN
74	Once I'm bit warmer, I go into Marie's room.	CN
75	I don't know how you can even speak of her in the same breath.	CN
76	Pineapple and gin have a truly spectacular affinity.	CN
77	You don't lock that shed?	CN
78	There can be debate and compromise.	CN
79	She was considering becoming a patron of his charity-registered foundation for Aids work.	CN
80	That weekend is also memorable for something Dana said to me.	CN
81	The Muslims and Sikhs have adopted the dowry system.	CN
82	I do not understand the motives, beliefs and values of terrorists.	CN
83	The major part is a recurrent grant paid by the government.	CN
84	The tricky part would come when they had to leave the straight track and turn down through a gate and a field.	CN
85	The question now is how much further they will push.	CN
86	The Royal Assent is today granted by Commissioners.	CN

87	The same thing happened to my father.	CN
88	The teaching is excellent.	CN
89	Then place them in a large saucepan with all the remaining ingredients.	CN
90	They were returning home from a factory in Germany.	CN
91	This happens when there is a voluntary tightening of the vocal cords.	CN
92	This has a spigot bearing the same size as the Range Rover but made of steel and not phosphor.	CN
93	This will put the Albanian economy back in the black.	CN
94	This, in over-simplified terms, is the precise function of drama.	CN
95	Time has passed you by.	CN
96	We have now agreed to produce a joint leaflet and hope to distribute it to BMH.	CN
97	What have we here?	CN
98	What major difference is there between the Yorkshire and Lancashire coalfields?	CN
99	Don't let him take me.	CN
100	Another Conservative, Terry Dicks, MP for Hayes and Harlington, condemned it as "filth".	CN
101	The new measures are in response to the public outcry over illegal rave parties such as the one at Castlemorton.	RC
102	I have no doubt this has been a terrible ordeal for you.	RC
103	Alcohol might have played some part in this serious miscalculation.	RC
104	By a narrow margin the draft was rejected on 5 May.	RC
105	He says it doesn't make sense to point the finger of blame, it's the law that is at fault.	RC
106	And as you are well aware, I have had precious little from the police on that case.	RC
107	I galloped into the hotel just as the most torrential downpour I have seen in my life cascaded from the sky.	RC
108	The food is ingested and the bacteria may lie dormant in the bird.	RC
109	The "Playmore" has proved an invaluable asset to many clubs.	RC
110	It is clear that we should not be led astray by glamorous starlets.	RC
111	And she twisted her arm sharply downwards and broke his grip.	RC
112	I'm sorry, but I cannot be at your beck and call all the time, my dear.	RC
113	Had he grown to think she would meekly do his bidding.	RC
114	We seem to have a prospective buyer and Liz is making frantic signals to me.	RC
115	He lit up a cigarette and looked at Kelly thoughtfully.	RC
116	He laughed so hard that he had a fit of coughing.	RC
117	Do you share a date when your paramour is unavoidably detained on business in another town halfway across the country.	RC
118	Through hard work and sheer determination, Dennis carved out a career in the building industry.	RC
119	She expects to find Minna and Becky sitting in a huff in the coffee lounge.	RC
120	Never a dull moment running a hotel in the Caribbean.	RC
121	The aftermath of an affair turned sour can be downright hostile.	RC
122	Now he had the nerve to go back on his word.	RC
123	Stimulation of the receptors in the inner ear gives rise to the experience of sound.	RC
124	Other information from the completed schedules does give cause for concern.	RC
125	It seems to me there is no other way to get the message over to some of the morons out there.	RC
126	But when he finally turned up, he no doubt got an earful like the rest.	RC
127	The listener is human, so allow him to get a word in edgeways; he too may have a point of view.	RC
128	Mrs Burrows and Joyce had guitars, Beverley flexed his muscles on the piano accordion, Walter had a saxophone.	RC
129	But, before you can find a solution, you first have to realise that a problem exists.	RC
130	The West German Chancellor also fell into line with his French allies.	RC

131	This unfortunate Empress has been so dragged through the mud that one feels almost compelled to defend her.	RC
132	The stranger closed the door and elbowed his way past the landlord.	RC
133	The death of his wife the following year drove him mad and he died in 1743.	RC
134	The evening was drawing to a close and everybody was dancing with new-found friends.	RC
135	Since the war, governments had by and large ducked the issue.	RC
136	He'll dive into a prolonged fit of the sulks again.	RC
137	If they acted properly, I did well by them.	RC
138	A cloud seemed to descend upon Lucy, causing her to remain silent.	RC
139	He had earlier declined to declare an interest in the country's leadership.	RC
140	My speculations were cut short by the appearance of Karen herself at my elbow.	RC
141	Behind closed doors Diana cried her eyes out with nervous exhaustion.	RC
142	The police are only now beginning to crack down on this type of smuggling.	RC
143	Although Sorley was to make light of the experience, he had been in considerable danger.	RC
144	No, a childminder cannot normally take care of a sick child.	RC
145	The retailer needs to pay attention to where and how he stores goods.	RC
146	In the autumn of 1987, Conran ran into a potentially more serious stumbling block to his plans.	RC
147	In a miscellaneous pile of documents, I came across the following essay.	RC
148	There was another uncomfortable silence, and Alyssia clicked her tongue impatiently.	RC
149	Her work is different and always tries to break new ground.	RC
150	Christopher Taylor blinked back tears as he described how he had arrived at the hospital.	RC
151	If she didn't clear her name then everything she had worked for would be wasted.	RC
158	They are certainly simple and that is one of the chief advantages of cash accounting.	LB
159	You can turn the television on if you want to.	LB
160	The success may be something to do with the positive attitude.	LB
161	Diana considered her wedding day to be one of the most emotionally confusing times of her life.	LB
162	When the hunters at the head of the column discover prey, they swarm all over it.	LB
163	Making a will is the only way to ensure your wishes are respected.	LB
164	So we are going to have to do some detective work to find out about Matthew.	LB
165	Scanning and the use of maps and diagrams may cause problems.	LB
166	I don't like it myself, but it won't go on for ever.	LB
167	By the end of the fifteenth century, sugar was a large export.	LB
168	"While I'm getting ready, would you like to come up and see my room?"	LB
169	A typical training programme usually takes the form of a series of workshops, sometimes in a residential setting.	LB
170	A referendum to determine the future of the island has been postponed indefinitely.	LB
171	There were dangerous jobs and little in the way of safety precautions.	LB
172	They found that the older children born in the first part of the year did better in reading and arithmetic tests.	LB
173	What is more, owing to various factors, the number of people living alone is growing.	LB
174	Drought conditions have led to the introduction of a new French water bill.	LB
175	It has been forced to slash prices, with the result that profits dropped 11 per cent to £62 million.	LB
176	For the life of me I can't remember what courses I took -- it was so boring.	LB
177	What we don't want to see is our own producers being undercut.	LB
178	If you crack one more joke I'm going to the supervisor.	LB

179	A comparable tendency is to be found in the theatre.	LB
180	It is interesting to see the gravestones of the numerous people who came here.	LB
181	People are more likely to be struck by lightning than bitten by an adder.	LB
182	He had championed the cause of the poor for many years in a series of investigative articles.	LB
183	And in fact it's not just the elderly who are affected.	LB
184	I don't personally believe that, but I think that's what is forced upon us by society.	LB
185	After the reduction in profits in the first half of the year, profits in the second half improved.	LB
186	I think you've got to be fair about how much time you give the family.	LB
187	This is a very tasty way of using up any left-overs after a meal.	LB
188	If it affects you so much, why don't you go inland yourself?	LB
189	Yet I don't think I used to be very naughty.	LB
191	We realise that this is not a very satisfactory state of affairs.	LB
192	Everything added together puts Mills very much on the side of guilt rather than innocence.	LB
193	In the nick of time, they discovered her in a state of hapless stage fright.	LB
194	If trouble does come, the police are going to be put under enormous pressure.	LB
195	I object to being misrepresented and I want to make sure that this sort of thing does not happen again.	LB
196	Rice is one of the most important foods in the world.	LB
197	It's not as bad now as it used to be.	LB
198	Do you need to go back to the house to change?	LB
199	Later in the same year he was able to compensate his brother in handsome style.	LB
200	Well it is awful but I don't know what you can do about it!	LB
201	Until suitable workshop facilities could be made, it was clear that little could be done on the Oxford's airframe.	LB
202	Silence punctuated with gasps from the other end of the line showed that Maurice was roaring with laughter.	LB
203	On the one hand, it can be seen as a discipline, indeed perhaps the oldest of all disciplines.	LB
204	Sometimes, though, the outcome of the advertisers' work has surprising results.	LB
205	He ran around with a gang of schoolfriends and was a member of the local swimming club.	LB
206	Mrs McGuire admitted the party had yet to fully come to terms with the election defeat last April.	LB
207	Yes. I saw her on the way to the doctor's and she'd thought she'd got chicken pox.	LB
208	You know I don't think you can get by without them really.	LB
211	I was desperately trying not to let the cat out of the bag.	ID
212	Journalists who refuse to toe the line will have to be sacked.	ID
213	It's inevitable that heads will roll.	ID
214	The penny dropped with a resounding clunk in her brain.	ID
215	And now she's trying to bite the hand that fed her," said a spokesman.	ID
216	It was raining cats and dogs and the teachers were running in and out helping us.	ID
217	Don't count your chickens, Dexter.	ID
218	But he was pretty down in the dumps and I felt terrible because I couldn't be there.	ID
219	Jenkinson blew his top after missing a penalty at Plymouth on Friday night.	ID
220	Course they were worried because ... they'd bit off more than they could chew.	ID
221	You broke her heart, you know you did!	ID
222	At first he kept saying that housekeeping was a piece of cake and that women made a big fuss about nothing.	ID
223	IBM never knows when to leave well enough alone.	ID
224	She hesitated, thought "In for a penny, in for a pound!" and began to walk.	ID
225	Constance knew the time had come to face the music and speak to Nora.	ID

226	The moment she said it she knew she had hit the nail on the head.	ID
227	And she had Uncle Geordie completely under her thumb.	ID
228	You're making a mountain out of a molehill, Dorothy.	ID
229	Lady, you're lucky you can't read my mind.	ID
230	He keeps his nose to the grindstone and thinks everyone else should.	ID
231	No, we're not trying to pull the wool over your eyes.	ID
232	He might as well make hay while the sun shone, he told himself.	ID
233	She just went haywire as soon as she saw the instructor.	ID
234	Yes, well you'd be cramping his style, wouldn't you?	ID
235	Mr Kanemaru's reputation took a dive from which it has not recovered.	ID
236	She had the strangest sensation he was toying with her as a cat would a mouse before a kill.	ID
237	The old boy had lost his marbles somewhere along the line.	ID
238	But then those two don't see eye to eye about anything these days.	ID
239	He needed someone to throw him a lifeline and I decided it might as well be me.	ID
240	Are they with me on this or am I boring the pants off them yeah?	ID
241	Rationalist philosophy paved the way for a reexamination of women's place in society.	ID
242	Family or others who went to her for help came away empty-handed.	ID
243	Keith is clutching at straws in an effort to win the argument.	ID
244	"And you're to scare the living daylights out of her, do you hear?"	ID
245	But Mr Kaifu is in no mood to let anyone steal his thunder.	ID
246	You may even throw caution to the wind and try one of our Mystery Trips.	ID
247	And he pointed the finger at a couple of his team mates.	ID
248	He had got under her skin, and after half an hour she went home alone.	ID
249	I think the papers made his day, perhaps he relished scandal.	ID
250	Ackroyd knew immediately where he meant and his blood ran cold.	ID
251	Sometimes even junior Libyan officials dug their heels in on particular points of practical application of rules.	ID
252	"I wanted to pick your brains," I said.	ID
253	And this could help Ballymena put their best foot forward.	ID
254	All my life I've had brothers breathing down my neck, watching my every move.	ID
255	When I decide to call it a day, they'll stop too.	ID
256	An injury to the Motherwell veteran Davie Cooper at the eleventh hour meant that Fleck was summoned to join the squad.	ID
257	It seems the collectors get a kick out of just looking at them.	ID
258	Anything bigger than this is too much like hard work.	ID
259	It was on the cutting edge of modern style.	ID
260	It's all right, Robbie, I was pulling your leg.	ID

Appendix B. Transparency of word strings questionnaire. (One of four versions with randomly-ordered stimuli.)

Dear Participant,

Thank you for very much for participating!

This questionnaire is about phrases we find within sentences. More specifically, it is about how transparent, or literal, the meanings of the phrases are. Each of the sentences in this questionnaire contains a phrase that may be literal, idiomatic or figurative. Idiomatic sentences are ones where the meaning of the phrase doesn't match the content of the individual words. Some examples of phrases like this are in bold below, where the Raptors don't literally run on steam (like a steam train would), and Janet is not literally walking or running a mile for her clients:

The Raptors ***ran out of steam*** *in the final quarter.*
Janet always ***went the extra mile*** *for her clients.*

Your task is to judge how transparent / literal these phrases are. That is, you have to **decide how much you can understand the meaning of each sentence from the meaning of the words in the sentence.**

Please judge every sentence by assigning a number to it. You can use any number that seems appropriate to you. Write down your rating in the box next to the first sentence (numbered 0). For each sentence after that, **assign a value to show how much more or less transparent that sentence is in proportion to the first one** (numbered 0).

You can use any range of positive numbers you want, including, if necessary, fractions or decimals. You may not use minus/negative numbers or zero, of course, because they are not multiples of positive numbers. Any convenient positive number will do for the reference. **Highly transparent (i.e., very literal) ratings should be at the low end of the scale** you use, and **completely opaque ratings (impossible to get to the meaning of the phrase through the meaning of each of the words) should be at the high end.**

You should not restrict your responses to, say, an academic marking scale. For example, if the first sentence was:

Mr Smith ***gave*** *his secretary* ***the boot****.*

and you gave it a 4, and you think the next example

Thomas was a real ***snake in the grass****.*

is three times less transparent/more opaque, you would give it a 12.

There are no 'correct' answers, so whatever seems right to you is a valid response. Nor is there a 'correct' range of answers or a 'correct' place to start. We are interested in your first impression, so don't spend too much time thinking about your rating.

Remember:

Use any value you like for the first sentence
Judge each sentence in proportion to the reference sentence (number 0)
Use any positive numbers you think appropriate.

I am a Native English speaker ☐ I am a non-native English speaker ☐

No.	Sentence	Rating
0	When you are planning a schedule you should also **bear in mind** the location of speeches and statements.	

1	Mrs Howard had **a ripple**.	
2	Don't **take birds' eggs or** nests.	
3	The success may be **something to do with** the positive attitude.	
4	Journalists have a professional obligation to **protect confidential sources** of information.	
5	If trouble does come, the police **are going to be** put under enormous pressure.	
6	It's all right, Robbie, I was **pulling your leg**.	
7	We could stop off at ... **at ... yeah... at road**.	
8	It was **raining cats and dogs** and the teachers were running in and out helping us.	
9	They found that the older children born in **the first part of** the year did better in reading and arithmetic tests.	
10	A referendum to determine **the future of the** island has been postponed indefinitely.	
11	In 1992, everything **(apparently) heats with a** D.	
12	And she **twisted her arm** sharply downwards and broke his grip.	
13	Do you share a date when your paramour is **unavoidably detained** on business in another town halfway across the country.	
14	He **keeps his nose to the grindstone** and thinks everyone else should.	
15	The listener is human, so allow him to **get a word in edgeways**; he too may have a point of view.	
16	He says it doesn't **make sense** to point the finger of blame, it's the law that is at fault.	
17	The death of his wife the following year **drove him mad** and he died in 1743.	
18	You're **making a mountain out of a molehill**, Dorothy.	
19	What **we don't want to** see is our own producers being undercut.	
20	**Don't count your chickens**, Dexter.	
21	His artistic fame rested primarily on the series of paintings **he did of** Queen Victoria.	
22	The same thing **happened to my** father.	
23	He **lit up a cigarette** and looked at Kelly thoughtfully.	
24	Never a dull moment **running a hotel** in the Caribbean.	
25	The retailer needs to **pay attention to** where and how he stores goods.	
26	They opened **the door**.	
27	My speculations were **cut short by** the appearance of Karen herself at my elbow.	
28	The Kiwis know that British is best and have said Klondyke's quality is **better than they don't obtain in the far** East.	

No.	Sentence	Rating
0	When you are planning a schedule you should also **bear in mind** the location of speeches and statements.	

29	She's changing; in **the apple, that is**.	
30	I don't know how you can even **speak of her in the same breath**.	
31	The major **part is a recurrent** grant paid by the government.	
32	That weekend is also memorable **for something Dana said** to me.	
33	Scouts were put **through their gazelles** at the weekend.	
34	If you crack one more joke **I'm going to the** supervisor.	
35	Using handholds in the wood to **steady herself, she spoke back** along the beam.	
36	In an interview with the Guardian, he attacked the Prime Minister for failing to **have a pleasant** strategy.	
37	But he was pretty **down in the dumps** and I felt terrible because I couldn't be there.	
38	JustText, the program in question, i**s a very high etiquette** tool.	
39	Then place them in a **large saucepan with all** the remaining ingredients.	
40	Christopher Taylor **blinked back tears** as he described how he had arrived at the hospital.	
41	The aftermath of an affair **turned sour** can be downright hostile.	
42	I mean ...some time ... in this next term **would be the metallic** thing really.	
43	Literary texts are not some static crystalline structure in which **we may glimpse a captured swimming** past.	
44	Making a will is **the only way to** ensure your wishes are respected.	
45	Behind closed doors Diana **cried her eyes out** with nervous exhaustion.	
46	I think **you've got to be** fair about how much time you give the family.	
47	Mrs Burrows and Joyce had guitars, Beverley **flexed his muscles** on the piano accordion, Walter had a saxophone.	
48	In the nick of time, they discovered her **in a state of** hapless stage fright.	
49	No, a childminder cannot normally **take care of** a sick child.	
50	Do you need **to go back to** the house to change?	
51	Family or others who went to her for help **came away empty-handed**.	
52	Although Sorley was to **make light of** the experience, he had been in considerable danger.	
53	We have now agreed to **produce a joint leaflet and hope** to distribute it to BMH.	
54	I do not understand the motives, **beliefs and values of** terrorists.	
55	Since the war, governments had **by and large** ducked the issue.	
56	Alice sat reticent, **watching the air** develop.	
57	He was an **Australian and** it sounded nasty the way he said it.	
58	The earth is **rapidly dried away** by the water.	

No.	Sentence	Rating
0	When you are planning a schedule you should also **bear in mind** the location of speeches and statements.	

59	Sometimes even junior Libyan officials **dug their heels in** on particular points of practical application of rules.	
60	The police are only now beginning to **crack down on** this type of smuggling.	
61	The old boy had **lost his marbles** somewhere along the line.	
62	When I decide to **call it a day**, they'll stop too.	
63	Once I'm a bit **warmer, I go into** Marie's room.	
64	"And you're to **scare the living daylights out of her**, do you hear?"	
65	But then those two **don't see eye to eye** about anything these days.	
66	Yet **I don't think I** used to be very naughty.	
67	The Royal **Assent is today granted by** Commissioners.	
68	She lived in a house **down the pipe at** Brampton.	
69	If the **future is anything to go by**, the position will change but little.	
70	If she didn't **clear her name** then everything she had worked for would be wasted.	
71	Keith is **clutching at straws** in an effort to win the argument.	
72	I don't personally believe that, but **I think that's what** is forced upon us by society.	
73	Course they were worried because ... they'd **bit off more than they could chew**.	
74	This could mean **an extra $6 cents.**	
75	She obviously feels **it cannot be** resolved without going to court.	
76	The tricky part would come when they had to leave **the straight track and** turn down through a gate and a field.	
77	Mr Kanemaru's reputation **took a dive** from which it has not recovered.	
78	I borrowed a **veil from my aunt**.	
79	An injury to the Motherwell veteran Davie Cooper **at the eleventh hour** meant that Fleck was summoned to join the squad.	
80	**This is a very** tasty way of using up any left-overs after a meal.	
81	She hesitated, thought "**In for a penny, in for a pound**!" and began to walk.	
82	He had **got under her skin**, and after half an hour she went home alone.	
83	You **don't need this**.	
84	Only use it to protect the contents **of the tree while** erasing.	
85	They **were returning home from** a factory in Germany.	
86	This unfortunate Empress has been so **dragged through the mud** that one feels almost compelled to defend her.	
87	I think the papers **made his day**, perhaps he relished scandal.	
88	He had championed **the cause of the** poor for many years in a series of investigative articles.	

No.	Sentence	Rating
0	When you are planning a schedule you should also **bear in mind** the location of speeches and statements.	

89	The new measures are **in response to** the public outcry over illegal rave parties such as the one at Castlemorton.	
90	I was desperately trying not to **let the cat out of the bag**.	
91	Fortunately this was **a case where I had felt very** confident of the remedy.	
92	It was **on the cutting edge** of modern style.	
93	A cloud seemed to **descend upon** Lucy, causing her to remain silent.	
94	A down-stuffed duvet in a printed-cotton **onion lay over them**.	
95	It's not as bad now as **it used to be**.	
96	It seems to me there is no other way to **get the message over** to some of the morons out there.	
97	This has a spigot bearing the same size as **the Range Rover but** made of steel and not phosphor.	
98	She just **went haywire** as soon as she saw the instructor.	
99	Pineapple **and gin have a truly spectacular affinity**.	
100	Then at 5 a.m. one **morning, the security eggs** kidnapped her.	
101	But Andrew Barker spotted the **advert in the sea**.	
102	And he **pointed the finger at** a couple of his team mates.	
103	We realise that **this is not a** very satisfactory state of affairs.	
104	At first he kept saying that housekeeping was **a piece of cake** and that women made a big fuss about nothing.	
105	Another Conservative, Terry Dicks, MP for Hayes and Harlington, **condemned it as "filth"**.	
106	On warm summer days ponds in South **Georgia are occasionally jogged with a grey** film.	
107	And in fact **it's not just the** elderly who are affected.	
108	Her frustration reached a pitch in which she screamed **and threw birds round the house**.	
109	It is clear that we should not be **led astray** by glamorous starlets.	
110	But, before you can **find a solution**, you first have to realise that a problem exists.	
111	And the town was said to be "**wet in piety, gifts and wealth**".	
112	"I wanted to **pick your brains**," I said.	
113	Rationalist philosophy **paved the way for** a reexamination of women's place in society.	
114	Sometimes, though, **the outcome of the** advertisers' work has surprising results.	
115	Peregrine's **attention was paid by** this.	
116	She had the strangest sensation he was **toying with her** as a cat would a mouse before a kill.	
117	Constance knew the time had come to **face the music** and speak to Nora.	

No.	Sentence	Rating
0	When you are planning a schedule you should also **bear in mind** the location of speeches and statements.	

118	By **a narrow margin** the draft was rejected on 5 May.	
119	IBM never knows when to **leave well enough alone**.	
120	Jenkinson **blew his top** after missing a penalty at Plymouth on Friday night.	
121	Melt **the remaining carrots in a wide-bottomed pan**, add the bamboo shoots.	
122	This, in over-simplified terms, **is the precise function** of drama.	
123	This has been **gilded in conversations with** Mr Keith Turner.	
124	The evening was **drawing to a close** and everybody was dancing with new-found friends.	
125	A typical training programme usually takes **the form of a** series of workshops, sometimes in a residential setting.	
126	I'm sorry, but I cannot be at your **beck and call** all the time, my dear.	
127	After the reduction in profits in the **first half of the** year, profits in the second half improved.	
128	Scanning **and the use of** maps and diagrams may cause problems.	
129	I have no doubt this has been a **terrible ordeal** for you.	
130	Now he **had the nerve** to go back on his word.	
131	So **we are going to have to** do some detective work to find out about Matthew.	
132	Her work is different and always tries to **break new ground**.	
133	The wind licked at the white **cloth and nailed up** one corner.	
134	People **are more likely to** be struck by lightning than bitten by an adder.	
135	On the one hand, it can **be seen as a** discipline, indeed perhaps the oldest of all disciplines.	
136	Rice is one **of the most important** foods in the world.	
137	He **was mayor of the wall in** 1552.	
138	And **banana used to go** in his car.	
139	You know I **don't think you can** get by without them really.	
140	I could see George in her, **I was sleepy of** it.	
141	But when he finally turned up, he no doubt **got an earful** like the rest.	
142	Beyond the Chindwin is a **restrained account of his** experiences as a Column Commander.	
143	So what exactly is gossip -- this **cat that is** so reviled.	
144	If it affects you so much, **why don't you go** inland yourself?	
145	Ten years later a state of emergency is **still in wood, renewed in** May 1991.	
146	"While I'm getting ready, **would you like to** come up and see my room?"	
147	She was considering **becoming a** patron of his charity-registered foundation for Aids work.	
148	It remained **as in Roman earth**.	

No.	Sentence	Rating
0	When you are planning a schedule you should also **bear in mind** the location of speeches and statements.	

149	The West German Chancellor also **fell into line** with his French allies.	
150	It has been forced to slash prices, **with the result that** profits dropped 11 per cent to £62 million.	
151	And this could help Ballymena **put their best foot forward**.	
152	A degree of **spitting is beneficial**.	
153	I object to being misrepresented and I want **to make sure that** this sort of thing does not happen again.	
154	Until suitable workshop facilities could be made, **it was clear that** little could be done on the Oxford's airframe.	
155	"Go to Paris", he advised, "there is **a painter there who is a** beautiful man".	
156	And she had Uncle Geordie completely **under her thumb**.	
157	One in five **will graze from a hip** fracture.	
158	Lady, you're lucky you can't **read my mind**.	
159	But Mr Kaifu is in no mood to let anyone **steal his thunder**.	
160	He laughed so hard that he had a **fit of coughing**.	
161	It would be much easier if **I could do that** informally.	
162	Girls have only the choice of continuing at school **or attending this coffee**.	
163	It seems the collectors **get a kick out of** just looking at them.	
164	When the hunters **at the head of** the column discover prey, they swarm all over it.	
165	I'm actually **in the carpet in Tower** Street.	
166	He's marvellous and **he was the curate**.	
167	No, we're not trying to **pull the wool over your eyes**.	
168	In the autumn of 1987, Conran ran into a potentially more serious **stumbling block** to his plans.	
169	He ran around with a gang of schoolfriends and was **a member of the** local swimming club.	
170	What is more, owing to various factors, **the number of people** living alone is growing.	
171	There was another **uncomfortable silence**, and Alyssia clicked her tongue impatiently.	
172	Diana considered her wedding day to be **one of the most** emotionally confusing times of her life.	
173	Alcohol might have **played some part** in this serious miscalculation.	
174	In response the Kremlin tightened **its reach** in the East.	
175	The Muslims and Sikhs have adopted **the dowry** system.	
176	Just don't include it in **the flower**, that's all.	
177	Anything bigger than this is **too much like hard work**.	

No.	Sentence	Rating
0	When you are planning a schedule you should also **bear in mind** the location of speeches and statements.	

178	Everything added together puts Mills very much **on the side of** guilt rather than innocence.	
179	All my life I've had brothers **breathing down my neck**, watching my every move.	
180	It is held when the baby **is seven days old**.	
181	There were dangerous jobs and little **in the way of** safety precautions.	
182	What major difference **is there between** the Yorkshire and Lancashire coalfields?	
183	**It is interesting to** see the gravestones of the numerous people who came here.	
184	You can turn the television on **if you want to**.	
185	Well it is awful **but I don't know** what you can do about it!	
186	More, he was almost the **last representative of Tory paternalism**.	
187	The **question now is how** much further they will push.	
188	By **the end of the** fifteenth century, sugar was a large export.	
189	I **know Freiburg**.	
190	For the life of me **I can't remember what** courses I took -- it was so boring.	
191	Mrs McGuire admitted the party had yet to fully **come to terms with** the election defeat last April.	
192	My husband would be a nature poet **if he didn't live now, here,** in England.	
193	A comparable tendency **is to be found** in the theatre.	
194	Silence punctuated with gasps from **the other end of** the line showed that Maurice was roaring with laughter.	
195	How good **is the course**?	
196	**I don't like it** myself, but it won't go on for ever.	
197	She expects to find Minna and Becky sitting **in a huff** in the coffee lounge.	
198	Had he grown to think she would meekly **do his bidding**.	
199	Yes, well you'd be **cramping his style**, wouldn't you?	
200	He'll **dive into** a prolonged fit of the sulks again.	
201	You **broke her heart**, you know you did!	
202	The stranger closed the door and **elbowed his way** past the landlord.	
203	Drought conditions have led to **the introduction of a** new French water bill.	
204	Other information from the completed schedules does **give cause for concern**.	
205	Freeze the mixture in a shallow **metal pie pan**.	
206	After the gig, we mooch around the **band's dressing room**.	

No.	Sentence	Rating
0	When you are planning a schedule you should also **bear in mind** the location of speeches and statements.	

207	The food is ingested and the bacteria may **lie dormant** in the bird.	
208	And as you are well aware, I have had **precious little** from the police on that case.	
209	If they acted properly, I **did well by them**.	
210	The **teaching is excellent**.	
211	Twenty readers plus their partners had ringside seats at **the Daily Mirror sponsored leaf**.	
212	Are they with me on this or am I **boring the pants off them** yeah?	
213	She was **alone in the concrete**.	
214	Please **sweep in the fun**, buy the T-shirt.	
215	**The penny dropped** with a resounding clunk in her brain.	
216	Yes. I saw her on **the way to the** doctor's and she'd thought she'd got chicken pox.	
217	Eyes sparkling, Jill gazes at a relaxed John in the **garden of their hideaway wallet**.	
218	The moment she said it she knew she had **hit the nail on the head**.	
219	He needed someone to **throw him a lifeline** and I decided it might as well be me.	
220	Jane Pargeter sniffed and prodded **with her hair at an imaginary** stain in the carpet.	
221	He had earlier declined to **declare an interest in** the country's leadership.	
222	We were going round and round in a weird dance and **nearly read over in the pee**.	
223	I galloped into the hotel just as the most **torrential downpour** I have seen in my life cascaded from the sky.	
224	He had **been gargled** at Westminster Abbey Choir School.	
225	We seem to have a **prospective buyer** and Liz is making frantic signals to me.	
226	There **can be debate and compromise**.	
227	Not surprisingly, Ken was apprehensive **about the position I** might inherit when I touched down.	
228	It's inevitable that **heads will roll**.	
229	The "Playmore" has proved an **invaluable asset** to many clubs.	
230	Later in the same year **he was able to** compensate his brother in handsome style.	
231	They are certainly simple and **that is one of the** chief advantages of cash accounting.	
232	Don't let **him take** me.	
233	He might as well **make hay while the sun shone**, he told himself.	
234	Our first serious **engine secret** was in Amman.	
235	He alleged that a West Ham player assaulted him, **in front of his wife**.	

No.	Sentence	Rating
0	When you are planning a schedule you should also **bear in mind** the location of speeches and statements.	

236	Time has **passed you** by.	
237	"Very ingenious," **he said**.	
238	This happens when there is a voluntary tightening **of the vocal cords**.	
239	You may even **throw caution to the wind** and try one of our Mystery Trips.	
240	What **have we** here?	
241	You don't **lock that** shed?	
242	Journalists who refuse to **toe the line** will have to be sacked.	
243	It **can mean greenish** death.	
244	Much later they lay in one another's arms, at peace **amidst a frenzied information** of bedclothes.	
245	**Through hard work and sheer determination**, Dennis carved out a career in the building industry.	
246	Stimulation of the receptors in the inner ear **gives rise to** the experience of sound.	
247	The pictures we form of the **places eaten in our** study are no exception.	
248	Ackroyd knew immediately where he meant and his **blood ran cold**.	
249	And now she's trying to **bite the hand that fed** her," said a spokesman.	
250	Will **there be paper employed** to actually fill in?	
251	In a miscellaneous pile of documents, I **came across** the following essay.	
252	This will put **the Albanian economy back** in the black.	

Thank you!

Please return your questionnaire to the experimenter.