On-line Processing of Multi-word Sequences in a First and Second Language: Evidence from Eye-tracking and ERP

Anna Siyanova, BA, MA

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

February 2010

Abstract

A view that has been gaining popularity is that humans are sensitive to frequency information at different levels, and that this information affects the processing of linguistic material, subsequently shaping our mental representations. Frequency effects have been reported extensively in word processing literature, but only a small number of studies have investigated frequency effects in units larger than a word. The question that the present thesis strives to answer is: Do units above the word level, both fully compositional and less so, exhibit frequency effects? In Study 1, using an eye-tracking paradigm, I investigate the comprehension of idioms used figuratively (*at the end of the day* – 'eventually'), literally (*at the end of the day* – 'in the evening'), as well as novel phrases (*at the end of the war*) in a first and second language. In Study 2, which also uses eye-tracking, native and nonnative processing of frequent binomial expressions, such as *bride and groom*, is compared to their infrequent reversed forms, such as *groom and bride*. Finally, three ERP experiments, which form Study 3, further investigate on-line processing of frequent binomial expressions versus novel phrases in a first language. The results of the studies point to the following. Frequent phrases are processed faster than novel ones by native speakers. Nonnative speakers, on the other hand, appear to have a "lexicon in transition", that is, their processing starts to approximate that of natives only with respect to very high frequency items. Overall, the processing of frequent multi-word sequences in a second language is more sequential than that in a first language (this is particularly the case with idioms). The processing advantage for binomials observed in the ERP study with native speakers also suggests that different neural correlates underlie the processing of familiar phrases when compared to novel ones. On the

whole, the findings reported in the thesis suggest that the units that language users attend to are not limited to single words, but extend to multi-word sequences as well.

Acknowledgments

Working on this thesis has been an extremely rewarding experience. I am infinitely grateful to a number of people who have contributed to it in so many different ways:

First and foremost, I would like to express my enormous gratitude to my supervisors, Norbert Schmitt, Kathy Conklin, and Walter van Heuven, whose wisdom and expertise have guided me during these years. Many thanks to Norbert for inspiring me to study multi-word speech and taking me 'under his wing' when I was an MA student. The interest I take in multi-word sequences I owe to the great teacher that Norbert is. A very special thanks to Kathy for always having time for me, for her incredible patience and timely advice, and for so many conferences where her presence and support were particularly needed. I am also indebted to Kathy for introducing me to the beauty of eye-tracking. And, of course, the ERP part of this thesis would have never materialised, had it not been for Walter, whose knowledge and experience have given me an infinite source of ideas. Many aspects of my technical skills, a more rigorous approach to data collection and analysis, as well as my interest in neurolinguistics are thanks to Walter. Finally, I owe a great deal to all my supervisors for always supporting me in all my endeavours and for simply being great human beings.

I would like to express gratitude to my friends and colleagues at the University of Nottingham, Phil Durrant, Phoebe Lin, Li Jie, Irina Dahlmann, Ron Martinez, Ana Pellicer Sanchez, Alice Doherty, Myq Larson, Suhad Sonbul, and many more, for their fruitful discussions, intellectual contribution, and rigour.

Although I did most of my research at the University of Nottingham, I also had the privilege to work under the supervision of Edith Kaan at the University of Florida.

Not only were the three months that I spent in Edith's lab exceptionally productive, but they also helped me look at my research from a different perspective.

Many thanks to Jiye of SR Research, who I have never met, but who has been so kind and generous with his time in helping me become a proficient user of Experiment Builder and Data Viewer. Without Jiye's help, my data collection and analysis would have been a much more painful experience. I also feel indebted to the SR Research team for lending me their equipment for two years and for being so patient and understanding. Thanks to them, I had the luxury of working in my study rather than spending nights in a windowless lab.

This thesis simply would not have been possible without the Economic and Social Research Council (ESRC), who financed my PhD throughout, and without the so many paid and unpaid volunteers who kindly agreed to participate in my numerous experiments. I am also grateful to the University of Nottingham Graduate School for sponsoring me on a number of occasions.

I would like to thank my close friends, Jakub Marecek and Klara Mareckova. Without Jakub's enormous help with the stimuli development and all the programming that he has done for me, the number of sleepless nights would have been far greater. And, of course, so many thanks to both Jakub and Klara for not letting me forget that life is not all about work!

My most special thanks I have saved till the end. I owe so much to Ali, who has been putting up with the workaholic that I am, for always being there for me and for lending me a shoulder to cry on in my hour of need. And I do not think I can possibly express my gratitude to my mum and dad, Liuba Siyanova and Yury Chanturia, my sister, Olga Siyanova, and my grandparents for their unconditional love and for never doubting me.

Contents

Chapter 1: Introduction

1.1. Setting the scene

The suggestion here is that the wonderful feats of the human intellect, such as the use of language, are based at least as much on memorisation as on any impromptu problem-solving (in this case, the generation of novel utterances).

Joseph, D. Becker (1975, p. 62)

Human language is thought to be original and highly creative (e.g., Chomsky, 1957, 1965; Pinker, 1995). However, while we undoubtedly *can* exercise its creative potential, we do not necessarily do so. In fact, we rarely do so. It has been previously proposed that a large part of our mental and motor behaviour is highly automatised (e.g., MacNeilage, 1970; Shallice, 1988). In the present thesis, it is argued that this observation equally applies to human language.

Most of the language we produce in everyday situations does not require novelty and creativity. Invariably, we wish someone to have *a good day*, but not *a pleasant, fine, or enjoyable day.* We offer them *a cup of tea*, but never *a mug of tea*, even though what we may end up giving them is likely to be a mug rather than a cup. And of course, at least in Britain, people love talking about *heavy rains*, *strong winds*, and *mild temperatures*, rather than *strong rains*, *heavy winds*, and *gentle temperatures.* In other words, the situations we find ourselves in on a daily basis demand an extensive use of phrases that we have heard and used many times before. Thus, the approach adopted in the present thesis, in the words of Becker (1975), is that an understanding of the use of familiar phrases is necessary to the understanding of the use of language as a whole.

The traditional view of the mental lexicon has been that, with the exception of idioms, the lexicon consists entirely of single words and morphemes. In this account, a lexical entry is something that cannot be explained or predicted by a rule (e.g., a word, a morpheme, a past form of an irregular verb, or an idiom whose meaning is unrelated to the meanings of its components). All regularities, on the other hand, are believed to be encoded in a set of rules.

Until recently, the phenomenon of multi-word speech has largely been ignored, marginalised, and delegated to the 'linguistic periphery'. Multi-word sequences have been looked down upon as is evidenced by some of the less than flattering terms, such as "nonintellectual speech" (Espir & Rose, 1970), linguistic "dead-end" (Bates, Bretherton, & Snyder, 1988), and a "lazy solution to linguistic selection" (Drew & Holt, 1988). However, in the past two decades, the attitude towards multi-word speech has been changing. The creation of large corpora of 'real' linguistic material, both written and spoken, has made it possible to explore the language as it is used by native speakers. Crucially, language corpora, among other things, have allowed us to explore important aspects of language use, such as linguistic patterns, or frequent co-occurrences of words. As Firth (1957, p. 11) famously said "You shall know a word by the company it keeps". Recent explorations of language have shown that speakers make use of a large number of ready-made or prefabricated chunks, which can be broadly defined as a combination of two or more words that co-occur more often than would be expected by chance alone (e.g., Manning & Schutze, 1999).

The present thesis argues against the traditional view of the lexicon, which is believed to consist only of single words, morphemes and highly idiosyncratic phrases (i.e., idioms). Rather, I adopt a view, according to which regular compositional phrases (such as frequent collocations, binomial expressions, speech routines, and so on) may also be represented in the lexicon of a native speaker. Thus, the underlying principle behind this thesis is the assumption that the mental lexicon consists of a large number of multi-word units, where a unit "is a structure that a speaker has mastered quite thoroughly, to the extent that he can employ it in a largely automatic fashion, without having to focus his attention specifically on its individual parts or their arrangement" (Langacker, 1987, p. 57).

While addressing the issues of how frequent phrases are processed and whether they are represented in the mental lexicon of one's first language, I am also interested in exploring these issues in relation to one's second language. This is driven by the fact that multi-word speech has been shown to play an important role in the acquisition of a first (e.g., Clark, 1974; Lieven, Pine, Dresner & Barnes, 1992; Nelson, 1973; Peters 1983), as well as a second language (e.g., Wong-Fillmore, 1976, 1979; Vihman, 1982). As argued by Tomasello (2003), a major part of mature linguistic competence (in a first or second language) involves the mastery and appropriate use of various kinds of prefabricated speech, such as formulae, fixed expressions, idioms, frequent collocations, and so on. If this is the case, then to be a successful language user, one has to acquire and use a wide array of such expressions at a native-like level.

Thus, the focus in the thesis is on frequent phrases; specifically, on their on-line processing and representation in the lexicon of native, as well as proficient nonnative, speakers.

1.2. Aims of the thesis

The present thesis is an empirical investigation of on-line processing of frequent multi-word phrases. As such, it has several goals. The first aim is to investigate how native and proficient nonnative speakers process frequent phrases on-line. Specifically, it will be explored how such phrases are processed in L1 and L2 in comparison with less frequent, novel phrases. Thus, the two key variables to be manipulated and explored in the thesis are the frequency of occurrence of frequent phrases (i.e., phrasal frequency) and proficiency (i.e., L1 and L2). In two eye-tracking and three ERP experiments that form the core of the thesis, the processing of frequent and infrequent strings of language in and out of sentence (or story) context will be investigated. The focus of these empirical investigations is on two specific types of phrases: idioms (e.g., *ring a bell*) and binomial expressions (e.g., *bride and groom*).

The second point of investigation regards the L1 and L2 mental lexicon and the issue of representation of frequent expressions. There are reasons to believe that such phrases may be encoded in the lexicon, as vast amounts of knowledge can be stored in long-term memory, but only relatively small amounts can be processed in real time. In effect, the brain may make use of a relatively abundant resource (long-term memory) to compensate for a relative lack in another (working memory) by processing recurrent phrases in a more unitary way. This means there is less demand on cognitive capacity because such units are 'ready to go' and require less cognitive processing than equally plausible novel phrases. While this may be the case when processing language in a first language, a further question is whether processing in a second language works in the same or a comparable way.

Finally, I wanted to shed more light on the issue of the acquisition and use of multi-word speech by late (i.e., adult) L2 learners. It has been proposed that L2 learners have a tendency to over-rely on linguistic creativity and to construct a large proportion of their language compositionally rather than use prefabricated chunks (e.g., Foster, 2001; Skehan, 1998; Wray, 2002; Wray, 2008). The results of the eye-tracking experiments with idioms and binomial expressions are discussed in the light of these claims.

1.3. Structure of the thesis

The aim of Chapter 2 is to introduce the concept of multi-word speech, define it and discuss some of its key properties, such as frequency, familiarity, predictability, fixedness, as well as phonological properties.

This is followed by a detailed discussion of one particular type of multi-word speech – idioms. Chapter 3 thus provides a theoretical and empirical account of idioms and argues that they are special in a number of ways. It draws on the existing empirical evidence from a number of sources, such as idiom processing in a first and second language, developmental studies with young children, as well as studies with language-impaired patients.

In Chapter 4, I talk about the acquisition of multi-word speech in the first and second language. With regards to the latter, the acquisition of multi-word sequences in child and adult L2 learners is discussed. Although the studies presented in the thesis do not directly address multi-word sequence acquisition, the issues raised here become relevant in the discussion of native and nonnative speaker data in the chapters to follow.

Two methodologies are employed in the thesis: eye-tracking and ERP. The aim of Chapter 5 is to cover them in some depth. Specifically, this chapter reports on some of the previous research relevant to the studies from the methodological standpoint, as well as introduces the key concepts, measures and models that are brought up in the eye-tracking and ERP studies.

Chapter 6 presents an empirical investigation of idioms (Study 1). It focuses on idiom comprehension by native and proficient nonnative speakers of English. Using an eye-tracking methodology, I look at the processing of idiomatic expressions used figuratively (e.g., *at the end of the day* – 'eventually') and literally (e.g., *at the end of the day* – 'in the evening'), as well as control phrases (e.g., *at the end of the war*).

As mentioned above, two types of multi-word speech are explored in the present thesis. One of them is idioms. The other type is binomial expressions (e.g., *bride and groom*). In Chapter 7, the second empirical investigation is presented (Study 2), where I explore the processing of frequent binomials in a first and second language. In an eye-tracking experiment, native speaker processing of frequent binomials (e.g., *bride and groom*) is compared with the processing of their infrequent reversed forms (e.g., *groom and bride*).

Chapter 8 is a follow-up to the study presented in Chapter 7. Study 3 looks at on-line processing of binomial expressions in a first language using an electrophysiological technique – event-related brain potentials (ERPs). The issues that are addressed in this study are representation of frequent phrases in the mental lexicon of a native speaker.

Finally, in Chapter 9, I sum up the results of the three studies and draw implications from them. This chapter also identifies directions for future work and raises a number of limitations specific to the investigations presented in the thesis.

Chapter 2: Multi-word Speech

2.1. Terminology and definitions

The use of multi-word sequences, such as collocations (e.g., *plastic surgery*), idioms (e.g., *a piece of cake*), speech formulae (e.g., *What's up!*), and binomials (e.g., *men and women*) is regarded as an essential part of native-like communication (e.g., Cowie, 1998; Langacker, 1987; Pawley & Syder, 1983; Sinclair, 1991; Tomasello, 2003; Wray, 2002, 2008). Although figures vary as to how much formulaicity is present in our everyday discourse, it is clear that multi-word speech is truly ubiquitous. According to Pollio, Barlow, Fine, & Pollio (1977) and Glucksberg (1989), about four multi-word sequences are produced by a native speaker in every minute of spoken discourse. Biber, Johansson, Leech, Conrad, and Finegan (1999) found that multi-word speech constituted 28% of the spoken and 20% of the written discourse they analysed. Erman and Warren (2000) and Howarth (1998) estimated that multi-word speech of various types amounted to 52.3% and 40%, respectively, of the written discourse they looked at.

In recognition of this linguistic phenomenon, researchers such as Langacker (1987) proposed that the lexicon of a native speaker consists of lexicalised units, which are familiar structures that can be employed in an automatic fashion without accessing the individual components. In a similar vein, Nattinger (1980, p.341) argues that for most part, language production "consists of piecing together the ready-made units appropriate for a particular situation", while language comprehension "relies on knowing which of these patterns to predict in these situations".

In the present thesis, the term *multi-word speech* will be used as an umbrella term, while the term *multi-word sequence* will refer to individual instances of multi-word speech (a number of other terms have also been used, e.g., see Schmitt (2010) and Wray (2002)). There have been a number of attempts to define multi-word speech. For example, Wray (2002, p. 9) defines a multi-word sequence as:

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

Although this is not the only definition of multi-word speech, it is by far the most inclusive one. It covers:

- phrasal verbs (e.g., *put up with*)
- prepositional verbs (e.g., *depend on*)
- other multi-word verbs (e.g., *make up one's mind*)
- idioms, which are expressions whose meaning cannot be deduced purely on the basis of the literal meanings of its constituents (e.g., *ring a bell)*
- proverbs (e.g., *a friend in need is a friend indeed*)
- discourse markers (e.g., *as a matter of fact)*
- a wide range of collocations, which are word combinations whose individual components cannot always (if ever) be substituted by semantically equivalent words without the expression becoming

unnatural (e.g., *plastic surgery, conduct research, statistically significant*)

- binomials, which are three-word combinations, connected by a conjunction *and*, that exhibit a clear word order preference (e.g., *bride and groom*)
- a wide range of speech routines (e.g., *What's up?*)
- famous quotes/titles/names (e.g., *Poor Yorick!*)
- various grammatical constructions (e.g., *the -er, the –er; X is Y*)
- frequent compositional phrases (e.g., *I don't know, don't worry about that*)
- unanalysed and partially analysed chunks in child's early L1 (e.g., *lemme, wanna, gimme)*
- various types of automatic speech, such as prayers, counting, singing, swearing, exclamations, etc.

From the above examples, it is clear that the phenomenon of multi-word speech is extremely broad. These strings of language differ vastly in their syntactic structure, semantic transparency, frequency of occurrence, register, function, and so on. In fact, it may even seem that they have nothing in common. However, all of these examples, and multi-word speech in general, possess a number of key features that unite them. They will be discussed next.

2.2. Properties of multi-word speech

2.2.1. Frequency

According to Ellis (2002), language processing is tuned to input frequency, because language users are highly sensitive to the frequencies of linguistic events in their experience. Indeed, frequency effects are one of the most robust in psycholinguistic research. As proposed by Bod, Hay, and Jannedy (2003), frequency effects are everywhere. Some researchers have even suggested that frequency may be the main factor responsible for the organisation of the lexicon (Forster, 1976). Frequency effects have been widely shown in the word processing literature (e.g., Balota, 1994; Jescheniak & Levelt, 1994; Rayner & Duffy, 1986), but less so in the case of units larger than a word (e.g., Bell et al., 2003) and syntactic processing (e.g., Bod, 2000, 2001; Frazier & Fodor, 1978; Reali & Christiansen, 2007).

Although Wray (2002) regards frequency to be one of the most salient and determining characteristics of multi-word speech, it has, nevertheless, received surprisingly little attention in psycholinguistic research and still remains, in the words of Jurafsky (2003), an important unsolved problem. As Jurafsky (2003) points out, the frequency of complex constructions (of any length or internal structure) is much lower than that of single words, and hence frequency effects in such larger units are harder to observe and to investigate. Although the evidence is rather scarce, it has nevertheless been proposed that the frequency with which multi-word units occur in language (as attested in written corpora) affects their on-line processing.

The processing benefit for multi-word sequences over novel and hence less frequent strings of language has been shown to manifest itself in a number of different ways, such as faster reading times in comprehension studies, and phonological reduction and faster articulation of frequent forms in production studies. With respect to comprehension, it has been shown that people are sensitive to the frequency of compositional phrases (i.e., those whose individual components contribute overtly to the meaning of the phrase), such as *don't have to worry* (e.g., Arnon & Snider, 2010; Bannard & Matthews, 2008; Bod, 2000, 2001) and frequently occurring collocations, such as *sort of* (Sosa & MacFarlane, 2002). These studies report that more frequent phrases are processed reliably faster than less frequent ones. A number of production studies have also reported that words within frequent multi-word expressions are more likely to be phonetically reduced (e.g., Bell et al., 2003, 2009; Bybee, 2000, 2002; Bybee & Scheibman, 1999; Gregory et al., 1999; Jurafsky et al., 2001).

Frequency within theories of language acquisition

Thus far, it has been proposed that multi-word sequences are processed more quickly than novel strings of language. If frequency affects the speed with which language is processed, then this information must be represented somewhere. In other words, the language processor must record frequencies of various linguistic events (Bod, Hay, & Jannedy, 2003). Thus, throughout their lifespan, language users must notice, accumulate, and use frequency information not just with regards to single words, but also phrases. If frequency effects are so pervasive in language, then they would logically also play an important role in language acquisition.

The traditional approach to language acquisition and use

With regards to frequency and language acquisition, there are two main views. On the one hand, according to a more traditional approach to language acquisition and use, frequency should not play a major role in language acquisition and use. As

Chomsky (1957) proposed, the statistical distributions of language are not central to the linguistic knowledge. According to this tradition, knowing a language means knowing a set of grammar rules, which can be used to produce (and understand) an infinite number of novel sentences. In the words-and-rules approach (e.g., Pinker, 1991, 1999; Pinker & Ullman, 2002), it is further argued that there is a clear distinction between the lexicon (a collection of memorised forms), and the grammar (a collection of rules that are applied to these memorised forms). Importantly, the processes involved in the formation of the mental lexicon and the mental grammar are rather different, because they are modulated by different cognitive abilities (e.g., Ullman, 2001; Ullman et al., 2005). This is evident with respect to the role of frequency attributed to linguistic forms. According to Pinker and Ullman (2002), it is possible to use frequency manipulations to distinguish between 'stored' and 'computed' lexical forms. They propose that frequency of occurrence should affect on-line processing of stored forms (such as words) but not computed ones (such as phrases). Thus, according to the words-and-rules approach, frequency should not play a role in the processing of compositional phrases, no matter the frequency of the phrase. In this respect, no such concept as 'phrasal frequency' should exist and no unit above the word level, irrespective of its frequency, can be represented, or encoded, in the mental lexicon (with the exception of highly idiosyncratic idioms, such as *kick the bucket*). However, because a number of studies have reported phrasal frequency effects (e.g., Arnon & Snider, 2010; Bannard & Matthews, 2008; Sosa & MacFarlane, 2002), it seems logical to propose that frequency effects do in fact play a role in language processing. An approach that supports this view is presented below.

Usage-based approach to language acquisition and use

In sharp contrast to the words-and-rules approach, a number of 'empiricist' views, such as usage-based (e.g., Bybee, 1998; Goldberg, 1995, 2006; Tomasello, 2003) and exemplar theories (e.g., Abbot-Smith & Tomasello, 2006; Bod, 1998; 2006; Pierrehumbert, 2001) propose that the basic unit of language acquisition is a *construction*. Borensztajn et al. (2009, p. 175) define a construction as "associations between a semantic frame and a syntactic pattern, for which the meaning or form is not strictly predictable from its component parts". Ellis (2002, p.167) further adds that a construction is "a conventional linguistic unit – that is, part of the linguistic system, accepted as a convention in the speech community". According to Ellis, constructions can be structurally complex (e.g., [Det Noun]), or simple (e.g., [Noun]); constructions may also represent complex structures above the word level (e.g., [Adj Noun]). Crucially, Ellis argues, constructions are independently represented in the speaker's mind, with some exhibiting certain unique, idiosyncratic properties, and others not. That is, absence of a unique property does not suggest that it cannot be represented independently. Ultimately, it is the frequency of occurrence that leads to independent representation of constructional patterns, both regular and highly idiosyncratic.

On this account, the task of a language learner is to gradually acquire a set of constructions that vary in size, complexity, and the level of abstractness (e.g., Goldberg, 2006; Tomasello, 2003). With respect to first language acquisition, it has long been noted that children first learn and make use of memorised multi-word constructions, or holophrases, before they can break them down into single words and apply linguistic rules on them (e.g., Lieven, et al., 2003; Peters, 1983; Tomasello, 2003). Thus, in line with usage-based and exemplar theories of language acquisition and use, first language learners start off not with single words, but with simple and concrete constructions gradually moving towards more complex and abstract ones (e.g., Borensztajn et al., 2009). Such theories further propose that language processing (both production and comprehension) operates with concrete linguistic experiences rather than a set of abstract linguistic rules. According to this view, the acquisition of grammar is "the piecemeal learning of many thousands of constructions and the frequency-biased abstraction of regularities within them" (Ellis, 2002, p. 144). Ellis further argues that language learning (both first and second) is "the associative learning of representations" with frequency being a key determinant of acquisition because the so-called rules of language, are in fact, regularities "that emerge from learners' lifetime analysis of the distributional characteristics of the language input" (p. 144).

Crucially, in line with the usage-based approach, *all* linguistic information (words as well as phrases) is represented and processed in a comparable way, and thus, it should be similarly affected by the frequency of occurrence. Any differences between less and more frequent units (of any length) will thus be informative of the way they are learnt and represented. This approach predicts faster reading times for frequent words, as well as frequent compositional phrases over less frequent ones, which the more traditional words-and-rules approach does not.

The usage-based approach to language acquisition thus highlights a special role of language use. It proposes that mental representations are determined purely by language use (e.g., Abbot-Smith & Tomasello, 2006; Bod, 2006; Bybee, 1985, 1995, 2006; Croft, 2001; Goldberg, 1995, 2006; Langacker, 1987; Tomasello, 2003, 2006). Every time a word or a phrase is used, it activates nodes in the lexicon, and subsequently, the frequency of activation of this unit affects its representation in the mental lexicon (e.g., Croft & Cruse, 2004). It is believed that new experiences with a word or a phrase are not decoded and then discarded; rather, they determine memory representations (Bybee, 2006). Thus, every linguistic token that is encountered by a language user is registered in memory, which leads to language processing operating with a vast set of exemplars (e.g., Bybee, 2006; Goldberg, 2006). Such exemplars could be words, phrases, grammatical contractions, and so on. In fact, evidence suggests that language users store all linguistic tokens they come across (e.g., Jurafsky, 2003; Tomasello, 2003). As Bod (2006) notes, without this seemingly massive storage of various exemplars, frequencies can never accumulate and, thus, conventional ways of speaking cannot be learnt. Importantly, in line with these models, there are no restrictions as to what can or cannot be represented in the lexicon, a single word, or a unit above the word level, such as a phrase.

Akin to the usage-based theory, exemplar-based models also propose that language processing (production and comprehension) is based on concrete linguistic experiences with language, rather than abstract linguistic rules (e.g., Bod, 2006). In his exemplar-based syntactic model, known as Data-Oriented Parsing, Bod (2006) proposes that the assignment of representations to linguistic events is done purely on the basis of statistics (in language acquisition and processing). In this account, the only rules are those that construct new representations out of already existing ones. Thus, language should be viewed not as a set of grammar rules, but as a statistical accumulation of experiences that changes every time a particular utterance is encountered.

Of course, as Ellis (2002) notes, frequency should not be viewed as the only explanation. Nonetheless, the role attributed to frequency in the usage- and exemplar-based theories does suggest that frequency may well be the most important factor that determines how language is learnt, processed, and used. To conclude, according to the 'empiricist' approach to language acquisition and use, there appears to be no clear distinction between the frequency of seemingly compositional phrases and that of single words. It is proposed that recurrent patterns, words or frequent phrases, can be represented in the lexicon in a similar way. As a result, phrasal frequency should modulate multi-word unit processing in the same way as lexical frequency affects reading times of single words.

2.2.2. Familiarity

It seems sensible to propose that frequency should correlate with familiarity. The more frequent a phrase is, the more likely it is to become familiar to a language user, and subsequently become encoded in their mental lexicon. However, familiarity does not equal frequency. There are multi-word sequences, which are undoubtedly formulaic and familiar to the linguistic community but are nevertheless extremely infrequent, for example *by kith and kin* and *raining cats and dogs* (0 and 3 occurrences in the British National Corpus (BNC), respectively). For comparison, the idioms *ring a bell* and *a piece of cake* are attested 75 and 70 times, respectively. Something very infrequent, like certain idioms, may still be thought of as highly conventional and formulaic because it is highly idiosyncratic. Thus, familiarity and frequency complement rather than duplicate each other.

Familiarity has been widely researched in developmental studies in children's acquisition and comprehension of idioms. Some researchers attribute a minor role to familiarity in children's comprehension of idiomatic expressions (e.g., Levorato & Cacciari, 1992). Others propose that familiarity is an important factor in idiom processing in young, as well as older children (e.g., Nippold & Rudzinski, 1993; Nippold & Taylor, 1995; Nippold, Taylor, & Baker, 1996). However, these researchers seem to suggest that familiarity is in fact frequency of occurrence. It may be the case with children, since they may have not yet had enough experience with infrequent idioms. However, when it comes to adults, most native speakers would agree that *raining cats and dogs* and *kick the bucket* are highly familiar expressions well known to the linguistic community, even though these idioms are rather infrequent as corpus evidence suggests (3 and 12 occurrences in the BNC, respectively).

2.2.3. Predictability

One of the key features of multi-word speech is that it is highly predictable. That is, upon hearing or reading the beginning of an idiom (e.g., *take the bull* ...), a multi-word verb (e.g., *put up* ...), a frequent collocation (e.g., *extenuating* ....), or a binomial (e.g., *fish and* ...), the hearer or the reader are very likely to finish it with the most likely completion (i.e., *by its horns, with, circumstances,* and *chips,* respectively).

According to probabilistic language models, statistical information about the co-occurrence of words is represented in the speaker's mind (e.g., Gregory et al., 1999; Jurafsky, 1996; McDonald & Shillcock, 2003a; McDonald & Shillcock, 2003b; Seidenberg & MacDonald, 1999). McDonald and Shillcock (2003a, 2003b) contend that the large amounts of language that a native speaker encounters on a daily basis represent a rich source of statistical knowledge about this language. Thus, the brain is capable of using this statistical information during language comprehension in order to estimate the probability of *Word 2* following *Word 1*. Importantly, it is pointed out that integrating a word into one's mental lexicon also involves encoding its surrounding context into the mental lexicon (e.g., McDonald & Shillcock, 2003b).

It is widely assumed that a word's predictability within a given context (sentential or phrasal) impacts the ease with which it is comprehended on-line (e.g., Balota, Pollatsek, & Rayner, 1985; Engbert, et al., 2005; Gregory, et al., 1999; Kutas & Hillyard, 1984; Levy, 2008; McDonald & Shillcock, 2003a, 2003b; Rayner & Well, 1996; Reichle, et al., 1998). This is because, upon seeing Word 1 comprehenders predict Word 2, thereby making it easier to recognise and/or integrate into their understanding of the sentence. The ease in processing and/or integration of highly predictable words has been accounted for in models of reading. For example, according to Reichle et al. (1998) and Engbert et al. (2005), there is a close link between eye-movement control and high-level cognitive processes. That is, eye-movement patterns are highly dependent on such properties of a word as frequency and word predictability because they represent readers' knowledge of and experience with language.

As is clear from the above, the role of predictability in language processing has been documented in L1 literature. How it is engaged in comprehension of a second language is less clear. However, it seems plausible to suggest that if rich exposure to language is necessary in order for statistical information to become encoded, then even highly proficient second language speakers are unlikely to perform at a native-like level due to their limited exposure. This issue will be addressed in the present thesis.

2.2.4. Fixedness

One of the features of multi-word speech is its relative fixedness. While novel propositional speech is characterised by full syntactic flexibility, most instances of multi-word speech are fixed or semi-fixed utterances. Although most idioms and other familiar expressions allow some variation, such as insertions, modifications, and passivisation, these changes, even when allowed, are limited and cannot be compared to the wide range of syntactic changes permitted in novel language production. However, some conventional phrases are so rigid that even minor changes are not allowed without the phrase losing its original meaning or without it sounding ungrammatical. For example, *by kith and kin, by and large, a piece of cake* do not allow any changes.

A number of researchers have proposed that idioms undergo syntactic analysis similar to more propositional speech (e.g., Cutting & Bock, 1997; Konopka & Bock, 2009; Peterson, Burgess, Dell, & Eberhard, 2001). Although it is undoubtedly true that some idioms are, at least to a certain degree, compositional, it is still possible to argue that idioms in general are rather fixed, when compared to novel strings of language. Namely, the syntactic flexibility that characterises propositional language by far exceeds that of idioms. Novel language can easily be subjected to numerous syntactic processes, such as passivisation and word order modifications. The changes permitted in the case of idioms are much more limited.

2.2.5. Phonology

An interesting area in multi-word speech research is its phonology, which is believed to differ from that of novel propositional speech. Fowler (1988) proposed that speakers can "get away with producing reduced versions of words in situations in which listeners have other sources of information about the words' probable identity" (p. 308). To test this assumption, researchers have looked at the role of such factors as reduction, stress, pauses, articulation and speaker fluency in the production of multi-word speech. Overall, the pronunciation of words in frequent word pairs and idioms have been found to be shorter in duration and phonologically reduced (e.g., Bell et al., 2003, 2009; Bybee, 2000, 2002; Bybee & Scheibman, 1999; Gregory et al., 1999; Jurafsky et al., 2001).

Van Lancker, Canter, and Terbeek (1981) analysed the phonological properties of frequent multi-word sequences and novel language. Instances of novel language were found to be longer in duration because they contained more and longer pauses, and, importantly, because lexical items within these expressions were spoken more slowly. They also contained a greater number of pitch changes. Van Lancker, Canter, and Terbeek report on a number of examples of a less precise articulation in the pronunciation of idioms, compared with novel language; for example, shorter initial consonants, shorter more neutral vowels in unstressed words, and diphthongs reduced to monophthongs.

In another production study, Bybee and Scheibman (1999) found that the word *don't* was phonetically reduced when it was part of a frequent (but fully compositional) phrase, such as *I don't know*. Bybee (2000) found that in very frequent word pairs, the word boundary between the two words behaved like word-internal segments. Finally, Bell et al. (2003) found that words were phonetically reduced when they were more predictable given both the previous and the following word (e.g., in the trigram *middle of the*).

2.3. Conclusion

The above serves as a brief overview of what multi-word speech is. Although this phenomenon is extremely broad and diverse, it is clear that idioms, collocations, binomials, phrasal and propositional verbs, speech formulae, and other types of multi-word speech all have a number of common properties. They key ones include frequency and familiarity, (semi-)fixedness, and predictability. I will now look in detail at one particular type of multi-word speech that will figure prominently in the present thesis – idioms. Of all the multi-word sequences, idioms have by far received the greatest amount of attention in psycholinguistic research, which has resulted in a wealth of studies. The following chapter will review the literature on idioms from a number of different perspectives: idiom processing in a first and second language, developmental studies on idiom processing in children, and idiom comprehension and production in language-impaired patients.

Chapter 3: Idioms

3.1. Introduction

Up to now, I have talked about multi-word sequences and their properties in general, without focusing on any particular type. However, as mentioned earlier, the core of the present thesis is an empirical investigation of two specific types of multi-word speech: idioms and binomial expressions. In this chapter, the focus will be on the former. In what follows below, idioms are discussed from a number of different perspectives.

Classically, idioms are broadly defined as phrases whose figurative meaning is distinct from their constituents (e.g., Cacciari & Glucksberg, 1991; Gibbs, Nayak, & Cutting, 1989; Titone & Connine, 1999). Idioms have long been a point of investigation for researchers from various disciplines. Linguists, applied linguists, psycholinguists, and speech pathologists have all studied idioms from a number of different perspectives: idiom description and theoretical frameworks (e.g., Fillmore et al., 1988; Langacker, 1987; Nunberg, Sag, & Wasow, 1994), idiom on-line processing in L1 (e.g., Bobrow & Bell, 1973; Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Gibbs, 1980; Konopka & Bock, 2009; Peterson, Burgess, Dell, & Eberhard, 2001; Swinney & Cutler, 1979) and L2 (e.g., Cieslicka, 2006; Conklin & Schmitt, 2008; Underwood, Schmitt, & Galpin, 2004; Van Lancker-Sidtis, 2003), idiom comprehension in children (e.g., Abkarian, Jones, & West, 1992; Cacciari & Levorato, 1989; Nippold & Martin, 1989; Nippold & Rudzinsky, 1993; Prinz, 1983), as well as idiom production and comprehension in speech-impaired patients (e.g., Mondini et al., 2002; Paul et al., 2003; Van Lancker & Kempler, 1987; Van Lancker Sidtis, Postman, & Glosser, 2004). The aim of the present chapter is to offer some insight into the way idioms are processed (produced and comprehended) by different populations. However, to offer a more comprehensive overview, a theoretically motivated approach to idioms will first be presented. Notably, three linguistic traditions, the universal grammar (UG), the construction grammar (CG) and the usage-based model (UBM), are discussed below due to their influential albeit rather contrasting views on treating idioms.

3.2. Idioms within the Universal Grammar tradition

A traditional view of idioms has been that expressions of the type *kick the bucket* and *by and large* are non-decomposable expressions, or dead metaphors, whose meaning cannot be inferred from the idiom's constituent parts (e.g., Chomsky, 1965; Fraser, 1970; Heringer, 1976; Katz, 1973). From the perspective of formal theories of grammar, idioms are problematic due to their non-generative and often non-compositional and fixed nature. Such expressions, being idiosyncratic in one way or another, defy the rules of syntax without being ungrammatical. Thus, idioms belong to a large group of expressions that appear to pose a problem for the compositional model of grammar, such as UG.

A number of linguists within the universal grammar tradition have attempted to provide an account of idiomatic expressions that exist in language (e.g., Katz, 1973; Katz & Postal, 1963; Fraser, 1970; Weinreich, 1969). In line with the UG approach, idioms are considered to exist outside of the rule-based linguistic system and thus belong to the linguistic periphery (rather than the core to which the grammar is said to belong to). In line with this tradition, idioms represent a limited set of learnable expressions. UG linguists posit that everything must be explained by and fall within the compositional approach to linguistic form. In other words, the individual components of any linguistic structure must contribute directly and unambiguously to the meaning of this expression and account for any grammatical relations that exist among them (e.g., *John ate a piece of cake* is nothing but John having consumed a slice of pastry). On this account, a language user should be able to produce an infinite number of rule-based grammatically correct sentences. Idioms, however, defy the rule-based approach to grammar in a number of ways. First and foremost, their individual components do not always contribute directly and fully to the overall meaning of the expression (e.g., *John kicked the bucket* has nothing to do with John hitting a pail with his foot). Further, among other things, idioms do not always permit syntactical transformations in the same way as novel language (e.g., *John ate a piece of cake* → *A piece of cake was eaten by John*, but *John kicked the bucket* → **The bucket was kicked by John*).

Because the UG approach cannot satisfactorily account for the presence of such idiosyncrasies in the language, idioms have been somewhat marginalised and delegated to the language periphery. According to UG, what a linguist should strive to study, describe and elucidate is linguistic *competence*, that is, the abstract "ideal" system of linguistic form based on a neat set of grammar rules ('the core'), rather than linguistic *performance*, that is, the "real" language, as it is acquired and produced by native speakers, abounding in irregularities and idiosyncratic features ('the periphery'). The latter has been mostly ignored by the UG grammarians due to many instances of linguistic performance not being able to be explained by the general rules and innate principles (idioms are a prime example of this).

3.3. Idioms within the Cognitive Grammar and Usage-based Model approach

A very different approach to idioms has been advocated by the CG approach (e.g., Croft & Cruse, 2004; Goldberg, 1995, 2006; Fillmore et al., 1988; Langacker, 1987; Nunberg, Sag, & Wasow, 1994) and the linguists within the usage-based tradition (e.g., Bybee 1985, 1995; Tomasello, 2003, 2006). The CG and UBM researchers have challenged the sheer notion of the universal grammar, the grammar-lexicon dichotomy, the marginalisation of a wide range of idiosyncratic expressions (such as idioms), and, crucially, the disregard of the role of language usage.

CG is an umbrella term, which covers a number of theories based on the assumption that the unit of grammar is a grammatical construction (a combination of form and meaning). CG was developed to accommodate a wide array of constructions and idiosyncracies that UG could not account for. CG postulates that the overall meaning of a construction or a sentence does not necessarily have to equal the meaning of its parts. According to CG, a language as such is a repository of constructions of various sizes and degree of abstractness. In their degree of abstractedness, constructions may vary from highly fixed expressions (e.g., idioms) to highly abstract ones (e.g., *X is Y*). Unlike UG, CG argues against the syntax-lexicon dichotomy. Crucially, it is further proposed that each construction is represented in the mental lexicon of a native speaker and that knowing a language presupposes knowing a finite (rather than infinite) set of grammatical constructions (e.g., *X is Y*; *the –er, the –er*; SVO; etc.).

Many cognitive linguists support a usage-based model (UBM) of language acquisition and language use (e.g., Bybee 1985, 1995; Tomasello, 2003, 2006). According to UBM, the primary factor, which determines the acquisition and use of word forms is the frequency with which these forms occur in language. That is, each time a word or a construction is used, it activates "a pattern of nodes in the mind", and thus the frequency of this activation affects the level of representation of that information (Croft & Cruse 2004, p. 292). In this view, the lexicon of a native speaker consists not only of single words, but also of thousands of conventional expressions consisting of more than one word, such as idioms, collocations, binomials, speech routines, formulae, and clichés (e.g., Langacker, 1987; Tomasello, 2003).

3.4. Idiom characteristics and classifications

Traditionally, idioms have been viewed as complex expressions whose meaning cannot be derived from the meanings of their individual parts. Among others, the main characteristics of idioms are considered to be their non-compositionality, grammatical deficiency (i.e., syntactic fixedness), and lack of substitutability (i.e., synonymous lexical items cannot be used with the same meaning) (Brinton & Traugott, 2005). However, a number of researchers have questioned these properties of idioms. In particular, Langacker (1987) questioned two major assumptions, (a) the assumption of unanalysability of idioms, and (b) the assumption of idiom fixedness. According to Langacker, the majority of idioms are analysable to at least some degree (e.g., *play with fire*), while some idioms are fully analysable. He further argues that to regard idioms as necessarily opaque, or as fully fixed phrases, is rather simplistic. For example, in the idiom *let the cat out of the bag*, *cat* is attributed a meaning roughly equivalent to 'information', *out of* is used in its usual sense, while *bag* conveys some notion similar to 'concealment'.

The problem of decomposability has also been raised by Fillmore et al. (1988) and Nurnberg, Sag, and Wasow (1994). Fillmore et al. (1988) put forward four idiom classifications. First, they categorise idioms as *encoding* and *decoding*. The meaning of a decoding idiom cannot be inferred by a language user unless they have learnt its meaning (e.g., *pull a fast one, everything but the kitchen sink*). The meaning of an encoding idiom, on the other hand, can be deduced by the language user even if they are unfamiliar with its meaning; however, they may not necessarily be able to produce and use this expression unless they have come across it before. Thus, similar to Langacker (1987), they acknowledge that idioms are, in fact, compositional in a number of ways. Second, Fillmore et al. categorise idioms as being *grammatical* or *extragrammatical*. The former ones are expressions formed in accordance with regular syntactic rules (e.g., *be left in the dark, spill the beans*). The latter ones are word combinations that are both syntactically and semantically idiosyncratic (e.g., *by kith and kin, by and large*). Further, they distinguish between idiomatic expressions *with* and *without pragmatic point*. Those with pragmatic point are normally related to certain pragmatic routines (e.g., *good evening, nice to see you*). Those without pragmatic point have a much wider usage (e.g., *at the end of the day, on the other hand*). Forth, Fillmore et al. (1988) classify idioms as *substantive* – "lexically filled", and *formal* – "lexically open". Substantive idioms are those which carry some concrete lexical meaning, like *spill the beans, kick the bucket, answer the door*, etc. Formal idioms are more abstract and appear to be 'shells', which need to be filled in with more concrete lexical items (e.g., the comparative construction *the –er, the –er*).

In another idiom classification system, Nunberg, Sag and Wasow (1994) distinguish between *idiomatic phrases* and *idiomatically combining expressions.* The meaning of the idiomatic phrase is not related to its overall figurative meaning (e.g., *saw logs* - snore), whereas the meaning of the idiomatically combining expression can be deduced from its parts (e.g., *spill the beans* – *spill* → reveal, *beans* → information). This classification is similar to Fillmore et al.'s distinction between encoding and decoding idioms. Idiomatically combining expressions, unlike idiomatic phrases, allow a number of syntactic modifications, such as adjectival modification (e.g., *leave no <u>legal</u> stone unturned*), quantification (e.g., *touch <u>a couple of</u> nerves*), negation (e.g., *spill <u>no</u> beans*), passivisation (e.g., *the law was laid*), and pluralisation (e.g., *drop hints*).

In sum, the sheer number of different types of idioms, as well as other fixed or semi-fixed familiar expressions that exist in language suggest that such phrases are not an insignificant peripheral part of our linguistic knowledge. Rather idioms are part of a rich and diverse family of expressions, which abound in language.

3.4.1. Idiom decomposability

In the previous section, idioms were discussed from a theoretical standpoint. Here, I turn to an issue that a number of linguists have raised, that of idiom decomposability. Decomposability is the extent to which the components of an idiom contribute to the overall figurative meaning. Langacker (1987), Fillmore et al. (1988), and Nunberg, Sag, and Wasow (1994) have proposed that idioms do not form a neat class of expressions that all fall under the definition of unanalysable and non-decomposable. Rather empirical evidence supports the classification of idiomatic expressions into two broad categories: decomposable and non-decomposable. In the section below, the research in this area is examined.

A number of psycholinguistic studies have dealt with the processing of decomposable versus non-decomposable idioms. In their *decomposition hypothesis*, Gibbs and colleagues propose that idioms differ in their degree of semantic decomposability and distinguish between *decomposable* and *non-decomposable* idioms (e.g., Gibbs & Nayak, 1989; Gibbs, Nayak & Cutting, 1989). Gibbs and colleagues maintain that when studying the issue of idiom on-line processing, one has to account for the degree of decomposability because the two types of idioms differ in the way they are processed on-line. They define a *decomposable* idiom as an expression whose constituent parts contribute directly to the idiomatic interpretation of the phrase (e.g., the idiom *be left in the dark* implies that one does not know or cannot see what is going on and therefore feels lost, literally or figuratively). In a *non-decomposable* idiom constituent parts do not make such a contribution (e.g., meanings of the individual components of the idiom *kick the bucket* are not in any way related to the figurative interpretation of the idiom). Gibbs and Nayak argue that many idioms are, in fact, analysable and that their constituent parts contribute to the utterance's figurative interpretation. Further, they suggest that the more decomposable an idiom is, the more likely it is to be syntactically productive, that is, subjected to modifications without any changes in its figurativeness (e.g., *John laid down the law* → *The law was laid down by John*). Non-decomposable idioms, on other hand, are more frozen in their syntactic behaviour (e.g., *John kicked the bucket* → **The bucket was kicked by John*). Thus, the idioms whose figurative meanings are closely related to the literal meanings of their constituents are more syntactically flexible than idioms whose literal meanings are unrelated. They conclude that the analysability of an idiom is the best predictor of its syntactic productivity.

In another study, Gibbs, Nayak and Cutting (1989) proposed a theory of idiom comprehension: initially, people process all idioms in a compositional manner, similar to comprehension of literal language. They hypothesised that readers' analysis of decomposable idioms (e.g., *pop the question, lay down the law*) will slow down their processing when compared to non-decomposable idioms (e.g., *kick the bucket*) that are believed to be processed in a more unitary way due to their non-decomposable nature. They proposed that since non-decomposable idioms are found in a smaller number of syntactic constructions (compared to decomposable phrases), they can be viewed as lexicalised units, and should therefore be accessed and understood faster than decomposable idioms. Interestingly, contrary to their expectations, it was found that decomposable idioms (e.g., *pop the question*) were read significantly faster than controls (e.g., *ask the question*). Non-decomposable idioms (e.g., *kick the bucket*), on the other hand, were read reliably slower than their control phrases (e.g., *fill the bucket*). On the basis of these findings, the authors concluded that when presented with a non-decomposable expression, people attempt to perform some compositional analysis. Comprehension of non-decomposable idioms, it is claimed, is more difficult precisely because the overall figurative meanings of these phrases cannot be determined through the analysis of their components.

Finally, Titone and Connine (1999) proposed a *hybrid model*. The authors suggested that the difference between the literal meanings of idiom constituents and the overall figurative meaning of a non-decomposable idiom should slow down idiom comprehension. This processing cost, however, should not happen for decomposable idioms whose figurative and literal meanings are related. They conducted a study in which a set of idioms of each type was embedded in a sentence context, which biased towards either the idiomatic or literal meaning; the biased context either preceded or followed the idiom. Their results showed that the reading speed for non-decomposable idioms was significantly slower when context preceded the idiom than when it followed it. Reading rates for decomposable idioms, on the other hand, exhibited no difference, whether the context preceded or followed the idiom. On the basis of these results, the authors suggested that due to their semantically distinct idiomatic and literal meanings, non-decomposable idioms are more difficult to process than decomposable idioms. In other words, the reader needs more time to integrate a contextually appropriate interpretation of the non-decomposable idiom. The authors conclude that in idioms, whose constituents contribute to both idiomatic and literal interpretations of the phrase (i.e., decomposable idioms), the selection between the phrase's idiomatic and literal meanings happens significantly faster than in idioms, whose constituents contribute only to the literal interpretation of the sentence (i.e., non-decomposable idioms).

The above empirical investigations have shown that the division of idioms into decomposable and non-decomposable holds true not only from the theoretical linguistic standpoint (e.g., Fillmore et al., 1988; Langacker, 1987; Nunberg, Sag & Wasow, 1994), but is also psychologically valid and can be empirically demonstrated (however, see Tabossi, Fanari, & Wolf (2009)). It thus appears misleading to speak of idioms as one homogenous class of unanalysable expressions, whose individual components make no contribution to the meaning of the idiom.

3.5. Idiom processing

The major part of the present chapter is dedicated to the processing (production and comprehension) of idioms. In the studies discussed below, I aim to show that idioms are processed differently from novel language. At the very least, the evidence suggests that idioms enjoy faster processing, with some researchers further proposing that idioms differ from novel literal language in terms of their hemispheric representation. The review below will cover the issue of idiom processing by a range of different populations, such as native speakers and second language learners, children and adolescents, as well as language-impaired patients.

3.5.1. Idiom processing in native speakers

Much of the research on idiom comprehension has been done with adult native speakers of a language. Specifically, this research has addressed the following two issues: (1) the activation of an idiom's figurative versus literal meanings for idioms that have two distinct interpretations, and (2) the processing of idiomatic expressions versus non-idiomatic novel phrases. Some researchers suggest that the activation of the figurative meaning happens in parallel with activation of its literal counterpart (e.g., Swinney & Cutler, 1979). Others propose that the figurative meaning is the first one to get activated and, hence, it enjoys a processing advantage (e.g., Gibbs, 1980) or, vice versa, that it is the literal meaning that is accessed first (e.g., Bobrow & Bell, 1973). One of the first theories of idiom comprehension, the *idiom list hypothesis*, holds that idioms are stored in a special idiom domain which is not part of the normal lexicon (e.g., Bobrow & Bell, 1973). According to this hypothesis, a literal interpretation of an idiom is always available before a figurative one. In their *lexical* *representation hypothesis*, Swinney and Cutler (1979) put forward the idea that literal and figurative meanings are activated in parallel. Namely, they argue that the computation of the literal meaning and access of the figurative one should happen simultaneously as soon as the first word of the expression is encountered. Thus, it is proposed that the individual words of the idiom are accessed in the mental lexicon and structural analysis is carried out on these words (literal interpretation), while at the same time the entire expression is accessed (figurative interpretation). In his *direct access hypothesis*, Gibbs (1980) claims that activation of the figurative meaning precedes the activation of the literal meaning. Thus, the literal meaning activation happens if, or when, the figurative sense is rejected as defective in the given context.

Another, more recent, theory of idiom processing was proposed by Cacciari and Tabossi (1988). According to their *configuration hypothesis*, activation of the figurative meaning occurs after a sufficient portion of the idiomatic expression has been read. They put forward the idea of the "idiomatic key", which refers to the place where the expression becomes recognisable as idiomatic. According to the configuration hypothesis, the individual words and their literal meanings are activated up until the point when the "key" has been reached. Once the idiomatic key is reached, the idiomatic configuration emerges and the figurative meaning is accessed, while the literal meaning is rejected as no longer viable. A crucial aspect of this hypothesis is the identification of the idiomatic key (the recognition point), which can be at the beginning, middle or end of the string. If the key is the last word of the idiom, the literal meanings of the idiom's constituent words will stay activated up until the last word. On the other hand, if the key occurs earlier on, by the end of the string, only the figurative meaning should be activated. If this happens, the final word of the idiom may not be accessed fully due to the idiom's high predictability, in which case we should observe a significant processing advantage for the figurative interpretation. However, Cacciari and Tabossi, as well as Tabossi and Zardon (1993) point out that the above only holds true in the absence of a biasing context, which prepares the reader for either a figurative or literal rendering. In the presence of such disambiguating context, idiom recognition may be anticipated and an idiom may be recognised before its uniqueness point has been reached (e.g., Cacciari & Tabossi, 1988; Tabossi & Zardon, 1993).

Another area of investigation has focused on whether idioms are processed more quickly than novel language. Swinney and Cutler (1979) found that idiomatic expressions, such as *break the ice* ('to facilitate social interaction') were processed reliably more quickly than non-idiomatic novel phrases (e.g., *break the cup*). Because idioms were processed more quickly than matched novel phrases, Swinney and Cutler proposed that idioms were retrieved as wholes. According to the authors, retrieving idioms as wholes means that each lexical item does not have to be activated and recognised, which thereby speeds idiom processing over that of a matched novel phrase, in which each lexical item must be retrieved and recognised. Similarly, Gibbs and Gonzales (1985) found that idioms are comprehended faster than literal control phrases. Such findings lend support to the lexical representation hypothesis, whereby idioms are stored in and retrieved from the mental lexicon akin to a lexical unit.

There is, however, another body of literature on idiom processing, which holds that idioms undergo syntactic analyses similar to novel language (e.g., Cutting & Bock, 1997; Konopka & Bock, 2009; Peterson, Burgess, Dell, & Eberhard, 2001). In an idiom recall experiment, Cutting and Bock (1997) presented participants with four types of pairs and asked them to repeat one member of the pair back: (1) idiom and its literal paraphrase (e.g., *hold you tongue* and *grab your lip*); (2) idiom and unrelated novel phrase (e.g., *hold your tongue* and *sign your name*); (3) idioms with same meaning (e.g., *hold your tongue* and *button your lip*); and (4) idioms with different meanings (e.g., *hold your tongue* and *flip your lid*). When compared to pairs with different meanings (Pairs 2 & 4), pairs with the same meaning (Pairs 1 & 3) generated significantly more blends. Importantly, when pairs had the same meaning (figurative or literal), there were similar numbers of errors (Pair 1 = Pair 3), suggesting a tendency for similar meanings to interact. This led Cutting and Bock to conclude that individual words and their meanings are active during idiom production. However, if we look more closely at the stimuli, we will see that for at least 10 out of the 32 items in Pair 1, the literal paraphrase is either nonsensical or must be construed as a metaphor (e.g., *shoot the breeze* and *fire the wind*). If such items are perceived metaphorically, it would explain why no differences were found in the error rates for Pairs 1 and 3, as both are similar types of non-literal language. If such problematic items are removed, the question is whether novel phrases (i.e., the literal paraphrase in Pair 1) are more prone to errors than the idioms in Pair 3. Without addressing issues with the stimuli, it is difficult to draw any strong conclusions from the findings.

Similarly, Konopka and Bock (2009) investigated priming patterns elicited by idiomatic (e.g., *pull off a robbery*) and non-idiomatic phrasal verbs (e.g., *pull off a sweatshirt*). They found similar priming patterns for phrasal verbs used both idiomatically and non-idiomatically. The authors took these results as an indication that all instances of language, conventional or completely novel, rely on generalised sentence building procedures in the same way. However, similar to the previous experiment, there are certain issues with the stimuli. Namely, the division between idiomatic and non-idiomatic items is unclear, as the ratings in one of their categories did not differ for the two types of stimuli. Even when the idiom/non-idiom contrast reached significance, the idiomaticity ratings for non-idioms were rather high in some categories (e.g., 3.85 for non-idioms versus 5.27 for idioms on a 7-point scale, where 1 = "not idiomatic", 7 = "highly idiomatic"). If the phrasal verbs used idiomatically and non-idiomatically were not different, it is unsurprising that the two yielded similar priming patterns. Crucially, this apparent lack of difference in the two types of phrases makes it difficult to draw strong conclusions about the processing of idiomatic and non-idiomatic language.

In a similar study, Sprenger, Levelt, and Kempen (2006) had participants produce previously memorised idiomatic or literal sentences upon seeing or hearing a prompt word. They found that both figurative and novel phrases could be primed by means of priming one of their individual words. Crucially, this priming effect was more pronounced in the case of idioms than novel phrases. This lead Sprenger et al. (2006) to propose a *hybrid* account of idiom processing, in which idioms have a holistic representation and are at the same time compositional. They hold that the individual words of an idiom are connected to the representation of the entire idiom. However, an alternative explanation for their finding, that idioms are primed more successfully than novel language, is that idioms due to their frequency are simply easier to remember and recall than novel phrases.

While the above research does indicate that idioms are processed compositionally and are thus subject to certain syntactic processes, this does not warrant the conclusion that they are processed just like any other instances of novel language. At the very least, the syntactic flexibility available in novel language by far exceeds that of idiomatic language. That is, while novel language can undergo a wide array of syntactic processes (e.g., passivisation, pluralisation, word order changes, aspect changes, etc.), it seems that when idioms *are* subject to syntactic processing, it is rather limited and some do not admit changes no matter how small (e.g., *by kith and kin, by and large, at the end of the day*). Finally, the last three studies discussed above address the issue of idiom processing from the language production perspective. Arguably, language production and comprehension are fundamentally distinct, and thus we cannot assume that findings and conclusions made in one domain (production) will be generalisable onto the other (comprehension). As Sprenger et al. (2006) argue, comprehension studies tell us little about the processes involved in idiom production, or the other way round.

Overall, research on idiom comprehension shows that, although idioms may be subject to syntactic processing similar to novel language, in terms of the speed of processing, they are, nevertheless, processed more quickly than novel language matched in individual word frequency and length. It is also evident that there is no consensus regarding figurative versus literal meaning activation as there seem to be three possibilities as to how the two meanings can be accessed: simultaneously, figurative preceding literal, or literal preceding figurative. The issue of idiom versus novel language processing, as well as comprehension of idiom's figurative versus literal meanings will be addressed in greater detail in Study 1.

3.5.2. Idiom processing in nonnative speakers

Similar to studies with native speakers, another important question is how figurative language is processed in a second language. A number of studies looked at nonnative processing of idioms versus novel language, as well as the processing of figurative versus literal idiom meanings (in ambiguous idioms). Van Lancker-Sidtis (2003) looked at whether prosodic cues were likely to help native and proficient nonnative speakers distinguish between an idiom's figurative and literal interpretations. Participants listened to tape-recorded sentences that contained idioms used either figuratively or literally (e.g., *He didn't know he was skating on thin ice, to skate on thin ice* = 'to do something dangerous'; *That's a real snake in the grass, snake in the grass* = 'hidden danger') and then had to identify the intended meaning. The native speaker group performed better than the high proficiency nonnatives. Results suggested that prosodic cues enabled native participants to successfully differentiate between idioms used figuratively and literally, whereas even highly proficient nonnatives were unable to use these cues at a native-like level.

In a cross-modal priming study by Cieslicka (2006), nonnative participants listened to sentences that contained familiar idioms. Sentence contexts did not bias towards the idiom's literal or figurative interpretation (e.g., *George wanted to bury the hatchet soon after Susan left, bury the hatchet* = 'make peace'). While listening to sentences, participants performed a lexical decision task on one of four targets: a word related to the idiom's figurative meaning (e.g., *forgive*), its control (e.g., *gesture*), a word related to its literal meaning (e.g., *axe*), or its control (e.g., *ace*). Faster response times to targets related to the literal meaning than to ones related to the figurative ones suggest that literal idiom interpretations were activated prior to figurative ones. Thus, according to Cieslicka, in nonnative idiom comprehension, the literal meaning enjoys a significant processing advantage over the figurative meaning, even when idioms are well known. However, perhaps, it is not surprising that upon hearing the word *hatchet* there is a strong facilitation for the word *axe* since the two words are semantically related.

Underwood, Schmitt, and Galpin (2004) used an eye-movement paradigm to investigate the on-line processing of idioms. They compared fixation count and fixation durations for the terminal word of an idiomatic phrase (e.g., *honesty is the* *best policy*) and a sentence containing the same lexical item (e.g., *It seems that his policy of* ...). They hypothesised that there should be a processing advantage for the word *policy* in the idiomatic context compared to the same word in the novel context. Indeed, a significant processing advantage was found for native participants; with fewer and shorter fixations made in the idiom condition compared to the novel one. For the nonnative speaker group, on the other hand, no such differences were observed: terminal words in and out of the idiomatic context were read with a similar speed, as evidenced by the same fixation durations and numbers of fixations. Although informative in terms of idiomatic versus novel language processing, the study does not deal with idiomatic versus literal idiom meaning processing.

Finally, Conklin and Schmitt (2008) conducted a self-paced moving-window reading experiment to investigate idiom comprehension by native and proficient nonnative speakers when a highly biasing story context preceded the idiom. The authors expected to find a processing advantage for idioms over matched novel phrases. Indeed, it was found that idioms (e.g., *hit the nail on the head* = 'to precisely capture the point') were read more quickly than novel phrases (e.g., *hit his head on the nail*) by both groups of participants. Further, they observed no processing differences between figurative and literal meaning processing for either natives or nonnatives. Because the same pattern of results was observed in both participant groups, the authors concluded that idiom comprehension in nonnative speakers is similar to that in native speakers. One downside of the study is that a within-subject design was used, which meant that the participants read idioms used figuratively, literally, as well as the novel phrase. Thus, any results obtained should be viewed with caution as they may have been influenced by repetition effects.

In the nonnative speaker domain, other types of non-literal language have also been investigated. For example, Matlock and Hereida (2002) looked at the processing of phrasal verbs with a figurative meaning and identical verb-preposition combinations used literally (e.g., *Paul went over the exam with his students* (a phrasal verb), *Paul went over the bridge with his bicycle* (a verb-preposition combination)). In an on-line reading experiment, it was found that monolinguals accessed phrasal verbs (with the idiomatic meaning) more quickly than identical verb-preposition combinations (with the literal meaning). Matlock and Hereida proposed that for monolinguals, the highly conventionalised figurative meaning of the phrasal verb is activated before the literal meaning of the verb-preposition combination. On the other hand, for the nonnative group, no significant differences were found in reading times for phrasal verbs versus verb-preposition combinations. Similarly, other research has demonstrated that nonnative speakers often have difficulties using phrasal verbs (e.g., Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Siyanova & Schmitt, 2007).

As can be seen from the above overview, research on whether nonnatives process idioms faster than matched novel strings is mixed. Unlike native speakers, previous findings suggest that even highly proficient nonnative speakers have difficulty processing idioms used figuratively.

3.5.3. Developmental studies on idiom processing in children

In the above sections, the issue of idiom processing in adult first and second language speakers was addressed. The current section focuses on similar issues with respect to children and adolescents.

Being a competent language user presupposes, among other things, the ability to appropriately use a wide range of idiomatic expressions. Nippold and Martin (1989) argue that the failure to master idiomatic phrases can negatively affect one's understanding of language in social, academic, and other settings. Idioms differ from novel language in that they are often rather opaque, which means that children may experience problems using them. To address this issue, a number of researchers have looked at idiom processing in children of various ages. Specifically, they looked at children's processing of literal and figurative meanings (i.e., ambiguous idioms).

Among the early developmental studies of idiom comprehension in children are those of Lodge and Leach (1975) and Prinz (1983). These studies deal with children's ability to interpret idioms used figuratively (i.e., intended meaning) or literally (i.e., not intended meaning). Overall, they suggest a developmental change in children's idiom comprehension. That is, younger children have a tendency to interpret idioms literally, while older ones (e.g., after the age of nine) appear to be more adult-like in their idiom interpretations. The same holds true for other types of non-literal language, such as metaphors, sarcasm, and indirect speech acts (e.g., Gardner, Winner, Bechofer & Wolf, 1978; Pollio & Pickens, 1980; Reynolds & Ortony, 1980). Overall, in the above studies, the following trend immerged: children below the age of nine had a tendency to interpret idioms and other types of non-literal language literally, rather than figuratively as they would be understood by adults. This implies that, first, children learn literal meanings and only then do they acquire figurative meanings of words and phrases (e.g., Cacciari & Levorato, 1989). In this, they are similar to second language learners discussed above for whom literal idiom renderings seem to have a processing advantage over figurative ones.

The question of idiom comprehension in young children is an interesting point of investigation due to, first, the abundance of idioms in language, and second, because of the relative inability of young children to grasp the concept of figurativeness.

Learning and processing the figurative meaning of an idiom may be difficult because it rarely equals the sum of the meanings of its constituent words. Thus, idiom comprehension may defy the child's existing linguistic knowledge (e.g., Lodge & Leach, 1975).

Developmental studies of idiom comprehension have looked at children aged five and above. Overall, it is generally believed that a figurative understanding of idioms develops gradually and that only by the late teens, do people master idioms fully (e.g., Ackerman, 1982; Gibbs, 1987; Lodge & Leach, 1975; Prinz, 1983). However, it is important to note that idioms vary greatly in their frequency of occurrence in child-directed speech, as well as their linguistic properties. A number of studies have looked at the role of context, familiarity, and metaphoric transparency in idiom comprehension.

With respect to *context*, research suggests that contextual information plays an important role in idiom comprehension in children. Ackerman (1982) found that six and eight year old children could understand idiom meanings mostly in the presence of idiomatically biasing context, whereas ten-year old children (as well as adults) were able to interpret idioms equally successfully in the presence of figurative or literal context. This suggests that younger children depend on context more than older children, and that for the latter group, idiom interpretations are fixed and not so strongly reliant on contextual bias.

The role of context in children's idiom comprehension was also investigated by Cacciari and Levorato (1989). The authors hypothesised that a rich context should facilitate children's comprehension of figurativeness even if the idiom itself is not familiar. Seven and nine year old children were presented with transparent idioms used figuratively and literally, in and out of biasing context. They were then required to perform a comprehension task. In another experiment, children's production of idioms was investigated. Although the data suggested a clear developmental trend (i.e., older children gave more idiomatic interpretations than younger ones), the authors also report that rich enough contexts can significantly improve idiom comprehension in children as young as seven. According to the authors, this argues against the idea that children are not able to understand idioms because they do not know them. Overall, the findings of this study suggest that even young children are aware of the fact that language can be used both figuratively and literally.

The role of context was further examined in a study by Nippold and Martin (1989). They used a large pool of adolescent participants (475) aged between 14 and 17 years old. Similar to Cacciari and Levorato (1989), they found that accuracy rates were higher for idioms presented in context, than for those presented in isolation. Again, there was a significant developmental trend: older participants performed better than younger ones. Despite this, however, it was further established that even the oldest group was not fully capable of performing the task perfectly either in, or out of context. Thus, these results suggest that an adult-like idiomatic competence is achieved relatively late.

In contrast to the above studies, Abkarian, Jones, and West (1992) failed to observe a significant effect of context in their experiment. They tested very young children's (three to six year old) idiom comprehension in and out of context. Their results showed that children did not find contextual information helpful.

The above studies suggest that while context has little effect on comprehension in children under the age of six (e.g., Abkarian, Jones, & West, 1992), its role becomes more prominent in older children, namely after the age of seven (e.g., Ackerman, 1982; Cacciari & Levorato, 1989; Gibbs, 1987, 1991; Levorato & Cacciari, 1992). It was also found to be an important factor in adolescents (e.g., Nippold & Martin, 1989; Nippold & Rudzinski, 1993; Nippold & Taylor, 1995; Nippold, Taylor, & Baker, 1996). Idiom comprehension in adult participants, on the other hand, was not found to be dependent on context (e.g., Ackerman, 1982; Cacciari & Levorato, 1989).

Familiarity of idioms has also been the focus of investigation in developmental studies. In Levorato and Cacciari's (1992) study, 264 seven to twelve year old children were tested to examine the role of idiom familiarity. It was found that familiarity played only a minor role in idiom processing. When it did play a role, it was evident in younger children (seven-year olds) who were not yet able to use context effectively to activate the appropriate figurative meaning, but not older ones. Thus, Levorato and Cacciari concluded that familiarity cannot adequately explain how young children comprehend idiomatic expressions. On the other hand, familiarity was found to play a bigger role in idiom production, where children were better able to provide correct idiomatic completions for familiar, but not unfamiliar idioms.

Idiom familiarity was further investigated in a recent study conducted by Laval (2003). Laval looked at idiom comprehension in six to nine-year old children. In the experiment, participants performed a story completion task, and were then asked to justify their chosen responses. Interestingly, unlike Levorato and Cacciari (1992), the familiarity effect was observed in older children (nine-year olds) but not younger ones (six-year olds). Laval suggests that the differences between the two studies may be due to the differing methodologies. She concludes that the relatively late emergence of the role of familiarity implies that the period of adolescence is crucial for the development of pragmatic knowledge.

Overall, while some studies attribute a relatively minor role to familiarity in children's comprehension of idioms (e.g., Levorato & Cacciari, 1992), others seem to give it more weight (e.g., Nippold & Rudzinski, 1993; Nippold & Taylor, 1995; Nippold, Taylor, & Baker, 1996). Such differences can be attributed to the different ages of the participants, as well as different experimental procedures. While some employed verbal explanation tasks (e.g., Nippold & Martin, 1989; Nippold & Rudzinski, 1993), other researchers used forced-choice tasks (e.g., Cacciari & Levorato, 1989; Levorato & Cacciari, 1992, 1999).

Finally, a third factor that has been shown to play a role in children's idiom processing is *semantic analysability*. Gibbs (1991) aimed to evaluate whether children had a harder time comprehending non-decomposable idioms than decomposable ones. A group of young children (five to nine year olds) listened to idioms in isolation, or at the end of a story context. They were asked to explain the intended meaning and then to choose their correct idiomatic interpretation. Idioms in the study varied in their degree of analysability, some were decomposable (e.g., *put down your foot*), while others were non-decomposable (e.g., *beat around the bush*). It was found that younger children's (six and seven year old) comprehension of decomposable idioms was significantly better than that of non-decomposable idioms. This held true when idioms were presented in isolation, or in a story context. Older children's idiom comprehension (eight and nine year olds) was found to be very similar for decomposable and non-decomposable idioms when items were presented within a context (which suggested that context aided these children). When given in isolation, on the other hand, the pattern resembled that observed in the younger group; namely, decomposable idioms were easier to comprehend than non-decomposable idioms. Thus, children were better able to understand idiomatic expressions if the idiom's components contributed overtly to the figurative meaning of the idiom than when they did not. For this reason, it is argued, children learn the meanings of decomposable idioms much earlier than those of non-decomposable idioms. Finally, the above findings imply that during idiom processing, children attempt to perform a compositional analysis on idiomatic expressions, which facilitates their understanding of decomposable idioms, but impedes that of non-decomposable ones.

Although, to the best of my knowledge, no study with second language learners has looked at the processing of decomposable versus non-decomposable idioms, some researchers have nevertheless also proposed that nonnative speakers have a tendency to analyse figurative speech compositionally, which may explain why they take more time to process idioms than literal language (e.g., Cieslicka, 2006). This 'literal-meaning-first' strategy may thus be shared by adult second language and child first language learners.

Overall, the above studies suggest that context, idiom familiarity, and semantic analysability play different roles in children's idioms processing. While context and semantic analysability have been shown to play a major role, the views regarding idiom familiarity appear to vary. The above studies further highlight the fact that idioms do not form a homogenous class but differ greatly in their properties.

3.5.4. Idiom processing in language-impaired patients

Thus far, I have talked about idiom processing in adult first and second language speakers, as well as children and adolescents. The findings presented above strongly suggest that figurative language is processed differently from literal language by these populations. Another strand of evidence for the special status of idiomatic expressions comes from studies with language-impaired patients.

There is a general consensus that familiar expressions are often selectively preserved in patients with language disorders, such as aphasia. In some cases, despite severe impairment in production of novel language, multi-word speech is produced with normal prosody and fluent articulation. For example, often patients with aphasia cannot name objects, but they nevertheless can do serial counting, as well as use expletives, swear words, and speech formulae (e.g., Van Lancker & Kempler, 1987). Fluency, articulatory precision and prosody of novel versus multi-word speech in aphasic patients has been shown to differ to such an extent that it has been proposed that different cerebral mechanisms may be involved (e.g., Van Lancker & Kempler, 1987). The division between novel and familiar language was pioneered by the English neurologist John Hughlings Jackson (1887), who studied the ability of aphasic patients to produce automatic multi-word speech (e.g., rhymes, speech formulae, etc.), while not being able to produce novel propositional speech. In his essay on the duality of the brain, Jackson proposed that the brain handles novel compositional speech and familiar multi-word speech differently, and that this is manifested in the way left- and right-brain damaged aphasic patients comprehend and produce different types of language.

According to Van Lancker (1988, 1990) and Van Lancker-Sidtis (2003) language ranges from completely novel at one extreme to over-learnt at the other. Newly-generated propositional speech entails the use of a range of grammatical and lexical rules. Familiar speech, on the other hand, due to being conventional and relatively fixed, does not need to be produced *de novo* every time it is used, nor are syntactic and lexical rules required to the same extent as for novel language processing.

A number of studies with aphasic speakers have investigated the proposition that multi-word speech is processed differently from novel, and that these two types of language may differ in their cerebral representation. Aphasia is an acquired language disorder characterised by an impairment of any aspect of the language faculty.

Patients with aphasia may experience difficulties producing or comprehending spoken or written discourse, or a particular aspect of discourse (depending on the extent and nature of the brain damage). Different accounts have been proposed. Some have suggested that while the left hemisphere is responsible for the novel language processing, the right one is strongly associated with familiar fixed expressions, also known as automatic speech (e.g., Van Lancker & Kempler, 1987). According to other accounts, multi-word speech may in fact be represented in both hemispheres, as it is not known whether the perseverance of familiar utterances is attributed to the undamaged right hemisphere, or intact areas in the left hemispheres (e.g., Van Lancker Sidtis, Postman, & Glosser, 2004). Below, I will review a number of studies that deal with the question of multi-word speech processing (production and comprehension) in patients with language impairment.

Van Lancker and Kempler (1987) investigated the comprehension of idioms and novel phrases by left- and right-brain damaged aphasic participants using a picture-matching auditory comprehension task. The authors proposed that if multi-word speech differs from novel language in terms of how it is represented in the brain, the pattern of processing should be different for the two types. Specifically, they predicted a larger role of the right hemisphere in multi-word speech processing. Van Lancker and Kempler tested left- and right-brain damaged patients, as well as a normal control group on comprehension of multi-word sequences (e.g., *He's turning over a new leaf; While the cat's away, the mice will play*) and novel phrases (e.g., *He's sitting deep in the bubbles; When the happy girl pushes, the angry boy swings*). The results for the two target groups revealed that despite impaired syntactic processing, the left-brain damaged group performed significantly better on familiar phrase comprehension than the right-brain damaged group. The latter group, on the other hand, performed better in the novel phrase comprehension task. Thus, familiar phrase recognition was found to be significantly less impaired than the ability to recognise novel expressions in the left-brain damaged population. The authors concluded that instances of multi-word speech, such as idioms, are represented in the brain differently from the newly generated language. It is noteworthy that the majority of the familiar phrases used in the study were idioms whose meanings are less transparent and more complex than those of novel phrases, and whose pragmatic load is higher than that of novel language. In spite of this, idioms were still recognised more easily than newly-generated phrases by the left-brain damaged group. The results of this study, it is argued, are suggestive of the special role played by the right hemisphere in the comprehension of idiomatic expressions by speech-impaired patients.

Following up on van Lancker and Kempler (1987), van Lancker Sidtis, Postman, and Glosser (2004) examined occurrences of multi-word expressions in the speech of normal, right- and left-hemisphere damaged participants. Their aim was to test the hypotheses about hemispheric processing of familiar expressions in the spontaneous speech of patients with unilateral brain damage and normal participants. In this study, the three groups of participants were required to describe their family and work. Subsequently, their discourse was analysed with respect to different types of multi-word sequences, such as idioms, proper names, and numerals employed. In line with van Lancker and Kempler (1987), they found that left-hemisphere damaged participants used significantly more multi-word expressions than either the normal control group or the right-hemisphere damaged group. Right-hemisphere damaged participants, on the other hand, produced fewer multi-word expressions than the control group or the left-hemisphere damaged group. It is argued that the finding that participants with the right-hemisphere damage use significantly fewer multi-word sequences than patients with the left-hemisphere damage in spontaneous unprepared speech offers further support to the view that an intact right hemisphere plays an important role in the processing of multi-word speech.

In a similar vein, in Kempler et al. (1999), left- and right-brain damaged patients performed a picture-matching task. They were auditorily presented with idioms (e.g., *she's got him eating out of her hands*) and four pictures for each of the idioms in the experiment. One of the pictures depicted a scene related to the correct idiomatic meaning of the idiom (e.g., a man showing affection towards a woman); another picture depicted a scene opposite to the intended meaning (e.g., a man paying no attention and ignoring a woman); the remaining two pictures were unrelated to the figurative idiom meaning but related to the literal meaning of idiom components. Thus, only one picture corresponded to the idiom and was accepted as a correct response. The results of the experiment showed that participants with the damage to the right hemisphere scored significantly worse than those with left-hemisphere damage. Similar to van Lancker and Kempler (1987) and van Lancker Sidtis, Postman, and Glosser (2004), the results were taken to suggest a unique role of the right hemisphere in non-literal language comprehension.

Speech disorders other than aphasia have also offered some insight into familiar language processing. For example, Paul et al. (2003) analysed familiar and novel language processing by patients with agenesis of the corpus callosum (ACC). ACC is a birth defect characterised by the absence of the corpus callosum (the part of the brain that connects the left and right cerebral hemispheres). Although the signs and symptoms of ACC have been shown to vary greatly among patients, a number of common symptoms have nevertheless been identified. Among them are vision impairment, bad motor coordination, sitting and walking related impairment, and swallowing difficulties. Individuals with ACC have also been shown to experience certain cognitive (e.g., problem solving) and social difficulties (e.g., impaired processing of pragmatic and paralinguistic cues). Despite the above symptoms, it is not uncommon for an individual with ACC to have a normal Intelligence Quotient (IQ).

Paul et al. (2003) examined the ability of patients with ACC and normal IQ to process non-literal speech, such as idioms and proverbs, as well as their ability to interpret prosodic cues of such non-literal language. Paul et al. administered three measures: LA Prosody Test (participants are required to match recordings with pictures denoting emotions), Formulaic and Novel Language Comprehension Test (participants are required to match literal and non-literal sentences with corresponding pictures), and the Gorham Proverbs Test (testing proverb comprehension). The results of the prosody test revealed that patients with ACC performed significantly worse than the control group, suggesting that their sensitivity to the emotional-prosodic cues (e.g., happy, sad, angry, surprised) was impaired. The results of the formulaic/novel language test showed almost identical scores in the literal language analysis for the ACC and control groups. In the non-literal language analysis, however, it was found that patients with ACC performed significantly worse than the control group. This implies that individuals with ACC had significant difficulties recognising the meaning of non-literal expressions, despite the fact that their processing of literal expressions was intact. Finally, the results of the proverb test revealed that the ACC group's interpretation of proverbs was significantly impaired compared to the control group. Thus, the results of this study suggest that individuals with ACC and normal IQ exhibit intact propositional language processing but are impaired in their processing of non-literal and emotional-prosodic meanings. Importantly, because none of the ACC participants had right hemisphere damage, the results were taken to suggest that language processing involves callosally-mediated integration of information from both the right and left hemisphere. Paul et al. thus conclude that for successful processing of non-literal language and paralinguistic cues, it is necessary for the two hemispheres to interact. In the case of the impaired interhemispheric integration, processing will diverge from normal despite the intact left, as well as right hemisphere.

The above studies with speech-impaired populations strongly suggest that multi-word speech, such as idioms, proverbs, compounds, and other types of familiar expressions, is processed differently from novel propositional speech. It has also been proposed that multi-word and novel language may differ in their cerebral representations. Another piece of evidence for literal/non-literal dichotomy in speech-impaired individuals comes from the research done in related areas, such as jokes (e.g., Bihrle et al., 1986; Brownell et al., 1983; Shammi & Stuss, 1999), indirect requests (e.g., Brownell & Stringfellow, 1999; Foldi, 1987; Stemmer, Giroux, & Joanette, 1994), and sarcasm (e.g., Kaplan et al., 1990). Unfortunately, it is beyond the scope of the current thesis to offer a comprehensive overview of these studies. However, it is noteworthy that the research done on jokes, sarcasm and irony, as well as indirect speech requests (all being instances of non-literal language akin to idioms) in speech-impaired populations points to similar findings as those outlined above.

3.6. Conclusion

A number of studies on idiom processing by child and adult first language speakers, proficient second language learners, as well as language-impaired patients were discussed above. Although these groups of populations vary greatly in their use of language, in particular, in their ability to produce and comprehend non-literal speech, they all have, nevertheless, made a unique contribution to our understanding of the phenomenon of idioms. The major proposition is that idioms are processed differently from novel language. What is meant by 'differently' depends on the study and the kind of participants. With regards to healthy native speakers, this implies quantitatively faster processing for idioms over novel language. For language-impaired patients, this suggests that different brain areas may be involved in the processing of idiomatic and novel propositional speech. For adult nonnative speakers, as well as children who are still in the process of acquiring their first language, the evidence is suggestive of the 'literal-meaning-first' strategy in the processing of ambiguous idioms (i.e., those that have figurative and literal rendering). Drawing on the above research, one of the aims of the present thesis is to address similar issues. Specifically, the first empirical study presented in Chapter 6 will look at the processing of ambiguous idioms used literally and figuratively, as well as novel strings of words by two groups of participants: adult first and second language speakers. Because research with children suggests that context plays an important role in idiom comprehension, the study presented in Chapter 6 will also address the role of context in native and nonnative processing of idiom literal and figurative meanings.

Chapter 4: Multi-word Speech in First and Second Language Acquisition

4.1. Introduction

The use of multi-word speech is regarded as an essential element of native-like communication (e.g., Cowie, 1998; Langacker, 1987; Pawley & Syder, 1983; Sinclair, 1991; Tomasello, 2003; Wray, 2002; Wray, 2008). Because our language abounds in various conventional expressions, it is important to study their acquisition and use, as well as the processes involved in their on-line comprehension. The present chapter addresses the issue of multi-word speech acquisition by first and second language learners. The latter will further be discussed with respect to child and adult second language learning.

4.2. Multi-word speech in first language acquisition

Studies into first language acquisition reveal quite clearly the interplay between holistic and analytic language processes (e.g., Locke, 1997; Peters, 1977, 1983). According to Bolinger (1975), first language learning is initially holistic, only later becoming more analytical. Researchers have noted that young children produce a large number of structures akin to unanalysed chunks, such as *Lemme see, I wanna do it, Gimme that* (e.g., Clark, 1974; Cruttenden, 1981; Lieven, Pine, Dresner & Barnes, 1992; Nelson, 1973; Peters 1983).

Thus, in L1 acquisition research, it has long been recognised that lengthy strings of language that correspond to several adult words can be treated as a single unit by a child (e.g., Bolinger, 1975; Plunkett, 1993). As Lieven (1987) argues, this is because segmenting words out of the speech stream is not an easy task for children. Thus for them, there is no one-to-one correspondence between a word and an acoustic signal, which may vary in accordance with the context. In other words, longer strings and single words may be perceived in a similar way.

It is generally agreed that children are capable of storing and using relatively complex strings of words before they are cognitively capable of analysing their internal structure. However, researchers have disagreed somewhat on the role that such chunks play in the child's early linguistic production. For example, Bates, Bretherton, and Snyder (1988) maintain that long strings are linguistic 'dead-ends'. However, they also point out that they can be useful for a child up until s/he is linguistically mature enough to apply the combinatorial rules of grammar. Brown and Hanlon (1970) also argue that because unanalysed chunks tend to resist segmentation, due to being over-learnt, they are unlikely to contribute to the child's linguistic development. Thus, children whose early vocabularies abound in such memorised chunks tend to be viewed as slow learners who are unable to analyse and segment adult speech (e.g., Bates, Bretherton, & Snyder, 1988; Bretherton et al., 1983).

Other researchers, on the other hand, hold that chunks produced by children play a crucial role in their early linguistic development (e.g., Clark, 1974; Peters, 1977, 1983; Pine & Lieven, 1993; Tomasello, 2003; Tomasello & Brooks, 1999). Specifically, Lieven, Pine, and Barnes (1992) challenge Bates, Bretherton, and Snyder's (1988) and Brown and Hanlon's (1970) views on the role of unanalysed chunks in children's early speech. In their study, Lieven and colleagues (1992) found that the use of frozen phrases correlated positively with general productivity. Peters (1977) further proposes that children *are* capable of breaking down multi-word strings into smaller components, the process, which is believed to contribute directly to the development of adult-like morphosyntax. In the same vein, Clark (1974), Cruttenden (1981), and Lieven, Pine, and Barnes (1992) suggest that the chunks that children have early on can be analysed for their internal structure and thus lead to their productive use.

According to Tomasello (2003), many children begin the language acquisition process by learning unparsed adult speech as holophrases (e.g. *I wanna do it, lemme see, where the bottle*), along with single words. Pine and Lieven (1993) add that almost all children have at least some instances of such frozen phrases in their early speech, and that it is not at all uncommon in early child language. What this means is that children learn meaningful linguistic structures of different shapes and sizes and of various degrees of abstraction (Tomasello, 2003).

In his work, Tomasello argues that if a child produced a construction, such as "wanna ride horsie", it is erroneous to assume that he or she has mastered complex grammatical concepts, such as an infinitival complement. More likely, it is an instance of a frozen phrase the child has previously heard from the parent. Or, it may be that the child knows how to say "wanna" and the activity or objects s/he wants. Thus, Tomasello (2003) proposes that children appear to create item-based constructions using:

- *intention-reading* (joint attention, understanding communicative intentions, cultural learning)

- *pattern-finding* (categorisation, schema formation, statistical learning, analogy)

That is, children are able to find patterns across various constructions they have been exposed to by means of schematising and making analogies (*Where's X* → *Where's Y* → *Where's Z*; *I wanna X* → *I wanna Y* → *I wanna Z*). Creating such schemas means learning concrete pieces of language for concrete functions and, crucially for the children's early linguistic development, forming abstract slots for abstract functions (Tomasello, 2003). Tomasello (2003, p. 306) concludes that the existence of holophrases (along with single words) that the child uses as single units is "of tremendous theoretical importance for theories of linguistic competence and performance". Thus, being "the major target of children's early language-learning efforts" (Tomasello, 2006, p. 310), utterance-level constructions of different complexity are an important point of investigation for psychologists and linguists alike.

4.2.1. Individual differences

Researchers distinguish between different types of language learners. Nelson (1973), for example, talks about *referential* vs. *expressive* children, Peters (1977) uses the terms *analytic* vs. *gestalt*, whereas Horgan (1980) – *noun-lovers* and *noun-leavers*. Despite these differences, the terms effectively mean the same: referential/analytic/noun-lovers refer to children favouring single words in their early language production. Expressive/gestalt/noun-leavers, on the other hand, are those with a strong presence of multi-word chunks in their early output. It is believed that different 'personalities', and other individual factors, such as immediate environment, may predispose a child to be one or the other style. However, it is also noteworthy that while some children exhibit characteristics of one particular style, many children appear to be using both styles.

According to Peters (1977, 1983), there is a continuum of children: those who are very analytic from the very beginning, through those who use both styles in different proportions, to those children who start with a gestalt approach and then adopt a more analytic one. Tomasello and Brooks (1999) also suggest that children can take either direction, part to whole or whole to part, and that most children use both styles. Nelson (1973) argues that referentiality requires individual word labels for objects and is thus supported by an analytic approach to language. Expressiveness, conversely, requires knowledge of longer units, as well as a means of employing them before they can be constructed from scratch. Referential/analytic children have little or no command of morphology and their vocabulary is mostly noun-based. Expressive children, on the other hand, have little knowledge of the word as a unit (Locke, 1997).

In her investigation, Clark (1974) reports on her son's usage of multi-word strings taken from adult utterances. She argues that such utterances are likely to be copied as unanalysed units, retained as such for some time, and only then do they become gradually analysed with some constituent parts substituted. Clark proposes that child language "becomes creative through the gradual analysis of the internal structure of sequences which begin as prepackaged routines" (Clark, 1974, p. 9).

In another study, Peters (1977) investigated a child's early language acquisition. While Peters expected approximation of words in the child's early production, what she also heard was an approximation to sentences. Peters reports that the child was producing at least two distinct kinds of speech:

- neat one-word utterances (analytic)
- phrases with a characteristic intonation contour, or 'melody'. The combination of syllables, stress, and intonation suggested that a longer unit, or even a whole sentence, was in fact intended (gestalt)

Crucially, the two types of speech appeared to be used for different communicative needs. Analytic, one-word-at-a-time speech was used in referential contexts, such as naming pictures and labelling. Conversely, the gestalt speech was used in more conversationally defined contexts, such as opening conversations, playing with the child's brother, discussing objects and events. According to Peters (1977), the child was making use of both styles. Likewise, Nelson (1973) proposes that referential children use language to name things, while expressive children use language to convey feelings, needs, and social forms. Thus, expressive children learn and use a large number of phrases and sentences early on in their ontogeny, while referential children do not. However, as Nelson (1981) and other researchers point out, most children are likely to exhibit features of both gestalt and analytic approaches.

In a similar vein, Pine and Lieven (1993) distinguish between two types of children: those who construct patterns by combining two or more items from their single-word vocabularies, and those who develop patterns by means of gaining some control over the slots in previously unanalysed memorised phrases. They further propose that variation in children's early speech can be explained in terms of different routes to multi-word speech. Similar to Peters (1977, 1983), Pine and Lieven argue that breaking down of initially unanalysed chunks is a common strategy that is used by most children to various degrees.

With respect to whether one approach is more advantageous than the other, Pine and Lieven (1993) propose that neither of them leads to any long-term advantage. However, they do point out that the presence of multi-word unanalysed chunks in the child's early speech may in fact facilitate their linguistic development by providing the child with 'slots-to-be-filled' templates. They thus argue against Bretherton et al.'s (1983) view that holistic speech in children's early production is a failure of a child to analyse his/her linguistic input into smaller component parts. Pine and Lieven's (1993) study showed that segmentation of unanalysed phrases from the child's input is as analytic as the segmentation of singe words. Thus, this phenomenon should not be viewed negatively. Pine and Lieven also argue against Bates, Bretherton, and Snyder's (1988) claim that the acquisition of unanalysed chunks is a 'dead-end'; in their study, they demonstrated the emergence of productive linguistic patterns from a range of unanalysed phrases. Thus, it is proposed, the acquisition of such phrases eases, rather than impedes, the child's transition from single- to multi-word speech.

In sum, the above accounts suggest that children remember utterances they are exposed to frequently, and that they are able to store such chunks along with single words and to subsequently unpack and use them as templates. In this way, frequent multi-word sequences may help children advance to productive syntax by generalising from them.

It is noteworthy that the majority of the studies mentioned above are based on naturalistic observations. Experimental evidence is rather scarce. One study, however, stands out in that it tested experimentally young children's processing of multi-word sequences. Bannard and Matthews (2008) claim to have found evidence that frequent multi-word units, such as *a drink of milk*, can become stored in the children's lexicon, suggesting that even very young children are sensitive to the frequency of multi-word sequences in their input. This study will be discussed in greater detail in Chapter 7.

4.3. Multi-word speech in second language acquisition

In the section below, I will turn to the role that multi-word speech plays in an L2 acquisition process, when learners' L1 is already (at least partially) in place. That is, the section will not deal with simultaneous L1-L2 acquisition. According to Ellis (2003), there are a number of fundamental differences between L1 and L2 learners. First, he argues, L1 learner's knowledge about the world around them and their linguistic knowledge develop in parallel. A more mature L2 learner, on the other hand, relies heavily on the notions and concepts already familiar to them in their mother tongue. More importantly, however, L2 learners possess analytical abilities, which L1 learners do not have. In the words of Ellis (2003, p. 72), L2 learners "can treat language as an object of explicit learning, that is, of conscious problem-solving and deduction, to a much greater extent than can children". While such analytical abilities will prompt the L2 learner to perform a compositional analysis on new forms and to obtain meanings for each of the components within a multi-word expression, their lack of these abilities in an L1 will result in such expressions remaining unanalysed until much later in the development of the L1. Further, L2 language learners normally have some knowledge of their L1, which may facilitate or hinder their L2 development. This, however, is not the case with L1 language learners. Finally, whereas an L1 learner has no other means of communicating other than using their developing L1 system, an L2 learner can, depending on the situation, bring in their L1 when felt necessary or appropriate (Wray, 2002). In what follows below, the acquisition of multi-word speech by child and adult second language learners will be discussed. It is worth noting that child learners covered below are those whose second language learning began in a naturalistic (i.e., untutored) setting long before the onset of puberty (roughly between the ages of 2 and 6). Adult learners in the review below are those whose second language acquisition took place in a classroom environment after the onset of puberty (e.g., after the age of 12).

4.3.1. Multi-word speech in child second language acquisition

Similar to what Bretherton et al. (1983) and Bates, Bretherton, and Snyder (1988) proposed with respect to a first language, Krashen and Scarcella (1978) proposed in the acquisition of a second language, namely, that long chunks are of little use either in real-life conversations or in the acquisition of grammar, and that they are effectively a dead-end for a learner. However, a lot of evidence has accumulated that shows that early second language acquisition is characterised by an extensive use of memorised strings of language (e.g., Hakuta, 1974; Huang & Hatch, 1978; Kenyeres, 1938; Vihman, 1982; Wong-Fillmore, 1976, 1979). Crucially, these studies suggest that multi-word speech offers great support in child second language acquisition.

Probably one of the earliest studies that looked at the use of multi-word speech in young second language learners is that of Kenyeres (1938). Kenyeres reports on her six-year old daughter Eva's naturalistic acquisition of French. The report spans over a year of Eva's family living in Geneva. Kenyeres found a considerable use of memorised sequences and other prefabricated patters in Eva's first months of acquisition of French. Kenyeres notes that some phrases became fixed in Eva's memory before she could understand their meaning, for example, *tout le monde à sa place*, *très joli*, *feuille d'où viens-tu?* Kenyeres further notes that Eva was eager to express herself in French even when she lacked the means to do so. For example, to make up with her mother after an argument, she says *maman, s'il vous plait, qu'est-ce que c'est, voulez-vous?* ('mum, please, what is it, would you like?'). She thus attempts to construct a sentence using three distinct multi-word expressions. However, around this time, Kenyeres reports, Eva exhibits her first attempts to construct a phrase from a previously learnt chunk. Addressing her father, she says *ou sont les mamans?* ('where are mothers?'). This phrase was constructed by analogy with a phrase learnt earlier at school *ou sont les ciseaux?* ('where are scissors?'). Eva clearly knew what the question meant, however, she was not aware of the plural form of the verb *sont* and the article *les*. In her essay, Kenyeres points out Eva's attempts to notice patterns and make use of analogies in her newly created constructions. Although Kenyeres' account offers an interesting insight into early second language acquisition in a naturalistic setting, it is unfortunately rather limited with regards to the use of multi-word speech, as Kenyere's study provides only a cursory mention of the use of multi-word sequences in her daughter's production.

In the 1970s, a number of longitudinal case studies were conducted with children acquiring a second language in a naturalistic setting. Hakuta (1976) reports on the acquisition of English by a five-year old Japanese girl. Hakuta points out that with advanced semantic development but with no means to express ideas and thoughts, learners' need to express various syntactic structures is particularly acute. He suggests that one way for a learner to meet these needs is to develop a strategy of using 'patterned' segments of speech (Hakuta, 1976). Hakuta identified and analysed the following patterns in his subject's production:

- patterns involving the use of the copula *be*

- the use of the construction *do you* as employed in questions
- the use of the construction *how to* in embedded *how*-questions

Hakuta (1976) argues that such patterns are prefabricated because they exhibit rigidity in their usage. The author reports that such prefabricated patterns accounted for a significant proportion of the child's utterances, namely around 50%. It is argued that by storing such prefabricated patterns as lexical items, the child was capable of producing common linguistic structures, which she would not normally be able to construct using her undeveloped language system. Thus, prefabricated patterns appear to enable learners to express a wide range of linguistic functions from the very beginning of L2 acquisition.

In another study, Huang and Hatch (1978) followed Paul, a five-year old Taiwanese boy in his acquisition of English. Huang and Hatch note that their subject used a large number of multi-word sequences, such as *get out of here*, not knowing or being able to use any of the components separately. The authors note that a lot of his early vocabulary was imitation of memorised chunks, for example, *let's go, don't do that, don't touch, it's time to eat and drink*, and so on. Importantly, the authors note that once the chunks were memorised, Paul used them in situations similar to the original ones. During Stage 1, his memorised sentences were grammatical even though he was totally unaware of the individual components within these utterances. During Stage 2, Paul is reported to have started substituting new nouns in questions (e.g. *where's pen? where's car? where's turtle*?). However, it is worth noting that while both studies (i.e., Hakuta (1976) and Huang and Hatch (1978)) are interesting accounts of multi-word speech use by early second language learners, they nevertheless suffer for lack of quantitative data. Furthermore, both Hakuta and Huang and Hatch investigated multi-word speech acquisition using only one subject, which makes it difficult to draw far-reaching conclusions and generalise onto larger populations.

Probably, one of the most detailed accounts of early L2 acquisition is given by Wong-Fillmore (1976). In a longitudinal study, Wong-Fillmore investigated the social and cognitive aspects of second language acquisition in five native speakers of Spanish. She argues that various "phrase-sized units" are among the first linguistic structures learnt in a new language, and that children are able to use a wide range of such constructions long before they know how to construct them. Similar to the observations made in the first language acquisition research, Wong-Fillmore argues that the form and meaning of the constituent parts of phrase-sized chunks are learnt only after the whole expression has been acquired and used a number of times by the learner. Wong-Fillmore collected natural speech samples from the learners as they interacted with their English-speaking peers. Although the children varied greatly in their abilities and attitudes, some constructions were used repeatedly by all children (e.g., *I dunno*, *I wanna + X*). Of the five children, one child in particular, Nora, made the greatest progress in learning English in a naturalistic setting. Not only did she use more multi-word speech in general, but also the constructions she used appeared to be more complex than those of other children (e.g., *I know how to do that*; *I gotta hurry up*; *Look, I have a better idea*). Similar to the studies discussed above, the author concludes that second language acquisition in children begins with the learning and using of multi-word sequences, and that it is the use of a range of such prefabricated patterns that gives learners their 'first grip' of the new language.

All of the above studies looked at a similar age group, namely, children between five and six. Vihman (1982), on the other hand, looked at her daughter's acquisition of English as a second language at a very young age – around the age of two. Vihman notes that such second language acquisition is of a particular interest due to still being within the period of intensive first language acquisition. At the age of 21 months Vihman's daughter (henceforth V.) started attending a day-care centre where only English was spoken. Vihman reports that in the first two months, most of V.'s lexical units were either complex words or multi-word strings, such as *happy birthday to you, thank you, what's that, come on, stop it, that's mine, my goodness, Happy New Year,* and so on. In the third month, V. began to construct simple sentences by substituting parts of previously learnt phrases. For example, V. produced *Linda out* on hearing *everybody out (of the car).* Vihman reports that there is no evidence that the complex units V. was producing were fully constructed; rather, some of them were fully memorised, or partly memorised and partly constructed. Among the formulae V. was producing during her third and fourth months, there were complex sequences, such as *I will be back, what happened, what you doing? what's the matter? I'm gonna come back to see you, I'm gonna bike, do it again, Jennifer,* and so on.

Finally, in a more recent study, Perera (2001) explored how three to five year old Japanese learners of English became linguistically creative by means of using prefabricated language. She found that the majority of the learners' novel expressions were constructed through the use of unanalysed, or partially analysed routines. Her study supports Wong-Fillmore's (1976) claim that prefabricated routines enable learners to construct their language with the help of the rules they elicit from the prefabricated templates. Perera's analysis showed that the most frequent types of utterances in her data were single-word utterances (45.7%), followed by productive multi-word utterances (27.4%) and prefabricated speech (14.6%). The least frequent categories in the data gathered were partially-analysed multi-word utterances (11.4%)

and freely-combined multi-word utterances (0.8%). Perera further explored developmental changes in the vocabulary of each subject and found that there was a general tendency for productive utterances to increase in number, whilst the number of prefabricated utterances decreased. Perera further notes that learners first seem to analyse (i.e., break down) prefabricated utterances that resemble the original one (e.g., *more cracker please → more apple please, more salad please*) and suggests that initially, learners tend to employ words from the same semantic field (e.g., food) before introducing words from other semantic fields (e.g., *more fork please*). Perera's overall observation is that learners start off with multi-word speech, then they gradually free up words within such strings creating slots and replacing them with new words, and thus produce novel expressions. She concludes that such multi-word phrases serve as "the basis for the learners' active analysis of linguistic rules" (Perera, 2001, p. 269).

The above overview suggests that the use of multi-word speech plays a crucial role in early second language acquisition since it allows the child to be an active participant, rather than a passive observer (Wong-Fillmore, 1979). Thus, children are able to use language before they know anything about its grammar or internal structure, and, more importantly, before they are capable of constructing a novel utterance from scratch. Crucially, it is claimed that such chunks constitute the linguistic material on which analytical activities could subsequently be carried out (Wong-Fillmore, 1979).

4.3.2. Multi-word speech in adult second language acquisition

The last part of the current chapter will deal with the role of multi-word speech in adult second language acquisition. Wong-Fillmore (1979) proposed that formulaicity plays a pivotal role in judgements about the speaker's degree of mastery of a language (i.e., when a native speaker judges a nonnative speaker as being native-like, or not). A number of longitudinal case studies have investigated multi-word speech acquisition and use in adult second language learners in a naturalistic setting (e.g., Hanania & Gradman, 1977; Huebner, 1983; Shapira, 1978; Schumann, 1978; Schmidt, 1983). Of these studies, only that by Schmidt (1983) documented an extensive use of multi-word speech by his subject. The results obtained in other studies do not seem to find evidence for an important role of prefabricated routines in untutored adult learners. Of greatest interest are thus studies with tutored adult second language learners.

Acquisition of grammatical constructions – evidence for learning chunks

A number of researchers have looked at the role of unanalysed chunks in the development of second language grammar. Bolander (1989) presents a study on the acquisition of syntactic rules, such as subject-verb inversion, by 60 adult learners of Swedish as a second language. For the purpose of the study, Bolander analysed object clauses of the following type: *det har jag last* ('it have a read'), *det kunde man gora* ('it/so could one do'), and *det tror jag* ('it/so think I') in spontaneous interviews and picture descriptions. The author maintains that although all elements, with the exception of *det*, vary, the clauses nevertheless give a 'stereotypical reading' (Bolander, 1989). She suggests that such patterns appear to be well integrated in the learner's language and that such constructions promote and facilitate the application of the inversion rule. It is proposed that creative language develops from previously memorised constructions. This proposition is similar to what has been reported in studies on first language acquisition. However, the major difference between first and second language learners is that, while the former may not necessarily know individual components of such complex strings, adult learners, on the other hand, have some knowledge of these components. Bolander concludes that syntactic rules begin to emerge when the number of memorised exemplars in learner language is large enough. However, Bolander's study suffers from a number of shortcomings. For example, the author provides very little quantitative data, such as frequency information, the total number of target constructions, or the number of correct responses. Thus, her data and conclusions appear to be rather impressionistic. Crucially, however, the results of the study are not incompatible with the view that these language learners were acquiring general grammar rules, rather than eliciting rules based on multi-word sequence segmentation, as argued by Bolander.

In another (longitudinal) study, Myles, Hooper and Mitchell (1998) examined data from a number of classroom learners of French and found that most learners gradually unpacked their early chunks, and were able to use them productively in new contexts. The focus in their study was on three constructions: *j'aime* ('I like', 'I love'), *j'adore* ('I love', 'I adore'), and *j'habite* ('I live'). The following examples, the authors claim, serve as the evidence for an unanalysed nature of such chunks: *j'aime le sp.. elle **j'aime** le sport* ('I like sp .. she I like sport' = 'she likes sport'); *un famille .. **j'habite** un maison* ('a family I live a house' = 'the family lives in a house'). The primary question that the authors wanted to answer was whether learners were able to unpack the initially unanalysed chunks and use them productively in new situations. The authors claim to have found that at first, the students kept the chunk intact but also added a lexical noun phrase to make reference, for example, *j'aime le sp- elle **j'aime*** *le sport (..) euh she likes euh elle .. **j'aime** la history museum* ('I like sp- she I like sport' [..] 'she likes she .. I like history museums'). Then, the process of segmentation took place, for example, *Richard tu n'aimes? .. Richard IL **n'aimes?*** ('Richard you don't like?' .. 'Richard HE doesn't like?'); *j'ai .. no oh .. Elle **habite** le (town)* ('I have .. no oh .. SHE lives in [town]'). Myles, Hooper and Mitchell report a great deal of variation in their subjects, with some showing very little or no segmentation, and others being able to analyse chunks and use their individual components in other constructions. The authors further acknowledge that even those learners who did show evidence of breaking down the chunks and subsequently using the parts productively in new situations (62.5%), were still using some of the formulae as unanalysed wholes by the end of the study.

Using the same data set as Myles, Hooper, and Mitchel (1998), Myles, Mitchell and Hooper (1999) investigated the relationship between unanalysed chunks and novel language in learner production of interrogatives, such as, *quel age as-tu?* ('how old are you?') and *comment t'appelles-tu?* ('what's your name?'). The aim of the study was to see how the students construct equivalent utterances with a third-person referent (e.g., *comment s'appelle-il?* 'What's his name?'). Despite the fact that only one learner ever used the correct third-person form *comment s'appelle-t-il*, the authors claim to have found a general route – starting with an inappropriately used, overextended chunk (e.g., second person *comment t'appelles tu*? referring to a third person), through chunks that started to break down with the subject omitted or replaced by a NP (e.g., *comment t'appelles (la fille)*), and finishing with the ultimate third-person pronoun question (e.g., *comment s'appelle-t-il*). Myles, Mitchell and Hooper (1999, p. 76) conclude that syntactic development and the process of chunk breakdown "go hand in hand", and that chunks serve as "a springboard for creative constructions" providing learners with linguistic data for further analysis.

Thus, Myles, Hooper and Mitchell (1998), as well as Myles, Mitchell and Hooper (1999) argue that their results indicate that learning of constructions and the appearance of productive syntactic rules are not two independent phenomena, but interact and feed into one another. However, from the results reported, it is not clear whether incorrect third-person forms, such as *comment t'appelles (la fille)* or *comment s'appelle?*, may have been the result of the process of active segmentation, as claimed by the authors, or some other process, such as, for example, overuse of a more frequent and hence over-learnt second-person form, from which these language learners may have been making overgeneralisations onto the third-person form. Further, given that so little progress was done in the course of the study (with very few learners mastering the constructions in question fully), the conclusions made by the authors do not seem to be very well justified. Finally, from the data reported in the two studies, it appears rather implausible that the target constructions ever managed to progress from memorised pieces of language to anything resembling creative grammar.

Acquisition of lexical constructions – evidence against learning chunks

Thus far, I have discussed studies that dealt primarily with the acquisition of grammatical constructions. However, another strand of evidence exists with respect to the acquisition of lexicalised routines, such as collocations. The use of lexical collocations (i.e., word combinations such as *heavy rain*, where *heavy* cannot be substituted by the semantically equivalent word *strong* without becoming unnatural as in **strong rain*) by adult second language learners has been investigated in a number of studies (e.g., Bahns & Eldaw, 1993; Biskup, 1992; Granger, 1998; Howarth, 1998; Nesselhauf, 2003, 2004; Siyanova & Schmitt, 2007; Siyanova & Schmitt, 2008).

In a cloze test and a translation task, Bahns and Eldaw (1993) tested second language speakers' knowledge of verb-noun collocations (e.g., *serve a sentence*). They found that such collocations accounted for a large number of errors (48%), despite the fact that the number of collocations constituted only 23% of the lexical items produced. Bahns and Eldaw concluded that collocational knowledge does not develop in parallel with general vocabulary knowledge.

Unlike Bahns and Eldaw (1993) who used elicitation tasks, Howarth (1998) extracted a number of verb-object collocations (e.g., *reach a conclusion*) from native and nonnative written corpora. Native speakers were found to use such collocations in their writing 50% more than proficient nonnative speakers. Further, nonnative writing was also characterised by the use of anomalous collocations (i.e., those that would not be normally deemed natural by native speakers). Similar to Howarth (1998), Nesselhauf (2003) analysed the use of verb-noun word combinations in nonnative speaker writing. She looked at free word combinations (e.g., *want a car*) and restricted collocations (e.g., *take a break*). Nesselhauf found significantly more errors in the use of restricted collocations (79%) than free combinations (23%).

Finally, Granger (1998) investigated the use of adverb-adjective collocations (e.g., *totally* + Adj.) in native and nonnative corpora. She found that such collocations were used less frequently by nonnatives than by natives. In a judgment task, nonnatives were further found to have a worse sense of appropriateness for adverb-adjective collocations. These findings made Granger conclude that even proficient second language speakers underuse native-like collocations and are more tolerant of atypical constructions. In a similar study, Siyanova and Schmitt (2008) also observed that nonnative speakers were more accepting of anomalous word combinations, such as **plastic operation* when compared to a group of native speakers.

Thus, the existing research suggests that second language speakers' use of appropriate collocations deviates from that of native speakers. One of the reasons for that is that they are difficult to translate across languages (e.g., Smadja, 1993). As Smadja (1993, p. 146) puts it, translating collocations "from one language to another requires more than a good knowledge of the syntactic structure and the semantic representation" (e.g., *plastic surgery* in English, but *plastic operation* in Russian; *heavy smoker* in English but *strong smoker* in German). Because many collocations are rather arbitrary word combinations, they must be readily available in both languages, that is, they should be memorised in their entirety (e.g., Smadja, 1993; also see Manning & Schutze, 1999).

Another reason why even proficient second language speakers experience difficulties with collocations is the fact that they tend to rely on linguistic creativity and make "overliberal assumptions about the collocational equivalence of semantically similar items" (Wray, 2002, pp. 201-202). Thus, if two expressions are synonymous (e.g., *heavy rain* vs. *strong rain*), L2 speakers may not be sensitive to the differences between the two phrases in the same way as native speakers. As Skehan (1998) and Foster (2001) suggest, unlike native speakers, nonnatives often construct and process a large proportion of their language compositionally, rather than using ready-made routines.

4.4. Conclusion

In the present chapter, a number of studies that deal with the acquisition and use of multi-word sequences in a first and second language have been reviewed. Despite the methodological differences between the studies, the findings point to the conclusion that multi-word speech plays an important role in the first and second language acquisition. L1 and early L2 learners have been shown to acquire grammatical knowledge by means of abstracting from previously learnt utterances. This implies that productive rules stem from unanalysed or partially analysed constructions, providing language learners with templates that enable them to participate in social interactions. Such chunks, thus, constitute the linguistic material on which analytical activities are carried out (e.g., Wong-Fillmore, 1979). With respect to late second language acquisition, it seems plausible to suggest that at the very least, the appropriate use of lexical collocations is an important factor in native speaker judgments about learners' mastery of the language. The above review further suggests that the use of lexical routines, such as collocations, is relatively poor in nonnative speakers. This has been largely attributed to the compositional nature of nonnative speech. This proposal, as well as nonnative findings discussed in the above studies, will become of relevance in two empirical investigations presented in the thesis (Study 1 and 2). With respect to grammatical constructions in late L2 learners, we have to remain cautious about the conclusions and propositions made in the above studies. Only future research will be able to show whether adult second language learners are indeed able to acquire grammatical knowledge by means of segmenting and abstracting from previously learnt chunks, as has been found in L1 acquisition.

Finally, it needs to be pointed out that as such, the vast majority of the studies reviewed above, both on first and second language acquisition, are naturalistic observations not extensively (if at all) supported by experimental evidence. In this respect, the area of multi-word sequence acquisition and use is very under-researched. Although the present thesis does not aim to address directly the issue of multi-word sequence acquisition in children or adults, it does, nevertheless, endeavour to provide substantial empirical evidence with respect to units larger than a single word, which will have important implications for the theories of first and second language acquisition, processing, and use.

Chapter 5: Experimental Techniques: Eye-tracking and ERP

5.1. Introduction

A number of researchers have suggested using different techniques on the same or similar stimulus material in order to obtain converging evidence on a particular topic (e.g., Altarriba et al., 1996). As Rayner (1998) argues, any differences or similarities across different paradigms can deepen our understanding and enrich our knowledge of the processes involved. With this in mind, it was decided to investigate on-line processing of multi-word speech using two methodologies: eye-tracking and event-related brain potentials (ERPs). Below, I will cover some of the key concepts and findings relevant to these techniques in general, as well as those more specific to the studies that will be presented later on in the thesis.

5.2. Eye-tracking

Eye-tracking has become an important tool in the study of language processing in real time. It can be broadly defined as a process of measuring *fixations* (what people look at and for how long) and *saccades* (very fast eye movements from one fixation point to another). Although the current review will focus on issues specific to reading (reading in a first language, to be precise), the eye-tracking technique is also commonly used in other areas of psychology and cognitive science (e.g., scene perception), as well as product design (e.g., advertising).

A French ophthalmologist Louis Emile Javal was the first to note in 1879 (see Rayner, 1998) that reading involves a series of fixations and saccades. Since then, a wealth of research has been presented with respect to eye movement behaviour during reading. As was first noted over a century ago, during reading (or looking at a scene), we repeatedly make very rapid eye movements (saccades). In between saccades, our eyes remain stationary for just about as long as needed to recognise a word (e.g., Rayner, 1998). Because saccades are so fast, it is believed that no information retrieval happens during saccadic movements (e.g., Liversedge, Paterson, & Pickering, 1998; Rayner, 1998). While saccades *per se* are not informative with respect to properties of words that are being read, fixations (namely, their duration as well as number) are highly representative of the information being attended to. It is a common finding that during reading (at least in English), fixation durations on an individual word are about 200-250 ms, with the mean saccade length being around 7-9 letter spaces (e.g., Rayner, 1998).

While the majority of words are fixated at least once, some words, especially shorter and more frequent ones, are skipped altogether. Carpenter and Just (1983) and Rayner and Duffy (1988) report that content words are fixated 85% of the time, while function words receive fixations only 35% of the time. However, this is not surprising because function words are among the most frequent words in language, and are also the shortest. Contextual constraints are believed to affect the amount of *skipping*. Balota, Pollatsek, and Rayner (1985) and Rayner and Well (1996) found that words that are highly predictable given the preceding context are skipped more frequently than words that are not constrained by the preceding context. Similarly, skipping rates are affected by word frequency and length: short and more frequent words are skipped significantly more often than long and less frequent words (e.g., Brysbaert & Vitu, 1998).

As Rayner (1998) points out, reading is not just about moving forward from left to right. Around 10-15% of all the saccadic movements are, in fact, *regressions*, or right-to-left movements (this assumes a left-to-right language like English). Regressions can be short, a few letters long, within the same word, suggesting processing difficulties specific to the word. Or, they can be over ten characters in length. Such long regressions imply processing difficulties and comprehension failures not with respect to an individual word, but to a longer stretch of language (e.g., garden path effects are often characterised by long regressions). Ambiguous words or problems with context integration can also lead to regressions. General text difficulty contributes to a large number of regressions being made (as well as longer fixation durations, shorter saccades, and fewer skippings) (e.g., Rayner & Pollatsek, 1989).

As has been mentioned above, individual properties of a word, such as length and frequency, affect fixation durations and the number of fixations made on a given word (e.g., Altarriba, et al., 1996; Inhoff & Rayner, 1986; Ryner & Duffy, 1986). The shorter and more frequent a word, the shorter and fewer fixations it will receive. Conversely, the longer and less frequent a word, the more likely it is to receive more and longer fixations. An interesting effect implicated in the processing of low frequency words is that of *spillover*. That is, the time spent on a low-frequency word 'spills over' onto the following word thus inflating this word's reading times (e.g., Rayner & Duffy, 1986; Rayner et al., 1989). With regards to the location of fixations, Rayner (1998) points out that the location of the first fixation (which is also likely to be the only fixation for shorter words) is roughly between the beginning and the middle of a word. Longer words, it has been shown, receive more than one fixation:

one towards the beginning of the word, and one towards the end (e.g., Rayner & Morris, 1992; Underwood, Bloomfield, & Clews, 1988).

One of the important questions raised in reading research regards the size of the *perceptual span*, namely, how much new information a reader can extract during a single fixation. A number of researchers have proposed that the size of the perceptual span (in alphabetical languages, such as English) is 3-4 characters to the left of a given fixation (e.g., McConkie & Rayner, 1976; Pollatsek, Rayner, & Balota, 1986; Rayner, Well, & Pollatsek, 1980) and 14-15 characters to the right of this fixation (e.g., McConkie & Rayner, 1975, Rayner, 1986; Rayner et al., 1981). The more difficult the text, the smaller the perceptual span.

The relatively large perceptual span to the right of a given fixation suggests that some information about the upcoming word may become available in the *parafovea* (i.e., the region to the right of the current fixation) (e.g., Rayner, 1998). Indeed, it has been shown that readers are able to extract some information about the word immediately to the right of a given fixation (e.g., Balota, Pollatsek, & Rayner, 1985; Henderson & Ferreira, 1990; Inhoff, 1989). For short words, Rayner (1998) argues that the information available in the parafovea allows the reader to identify the word and 'decide' if it can be skipped altogether. For longer words, on the other hand, the partial-word information in the parafoveal is unlikely to allow full identification of the word; however, it may still facilitate its processing. Similarly, frequency and predictability of words in the parafovea seem to play an important role. Inhoff and Rayner (1986) observed a larger role of the parafoveal view for more frequent words than less frequent ones.

With respect to *predictability*, the results of Balota, Pollatsek, and Rayner (1985) are of interest. In their study, Balota and colleagues manipulated the predictability of a word given its context. They found that highly predictable words were more likely to be skipped than less predictable ones, and when they were not skipped, such words were read significantly faster. This is an interesting observation that may also play an important role in the processing of multi-word sequences, such as idioms. Because idioms are familiar expressions, they are also highly predictable. That is, upon encountering *take the bull by ...* or *early bird catches ...*, comprehenders will automatically expect to hear or see *its horns* and *the worm*. Thus, the reading of familiar and hence predictable phrases may be facilitated in terms of the number of fixations and/or their durations.

Overall, eye-movement data, namely, fixations and saccades together with the associated events such as regressions and skippings, provide one of the richest accounts of how people read text in real time. Eye-movement recordings can tell us what has been fixated or re-fixated and for how long. Importantly, they can provide a millisecond-precise report of a reader's syntactic and semantic processing (e.g., Frenck-Mestre, 2005; Rayner, 1998). Another advantage of this methodology is that no secondary tasks are necessary (i.e., decisions requiring a button press). Thus, readers are engaged in the task of normal reading and can proceed entirely at their own pace (e.g., Rayner, 1998). Eye-tracking is therefore believed to permit reading which is as close to normal as possible in an experimental setting (e.g., Duyck, van Assche, Drighe, & Hartsuiker, 2007).

By far the greatest advantage of the eye-tracking paradigm is, however, the possibility to tease apart early and late processes of on-line reading. This means that both early and late effects of the experimental manipulation can be detected and examined separately. For example, it is possible to look at fixations made during first-time reading and then those that may have been the result of a certain processing difficulty. It is generally assumed that early measures (e.g., *first fixation duration* and *first pass reading time*, see below) are sensitive to early processes in the comprehension of a text, such as early integration of information. Late measures (e.g., *total reading time* and *fixation count*, see below), on the other hand, are believed to be sensitive to later processes associated with comprehension of a text, such as information re-analysis, discourse integration and recovering from processing difficulties (e.g., Paterson, Liversedge, & Underwood, 1999; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989).

Because early and late measures are thought to tap into different processes, it is important to analyse both. As Rayner (1998) argues, any single measure is a poor reflection of the reality of cognitive processing. It has, therefore, been proposed that in order to obtain a more complete picture of the cognitive processes involved in reading, one should examine a number of different measures (e.g., Rayner, et al., 1989). Liversedge, Paterson, and Pickering (1998) further suggest summing up fixation durations that are *spatially* and *temporally contiguous* in the text. Reporting both spatially and temporary contiguous measures, they argue, minimises the possibility that an effect may not be detected. Spatially contiguous fixations are those that "neighbour each other in a specified region of space" (e.g., total reading time and first pass reading time); while temporally contiguous are those fixations that "occur in a sequence over a specified period of time" (e.g., regression path duration and rereading) (Liversedge, Paterson & Pickering, 1998, p. 55). Liversedge, Paterson and Pickering argue that both approaches are needed to fully understand the influence of a linguistic variable on readers' processing of text, in particular, those effects attributed to processing recovery.

Below, some of the most common eye-tracking measures used in reading research are listed. The first two are considered to be an early measure (although first pass reading time is sometimes refereed to as 'mid' measure (e.g., McDonald & Shillcock, 2003b), while the last four are late measures:

1. *First fixation duration* – the duration of the first fixation within the area of interest regardless whether it is the only fixation or the first of multiple fixations within this region (represented by 3 in Figure 5.1). First fixation duration is the most commonly used technique in (single) word recognition research. This measure is taken to be the earliest point when one might expect to observe an effect due to the experimental manipulation (e.g., Liversedge, Paterson, & Pickering, 1998).

2. *First pass reading time* – the sum of all fixation durations made within a region of interest until exiting either to the left or to the right (also, known as *gaze duration*). This measure tells us how long the reader fixated the target the first time it was encountered (represented by 3 + 4 in Figure 5.1). According to Inhoff (1984), first fixation duration is a measure of lexical access, while first pass reading time also reflects text integration processes. Rayner (1998) points out that for much of the time, first fixation duration and first pass reading time yield very similar results. However, it is noteworthy that this only holds true for single words, which are likely to receive only one fixation. With respect to larger stretches of languages (such as phrases), the two measures are rather distinct. Rayner (1998) proposes that if the unit of analysis is larger than a word, then the total first-pass fixation time on that segment should be used as the primary eye-tracking measure. Because the aim of the present thesis is to look at on-line

processing of multi-word sequences (i.e., idioms and binomial expressions) that are made of at least three words, this measure was included in the analysis of the idiom, as well as binomial eye-tracking data.

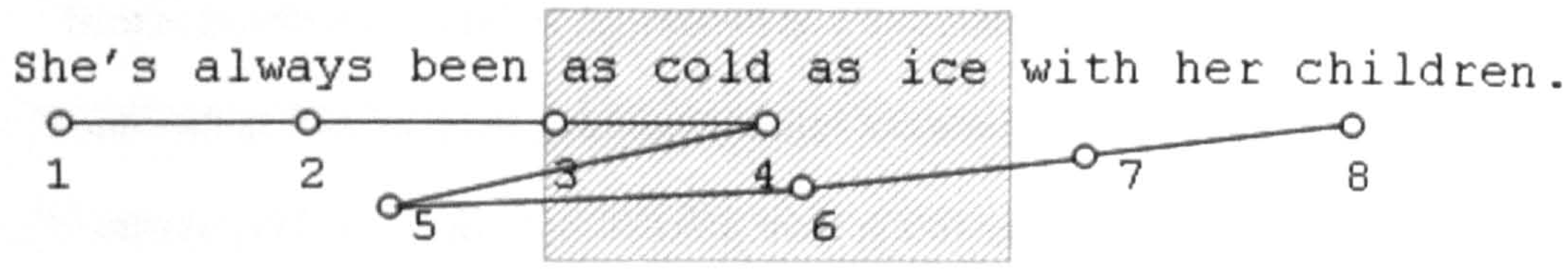
3. *Total reading time* – the sum of all fixation durations made within a region of interest. This measure includes all fixations that landed on the target and indicates how much time the participant spent reading the target (represented by 3 + 4 + 6 in Figure 5.1). Liversedge, Paterson, and Pickering (1998) propose that the total reading time measure is a mixture of initial processing time, as well as the time that may have been spent recovering from processing difficulties. They further argue that if an effect is observed for this measure, but not for an earlier one, such as first pass reading time, then this may be indicative of the manipulation having a late effect on processing.

4. *Fixation count* – the number of all fixations made within a given region of interest. This measure indicates how many times the target was fixated (represented by 3 + 4 + 6 in Figure 5.1).

5. *Regression path duration* – the sum of all fixation durations starting with the first fixation within a region of interest up to but excluding the first fixation to the right of this region. This measure gives us the durations of all fixations that were made on the target, plus all later regressions to the left of the target (represented by 3 + 4 + 5 + 6 in Figure 5.1)

6. *Rereading* – regression path duration for the region of interest minus first pass reading time for this region. Rereading time gives an indication of the time the participant spent rereading the text after having encountered a problem (e.g., Liversedge, Paterson, & Pickering, 1998) (represented by 5 + 6 in Figure 5.1).

Figure 5.1. Hypothetical eye movement record. Shaded area represents the region of interest.



1. First Fixation Duration = 3
2. First Pass Reading Time = 3 + 4
3. Total Reading Time = 3 + 4 + 6
4. Fixation Count = 3 + 4 + 6
5. Regression Path Duration = 3 + 4 + 5 + 6
6. Rereading = 5 + 6

As is clear from the above, not only does eye-tracking allow us to separate early and late processing stages as broadly defined, but it also enables us to look at a number of different early and late measures, which can shed further light on the processes involved in language comprehension. This is very unlike other reading techniques, such as self-paced reading or rapid serial visual presentation (RSVP), which can only provide one measure – total reading time of a particular segment, such as a word or a phrase. To sum up, eye-movement data are believed to reflect the moment-to-moment cognitive processes engaged during reading (e.g., Just & Carpenter, 1980; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989; Rayner, 1998), while the different eye-tracking measures (and the variability between them) are able to provide a multi-dimensional picture of reading. All this makes eye-tracking an invaluable tool in the investigation of on-line language comprehension.

5.2.1. Eye movement models of reading

Reading is thought to be the most complex cognitive activity that humans engage in on a daily basis (e.g., Rayner & Pollatsek, 1989). To provide a better account of the processes involved in normal reading, a number of computational models of readers' eye movements have been developed (e.g., Engbert, et al., 2005; McDonald, Carpenter & Shillcock, 2005; Reichle, et al., 1998). Such models can be divided into three broad categories: serial attention models, attention gradient models, and oculomotor-control models (e.g., Reichle et al., 2009). The most developed of the serial attention models is the *E-Z Reader model* (e.g., Reichle, Rayner, & Pollatsek, 2003), which will be discussed below. Of the attention-gradient models, I will focus on the *SWIFT model*. Because attention allocation is not the focus of oculomotor models, they will not be discussed. According to E-Z Reader and SWIFT, attention, which is necessary for lexical processing, plays a major role in guiding eye-movements. However, an ongoing debate exists regarding the nature of attention allocation during reading. Proponents of serial attention models argue that attention is distributed serially (i.e., only one word at a time can be processed), while proponents of attention-gradient models argue for parallel distribution (i.e., more than one word at a time can be processed). In oculomotor models, on the other hand, attention is thought to make no contribution to eye movements.

There is a close link between the nature of eye movements and cognitive processing. That is, eye movements are influenced by a number of variables on a moment-to-moment basis (e.g., Reichle et al., 1998). Characteristics of a given word, such as length and frequency, are reflected in the amount of time needed to process this word (e.g., Rayner & Duffy, 1986). Word predictability has also been shown to affect reading times. As mentioned in the previous section, if a word is highly constrained by the preceding context (sentential or phrasal), fixation durations on this word tend to decrease (e.g., Balota, Pollatsek, & Rayner, 1985; Rayner & Well, 1996). Crucially, Balota, Pollatsek, and Rayner (1985) and Rayner and Well (1986) showed that word predictability impacts the ease of processing even when frequency and length are kept constant. Because frequency, length, and predictability effects have been widely shown to affect language processing during reading, they are accounted for in most models of reading.

One of the most critical and highly debated questions raised in models of reading, and one that polarises the abovementioned serial attention and attention gradient models, is whether during normal reading, words are processed in a serial (i.e., strictly one word at a time) or parallel (i.e., two or more words at a time) manner. This issue will be discussed below.

E-Z Reader

The E-Z Reader model distinguishes three major stages involved in reading: visual processing, word identification, and attention allocation (e.g., Reichle, Rayner, & Pollatsek, 2003). With respect to early visual processing, it is proposed that word identification is most rapid if the word is fixated near the centre, and that longer words are processed slower than shorter ones. Visual processing is believed to involve low-level processing when word-boundary information is obtained, which necessarily precedes the following stage of word identification (high-level processes). The second stage, that of word identification, includes the early and late stage of lexical processing. The early stage is the identification of the orthographic form of the word, while the late one is the stage of lexical access (e.g., Reichle, Rayner, & Pollatsek, 2003). It is argued that the completion of the early stage of word identification makes the oculomotor system start to programme the next saccade, while the completion of the late stage shifts attention to the following word. The stage of lexical access is then followed by attention allocation, which is believed to occur serially. This is the key assumption of the model: words are processed in a strictly sequential manner, which is crucial to keep word order 'straight' (e.g., Reichle, Rayner, & Pollatsek, 2003; Reichle et al., 2009). Thus, even if a long and infrequent word (e.g., *marsupial*) is followed by a short and frequent word (e.g., *and*) as in *marsupial and*, the order in which they are accessed will not be affected: the first word will always be processed first, and second word second, irrespectively of their frequency and length. Thus, in line with the E-Z Reader model, the meanings of words should always be accessed incrementally, allocating attention sequentially, and never in parallel.

Criticisms of E-Z Reader come from a number of studies that report on *parafoveal-on-foveal* effects. It has been shown that information available in parafovea plays an important role in reading (e.g., Morris, Rayner, & Pollatsek, 1990). An interesting question is how a parafoveal word (word n+1) affects the processing of the currently fixated word (word n). It has been shown that fixation durations on word n were shorter when word n+1 was a low frequency long word (e.g., Kennedy, 1998). Similarly, Brysbaert, Desmet, and Drieghe (2005) report an effect of parafoveal word length. That is, a long parafoveal word led to shorter and fewer fixations on the preceding foveal word. These findings have been explained as supportive of the models of eye movements in which word n and word n+1 can be processed in parallel, rather than serially. Such effects, known as parafoveal-on-foveal, have become a major issue in reading research because they can help elucidate the question of serial versus parallel language processing (e.g., Brysbaert, Desmet, & Drieghe, 2005).

Because parafoveal-on-foveal effects seem to support parallel word processing (e.g., Hyönä & Bertram, 2004; Kennedy, 1998; Vitu et al., 2004), they have given rise to parallel models of eye movements, the most prominent and well developed being the SWIFT model (e.g., Engbert et al., 2005).

The SWIFT model

The SWIFT model has borrowed many of the features of the E-Z Reader. For example, similar to E-Z Reader, SWIFT assumes that lexical access happens in two stages and that processing difficulty is related to word frequency and predictability. Contrary to E-Z Reader, however, in the SWIFT model, the assumption is that attention is simultaneously distributed across more than one word at a time, suggesting parallel lexical processing of words during reading (e.g., Engbert et al., 2005). It is assumed that the processing rate is highest for the word currently being fixated (i.e., foveal word) and decreases on parafoveal words to the left and to the right of the fixated word (e.g., Engbert et al., 2005). Thus, the central idea of the SWIFT model is that a few words can be accessed in parallel. Going back to the *marsupial and* example, the idea of parallel processing would imply that the second word (i.e., *and*) should be accessed first because it is shorter and more frequent, while the first word (e.g., *marsupial*) will be accessed second because it is longer and less frequent. In other words, the SWIFT model seems to allow processing where word n+1 can be identified before the preceding word (word n). However, if this were the case this would suggest that during reading, words can be accessed and processed out of order. Because this is unlikely to be the case (as this would disrupt reading), this argument is used as a major criticism of the SWIFT model (e.g., Reichle et al., 2009).

Finally, as stated above, both E-Z Reader and the SWIFT models acknowledge the role or predictability (i.e., predictable words are processed more quickly and are skipped more often than less predictable words). However, due to the fundamental differences with respect to attention allocation, their specific predictions regarding word predictability will differ. Namely, in E-Z Reader, it is assumed that the information that constrains the identity of a predictable word will become available only *after* the preceding word has been fully identified and processed. In SWIFT, on the other hand, the processing of a particular word is facilitated even if the preceding word has not yet been fully processed (e.g., Reichle et al., 2009). Despite these differences, however, both models predict faster reading times for predictable words given the context.

5.2.2. The concept of a word

As is clear from the above discussion of the two models, the major difference between the two models is the number of words that can be processed at a time: one (E-Z Reader) or more than one (SWIFT). This makes a word the main unit of measurement in both models. With regards to this, Elman (1990) raises an interesting question: what should be considered to be a word? He argues that despite the fact that it is common to speak of the basic units of language being words and morphemes, such units are yet to be clearly defined because a large number of instances appear to be ambiguous. Elman (1990) argues that languages differ vastly in what they treat as a word. For example, what would be considered to be a word in the Eskimo-Aleut family of languages would more likely be called a phrase or even a sentence in English. In non-alphabetic languages, such as Chinese, there is often ambiguity about which characters constitute a word; while in some alphabetic languages, such as Thai, word boundaries are not indicated (e.g., Reichle et al., 2009). Even in English, Elman (1990) proposes, there is no clear distinction between monomorphemic words, such as "apple", compounds "apple pie", and frequent phrases "Library of Congress". He concludes that the key concepts of linguistic enquiry are thus rather fluid, which is likely to have implications for language processing.

In their E-Z Reader model, Reichle et al., (2009) acknowledge that their definition of a word – a sequence of letters that is separated by spaces – is problematic. Although no such definition is given in the SWIFT model, we can assume a similar stance. This definition, albeit the most obvious with regards to the English language, is not without limitations. For example, there is a class of words, known as *compounds* that vary greatly in their orthography in English. The same compound can be spelled as one word, a hyphenated word, or as two separate words without violating English orthographic rules (e.g., *lifestyle*, *life-style*, and *life style*). According to their definition of a word, *life style* should be treated by the language processor as two separate words in both E-Z Reader and the SWIFT model; while *lifestyle* should be read as one word. A hyphenated word (e.g., *life-style*), on the other hand, may be perceived either as one, or two words. However, in reality, it is unclear whether *lifestyle, life style,* and *life-style* will be read in the same or different way.

A somewhat similar issue may arise in the case of highly frequent phrases which are always spelled as two words, but which may be treated by a reader as a singe unit.

For example, in a reaction-time study, Sosa and MacFarlane (2002) found that access to the preposition *of* within very frequent collocations (e.g., *kind of* and *sort of*) was severely disrupted (as evidenced by slower reaction times, as well as poorer accuracy) when compared to less frequent collocations, suggesting that readers may have treated such collocations not as two words, but one.

In the present thesis, the question of what constitutes a word is not directly addressed. However, it was deemed necessary to raise this issue due to the nature of multi-word sequences, characterised (among other things) by spanning over one word, high frequency of occurrence, relative fixedness, and, often, non-compositionality. The above discussion merely serves as an indication that it may not always be accurate to consider spaces as word boundaries, because our mental representations of certain linguistic material (e.g., frequent phrases) may span more than a single word. If this is the case, then the question of how many "words" can be processed at a time, one, two, or three, may become invalid.

5.3. ERP

Electroencephalography (EEG) is the recording of electrical activity produced by neurons in the brain. EEG is recorded using electrodes placed on a participant's scalp, and can vary in number from 16 to 256. The observed EEG is believed to reflect the activity of a number of functionally distinct neuronal populations (e.g., Van Petten & Kutas, 1991). EEG is a common technique used in clinical research to diagnose various conditions, such as, epilepsy, coma, strokes, and other brain disorders. It is also widely used in non-clinical research. Specifically, *event-related brain potentials* (ERP), which are EEG responses time-locked to a particular stimulus and averaged over a large number of trials, are commonly used in cognitive science.

ERPs plotted against post-stimulus time are represented by a series of positive and negative peaks (e.g., Van Patten & Kutas, 1991). Such positive and negative waves are associated with different ERP components. A component is a reflection of the neural mechanisms associated with particular cognitive or perceptual processes (e.g., Kaan, 2007). A number of ERP components have been documented in literature (e.g., LAN, N100, P200, P300, N400, and P600).

One of the greatest advantages of the ERP methodology is its high temporal resolution. ERP recordings have a temporal resolution up to a millisecond, which is the precision often required in language research but not achievable in experiments employing behavioural measures, such as self-paced reading. According to Kutas and Van Petten (1994), ERP measures are as close to immediate and on-line processing as is technically possible. Another important benefit of the ERP, and the reason why it has been so widely used in language research, is that not only can it tell us *when* something happened, but it can also inform us about the very nature of the cognitive or perceptual process involved, such as semantic or syntactic processing difficulty. As such, the ERP methodology, unlike eye-tracking which reflects the pattern of eye-movements and provides reading times, is a direct reflection of the brain activity.

Another advantage of the ERP is that no secondary task is necessary in order to obtain data (Kaan, 2007). In many behavioural experiments, participants are required to make a secondary button press. No such tasks are required in ERP experiments as the recorded brain waves reflect all the cognitive and perceptual processes that the participant is going through at that very moment. However, this is not to say that no secondary tasks are needed at all. As in behavioural experiments, some tasks (e.g., categorisation or Go-no Go task) are still required in order to ensure that participants stay engaged and alert at all times. Finally, as Kaan (2007) points out, ERP is one of the few techniques that enable researchers to investigate on-line processing of spoken discourse.

ERP is, thus, an extremely valuable and informative tool. Crucially, it has been used extensively in language research. Below, I will review some of the well-documented ERP components that have been studied with respect to language processing. Specifically, I will focus on a number of components normally associated with two types of linguistic processing: syntactic and semantic. Because the present thesis deals primarily with the latter, the syntactic processing will be discussed briefly.

5.3.1. Syntactic processing

Two ERP components have been shown to be sensitive to syntactic violations: the left anterior negativity (LAN) and the P600. A left anterior negativity (LAN) has been shown in sentences with grammatical violations, for example, where the verb does not agree with the noun (e.g., Kutas & Hilyard, 1983; Coulson, King, & Kutas, 1998), and in sentences with garden paths, such as *John painted the table and the chair was already finished* (e.g., Kaan & Swaab, 2003). LAN is negativity most prominent in the left-anterior area of the scalp. Two LAN effects have been documented in the literature, an early and late LAN. An early LAN (ELAN) peaking around 100 and 200 ms after stimulus onset has been shown to be sensitive to word category violation. For example, when the expected completion is a noun, but the reader encounters a verb, then this elicits ELAN (e.g., Friederici et al., 1993; Hahne & Friederici, 1999). A late LAN, which peaks around 300 and 500 ms after stimulus onset, has been found to show sensitivity to morpho-syntactic agreement (e.g., Friederici et al., 2003).

Another component associated with syntactic processing is the P600. It is a positive-going wave peaking between 500-900 ms after the onset of the critical stimulus. This component has been found to show sensitivity to syntactically incorrect sentence completions (e.g., Osterhout & Holcomb, 1992; Hagoort, Brown, & Groothusen, 1993), or grammatically correct ones that are difficult to process (e.g., Kaan et al., 2000). In certain cases, the P600 is preceded by the ELAN.

5.3.2. Semantic processing

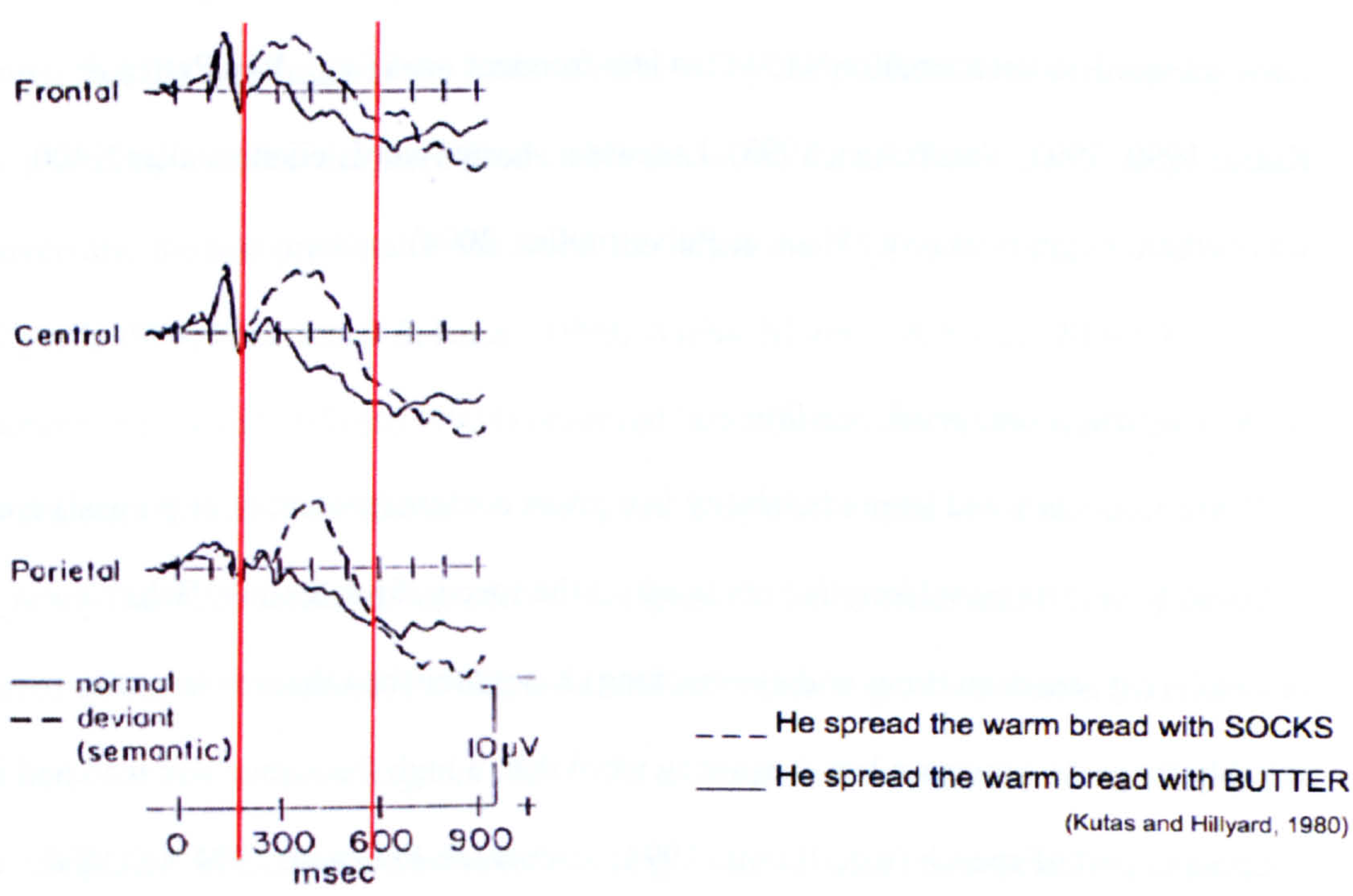
As stated above, the ERP investigations presented in the current thesis will deal with semantic processing. Thus, the two components that will be of most interest and relevance are the P300 and the N400, which are normally associated with semantic processes. Below, they will be discussed in some detail.

N400

The N400 is a negative-going wave peaking between 300 and 500 ms after the onset of the stimulus, which can be a word or a picture. The N400 has a widespread topographic distribution, but is most prominent in the centro-parietal area of the brain, with the maximal over the vertex (e.g., Curran, et al., 1993). All words elicit the N400 component, both semantically congruent and incongruent. The N400 component should not be confused with the N400 effect, the latter being the actual difference between the waveforms produced by two conditions (i.e., semantically legal vs. semantically anomalous). In literature, the N400 has been shown to be sensitive to the processing of lexical-semantic information and frequency, as well as real-world knowledge (e.g., Kutas & Hillyard, 1980; Hagoort et al., 2004).

The N400 component was first described by Kutas and Hillyard (1980) and Kutas, Van Petten, and Besson (1988). In a set of experiments, ERPs evoked by semantically congruent sentence completions (e.g., *He spread the warm bread with butter*) were compared with those evoked by semantically incongruent ones (e.g., *He spread the warm bread with socks*). It was found that incongruent completions elicited a negative wave most prominent over posterior scalp locations and larger over the right than the left hemisphere. Congruent sentence completions, on the other hand, elicited a positive-going wave (Figure 5.2).

Figure 5.2. N400 effect observed in normal and semantically deviant conditions (from Kutas & Hillyard, 1980).



Kutas and Van Petten (1994) note that the latency and amplitude of the N400 depend on experimental manipulations, with the largest N400 being elicited by semantically anomalous content words. When presented in lists or pairs rather than sentences, words that are unrepeated and semantically unrelated to the previous stimulus have been found to elicit the largest N400. Thus, as the incongruity of a word within the sentence context increases the amplitude of the N400 also increases. Similar effects have been observed in word pairs. Semantically unrelated words elicit larger N400 amplitudes than semantically related ones. It has been proposed that this ERP component serves as an index of semantic priming, a process where identification of a word is easier if preceded by a related word (e.g., Steinschneider & Dunn, 2002).

The N400 has also been shown to be sensitive to lexical properties of a word. More frequent words elicit a smaller N400 than less frequent ones (e.g., Van Petten & Kutas, 1990, 1991; Van Petten, 1993). Likewise, shorter words elicit smaller N400 waves than longer ones (e.g., Hauk & Pulvermuller, 2004).

N400, frequency, and predictability

Word frequency and its predictability in a given context (sentential or phrasal) are believed to be two most likely factors to affect the speed of processing. Behavioural research (e.g., reaction times and eye-tracking) has established that readers take reliably longer to process a low frequency word than a high frequency one matched in length and part of speech (e.g., Balota, 1994; Jescheniak & Levelt, 1994; Inhoff & Rayner, 1986; Rayner & Duffy, 1986). With regards to ERP, frequency effects have most commonly been reported around 400 ms after the onset of the stimulus and are thus associated with the N400 component (e.g., Van Petten & Kutas, 1990). However, it is also worth pointing out that because a word can be read in under a quarter of a second, some frequency effects have been obtained in a much earlier window, namely, around 130 - 190 ms after the stimulus onset (e.g., Sereno et al., 1998; Hauk & Pulvermuller, 2004).

In general, it is believed that language comprehension relies heavily on the predictive mechanisms based on the information that has already been processed and the information that is currently being processed (e.g., Roehm et al., 2007). Word probability (also known as *cloze probability*), which can be broadly defined as the reader's ability to predict the upcoming word(s), has been shown to influence word recognition in reaction time and eye-tracking studies (e.g., Kleiman, 1980; Kliegl et al., 2004; Kliegl, Nuthmann, & Engbert, 2006; Rayner & Well, 1996). In the ERP research, word predictability has been linked to the N400 component. It has been shown that the amplitude of the N400 is affected by the predictability of a word given the preceding context: the more predictable the word, the smaller its N400 amplitude; conversely, the less predictable the word, the larger the N400 wave (e.g., Kutas & Hillyard, 1984; Federmeier & Kutas, 1999; Wicha, Moreno, & Kutas, 2004). For example, Kutas and Hillyard (1984) observed larger negativity on the word *hour* in the sentence *The bill was due at the end of the* <u>*hour*</u> than on the word *month* in the sentence *The bill was due at the end of the* <u>*month*</u>. In Federmeier and Kutas (1999), a similar effect was observed. Participants read sentences (e.g., *They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of* ...) completed with expected completions (e.g., *palms*), with unexpected but plausible ones of the same category (e.g., *pines*), or of different category (e.g., *tulips*). Similar to Kutas and Hillyard (1984), the expected completion elicited a smaller N400 than either of the two unexpected but plausible ones. Interestingly, despite their similar cloze probability, the same category word (*pines*) was found to elicit a smaller N400 than the word from a different category (*tulips*). Thus, the N400 effect is believed to reflect semantic integration of a word into the unfolding (sentential or phrasal) context.

P300

The P300 component, first discovered by Sutton, Braren, Zubin, and John (1965), is a positive deflection in voltage observed between 250 and 400 ms following the stimulus presentation. The signal, peaking around 300 ms, is strongest in the parietal area. The most common interpretation of the P300 is that it is the result of unexpected stimuli, and that it reflects the updating of the working memory (e.g., Verleger, 1988). However, as we will see below, this view has been challenged.

The P300 encompasses a number of distinct components, of which the P300a and the P300b are most common. The P300a (also known as 'novelty P300') is associated with unexpected events. It is more anterior in its topography. The P300b, which is more posterior, is known to be elicited by infrequent task-relevant events. The studies presented below deal with the P300b effect.

Two major accounts have been proposed to account for the P300 effect: a *context-updating theory* (e.g., Donchin, 1981; Donchin & Coles, 1988; Donchin & Fabiani, 1991) and a *context-closure account* (e.g., Verleger, 1988). Both accounts relate the P300 effects to expectancies that may arise during stimulus processing. However, while the context-updating theory predicts larger effects for unexpected events, the context-closure theory accounts for the larger P300 in terms of the closure of certain expectations (i.e., event $_{n-1}$ implies that event $_{n}$ will follow). The major difference between the two theories is that in the context-updating theory, the P300 reflects an *expectancy violation*, while in the context-closure theory, the P300 reflects an *expectancy confirmation* (e.g., Riess Jones, 1988). According to the former, target stimuli are compared against the content of working memory and then updated with respect to incoming information. Closure, on the other hand, has been described in terms of post-stimulus activities that lead to the decision that a signal belongs (or does not) to a particular class (e.g., Desmedt, 1980; Verleger, 1988). Verleger 's account further predicts that the P300 waveform should be related to the closeness of the match, and inversely related to the difficulty of the task. Thus, the P300 is evoked by stimuli that are awaited when expectancies have been fulfilled (e.g., Verleger, 1988).

A number of researchers have linked the P300 effect to "template matching" (e.g., Chao, Nielsen-Bohlman, & Knight, 1995; Ford, 1978). That is, participants may develop a neural representation or a template of the stimulus. The closer the match between the incoming information and the template, the larger the amplitude of the P300 (e.g., Kok, 2001). Thus, according to Kok (2001, p. 573), the P300 reflects "the awareness that a stimulus belongs or does not belong to the category of a certain memorised target event" (Kok, 2001, p. 573).

P300 versus N400

Duncan-Johnson and Donchin (1977) have proposed that the P300 component is influenced by the *probability* that a given stimulus will appear given the previous one. The N400 component, on the other hand, is believed to be associated with semantic processing under *unexpected* conditions (e.g., Finnigan et al., 2002).

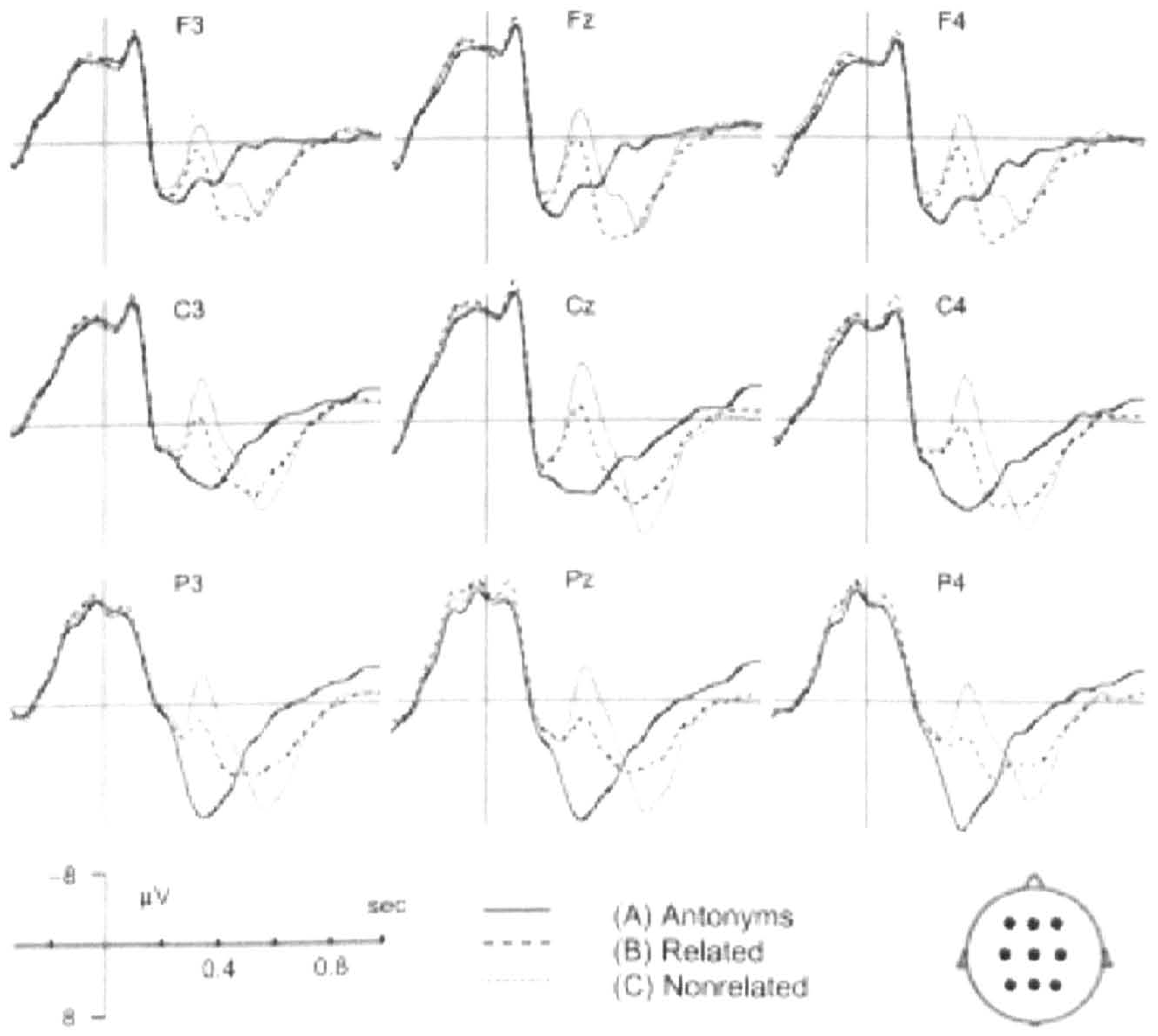
Both the P300 and the N400 are late visual evoked potentials (VEP). According to Luck (2005), one of the important issues in the ERP research is that of establishing whether a particular effect was caused by a single component or by two different components. Vespignani et al. (2009) further highlight this issue with respect to distinguishing between a diminished N400 and a larger P300. A number of studies showed that P300 effects may be observed within the N400 time range (e.g., Roehm et al., 2007; Vespignani et al., 2009). Thus, the latencies and peaks of the two components may overlap. There are, however, important differences between the P300 and the N400. First, the two components have different latencies. The N400's latency is between 300 and 500 ms, with the peak around 400 ms after the onset of a critical stimulus. The P300's latency is between 250 and 400 ms, peaking around 300 ms after the onset of the stimulus. Further, as mentioned above, the N400 component is most prominent in the centro-parietal area of the brain, with the maximal over the vertex (Cz). The P300 component, as reported in the most recent studies (e.g., Roehm et al., 2007; Vespignani et al., 2009), has a more posterior distribution with the maximal over the parietal sites (Pz).

P300 and expectancy

The P300 component is believed to be influenced by the probability that a given stimulus will appear and is a measure of attention allocation (e.g., Duncan-Johnson & Donchin, 1977). Roehm et al. (2007) focused on the P300 and the N400 components (Experiment 1) in the processing of antonymous adjectives (e.g., *black* and *white*). Participants read sentences like *The opposite of black is ...* which ended in with *white* (the correct completion), *yellow* (related), or *nice* (nonrelated). As predicted, the N400 was observed when the sentence was completed with the nonrelated adjective (e.g., *nice*). Crucially, the expected completion (e.g., *white*) elicited the P300 (Figure 5.3). The authors proposed that the P300 indexes "functionally distinct levels of predictive processing via distinct electrophysiological characteristics" (Roehm et al., 2007, p.

1260). Roehm et al. (2007, p. 1272) further argue that the highly expected antonymous adjective (*white*) elicited the P300 component precisely because “the correct identification of the predicted word does not require a lexical search (there is a unique prediction that may either be fulfilled or not)”. However, this P300 effect was not replicated when the antonymous word pairs were presented out of sentence context in a lexical decision task (Experiment 2). This led authors to conclude that the P300 was task dependent.

Figure 5.3. P300 effect observed in the antonymous condition and N400 observed in the nonrelated condition (from Roehm et al., 2007).



The P300 component and how it is modulated by the expectancy factor will be further brought up in Study 3, where on-line processing of frequent multi-word sequences is investigated by means of using the ERP methodology.

5.4. Conclusion

As is clear from the above review, eye-tracking and ERP have been shown to be invaluable techniques in the exploration of frequency and predictability. This makes them particularly useful in the investigation of multi-word speech, which is both frequent and highly predictable. Crucially, the two techniques complement, rather than replicate, each other. Thus, using both eye-tracking and ERP will enable me to provide a clearer picture of on-line processing of multi-word sequences. In what follows, two studies (Study 1 and 2) will be presented that make use of eye-tracking to address the issue of on-line processing of idioms (e.g., *ring a bell*) and binomial expressions (e.g., *bride and groom*) in native and nonnative speakers. Further, a series of experiments (Study 3) will be presented, which use the ERP methodology to address the issue of mental representations of frequent phrases in native speakers. Specifically, they will focus on the P300 and N400 components. Taken together, these studies will provide further evidence in support of the view that multi-word speech is processed differently from novel speech, and that due to their frequency, relative fixedness and high predictability, such phrases are represented in the mental lexicon.

Chapter 6: Processing of Idioms in a First and Second Language: Evidence from Eye-tracking

6.1. Introduction

It has been suggested that multi-word speech is processed and represented in the brain differently from novel language (e.g., Jurafsky, 2003; van Lancker & Kempler, 1987; Wray, 2002). It has also been proposed that due to its high frequency, relative fixedness and limited compositionality, various instances of multi-word speech may be stored in the speaker's long-term memory (e.g., Bybee, 2007; Croft & Cruse, 2004; Jackendoff, 2002; Wray, 2002). A number of studies have addressed the issue of storage and representation of a range of frequent phrases (e.g., Arnon & Snider, 2010; Bannard & Matthews, 2008; Mondini et al., 2002; Sosa & MacFarlane, 2002), and suggest that there may be differences in the way we store, retrieve and produce novel and multi-word speech. At the very least, multi-word sequences seem to enjoy faster processing, require less working memory, and, furthermore, may be represented differently in the brain. In order to better understand how such units are processed in a first and second language, the present chapter will focus on one particular type of multi-word speech – idioms.

6.2. The present study

As is clear from the literature review on idioms (Chapter 3), idiomatic expressions have received a fair amount of attention. One of the reasons why idioms have been widely studied is the availability of two distinct interpretations, figurative and literal.[1] This idiom ambiguity has led to a wealth of research aiming to answer the question of which of the two idiom meanings, figurative or literal, is activated first. Most of the research to date has focussed on idiom processing in a first language; however, more recently, a number of studies have also looked at the processing of idiomatic expressions in a second language. In the present study, I will use an eye-tracking paradigm that approximates natural reading in an experimental setting as far as possible to further investigate how natives and proficient nonnatives process idioms in a highly biasing story context.

Based on previous findings in the literature, I have a set of predictions regarding the processing of idioms used figuratively and literally, and novel phrases. With regards to the native speaker group, it was hypothesised that, first, native participants should show a processing advantage for idioms over novel phrases, as previous research showed that familiar expressions are read faster than novel strings. Second, on a purely frequency-based account, it was further hypothesised that native participants should read idioms more quickly when they are used figuratively than when they are used literally, as idioms' figurative uses are more frequent than literal ones.

The second set of hypotheses regards nonnative speakers. First, if idioms are represented in the lexicon of nonnative speakers in a similar way to how they are represented in the lexicon of native speakers, then they too should be processed more quickly than novel strings. If, however, no processing advantage is found for idioms over novel phrases, this will imply that idioms are less strongly represented in the nonnative lexicon. Second, because L2 learners are likely to have learnt the literal meaning of idioms' components before learning the figurative meaning of the idiom itself, idioms' individual parts and their literal renderings are likely to be more salient and more easily accessible to L2 speakers than idioms' figurative interpretations. If this is the case, a processing advantage should be found for idioms' literal renderings over their figurative counterparts.

6.2.1. Experiment

The purpose of the current study was to investigate on-line processing of idiom figurative and literal uses, as well as matched novel phrases in a biasing story context by native speakers and proficient nonnative speakers of English. Another goal was to explore idiom processing before and after the recognition point (i.e., the point when the expression becomes recognisable as idiomatic). Using an eye-tracking paradigm, eye movements were monitored while participants read a series of short stories presented one by one. The stories contained one of the following types of stimuli: an idiom used figuratively (e.g., *at the end of the day* – 'eventually'), an idiom used literally (e.g., *at the end of the day* – 'in the evening'), or a novel phrase (e.g., *at the end of the war*).

Materials

The idioms used in the study were chosen using the following criteria. First, they had to be frequent English expressions. Second, it was necessary for the idioms to be able to be used figuratively, as well as literally, and sound plausible in both conditions. Third, matched novel phrases should be as close to the idiom (in form) as possible. To do this, the novel phrases had changes of the following types: substitution of function words matched for frequency and length as closely as possible (e.g., *under your nose* and *below your nose*); replacement of one of the content words by another content word matched in frequency and length as closely as possible (e.g., *at the end of the day* and *at the end of the war*); word order change (e.g., *sick and tired* and *tired and sick*). Following the above three criteria, a pool of 53 idioms was selected.

Norming study 1. Since one of the aims of the study was to investigate the way proficient nonnative speakers comprehend idioms, it was essential to make sure that the target idioms were, in fact, known to nonnative participants. It is noteworthy that the majority of the idioms used in this study were frequent word combinations, such as *on the other hand, at the end of the day, sick and tired, a piece of cake*, and so on. Every effort was made not to use rare or unusual idioms, such as *kick the bucket* or *spill the beans*, as they may have not been known by the participants. To ensure that potential participants knew the idioms, a test was compiled with 77 idioms. It was given to a group of 20 nonnative participants who were full-time students at the University of Nottingham. These nonnative speakers met English language requirements prior to commencing their degree (minimum IELTS score of 6.0 or TOEFL score of 550). Of these, 53 idioms were those described above, whereas the remaining 24 were low frequency filler idioms (e.g., *egg on your face*). The participants were asked to indicate how familiar they were with the idioms by rating their knowledge on a four-point scale, ranging from 1 – 'I don't know the idiom', to 4 – 'I know the idiom'. On the basis of the results obtained, 21 idioms with an average rating of 3.5 were selected for the use in the study.

Norming study 2. Previous research has shown that the status of an idiom as decomposable or non-decomposable plays an important role in its processing (e.g., Abel, 2003; Gibbs & Nayak, 1989; Gibbs, Nayak & Cutting 1989; Titone & Connine, 1999, but see Tabossi, Fanari, & Wolf (2009)). Unlike decomposable idioms, non-decomposable idioms' syntactic behaviour is more "frozen" because their individual components do not relate to their figurative meaning. Since the question of idiom decomposability was not directly addressed in this study, it was important to control for this factor. Following the procedure established by Gibbs and Nayak (1989), I asked 14 native speakers (who did not participate in the on-line reading experiment) to judge whether the individual components of the idiom made some unique contribution to the phrase's figurative meaning (the instructions and results can be found in Appendix 1). Out of the 21 idioms, 12 were judged as decomposable and nine as non-decomposable.

Norming study 3. Because of the contention of Cacciari and Tabossi (1988) that idioms have an idiomatic key, one of the aims of the study was to explore idiom processing before and after the idiomatic key, or the recognition point. Once a recognition point was established for the idiomatic expressions, it was then possible to explore whether the number and duration of fixations differed before and after the recognition point for the literal and figurative uses. Because the figurative meanings of the idioms used in the study are more frequent than their literal equivalents, it would seem logical to predict that after the recognition point, there should be fewer and shorter fixations made to the figurative use than to the literal one. Thus, the recognition point analysis should shed light on when an idiom's literal and figurative meanings are activated. The eye-tracking paradigm is ideal for this purpose, as it will allow me to separate fixations made before and after the recognition point. If the fixations made for figurative uses after the recognition point are shorter and fewer than those for literal uses, this will provide evidence that due to their high frequency and predictability, idiom figurative meanings are processed faster than their literal equivalents. If no difference is found, this will support the claim that an idiom's literal and figurative uses are activated simultaneously. If, before the recognition point, only the literal meaning of idiom components is activated, then longer reading times will be expected when such a meaning does not fit with the figurative context.

To determine the point at which the expressions are recognised as idiomatic, five versions of a sentence completion task were created, which included 65 sentence fragments presented out of context, 21 of which were target idioms while the rest were novel distractors. The large number of distractors was meant to prevent participants from noticing the idioms and adopting 'an idiom completion strategy'. Since the aim was to find the point where the expression becomes recognisable as an idiom, the test only included a fragment of each phrase. Thus, for the idiom *leave a bad taste in your mouth*, Version 1 contained the shortest fragment – 'leave'. Version 2 had a slightly longer fragment – 'leave a bad'. Version 3 had longer still – 'leave a bad taste', and so on. The test was given to 50 native speakers of British English (ten people per version) who were asked to complete the phrases. According to McFalls and Schwanenflugel (2002), a sentence is considered to be high constraint if the probability of its expected completion is 70% or more. Therefore, the threshold of 70% was adopted; that is, if seven out of ten people completed the phrase correctly, it was taken to be the recognition point. For example, no participant provided the full idiom having read 'leave' in Version 1. Only one participant completed the idiom having read 'leave a bad' in Version 2. Nine out of ten people completed the idiom correctly after seeing 'leave a bad taste', which was thus taken to be a recognition point. A recognition point for each of the 21 idioms used in the study can be found in Appendix 2. Although the threshold of 70% was adopted, many idioms were completed correctly by more than seven people. It is noteworthy that out of the 21 idioms used in the study, seven did not reach the threshold of 70% and hence were excluded from the recognition point analysis. The mean probability of the remaining 14 items to be completed idiomatically was found to be 86.5% with the completion range being 70% - 100% (see Appendix 2).

As pointed out by Cacciari and Tabossi (1988), figurative meaning activation happens only after the idiomatic key, or the recognition point, has been reached. If the recognition point happens to be at the beginning of the idiom, the figurative meaning will become activated early during idiom processing. However, if the recognition point occurs at the end of the idiom, then the figurative meaning activation will be delayed. Thus, it was important to identify where the recognition point was for the target idioms. The average length of the target idioms was 4.8 words (ranging from 3 to 8 words), and the recognition point was found to be either after the second, third or fourth word (average 2.6). Thus, the recognition point occurred in the middle of the string (2.6 out of 4.8).

The results of the norming studies indicate that the idioms used in the study were easy to use both figuratively and literally, had a recognition point, and were well known to both native and nonnative participants (frequencies given in Appendix 2). It is worth noting that idioms used literally were identical in form to idioms used figuratively. Each of the 21 idioms selected for the experiment was then embedded in a story context. It was deemed necessary to write different story contexts for each of the three stimulus types (examples given in Appendix 3). First, figurative and literal idiom uses have different meanings. Second, they had to be preceded by a biasing story context, which made it impossible to use the same context. Finally, the target idioms differed from novel phrases in the form and meaning. Therefore, different story contexts were written for the three stimulus types.

Participants

Thirty-six native speakers of British English and 36 proficient nonnative speakers took part in the study. None of these participants had taken part in any of the above norming studies. All participants were full-time students at the University of Nottingham. The native participants were given course credit, whereas the nonnatives were paid a small fee. The nonnative participants had learnt English in a classroom setting and came from different language backgrounds. At the time of the experiment, they all had lived in the UK between a few months and a few years. As full-time students at the University of Nottingham, they were required to meet English language requirements prior to commencing their degree (minimum IELTS score of 6.0 or TOEFL score of 550). Their self-rating of English language proficiency is summarised in Table 6.1. All participants had normal or corrected-to-normal vision.

Table 6.1. English language proficiency for nonnative speakers (Means), n. = 36.

Age	Time in UK	1st exposure	Speaking [a]	Reading [a]	Writing [a]	Comprehension [a]
22.5	20 months	7 yrs	3.7	4.1	3.7	3.9

[a] Self-rating task: 1– very poor; 2–weak; 3–ok; 4–good; 5–excellent.

Apparatus and procedure

Stories were presented across three presentation lists. Each list contained 21 items: seven idioms used figuratively, seven idioms used literally, and seven novel phrases. It is noteworthy that decomposable and non-decomposable idioms were evenly distributed across the three presentation lists (each list contained exactly four decomposable and three non-decomposable idioms). We also ensured that no participant saw more than one version of the same phrase.

The participants were asked to read the stories quickly but for comprehension and were advised that each story would be followed by a comprehension question. Following this, a nine-point grid calibration procedure was completed. The first three trials were always practice trials. The eye-tracker was calibrated at least four times during the experiment. The stories were presented in a pseudorandomised order. Before each trial, a fixation point appeared in the middle of the screen. After participants fixated it and a calibration check was done, a story appeared on the screen. Once participants finished reading each story, they pressed a key to proceed to the comprehension question. Eye movements were monitored using an EyeLink I eye-tracker.

After the experiment, the nonnative participants were asked to rate their knowledge of the 21 idioms on a four-point scale, resulting in the same familiarity rating as in the norming study described above.

Analysis and results

Prior to the analysis, all trials where track loss occurred were removed. The missing data accounted for 0.2% of the total data and were equally distributed across the conditions. The participants had no difficulty answering comprehension questions, with an overall accuracy rate of 91.4% for native, and 90.4% for nonnative speakers. One nonnative participant was excluded from the analysis due to a high number of incorrect answers. For each target, the following measures were examined (see Chapter 5 for a more detailed description of the eye-tracking measures):

- *First pass reading time* – the sum of all fixation durations made within a region of interest before exiting the region either to the left or to the right.

- *Total reading time* – the sum of all fixation durations made within a region of interest.
- *Fixation count* – the number of all fixations made within a region of interest.

It is generally assumed that early measures (first pass reading time) are sensitive to early processes in the comprehension of a text, such as early integration of information. Late measures (total reading time and fixation count) are believed to be sensitive to later processes associated with comprehension of a text, such as information re-analysis and discourse integration (e.g., Paterson, Liversedge, & Underwood, 1999; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989).

The data for native and nonnative speakers were analysed in two different ways: analysis of the entire phrase (full idiom analysis), and analysis with regards to the idiom's recognition point (analysis before and after the recognition point).

Full idiom analysis

In the full idiom analysis, I looked at an idiom's figurative and literal uses, as well as novel phrases. The data were analysed using repeated measures ANOVAs treating participants and items as random variables. The dependent variables were mean total reading time, first pass reading time, and fixation count, which can be found in Table 6.2 for native and nonnative speakers. Statistical comparisons for the two participant groups are illustrated in Tables 6.3 and 6.4.

Table 6.2. Native and nonnative fixation durations (in milliseconds) and fixation count in the full idiom analysis with Standard Error (SE) in parenthesis.

	First Pass Reading Time		
	Figurative	**Literal**	**Novel**
Natives	447 (25)	454 (21)	497 (30)
Nonnatives	743 (44)	705 (38)	720 (42)
	Total Reading Time		
	Figurative	**Literal**	**Novel**
Natives	514 (32)	507 (25)	628 (37)
Nonnatives	937 (52)	817 (37)	880 (44)
	Fixation Count		
	Figurative	**Literal**	**Novel**
Natives	2.8 (0.2)	2.7 (0.1)	3.2 (0.1)
Nonnatives	4.2 (0.1)	3.7 (0.1)	3.9 (0.1)

Table 6.3. Analyses of Variance and planned comparisons with participants (F_1) and items (F_2) as random variables for native speakers in the full idioms analysis.

	by participants			by items		
	df	F_1	*p*	*df*	F_2	*p*
First Pass Reading Time						
Phrase Type	2,70	2.0	ns	2,40	1.4	ns
Total Reading Time						
Phrase Type	2,70	9.6	**	2,40	6.3	**
figurative vs. novel	1,35	14.1	**	1,20	8.0	*
literal vs. novel	1,35	12.8	**	1,20	10.1	**
figurative vs. literal	1,35	.05	ns	1,20	.032	ns
Fixation Count						
Phrase Type	2,70	5.6	**	2,40	3.5	*
figurative vs. novel	1,35	8.2	*	1,20	4.3	*
literal vs. novel	1,35	8.8	**	1,20	6.9	*
figurative vs. literal	1,35	.16	ns	1,20	.09	ns

* significant at $p \leq .05$, ** significant at $p \leq .005$, ns – non-significant

Table 6.4. Analyses of Variance and planned comparisons with participants (F_1) and items (F_2) as random variables for nonnative speakers in the full idioms analysis.

	by participants			by items		
	df	F_1	*p*	*df*	F_2	*p*
First Pass Reading Time						
Phrase Type	2,68	.45	ns	2,40	.29	ns
Total Reading Time						
Phrase Type	2,68	3.5	*	2,40	3.4	*
figurative vs. novel	1,34	1.3	ns	1,20	0.7	ns
literal vs. novel	1,34	2.4	ns	1,20	3.8	ns
figurative vs. literal	1,34	6.7	*	1,20	8.4	*
Fixation Count						
Phrase Type	2,68	4.2	*	2,40	3.5	*
figurative vs. novel	1,34	2.3	ns	1,20	1.2	ns
literal vs. novel	1,34	2.0	ns	1,20	1.0	ns
figurative vs. literal	1,34	8.0	*	1,20	11.8	**

* significant at $p \leq .05$, ** significant at $p \leq .005$, ns – non-significant

Native speakers

There was no effect of Phrase Type in the early measure, first pass reading time, and thus, no planned comparisons were conducted. There was a significant main effect of Phrase Type in total reading time and fixation count. Planned comparisons for these measures, revealed that idioms used figuratively and literally were read significantly faster and elicited fewer fixations than novel phrases. No significant difference was found in the figurative versus literal comparison in either of the two late measures.

Overall, these results indicate a processing advantage for idiomatic expressions over novel strings. More importantly, they show that an idiom's two meanings, literal and figurative, were processed with a similar speed. The latter finding is particularly robust, given that none of the measures showed any processing advantage for one meaning over the other. One interesting finding is that the difference between idioms and novel language emerged relatively late. This might be because when reading longer strings of text, such as idioms, early measures may not be sensitive to potential differences. I will come back to this finding in the general discussion.

Nonnative speakers

There was no significant main effect of Phrase Type in the first pass reading time analysis (Table 6.4). Because no significant main effect was found, no further comparisons were conducted. However, in the total reading time and fixation count analyses, a significant main effect of Phrase Type was observed across participants and items. Planned comparisons revealed no differences in figurative versus novel or literal versus novel processing, suggesting that both meanings of idioms were read with the same speed as novel language. More importantly, planned comparisons showed that an idiom's figurative meaning was processed significantly slower than the literal one even though it was supported by the context.

The above results indicate that for nonnative speakers, in contrast to the native group, novel phrases are not processed any slower than figurative or literal uses of idioms. Further, there is clear evidence that the figurative meaning of an idiom is processed more slowly than the literal one. Interestingly, this difference was observed in the late but not early measures (again, I will come back to this in the general discussion).

Comparing native and nonnative speaker reading times

In order to assess the role of proficiency on idiom processing more directly, further ANOVAs were conducted on combined native and nonnative data with Participant Proficiency as a between factor. Overall nonnative speakers took significantly more time to read all three types of stimuli, which was evidenced by a highly significant main effect of Proficiency in both early and late measures ($F1s$ & $F2s$ $p < .005$). However, it is hardly surprising that native speakers are faster readers than nonnative speakers. Importantly, a significant main effect of Phrase Type, as well as a significant interaction between Phrase Type and Proficiency were observed in the two late measures ($F1s$ & $F2s$ $p < .05$). This suggests that not only are nonnative speakers overall slower than natives, but that the nature of their processing differs. Namely, where native speakers tend to slow down (reading novel strings compared to idioms), nonnative speakers do not. On the other hand, where nonnatives show a significant processing cost (figurative renderings vs. literal ones), natives do not.

Recognition point analysis

The aim of the recognition point analysis was to investigate the figurative and literal meaning processing in natives and nonnatives before and after recognition point. With respect to nonnative speakers, there was a secondary goal. In the full idiom analysis, it was established that these participants slow down when reading idioms' figurative meanings. However, where exactly this slow-down happens was unclear. Thus, another aim of the recognition point analysis was to establish whether nonnatives' slow-down occurred before or after the recognition point.

The analyses were performed separately on two idiom portions, before and after the recognition point. Because the novel phrases had no recognition point, they were not included in the recognition point analysis.[2] Thus, the analyses reported below include two types of stimuli: idioms used figuratively and literally.[3]

The data were analysed using repeated measures ANOVAs. The mean total reading time, first pass reading time, and fixation count for idioms used figuratively and literally before and after the recognition point for both groups of participants are given in Table 6.5. Statistical comparisons for native and nonnative speakers before and after the recognition point are illustrated in Tables 6.6 and 6.7, respectively.

Table 6.5. Native and nonnative fixation durations (in milliseconds) and fixation count before and after the recognition point with Standard Error (SE) in parenthesis.

	First Pass Reading Time			
	Before Recognition Point		***After Recognition Point***	
	Figurative	**Literal**	**Figurative**	**Literal**
Natives	284 (19)	270 (13)	220 (8.6)	221 (11)
Nonnatives	413 (27)	399 (22)	385 (20)	375 (15)
	Total Reading Time			
	Figurative	**Literal**	**Figurative**	**Literal**
Natives	299 (22)	299 (15)	200 (12)	214 (13)
Nonnatives	526 (38)	444 (22)	424 (27)	371 (17)
	Fixation Count			
	Figurative	**Literal**	**Figurative**	**Literal**
Natives	1.6 (.10)	1.5 (.07)	1.1 (.06)	1.2 (.05)
Nonnatives	2.3 (.12)	2.0 (.09)	1.9 (.11)	1.7 (.06)

Table 6.6. Analyses of Variance and planned comparisons with participants (F_1) and items (F_2) as random variables for native speakers in the recognition point analysis.

	by participants			by items		
	df	F_1	*p*	*df*	F_2	*p*
First Pass Reading Time						
Before the recognition point						
figurative vs. literal	1,35	.492	ns	1,13	.832	ns
After the recognition point						
figurative vs. literal	1,35	.003	ns	1,13	.048	ns
Total Reading Time						
Before the recognition point						
figurative vs. literal	1,35	.000	ns	1,13	.178	ns
After the recognition point						
figurative vs. literal	1,35	.952	ns	1,13	.404	ns
Fixation Count						
Before the recognition point						
figurative vs. literal	1,35	.558	ns	1,13	.627	ns
After the recognition point						
figurative vs. literal	1,35	.744	ns	1,13	.267	ns

ns – non-significant

Table 6.7. Analyses of Variance and planned comparisons with participants (F_1) and items (F_2) as random variables for nonnative speakers in the recognition point analysis.

	by participants			by items		
	df	F_1	*p*	*df*	F_2	*p*
First Pass Reading Time						
Before the recognition point						
figurative vs. literal	1,34	.243	ns	1,13	.485	ns
After the recognition point						
figurative vs. literal	1,34	.230	ns	1,13	.007	ns
Total Reading Time						
Before the recognition point						
figurative vs. literal	1,34	5.9	*	1,13	6.4	*
After the recognition point						
figurative vs. literal	1,34	5.2	*	1,13	2.5	ns
Fixation Count						
Before the recognition point						
figurative vs. literal	1,34	6.8	*	1,13	6.4	*
After the recognition point						
figurative vs. literal	1,34	3.5	=.07	1,13	1.1	ns

* significant at $p \leq .05$, ns – non-significant

Native speakers

None of the three measures that were analysed showed any processing differences in figurative versus literal idiom interpretations before or after the recognition point. This finding does not support the proposition of Cacciari and Tabossi (1988). However, this is not surprising given that in the current study, the context biased the reader to the upcoming idiom interpretation, while Cacciari and Tabossi's theory makes predictions for idioms *in the absence* of a biasing context.

Overall, it is clear that the native speaker group processed the two idiom meanings in a very similar way. In this, these findings replicate those obtained in the full idiom analysis discussed above. Crucially, it was observed that there was no speed-up for the figurative rendering after the recognition point, where the difference was most likely to occur.

Nonnative speakers

Similar to the full idiom analysis, the early measure revealed no reliable differences before or after the recognition point for figurative and literal uses of idioms. Before the recognition point, both late measures showed that figurative uses were read reliably slower than literal ones (Table 6.7). After the recognition point, the only significant difference observed was that for the total reading time measure in the analysis by participants but not items. Fixation count data suggested marginally significant differences between the two idiom meanings in the analysis by participants but not items.

Taken together, the nonnative speaker recognition point results confirmed what was suggested previously in the analysis of the entire idiom. Namely, the idiom's figurative meaning incurs a significant processing cost when compared to its literal equivalent. Importantly, with the help of the recognition point analysis, it became possible to establish where exactly nonnative speakers slow down when encountering an idiom in a story context. Both late measures strongly suggest that nonnative speakers make more and longer fixations when reading an idiom's figurative meaning *before* the recognition point.

Comparing native and nonnative speaker reading times

In order to explore the role of participant proficiency in a more direct way, further ANOVAs with Participant Proficiency as a between factor were conducted on combined native and nonnative data. This analysis showed that, for all three measures, native speaker processing was significantly faster than that for nonnatives before and after the recognition point, as evidenced by a significant main effect of Proficiency ($F1s$ & $F2s$ $p < .005$). Again, it is hardly surprising that native speakers read more quickly than nonnative speakers. What is of greater interest is that before the recognition point, both late measures showed a significantly different processing pattern for the two participant groups ($F1s$ & $F2s$ $p < .05$). Specifically, the literal versus figurative meaning contrast for natives was significantly different from that for nonnatives. Nonnatives were more likely to read figurative meanings more slowly than literal ones. The early measure analysis showed a comparable pattern of idiom processing for both participant groups ($F1s$ $F2s$ $p > .05$). After the recognition point, no reliable differences were found in any of the three measures ($F1s$ & $F2s$ $p > .05$).

6.3. General discussion

In this study, I looked at how native and proficient nonnative speakers process idioms in a story context that encouraged either a figurative or literal interpretation. This was compared to the processing of matched novel phrases. The study had four aims. First, in native speakers, the goal was to confirm previous findings that idioms are processed faster than matched novel phrases. The second aim was to explore whether there are any processing differences between figurative and literal idiom renderings encountered in a biasing story context. The third goal was to compare the processing of idioms' literal and figurative meanings before and after the recognition point. Finally, I aimed to compare idiom comprehension in a first and second language.

Although in native speakers, no significant differences were observed in the early measure, a processing advantage was found for idioms like *at the end of the day* over novel phrases such as *at the end of the war* in the two late measures. This indicates that compared to novel phrases, idiomatic expressions are read faster and require less rereading and re-analysis. Although this finding is highly compatible with the existing research (e.g., Gibbs, 1980; Gibbs & Gonzales, 1985; Swinnery & Cutler, 1979; Tabossi, Fanari, & Wolf, 2009), it is not entirely clear why idioms are read faster than their controls. In a recent study, Tabossi, Fanari, and Wolf (2009) investigated the processing of familiar expressions (decomposable idioms, non-decomposable idioms, and compositional clichés) versus novel control phrases. It was shown that all familiar expressions were recognised more quickly than their controls. The authors proposed that familiarity of these expressions, more than anything else (e.g., holistic representation, compositionality, idiomaticity, or predictability), was likely to explain their fast recognition. Although the present study did not aim to explore the *reasons* behind the idiom fast recognition, I am inclined to agree with Tabossi, Fanari, and Wolf's standpoint.

Much of the idiom research in the past has focused on how the figurative and literal meanings available in ambiguous idioms are activated in relation to each other. For native speakers, none of the measures, early or late, showed a processing advantage for figurative idiom uses over their literal equivalents. This suggests that the preceding disambiguating context was sufficient to resolve the ambiguity that may have arisen during the processing of idioms that have both literal and figurative interpretations.

Because none of the theories of idiom processing make specific claims about the effect of a biasing context, it is difficult to compare the current study with the existing models. The fact that for native speakers, no differences were observed in terms of fixation durations or fixation count for the literal and figurative meanings indicates that in a biasing context, the activation of both meanings occurs comparably quickly. As this pattern of activation is in part driven by the presence of a preceding disambiguating context, further study is needed to investigate how idiom activation is modulated by the presence or absence of the disambiguating region. It is possible that when disambiguating context is not provided, the figurative meaning will get activated prior to the literal one because it is more frequent.

Because the figurative meaning of an idiom is of a higher frequency than the literal one (the context being equally biasing), it seemed logical to expect a processing advantage for the figurative use before the recognition point is reached. After the recognition point (i.e., after the expression has been recognised as idiomatic), it seemed even more probable that the figurative interpretation might be read more quickly than the literal one. However, the results clearly indicated that before, as well as after the recognition point, both idiom uses were read with the same speed by the native participants.

With respect to contextual constraints, one particular study is of relevance. Colombo (1993) investigated the role of context in the activation of figurative and literal idiom meanings. In a series of lexical decision tasks, it was found that the idiomatic meaning of an ambiguous idiom became activated *only* following the context that biased the figurative interpretation. In the absence of a figuratively-biasing context (i.e., when a neutral or literally-biasing context was provided), only literal computations were observed. It is important to note, however, that the two meanings of the idioms used in Colombo's study were equally frequent. Thus, when the frequency factor is eliminated, it is ultimately the context that determines which of the two meanings will be activated. If, similar to the present study, idioms with the dominant figurative meaning were used, it is possible that a different pattern of results would have been observed in the absence of a figuratively-biasing context. In future work, I will manipulate the relative frequency of the two idiom meanings, as well as the context (figurative, literal, or neutral) in order to obtain a better picture of the role that these two factors may play in on-line idiom comprehension.

Due to a range of findings in the literature on nonnative speakers, it was unclear whether, like natives, they would too process idioms faster than novel language. What was found was that, unlike native speakers, the nonnative group's processing of idioms and novel phrases was very similar. Both early and late measures showed that idioms were processed with the same speed as novel phrases; no significant differences were found in the figurative versus novel, or literal versus novel comparisons. This is suggestive of the fact that idioms are not represented in the mental lexicon of a nonnative speaker in the same way they are represented in the lexicon of a native speaker. The nonnative results are in contrast with those of Conklin and Schmitt (2008) who found that both figurative and literal meanings had a robust processing advantage over novel phrases. However, as was mentioned in the idiom literature review in Chapter 3, the difference in the results between the two studies may be due to problems with Conklin and Schmitt's experimental design. The nonnative speaker results seem to be in agreement with those reported in Underwood, Schmitt, and Galpin (2004). Similar to the present study, they did not observe any processing advantage for idioms over novel phrases for nonnative speakers. However, because Underwood, Schmitt, and Galpin only measured reading times for the terminal word of idioms and novel phrases, their results are not directly comparable with those reported in the present study, where reading times of the entire phrase were analysed.

As has been discussed throughout, one of the key issues in idiom processing is when the two meanings of idioms are activated relative to each other. Second language learners are more likely to learn literal meanings of idioms' constituents before learning idioms' overall figurative meanings. Thus, one might expect a processing advantage for literal uses over their figurative counterparts, even though the literal uses are less frequent. The eye-tracking measures reveal that in nonnative speakers, figurative meanings required more rereading and re-analysis than literal ones. These findings are in line with those reported in Cieslicka (2006), who observed that literal meanings were activated prior to figurative ones by her nonnative participants. The current findings also support those of Matlock and Hereida (2002), who looked at the processing of phrasal verbs with a figurative meaning (e.g., *Paul went over the exam with his students*) versus identical verb-preposition combinations used literally (e.g., *Paul <u>went over</u> the bridge with his bicycle*). Matlock and Hereida found that native speakers accessed idiomatic phrasal verbs more quickly than identical verb-preposition combinations used literally. For the nonnative group, on the other hand, no differences were observed in reading times for phrasal verbs used figuratively versus verb-preposition combinations used literally.

The main rationale behind the recognition point analysis was to find out where, in the course of idiom comprehension, the processing cost associated with the figurative meaning is greatest for nonnative speakers. Both late measures showed that nonnative speakers spent significantly more time reading the figurative meaning of an idiom than the literal one before the recognition point. After the recognition point, the total reading time and fixation count measures revealed a significant and marginally significant processing cost, respectively, for the figurative meaning. However, because this was observed in the analysis by participants but not items, further work is needed to ascertain if the figurative meaning continues to cause processing difficulty after the recognition point has been reached.

It is important to consider why nonnative speakers require more processing effort when reading the figurative meaning of idioms even in the presence of a biasing context. Researchers agree that a fundamental task in second language vocabulary acquisition is building connections between a form and meaning (e.g., Schmitt, 2008; van Patten, Williams & Rott, 2004). If the nonnative speakers have not yet developed strong form-meaning connections between an idiom and its figurative meaning, they will not show the same pattern of idiom processing as the native speaker group. Let us consider the idiom *at the end of the day*. The nonnative speakers have connections between the individual lexical items and their meanings, and norming showed that they knew that these items occur together in an idiomatic phrase *at the end of the day*. The finding of slow reading times for *at the end of the day* when used figuratively suggests that the link between the idiom and the meaning 'eventually' is not as strong as the link between the form and the meaning of the individual lexical items. As a result, the meaning 'eventually' is not activated as quickly as the meaning 'in the evening'. Thus, the meaning with the highest level of activation is the incorrect one in a context where the figurative meaning is the appropriate interpretation.

In order to better understand the overall pattern of results observed in the current study, one need to consider three factors: frequency, predictability, and context. I will look at each of these in turn. First, an idiom's figurative meaning is almost always more frequent than its literal counterpart. As evidenced by the BNC, all of the idioms used in the present study appear figuratively much more frequently than they do literally (approx. 83% vs. 17%).[4] Researchers have suggested that (at least in native speakers) idioms are more readily understood figuratively than literally because they occur figuratively with much higher frequency in everyday discourse (e.g., Gibbs, 1986; Popiel & MacRae, 1988; van Lancker-Sidtis, 2003). Thus, based on a frequency account alone, processing should be faster for the figurative use of an idiom than for its literal counterpart. However, the results of the full idiom analysis, as well as the recognition point analysis showed that this was not the case.

The second factor that may be implicated is predictability. Idioms, or at the minimum the words after the recognition point, can be considered to be highly predictable, as indicated by their high cloze probability in Norming Study 3. What this means is that readers can predict *day* after having seen *at the end of the*. On a predictability account, idioms used both figuratively and literally should be processed faster than novel language. However, because an idiom's completion is equally predictable in both literal and figurative phrases, this factor cannot be used to hypothesise which meaning of an idiom should activated more quickly.

It is clear from the above discussion that frequency and predictability cannot satisfactorily account for the pattern of results observed in the study and therefore a third factor – context – may be implicated. A number of word recognition studies have shown that if a preceding context creates strong enough expectancies, then the processing of the low frequency form of a word that has multiple interpretations may be processed equally as fast as its high frequency equivalent (e.g., Martin, et al., 1999; Vu, Kellas, & Paul, 1998). In a self-paced reading task, Martin et al. (1999) showed that in the presence of a strongly biasing context, reading times for the less frequent meaning of a homophone (e.g., *bulb* – 'the root of a plant') did not differ from those of the more frequent meaning (e.g., *bulb* – 'light bulb'). In the absence of a strongly biasing context, the more frequent meanings were read faster than less frequent ones. Because the story contexts used in the present study biased readers towards either literal or figurative renderings, the results obtained appear to suggest that in native speakers, if a preceding context is strong enough it facilitates the processing of the less frequent literal form. Nonnative speakers, on the other hand, were not conferred a similar bias for interpreting the figurative meaning of an idiom.

Finally, one last issue merits attention. Although it is apparent that the patterns of idiom activation in native and nonnative speakers are rather different, there appears to be one thing that the two groups have in common – the absence of significant differences in the early measure (when the differences *are* significant in the late measures). As such, this finding has important implications in terms of the nature of eye-tracking measures and their significance for (long) multi-word sequences, such as idioms. It appears that, upon initial reading (i.e., during the first pass reading time), figurative and literal idiom uses, as well as novel strings are all read in a comparable way. However, there seems to be a need for a reader to exit the region of interest (to the left or to the right) and then come back to it, resulting in significant differences across conditions in late measures, because some items require longer re-reading and/or re-analysis than others. It is thus possible that when reading longer strings of language (as opposed to single words or shorter multi-word units (the average length of the idioms was 4.8 words)), early measures may not be particularly sensitive to potential differences. Previous research (e.g., Hyona, 1993; Rayner & Well, 1996) showed that length manipulations for individual words affect early measures differently from late ones. Unfortunately, the present study does not allow me to be more certain or specific with regards to the differences between early and late measures. However, the fact that this trend was apparent in both native and nonnative speakers (whose idiom comprehension was otherwise found to be rather distinct) does suggest that early eye-tracking measures may not be adequate for investigating long multi-word sequences.

To conclude, the analyses of figurative and literal idiom uses, as well as novel phrases revealed a number of findings. First, proficient nonnative speakers do not process idioms more quickly than novel phrases. This suggests that idioms are less strongly represented in the mental lexicon of nonnative speakers than in that of native speakers. Crucially, nonnatives require more time to retrieve figurative senses of idioms than literal ones, even when the context biases the reader towards the figurative interpretation. This slow-down was largely observed before the recognition point. With respect to native speakers, the present study further confirmed previous findings that idioms are read faster than novel language. Finally, in the presence of a preceding disambiguating context, native speakers do not process the low frequency literal meaning of an idiom any differently from the high frequency figurative one.

Chapter 7: Processing and Representation of Binomial Expressions in a First and Second Language: Evidence from Eye-tracking

7.1. Introduction

There is widespread agreement that words are encoded in our mental lexicon. An open question is whether units larger than a word can also be represented in the lexicon. Research on lexical storage has for the most part disregarded phrases on grounds that they are necessarily derived via general rules from individual words. That is, the meaning of a sentence, such as *I play football* or a phrase, such as *bride and groom* can be derived from the individual words that compose them. Such a view is supported by the observation that encoding every possible utterance one has ever heard is clearly not feasible. However, a handful of recent psycholinguistic studies report reduced processing loads for very frequent phrases (e.g., Arnon & Snider, 2010; Bannard & Matthews, 2008; Sosa & McFarlane, 2002; Van Lancker & Kempler, 1987). Such results support the view that frequently used phrases may be represented in the mental lexicon (along with single words), and thus one might conclude that, by analogy to Hebb's (1949) law of neural plasticity, *words used together wire together.*

If it is the case that frequency of exposure plays an important role in what is *wired together*, or represented, in the mental lexicon, one would expect that native English speakers, who have accumulated a sufficient amount of experience with frequent phrases, will show a processing advantage for them. In contrast, nonnative speakers, who will have had less exposure to English, may exhibit a *lexicon in transition*. In other words, only the most frequently occurring expressions may be represented in their mental lexicon. Thus, the inclusion of nonnative speakers in the study will allow a researcher to more explicitly investigate the role of frequency of occurrence to establish how frequent a phrase has to be for it to be processed differently from a matched novel one.

The current investigation explores the issue of multi-word sequence processing and representation by native and proficient nonnative speakers by looking at one particular type of multi-word speech – binomial expressions (e.g., *bride and groom*). Binomial expressions are ideal for studying frequent phrase comprehension for a number of reasons. First, they are much more frequent and ubiquitous than idioms. Second, unlike idioms, binomial expressions are transparent; that is, their individual components contribute fully and overtly to the overall meaning of the expression.[5] While readers cannot compute the meaning of the idiom *ring a bell* ('sound familiar'), they can compute the meaning of the binomial *bride and groom*. Finally, in more idiosyncratic expressions, such as idioms, changes are rarely permitted without the expression losing its figurative meaning. Thus, *kick the bucket* is no longer considered to be a figurative expression, if it is changed to *the bucket was kicked*. Because the word order in binomial expressions can be reversed without any meaning change (*bride and groom* means the same as *groom and bride*), it will be possible to investigate whether such 'fixed' expressions have a processing advantage over matched reversed forms, which only differ in phrasal frequency.

7.2. Frequent phrases and the mental lexicon

According to Van Lancker (1988, 1990) and Van Lancker-Sidtis (2003), language ranges from completely novel at one extreme to highly familiar at the other. Newly-generated propositional speech entails the use of a wide range of grammatical and lexical rules. Familiar speech, on the other hand, is highly conventional and relatively fixed. It does not need to be produced *de novo* every time it is used, and thus syntactic and lexical rules are not required to the same extent. Idioms, which are an example of conventional language, are often hypothesised to be lexicalised units represented in long-term memory (e.g., Bybee, 2006, 2007; Croft & Cruse, 2004; Jackendoff, 1995, 2002). Thus, the lexicon of a native speaker may encompass morphemes and single words at one end, and highly idiosyncratic items, like idioms, at the other. This proposition would be in line with a *usage-based model* of lexical storage and processing (e.g., Bybee, 1985, 1995; Croft, 2001; Goldberg, 1995, 2006; Langacker, 1987; Tomasello, 2003, 2006). According to a usage-based model, what is represented in the mental lexicon is determined solely by language use. Each time a particular word or a linguistic structure is used, it activates a pattern of nodes in the lexicon, and the frequency of activation of this word or phrase affects the representation of this information, which eventually results in its representation as a conventional unit (Croft & Cruse, 2004). In this view, the lexicon of a mature speaker consists of thousands of multi-word conventional expressions. Although they may appear to be structurally complex, such units constitute for a native speaker a 'pre-packaged' assembly. Importantly, with its focus on utterances and phrases, not isolated words and morphemes, a usage-based model postulates that there are no restrictions as to what can, or cannot be stored in the lexicon – a morpheme, a word, or a multi-word unit.

Outside of the domain of idioms, only a small number of studies have addressed the issue of processing and representation of units above the word level in the lexicon of a native speaker. One example of this is a study by Sosa and MacFarlane (2002), which uses an auditory word-monitoring task for the function word *of* in two-word collocations varying in frequency (e.g., *sort of* and *kind of*). They found that reaction times to *of* in higher frequency combinations were significantly slower than in lower frequency ones, indicating that very frequent combinations were treated as units. Sosa and MacFarlane maintain that there is no direct access to *of* when the stimulus is a high-frequency multi-word sequence because it is stored as a unit without links to the constituent parts. Further, the number of correct responses was very low for high frequency collocations (37%) when compared to low frequency ones (60%). Sosa and MacFarlane argue that their results indicate that when phrases are used frequently, they become chunked and may subsequently be stored as a unit. One downside of the study, however, was that the frequency and length of the words within target phrases were not matched, which may have affected the pattern of results.

Bod (2000, 2001) tested the hypothesis that frequently occurring compositional sentences are stored in long-term memory. Bod's participants read high frequency three-word SVO sentences (e.g., *I like it*) and low frequency control sentences (e.g., *I test it*), whose individual components were matched in lexical frequency and length. Participants responded faster to high-frequency sentences than to low frequency ones. According to Bod, these results suggest that frequent sentences may also be represented in long-term memory. However, it is possible that the processing cost found for the less frequent sentences was due to these phrases being less natural.

Because of the tense and aspect of some of the low frequency experimental items, they may have sounded less natural to the participants than high frequency ones.

In a similar study, Arnon and Snider (2010) investigated the role of frequency in the comprehension of compositional four-word phrases (e.g., *don't have to worry*). They compared reading times for phrases, which differed in phrasal frequency but whose individual components were controlled for length and frequency. They found that the more frequent phrases were processed reliably faster than the less frequent ones. The authors concluded that language users appear to notice, learn, and subsequently store frequency information not only about words, but also with regards to multi-word phrases, even when they are entirely compositional. Although informative with respect to the role of phrasal frequency, Arnon and Snider's study is limited to highly compositional phrases that are rather different from highly familiar fixed or semi-fixed multi-word expressions, such as frequent collocations and compounds discussed above, or binomial expressions and idioms that the present thesis focuses on.

Mondini et al. (2002) investigated the processing of two-word compounds of the type Adj + N and N + Adj (e.g., *natura morta* 'still life') and matched novel combinations (e.g., *natura bella* 'beautiful nature') by two non-fluent aphasic patients. In Italian, adjectives agree with the grammatical gender of the noun in both compounds and novel combinations. Mondini and colleagues hypothesised that if compounds are represented in the brain as unit, the participants should have difficulty making noun-adjective agreement for novel combinations, but not compounds. They found that both participants performed significantly better on compounds than on novel noun-adjective combinations. This suggests that for novel combinations the participants retrieved the adjective and noun separately and then applied agreement rules. Compounds, on the other hand, were retrieved as wholes and, therefore, no morphosyntactic operations were necessary. Interestingly, one of the participants was also able to repeat compounds significantly more accurately than non-compounds. According to the authors, this implies that compounds require less working memory than novel language. Such results suggest that compounds may be stored and processed as wholes, rather than computed on-line word-by-word. Because the study only investigated two brain-damaged participants, it is difficult to draw any far-reaching conclusions.

Finally, as mentioned in Chapter 2, a number of production studies suggest phonological differences in the production of frequent multi-word sequences versus novel ones. The overall finding is that words within frequent utterances tend to be phonologically reduced compared to words within novel phrases (e.g., Bell et al., 2003; Bybee & Scheibman, 1999; Van Lancker, Canter, & Terbeek, 1981).

Thus far, it has been proposed that frequent phrases are processed (comprehended and produced) differently from less frequent ones. If frequency affects the way language is processed, then frequency information must be represented somewhere and the language processor must register each and every occurrence of a particular linguistic event (e.g., Bod, Hay, & Jannedy, 2003). In other words, throughout their lifespan, language users must notice, accumulate, and use this frequency information. If this is the case, then frequency effects should be observable in native speakers. Nonnative speakers, whose exposure to a second language is not as rich as that of adult native speakers, may also exhibit sensitivity to frequent linguistic patterns. However, this might be mediated by the frequency of the expression and how much exposure a speaker has had to the language.

With respect to first language acquisition, it has been proposed that children learn not only single words, but also longer stretches of language, which differ in shape, size, and degree of abstraction (e.g., *where the bottle*) (e.g., Tomasello, 2003). As was mentioned in Chapter 4, the use of such multi-word utterances is believed to contribute to the development of the adult-like morphosyntax (e.g., Clark, 1974; Cruttenden, 1981; Lieven, Pine, & Barnes, 1992; Locke, 1997; Peters, 1977). One study in particular tested experimentally whether children were able to store and reuse sequences whose individual components were already known to the participants. Bannard and Matthews (2008) compared children's production of phrases that differed in the frequency with which they appeared in the child-directed speech (e.g., *a drink of milk* vs. *a drink of tea*). Two and three-year-old children were found to be reliably faster and more accurate at repeating higher frequency phrases than lower frequency ones. Bannard and Matthews concluded that frequent multi-word utterances, such as, *a drink of milk*, are stored in young children's lexicon. This shows that children as young as two are sensitive to the frequency with which multi-word strings occur in their input.

As the above review suggests, phrasal frequency effects have been shown to manifest themselves in studies with adult native speakers and first language learners. However, very little evidence exists with respect to the role of phrasal frequency in second language processing. Nonnative speakers, who start learning a foreign language, will not have any multi-word phrases in their mental lexicon. However, as they become more proficient and have more exposure to the language, they will have not only single words in their lexicon, but also instances of frequent multi-word sequences. If it *is* the case that frequency of exposure determines what is represented in the mental lexicon, one should expect native speakers, who have accumulated a sufficient amount of experience with frequent expressions, to show a robust processing advantage for them. Nonnative speakers, who will have had significantly less exposure to English, may show a processing advantage only for the most frequently occurring phrases. Thus, the inclusion of nonnative participants in the present study will allow me to investigate the scale of frequency of exposure in order to establish *how* frequent a multi-word sequence has to be for it to be processed differently from a matched novel one. Second language learners are likely to have a lexicon in transition, and are ideal candidates to explore the role of phrasal frequency.

Thus, the aim of the current investigation is to shed more light on the issue of multi-word speech processing and representation in native and proficient nonnative speakers by looking at binomial expressions, such as *bride and groom.*

7.2.1. Binomial expressions

Of all the types of multi-word speech, idioms have by far received most attention in the psycholinguistic literature (e.g., Bobrow & Bell, 1973; Cacciari & Tabossi, 1988; Gibbs & Gonzales, 1985; Gibbs & Nayak, 1989; Swinney & Cutler, 1979; Titone & Connine, 1999). Other types, such as binomials, which are far more frequent in English, have received little or no attention. I define binomials as recurrent (i.e., frequent), familiar (i.e., conventional) expressions formed by two words from the same lexical class connected by a conjunction, where one word order is always more frequent and considered more acceptable than the other. For the purpose of the study, novel word combinations with no word-order preference (e.g., *green and yellow* and *tired and bored*) are not considered to be binomial expressions. Binomials come in a variety of forms. A small number of such expressions are *trinomials* (e.g., *cool, calm and collected*). Some binomials contain words that in Modern English can only be used within a particular binomial, and never on their own (e.g., *kith and kin*). Neither trinomials, nor expressions like *kith and kin* are part of the current investigation. Most binomials only have a literal meaning (just like any other novel word combination), while some, similar to ambiguous idioms, can be used both literally and figuratively (e.g., *bread and butter*). It is important to note that the binomial expressions used in the current study are completely transparent word combinations. That is, their individual components make a direct and unambiguous contribution to the meaning of the expression. In that, they are akin to novel language. What makes them interesting is that, unlike novel word combinations but similar to idioms, they are frequent, familiar, relatively fixed, and highly predictable. A more detailed overview of binomial expressions and their linguistic properties is beyond the scope of this paper and an interested reader should consult Benor and Levy (2006), Bolinger (1962), Malkiel (1959), Lambrecht (1984), Cooper and Ross (1975), Fenk-Oczlon (1989), and McDonald, Bock, and Kelly (1993).

As mentioned above, in binomials (also known as *irreversible binomials*) one word order is always more frequent than the other. In the binomial expressions on which the current research is focused, whatever the word order is, the more frequent *bride and groom* or the less frequent *groom and bride*, the meaning is the same.[6] Thus, the binomial expressions that are investigated in the present study are, first, transparent and, second, have only (one) literal meaning, which does not change if the word order is reversed.

Word order in binomial expressions

A number of studies have investigated the word order in binomial expressions (e.g., Benor & Levy, 2006; Bolinger, 1962; Lambrecht, 1984; Malkiel, 1959; McDonald, Bock, & Kelly, 1993). According to Benor and Levy (2006), a few factors, such as semantics, metrics, frequency and phonology, can account for word order in binomial expressions (Appendix 4). It is undoubtedly true that such constraints can explain the preferred word order in many expressions. However, it would be erroneous to suggest that they can fully account for the word order in all binomials. For example, in the case of the binomial *knife and fork*, one might assume that *knife* precedes *fork* because we hold it in the right hand, which is the dominant hand for most humans. However, one might also argue that it is possible to eat without a knife but not without a fork (and, in fact, many people do exactly that), which should make a fork a more central or salient entity than a knife. Further, for such binomials as *alive and well* and *church and state*, most of the constraints proposed would predict the opposite word order, namely *well and alive* and *state and church*. Similarly, for *bride and groom*, the semantic constraint predicts the order *groom and bride*, since a masculine entity is said to precede a feminine one (as in *men and women*, *male and female*, *husband and wife*, and *brothers and sisters*). Finally, it would be logical to assume that some constraints, for example, semantic-pragmatic ones, should hold true across different languages. However, as the following examples illustrate, this is not always true (e.g., Russian: *demand and supply, sour and sweet*; Welsh: *pepper and salt*; Spanish: *white and black*; Czech: *forwards and backwards*).

The purpose of the current study is not to call into question the set of constraints that have been proposed in literature. I would simply like to point out that while these constraints *can* account for why one word order may be preferred over the other in some binomials, these constraints do not account for all binomials, nor are they always an accurate predictor of the preferred word order. A larger discussion of these constraints is outside of the purview of this study and will not be discussed further.

7.3. The present study

In order to investigate the processing of binomial expressions by native and proficient nonnative English speakers, the current study uses an eye-tracking paradigm. If it *is* the case that frequency of exposure plays an important role in what is *wired together*, or represented, in the mental lexicon, then one would expect native English speakers, who have accumulated a sufficient amount of experience with frequent expression, to show a processing advantage for binomials over their reversed forms. In contrast, nonnative speakers, whose exposure to English will not have been as rich, may exhibit a lexicon in transition. In other words, only the most frequently occurring expressions may be represented in their mental lexicon, while less frequent ones will be processed compositionally.

The current study will address the following questions. First, are native and proficient nonnative speakers sensitive to phrasal frequency? Second, the inclusion of nonnative speakers will allow me to more explicitly investigate the role of frequency and to establish *how* frequent a multi-word sequence has to be for it to be processed differently from novel language.

7.3.1. Experiment

Materials

The British National Corpus (BNC) was used to find a set of binomial expressions and their reversed forms. First, the target binomials had to be frequent word combinations. Second, because some binomials can be used both literally and figuratively, only those items were chosen which have only one literal meaning.

Finally, experimental items had to have identical meaning if the expression is reversed (e.g., *bride and groom* means the same as *groom and bride*, while *cut and paste* is not the same as *paste and cut*). Having these criteria in mind, 34 binomial expressions were selected. By default, binomials and their reversed forms are matched in individual word frequency and length. The lexical properties of the experimental items can be found in Table 7.1.

Table 7.1. Means for binomials and their reversed forms for phrasal frequency (absolute frequency, i.e., per 100 mil words of the BNC), semantic association strength, and frequency of the initial word.

	Binomial	**Reversed**
Phrasal frequency	240.4	27.4
Semantic association strength	27.2	21.4
Initial word frequency	15325	15549

Further, to ensure that any processing advantage for *bride and groom* over that of *groom and bride* could not be entirely due to *bride* serving as a better prime for *groom* than *groom* for *bride*, the Edinburgh Associative Thesaurus database [7] was used to check that binomials and their reversed forms were matched in semantic association strength as closely as possible. Table 7.1 gives the association strength for both the forward association ($Word_{n+1}$ responses (*groom*) to $Word_n$ (*bride*)) and backward association ($Word_n$ responses (*bride*) to $Word_{n+1}$ (*groom*)). Although we can see that in the binomial condition, $Word_n$ is numerically more strongly associated with $Word_{n+1}$ than $Word_{n+1}$ with $Word_n$ in the reversed condition, no statistically significant differences were found in the binomial versus reversed comparisons ($t(32) = 1.9, p = .07$). It is noteworthy that some constituent words (e.g., *groom* as in *bride and groom*) were not used as prime words in the Edinburgh Associative Thesaurus database, and hence they were not included in calculation of the mean (in such cases, a missing value was used).

Participants

Twenty-eight native speakers and twenty-eight proficient nonnative English speakers took part in the study. All participants were students at the University of Nottingham. Nonnative speakers met English language requirements prior to commencing their degree (minimum IELTS score of 6.0 or TOEFL score of 550). Native participants received course credit, while nonnative speakers received a small fee for their participation. Nonnative speakers came from various L1 backgrounds. On average, they spent 24.3 months studying in the UK, while their first contact with English was at the age of 6.9 years. Their self-rated proficiency for speaking, reading, writing, and listening comprehension on a 5-point Likert scale (1 = 'very poor', and 5 = 'excellent') was 3.8, 4.1, 3.8, and 4.1, respectively.

Apparatus and procedure

Thirty-four binomials and their reversed forms were presented across two presentation lists. Thus, no participant saw both versions of the same phrase (i.e., the binomial and its reversed form). Experimental items were intermixed with 42 filler sentences, which contained low frequency novel but entirely plausible sequences of the type '*Noun* and *Noun*', '*Adjective* and *Adjective*', or '*Verb* and *Verb*' (e.g., "tennis and badminton" and "determined and ambitious"). The purpose of these fillers was to prevent the participants from noticing a large number of binomial expressions and their reversed forms, which may have stood out.

Eye movements were recorded from the left eye using a SMI EyeLink I apparatus. Participants were given a verbal explanation of the eye-tracking procedure. A nine-point grid calibration procedure was done before the experiment. Participants first completed a short practice session. Each trial started with a fixation point that appeared in the middle of the screen. After participants fixated it and a calibration check was done, a sentence appeared in full always across one line in the middle of the screen. Participants were instructed to read the sentences as quickly as possible for comprehension and press a button on a response box to go from one trial to another. One quarter of the sentences in the experiment were followed by a comprehension question. Those trials that did not have a comprehension question were followed by a 'Ready?' question. The eye-tracker was calibrated at least four times during the experiment. After the experiment, nonnative participants completed a short language background questionnaire, assessing their self-reported English speaking, reading, writing, and comprehension on a five-point Likert scale (reported above).

Analysis

Because early and late measures are thought to tap into different processes, it was decided to analyse one early (first pass reading time) and two late measures (total reading time and fixation count). Two nonnative participants were excluded from the analysis due to very slow reading times. The participants had no difficulty answering comprehension questions, with the overall accuracy rate of 94.5% for natives, and 89.9% for nonnatives. Fixation durations shorter than 100 ms and longer than 800 ms were excluded from the analysis, because short fixations reflect oculomotor programming (e.g., Morrison, 1984), and fixations longer than 800 ms are due to momentary track loss or blinks. The missing data accounted for 1.2% of the native and 2.6% of nonnative speaker data. The lost data were equally distributed across the conditions.

Results

Following the data removal procedure, means were calculated for binomials and their reversed forms for native (Table 7.2) and nonnative participants (Table 7.3). ANOVAs were conducted on the critical region, treating participants and items as random variables. The dependent variables were mean first pass reading time, total reading time, and fixation count.

Native speakers

Table 7.2. Native speaker mean reading times (in milliseconds) for binomials and reversed forms for three eye-tracking measures with Standard Error (SE) in parenthesis.

	Binomials	**Reversed**	**Difference**
First Pass Reading Time	322 (12.1)	359 (17.2)	37 **
Total Reading Time	343 (13.7)	403 (18.6)	60 ***
Fixation Count	1.8 (.05)	2.0 (.07)	.2 ***

Note: ** $p < .01$, ***, $p < .001$

A significant main effect of Phrasal Frequency was observed in all three eye-tracking measures (first pass reading times: $F1(1,27) = 11.0, p < .005$; $F2(1,33) = 19.0, p < .001$; total reading time: $F1(1,27) = 28.6, p < .001$; $F2(1,33) = 34.2, p < .001$; fixation count: $F1(1,27) = 17.0, p < .001$; $F2(1,33) = 26.1, p < .001$).

Discussion

The results indicate that native English speakers are sensitive to the frequency with which linguistic patterns occur in language. Despite the fact that binomials (e.g., *bride and groom*) and their reversed forms (e.g., *groom and bride*) mean the same thing and are matched for meaning, lexical frequency and length, native speakers read frequent binomials reliably faster than their less frequent reversed forms, not only in early (first pass reading time) but also in late (total reading time and fixation count) eye-tracking measures. As such, these results highlight the role of phrasal frequency in language processing. Implications of these findings will be discussed in the general discussion.

Nonnative speakers

Table 7.3. Nonnative speaker mean reading times (in milliseconds) for binomials and reversed forms for three eye-tracking measures with Standard Error (SE) in parenthesis.

	Binomials	**Reversed**	**Difference**
First Pass Reading Time	550 (26.4)	573 (26)	23 +
Total Reading Time	592 (28.6)	610 (27.7)	18
Fixation Count	2.4 (.09)	2.5 (.09)	.1

Note: + $p < .10$

The first pass reading time measure revealed a trend towards shorter reading times for binomials over their reversed forms (550 ms vs. 573 ms; $F1(1,25) = 2.9, p = .099$, $F2(1,33) = 3.1, p = .086$). In the other eye-tracking measures, no such trend was observed (total reading time: $F1(1,25) = 1.9, p = .174$, $F2(1,33) = 1.9, p = .179$; fixation count: $F1(1,25) = 1.2, p = .276$, $F2(1,33) = .72, p = .401$). This trend towards shorter reading times for binomials suggests that participants are sensitive to phrasal frequency. However, the effect of phrasal frequency in nonnative speakers may be small because they will have had less exposure to English and may therefore be sensitive only to very high frequency phrases. To investigate this, a post-hoc analysis was conducted in which the experimental items were divided into two frequency groups: low and high (Appendix 5). Frequency and length information for these items is given in Table 7.4. Following this, means were calculated for low and high frequency binomials and their reversed forms (Table 7.5).

Table 7.4. Mean phrasal frequency (absolute frequency, i.e., per 100 mil words of the BNC) and length for low and high frequency binomials and reversed forms.

	Low		**High**	
	Binomial	**Reversed**	**Binomial**	**Reversed**
Phrasal frequency	79.5	4.5	401.2	50.3
Length	13.9	13.9	14.5	14.5

Table 7.5. Nonnative speaker mean reading times (in milliseconds) for low and high binomials and reversed forms for three eye-tracking measures with Standard Error (SE) in parenthesis.

	Low		**High**	
	Binomial	**Reversed**	**Binomial**	**Reversed**
First Pass Reading Time	566 (34.0)	563 (25.9)	538 (24.9)	586 (31.0)
Total Reading Time	627 (35.4)	608 (26.6)	562 (26.2)	615 (33.4)
Fixation Count	2.54 (.10)	2.50 (.10)	2.35 (.10)	2.50 (.09)

The results of this post-hoc analysis revealed a significant interaction between Phrase Type (binomial vs. reversed) and Frequency (high vs. low) in the total reading time measure ($F1(1,25) = 4.5$, $p < .05$, $F2(1,16) = 6.3$, $p < .05$). This interaction was found because total reading times in the low frequency group were similar for binomials and their reversed forms (binomials 627 ms, reversed 608 ms; $F1(1,25) = .868, p = .351, F2(1,16) = .182, p = .676$); whereas, in the high frequency group, binomials were read significantly faster than their reversed forms (binomials 562 ms, reversed 615 ms; $F1(1,25) = 5.5, p < .05, F2(1,16) = 4.9, p < .05$). The other two eye-tracking measures revealed no significant interaction between Phrase Type and Frequency (first pass reading time: $F1(1,25) = 1.9, p = .183, F2(1,16) = 3.5, p = .080$; fixation count: $F1(1,25) = 1.9, p = .183, F2(1,16) = 3.4, p = .082$).

Discussion

These results suggest that nonnative speakers are sensitive to phrasal frequency, in particular, when phrases are of very high frequency. Thus, binomials with a high phrasal frequency were processed faster than their reversed forms. It is important to note that the phrasal frequencies were obtained from the BNC corpus and only provide an estimate of exposure to phrases for native English speakers. Because these participants were all nonnative speakers and thus they would have had less exposure to English than native English speakers, it may be unsurprising that they were only sensitive to the very high frequency binomials. Crucially, this finding highlights the fact that language processing in nonnative speakers is also affected by phrasal frequency.

7.4. General discussion

In the present study, I investigate the on-line processing of binomial expressions by two groups, native and proficient nonnative speakers of English. Because the binomials and their reversed forms used in the study mean the same thing and are both grammatically correct and plausible word combinations, there is no *a priori* reason to treat them differently. However, if things that occur together frequently are also *wired together*, then native speakers should show a processing advantage for binomials. In contrast, nonnative speakers may display a lexicon in transition, in which only the most frequently occurring binomials will be represented and the less frequent ones will be processed compositionally.

It was hypothesised that binomials, being frequent multi-word sequences, would be good candidates for being represented in the mental lexicon, and that a processing advantage for binomials over their reversed forms would support such a claim. The analysis of the native speaker data revealed that the frequency with which multi-word sequences occur in language affects the speed of their processing during reading. This finding is particularly robust given that the meaning, syntactic structure, lexical frequency and length of the component words were identical in the binomial and reversed conditions, and the semantic association strength was matched. The finding that native speakers process frequent multi-word sequences faster than low frequency ones is consistent with previous research (e.g., Arnon & Snider, 2010; Bannard & Matthews, 2008; Sosa & MacFarlane, 2002).

However, the key finding of the present study regards the nonnative speaker group. If the frequency of occurrence of a particular form (e.g., a word or a phrase) leads to its representation in the mental lexicon, as well as to its resistance to morphosyntactic changes (e.g., word order changes), then it appears that the nonnative participants may have not had enough experience with lower frequency binomial expressions. The finding that low frequency binomials and their reversed forms were read with a similar speed appears to be in line with the view according to which nonnative speakers construct and process a large proportion of their language compositionally rather than using frequent routines (e.g., Foster, 2001; Skehan, 1998; Wray, 2002). What is noteworthy is that high frequency binomials, that is, those that a nonnative speaker will have come across a sufficient number of times, were read significantly faster than their reversed forms (as shown in the total reading time measure). If, indeed, the frequency of occurrence of a linguistic form leads to its representation in the lexicon, then it is plausible that highly frequent phrases are represented in the lexicon of a second language speaker, similar to how they are encoded in the lexicon of a native speaker.

The finding that phrasal frequency affects the ease of processing in both native and nonnative speakers is of importance for a number of models of language use and processing. According to the traditional view, knowing a language presupposes knowing a limited set of grammar rules, which can be used to produce and comprehend an infinite number of novel utterances. As mentioned in Chapter 2, according to the words-and-rules approach (e.g., Pinker, 1991, 1999; Pinker & Ullman, 2002), there is a distinction between the mental lexicon and the mental grammar. For example, with respect to the role of frequency, it is believed that while the frequency of occurrence is expected to affect on-line processing of memorised forms, no such effect is predicted for compositional phrases. Because this approach does not predict faster reading times for frequent compositional phrases over less frequent ones, this view is incompatible with my results.

On the other hand, 'empiricist' theories, such as usage-based (e.g., Bybee, 1998; Goldberg, 1995, 2006; Tomasello, 2003) and exemplar-based models (e.g., Abbot-Smith & Tomasello, 2003; Bod, 1998, 2006; Pierrehumbert, 2001) propose that all linguistic material should be similarly affected by the frequency factor. That is, new experiences with a word or a phrase are not decoded and then discarded; rather, they determine memory representations (e.g., Bybee, 2006). As Bod (2006) notes, the assignment of representations to linguistic events is accomplished solely on the basis of statistics (both in language acquisition and processing). In this account, language should be viewed not as a set of grammar rules, but as a statistical accumulation of experiences that changes every time a particular utterance is encountered (Bod, Hay, & Jannedy, 2003). This view predicts faster reading times for frequent words, as well as compositional phrases, over infrequent ones. Thus, I take the results of the present study to support the usage-based approach to language processing and use.

Thus far, it has been argued that due to their relative fixedness and high frequency, multi-word units may have a special status in the mental lexicon and are therefore processed faster than novel language. However, the above results are also in line with an alternative possibility, according to which the processing advantage observed for *bride and groom* is the result of a very quick, almost simultaneous activation of *groom* upon encountering *bride*. In line with probabilistic models of language processing, probabilistic information about co-occurrences of words forms an integral part of speakers' knowledge of language (e.g., Gregory et al., 1999; Jurafsky, 1996; McDonald & Shillcock, 2003a, 2003b). Reichle et al. (1998) and Engbert et al. (2005) hold that eye-movement patterns reflect a reader's experience with language and are thus influenced by such factors as frequency and predictability. Crucially, the probability of $Word_{n+1}$ (e.g., *groom*) occurring after $Word_{n\ and}$ (*bride and*) is about six times as high as the probability of $Word_{n}$ (e.g., *bride*) appearing after $Word_{(n+1)\ and}$ (e.g., *groom and*). Because *bride and groom* is a frequent expression, while *groom and bride* is not, *bride* serves as a better prime for *groom*, than *groom* is for *bride*. In this account, the processing difference between binomials over their reversed forms is due to the difference in their predictability, rather than one being represented in the lexicon and the other one not.

So, the critical question to ask is: Is the processing advantage observed for *bride and groom* over *groom and bride* the result of the high probability of seeing *Word* $_{n+1}$ after *Word* $_{n}$, or is it due to frequent binomial expressions being represented in the lexicon? Unfortunately, the present study is unable to answer this question. If a very fast, almost instantaneous activation of one constituent that frequently occurs with another constituent underlies what has been referred to in the literature as a 'single representation', it would be very difficult, indeed, to distinguish between the two possibilities. However, it is an important question that should be addressed in future work. It is finally worth pointing out that the 'predictability story' *per se* does not by any means go against the representation account. Each and every instance of a particular multi-word sequence, be it a pure idiom (e.g., *kick the bucket*), a binomial (e.g., *bride and groom*), a collocation (e.g., *extenuating circumstances*), a prepositional verb (e.g., *depend on*), or a speech formula (e.g., *Good morning!*), is a highly predictable word combination where *Word* $_{n+1}$ can be easily predicted from *Word*$_{n}$. Thus, I would like to argue, being highly predictable is an intrinsic characteristic of multi-word speech.

In sum, the results of the current study show that both native and nonnative speakers are sensitive to the frequency with which units larger than a traditional word occur in language, albeit as one would expect, this sensitivity is more robust in the native population. In nonnative speakers, the effect of phrasal frequency was observed only with regards to very high frequency forms. The above findings provide further evidence for the important role of phrasal frequency (on a par with lexical frequency) in first language processing, and, crucially, suggest that similar processes are in place for nonnative speakers. As such, these results offer support to a usage-based theory that postulates that frequency of occurrence shapes memory representations not only with regards to single words, but also units larger than a single word.

Chapter 8: Processing and Representation of Binomial Expressions in a First Language: Evidence from ERP

8.1. Introduction

Chapter 7 presented an eye-tracking experiment that investigated on-line processing of frequent binomial expressions (e.g., *bride and groom*) and their less frequent reversed forms (e.g., *groom and bride*) by native and proficient nonnative English speakers. The results of the study clearly showed that both native and nonnative speakers are sensitive to frequency manipulations not only at the word level (e.g., Balota & Chumbley, 1984; Monsell et al., 1989; Rayner & Duffy, 1986), but also at the phrase level. The results support the view according to which each and every occurrence of a linguistic form contributes to how this information is represented in the speaker's memory. The findings presented in Chapter 7 highlight the role of phrasal frequency in language processing and, as such, suggest that recurrent phrases may be represented in the mental lexicon along with single words.

The present chapter further investigates the processing of frequent binomial expressions in native English speakers using a neurophysiological technique, namely, electroencephalogram (EEG). Similar to the eye-tracking study, I aim to investigate native speaker sensitivity to frequent linguistic patterns versus infrequent ones. The main goal of the present investigation, however, is to show that the language processor treats frequent phrases in a *unitary* way. While Chapter 7 showed that frequency plays an important role in phrasal processing, in the present chapter, I aim to demonstrate that high frequency phrases are not only processed faster than low frequency ones, but that they are, in fact, processed akin to a unit, or a chunk. The eye-tracking methodology used in Chapter 7 is informative with respect to the speed of processing; the use of ERP will allow me to investigate the very nature of the cognitive processes involved in phrasal processing. Thus, it is hoped that the ERP findings will complement rather than replicate the eye-tracking ones.

As was mentioned in the previous chapter, binomial expressions are frequent, familiar, and highly predictable expressions. These linguistic properties, having been the focus of a number of ERP studies, are often associated with two ERP components: the N400 and the P300. In literature, the N400 has been shown to be sensitive to the processing of lexical-semantic information, such as frequency and predictability, as well as real-world knowledge (e.g., Hagoort et al., 2004; Kutas & Hillyard, 1980; Kutas & Hillyard, 1984). In Kutas and Hillyard (1980) and Kutas and Hillyard (1984), semantically incongruent sentence completions elicited larger N400 than semantically congruent ones. These findings highlight the fact that speakers form strong expectations of the upcoming information given the preceding context, and when the expectations are not met, the N400 effect is observed. Hagoort et al. (2004) observed comparable N400 effects with respect to world knowledge. Sentences that violated participant's world knowledge (e.g., *The Dutch trains are white and very crowded*) resulted in a larger N400 effect than sentences in which this knowledge was not violated (e.g., *The Dutch trains are yellow and very crowded*).

Unfortunately, very few ERP studies have addressed the issue of multi-word sequence processing. Of the many types of multi-word speech, only one type has been investigated using evoked potentials – idioms (e.g., Laurent et al., 2006; Strandburg et al., 1993; Vespignani et al., 2009). This is hardly surprising, as idioms are by far the most archetypal type of multi-word speech. Although binomial expressions are different from idioms in that they are most commonly used literally and their constituent words contribute fully and overtly to the meaning of the entire phrase, they are akin to idioms in terms of their familiarity and predictability. A number of ERP studies have looked at the processing of figurative phrases versus matched novel phrases. For example, in a recognition task, Strandburg et al. (1993) recorded ERPs on the final word of unambiguously figurative, literal (novel), and nonsensical phrases. Smaller N400 amplitudes were observed for figurative phrases compared to literal and anomalous ones, suggesting that familiar phrases were processed more easily than matched novel phrases. This finding is of particular relevance to the present study because the figurative phrases used in Strandburg et al.'s study were familiar figurative expressions (e.g., *square deal* and *vicious circle*).

In another study, Laurent et al.'s (2006) participants performed a semantic relatedness task on French idioms that varied in their degree of salience. The authors found that N400 amplitudes were smaller for the last word of the strongly salient (e.g., highly conventional) idioms than for the weakly salient ones (e.g., new metaphors). Similar to Strandburg et al. (1993), reduced N400s on highly salient (conventional) idioms suggest that idiomatic expressions are easier to integrate and process.

Finally, Vespignani et al. (2009) investigated on-line processing of Italian idioms. The authors proposed that due to their frequency and relative fixedness, idioms are ideal for investigating predictive mechanisms using ERP. In the study, a number of idioms were selected, and their recognition points (i.e., the point at which the expression becomes recognisable as idiomatic, rather than novel) identified. Three conditions were investigated in the study: one idiomatic and two literal. Vespignani and colleagues found that idiomatic phrases elicited smaller N400s than matched literal phrases on the word that represented the recognition point of the idiom. As above, the diminished N400s on idioms were taken to suggest a processing advantage for multi-word sequences.

The above studies link predictability and familiarity to the N400 component. However, the N400 is not the only ERP component associated with the issue of predictability of the upcoming information. Roehm et al. (2008) showed that in a highly predictable context, such as, *The opposite of black is white*, where only one possible continuation is possible (i.e., *white*), the P300 effect was elicited on the highly predictable word (i.e., *white*). The N400 effect was observed for the related word *yellow* and the nonrelated word *nice* in the same context. With regards to multi-word speech, one study in particular is of interest. In the above-mentioned study, Vespignani et al. (2009) compared ERP waveforms for idioms and literal phrases before and after the recognition point. Before the recognition point, the authors propose, the difference between the literal and idiomatic conditions is attributed to the N400 (the finding reported above). After the recognition point, on the other hand, the idiomatic condition elicited a pronounced P300. Because after the recognition point, only one idiom completion is possible, Vespignani and colleagues drew a parallel between their results and those of Roehm et al. (2007), where sentences like *The opposite of black is …* could be completed only by *white*, and concluded that the observed P300 effect is the result of categorical template matching. I will come back to Vespignani et al.'s results in the general discussion.

The above literature review suggests that the processing of idiomatic expressions differs from novel language not only in terms of the speed of processing (as was suggested in Chapters 3 and 6), but also with respect to the neural correlates that underlie their comprehension. In a series of experiments, the present ERP study aims to shed more light on the nature of the brain activity elicited during familiar and novel language comprehension.

8.2. The present study

The current study encompasses three ERP experiments with native speakers of English. In Experiment 1, the processing of frequent binomial expressions is compared with that of infrequent reversed forms in a sentence context (e.g., *Despite the crises the <u>king and queen</u>/ <u>queen and king</u> are still popular among the people*). Experiment 2 looks at the processing of frequent *and* strongly associated phrases (binomials) versus infrequent but equally strongly associated phrases presented out of sentence context (e.g., *knife and fork* vs. *spoon and fork*). Experiment 3 uses the same set of materials as Experiment 2, however, the binomials and the semantic associates are presented without the conjunction 'and'. By investigating the processing of binomials in and out of sentence context, as well as when they are and are not in their formulaic form, it will be possible to obtain a clearer picture of multi-word speech processing and representation in native speakers.

8.2.1. Experiment 1

In Experiment 1, I am looking at the processing of frequent binomials (e.g., *king and queen*), their infrequent reversed forms (e.g., *queen and king*), and phrases with a semantic incongruity (e.g., *king and cloud*) presented in a sentential context. If frequent binomial expressions are processed differently from novel language, one should observe different waveforms for the two critical conditions (i.e., binomial vs.

reversed). This expectation is driven by the different reading pattern observed in the eye-tracking study (Chapter 7).

Because binomial expressions are frequent and highly predictable, such that the final word of the binomial is more predictable than the final word of its reversed form, it was plausible to assume two possible patterns of results. The first possibility is that the infrequent *queen and king* should result in larger negativity than the frequent *king and queen*. However, because the P300 component has also been observed in highly constraining context (such as those described above), it was possible that the binomial condition (e.g., *king and queen)* would result in larger early positivity than the less frequent and thus less predictable reversed condition (e.g., *queen and king*). With respect to the semantically incongruent condition (e.g., *king and cloud*), a larger N400 was expected in this condition than in the binomial and reversed condition.

Materials

In the experiment, 180 matched sentence triplets that contained binomials, their reversed forms, and phrases with semantic incongruity were used (Figure 8.1) (experimental items can be found in Appendix 6). Binomials and their reversed forms are by default matched in lexical frequency and length. Content Word 2 in the semantic incongruity condition (e.g., *cloud*) was matched in length, lexical frequency and word class with Content Word 2 in the binomial condition (e.g., *queen*). The properties of the experimental items in the three conditions can be found in Table 8.1. Phrasal frequencies in this study are from the British National Corpus (BNC), while lexical frequencies were obtained from CELEX.[8]

Figure 8.1. Sentence triplets used in the experiment (target phrase underlined).

Binomial: Despite the crisis the <u>king and queen</u> are still popular among the people.

Reversed: Despite the crisis the <u>queen and king</u> are still popular among the people.

Incongruity: Despite the crisis the <u>king and cloud</u> are still popular among the people.

Table 8.1. Mean phrasal frequency (absolute frequency, i.e., per 100 mil words of the BNC corpus), length, and semantic association strength means for binomials, their reversed forms, and semantic incongruity.

	Binomial	Reversed	Incongruity
Phrasal frequency	185.2	19.5	0
Phrasal length (characters)	14.1	14.1	14.1
Semantic association strength	22.9	19.2	0

Semantic association norming

As in Study 2 (Chapter 7), the Edinburgh Associative Thesaurus database was used to measure the strength of association between the two content words of the target phrase. Table 8.1 gives the association strength for both the forward association (Word 2 responses (*queen*) to Word 1 (*king*)) and backward association (Word 1 responses (*king*) to Word 2 (*queen*)). Statistical analysis showed that the forward association (i.e., in the binomial condition) was significantly stronger than the backward association (i.e., in the reversed condition) ($t(166) = 2.8, p < .05$). Because of the large number of items required for ERP experiments, it was impossible to completely control for association strength. However, Experiments 2 and 3 discussed further in the chapter will address the issue of association strength. It is worth pointing out that some constituent words (e.g., *groom* as in *bride and groom*) were not used as prime words in the Edinburgh Associative Thesaurus database, and hence they were not included in the calculation of the mean (in such cases, a missing value was used). The two content words in the semantic incongruity condition (i.e., *king* and *cloud)* were not associated.

As was noted in Chapter 5 where the ERP methodology was discussed, ERP requires a large number of experimental items. Thus, in order to have a sufficient number of items, in the present experiment, binomials like *black and white* and *night and day,* which can be used both literally and figuratively, were included. However, only a very small number of such binomials were used, and, importantly, only the literal interpretation was supported by the sentence context.

All three types of stimuli were embedded in identical sentence context (Figure 8.1). Thus, any differences observed between the three conditions cannot be due to context.

Participants

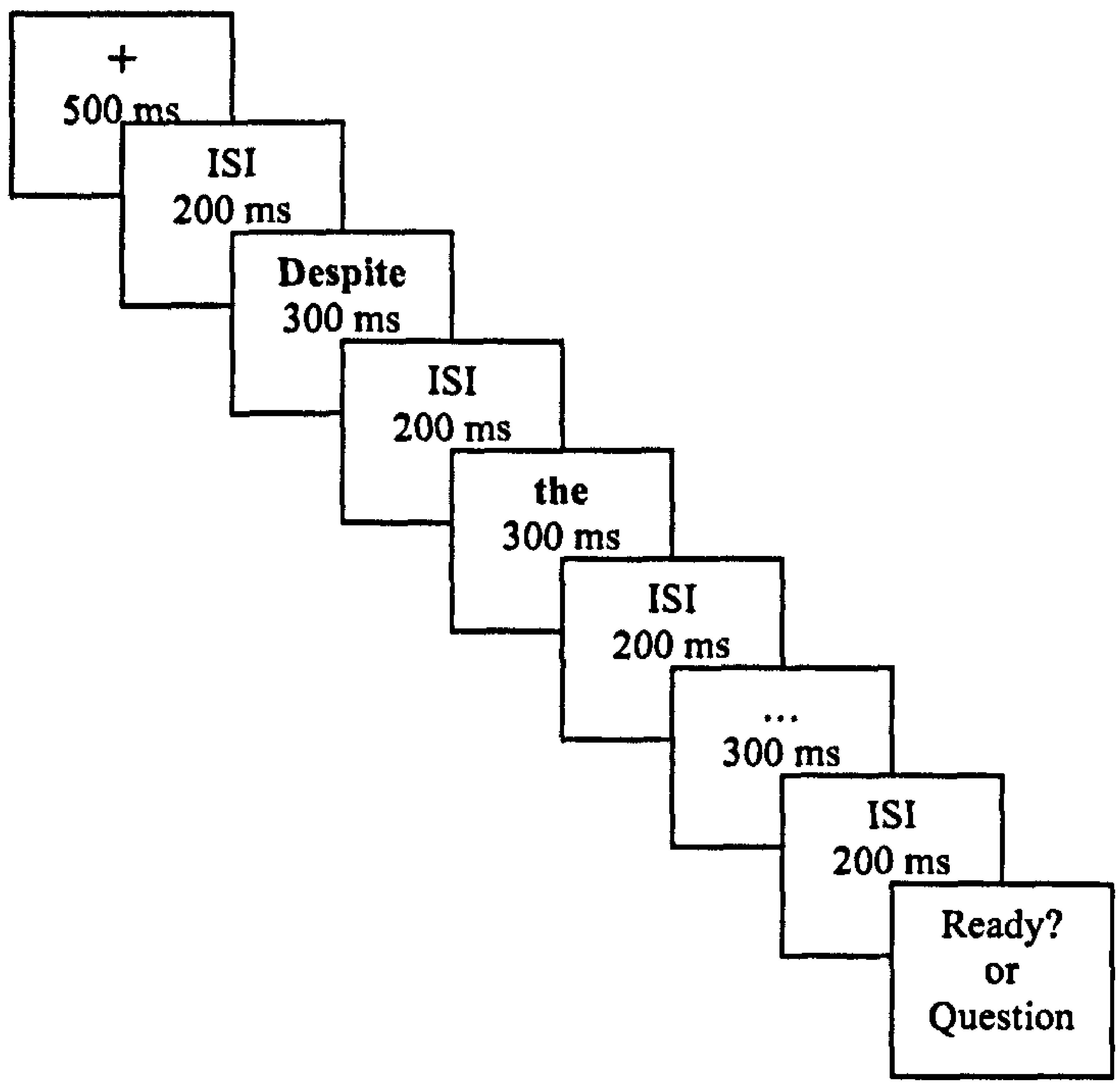
Forty-five healthy native speakers of British English participated in a two-hour long experiment. They were drawn from the undergraduate and graduate student population at the University of Nottingham (15 females, 30 males; age 18–36; mean age 21.6). All participants were right handed as assessed by the Edinburgh handedness inventory (Oldfield, 1971), and none had suffered from any neurological conditions, or language-related impairments, as indicated through self-report. All participants had normal or corrected-to-normal vision. They gave informed consent before the experiment and read an information sheet describing the EEG methodology, experimental procedure and instructions. After the experiment, they were paid £12.

EEG Procedure

Each participant first completed a practice session that consisted of four trials. In the actual experiment, 180 triplets were pseudorandomised across three lists so that each participant only read one version of each triplet. Thus, each list contained 60 exemplars of each of the three conditions. The 180 target items per list were intermixed with 120 filler sentences. Filler sentences contained low frequency novel but grammatically correct and meaningful sequences of the type '*Noun* and *Noun*', '*Adjective* and *Adjective*', or '*Verb* and *Verb*' (e.g., *I forgot my umbrella and glasses on the bus*). The target phrase was never at the very beginning or end of the sentence. To encourage participants to read the sentences for comprehension, 25% of the trials were followed by a Yes/No comprehension question. Those trials that did not have a comprehension question were followed by a 'Ready?' question.

Each trial started with a 500 ms fixation cross. Sentences were presented using rapid serial visual presentation (RSVP), that is, one word at a time. The inter-stimulus interval (ISI) was 200 ms, while each word remained on the screen for 300 ms (Figure 8.2). Comprehension questions and "Ready?" stayed on the screen for as long as participants needed to answer the question, blink, and prepare for the next trial. They were asked to press a button on the keyboard when they were ready to proceed to the next trial. They were urged to stop moving and blinking as soon as a fixation cross appeared on the screen.

Figure 8.2. Experiment 1 presentation mode.



The words were presented in white lower case Courier New letters against a black background in the center of a VGA computer screen. The viewing distance was approximately 100 cm. The participants' task was to read the sentences for comprehension and to answer questions when prompted. Participants were seated comfortably in a dimly lit lab. They were asked to read silently the words appearing one-by-one and to understand them as well as possible. The experimental session consisted of three blocks. In between the blocks, participants had a break during which an impedance check was performed.

EEG recording

High density event-related electrical potentials (ERPs) were recorded from each participant using a 128-channel EGI geodesic sensor net coupled to a high input impedance amplifier (Electrical Geodesics, Inc.; Tucker et al., 1994). EEG was continuously recorded and digitised at 250 Hz, with a hardware bandpass filter of 0.01–100 Hz. Wherever possible, impedances were reduced to <70 KΩ prior to recording. Segmentation was carried out target-locked into 1000 ms epochs starting with 100 ms prior to the onset of the target stimulus. Samples were low-pass filtered with a cut-off frequency of 40 Hz. Epochs were baseline corrected using the data from the first 100 ms of the epoch.

Analysis and results

The EEG data were screened for eye movements, electrode drifting, and other artifacts. Trials containing such artifacts were rejected (6.4%). The lost data were equally distributed across conditions. Five subjects were excluded from the analysis because of a large number of artifacts. Thus, the data from 40 participants were included in the analysis. For each participant, average waveforms were computed across all remaining trials per each of the three conditions. Although the critical word was always Content Word 2 of the phrase, ERPs were recorded, analysed, and are reported below on Content Word 1, the conjunction 'and', as well as Content Word 2.

Content Word 1

In the binomial (e.g., *king and queen*) and semantic incongruity condition (e.g., *king and cloud),* Content Word 1 was the same (e.g., *king*) and hence length and phrasal frequency were matched. However, in the reversed condition, Content Word 1 (e.g., *queen*) was not matched with the Content Word 1 of the other two conditions (e.g., *king*). Although the critical word was the same in all three conditions, the waveforms produced by Content Word 1 across the conditions were also compared. In the latency range of -100 to 1000 ms after the onset of the word, a series of running t-tests was performed in 24 ms bins, which shifted in time by 12 ms (e.g., 100-124 ms, 112-136 ms, etc.). The difference between any two waveforms (i.e., conditions) was taken to be significant if five or more consecutive windows were shown to be significant ($p < .05$). The analysis revealed no significant differences for Content Word 1 waveforms in the binomial, reversed, or semantic incongruity condition (*Fs* $p > .05$).

The conjunction 'and'

The middle word (the conjunction 'and') was identical across the three conditions. However, because in the frequent binomial condition, *king and* is more predictive of the upcoming *queen* than *queen and* is of *king* in the infrequent reversed condition, I wanted to compare the waveforms across the conditions, in order to see whether any possible differences between the two critical conditions (binomial vs. reversed) could show up as early as after the onset of the word 'and'. Similar to the above, in the latency range of -100 to 1000 ms after the onset of the word, a series of t-tests in 24 ms bins that shifted by 12 ms was run. The analysis revealed no significant differences for the conjunction 'and' (*Fs* $p > .05$) in any of the three conditions in the -100-500 ms.

Content Word 2

Content Word 2 (the last word of the pair) was the critical word and where differences across the conditions were expected. Visual inspection of the waveforms revealed a more negative potential (peaking around 400 ms) for the incongruous conditions relative to the to other two conditions. Furthermore, a larger posterior positivity was observed in an earlier window for the binomial condition. In the same latency range, a series of running t-tests in 24 ms bins that shifted by 12 ms was performed. The results of the running t-tests showed that the waveforms for the binomial and reversed conditions deviated significantly across a number of parietal electrodes (Table 8.2). Figure 8.3 shows the significant electrodes in the binomial versus reversed comparison and their topography, while Figure 8.4 shows the waveforms in a selection of electrodes. The difference in the other two comparisons (binomials vs. incongruity and reversed vs. incongruity) was apparent across a large number of electrodes in the 300-500 ms window peaking around 400 ms after the onset of the word (Figure 8.4).

Table 8.2. Significant electrodes in the binomial versus reversed comparison.

EGI 128 system	10-10 system	significant window (ms)	duration (ms)
53	P3	272 - 368	96
54	-	260 - 380	120
55	CPz	284 - 356	72
61	P1	272 - 368	96
62	Pz	272 - 344	72
66	-	284 - 356	72
67	-	272 - 356	84
68	POz	272 - 356	84
79	P2	272 - 392	120
80	-	272 - 356	84
85	-	272 - 452	180
86	PO4	260 - 440	180
87	P4	272 - 380	108

Figure 8.3. Binomials versus reversed forms in 300-350 ms window (red circles represent p < .05).

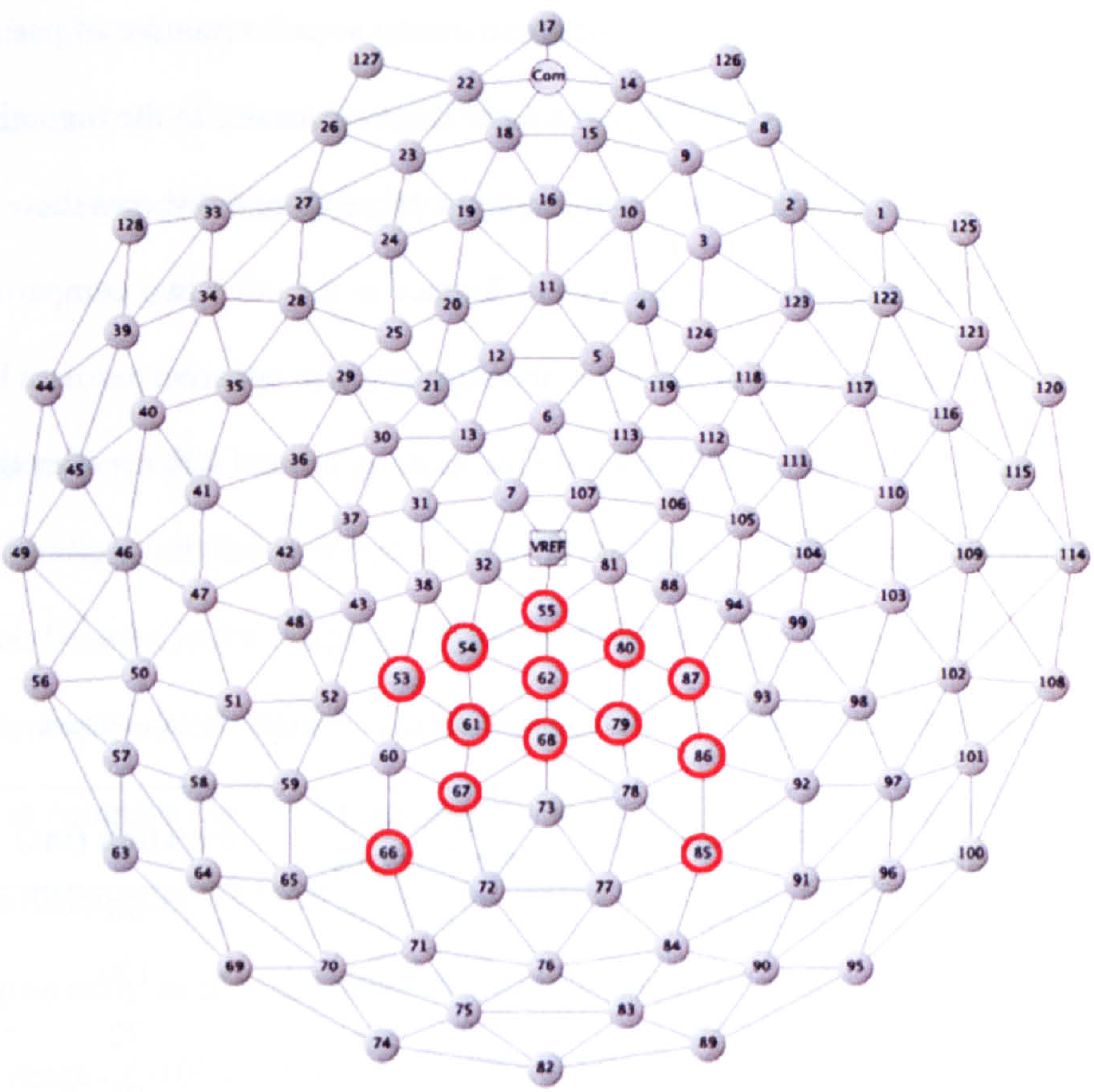
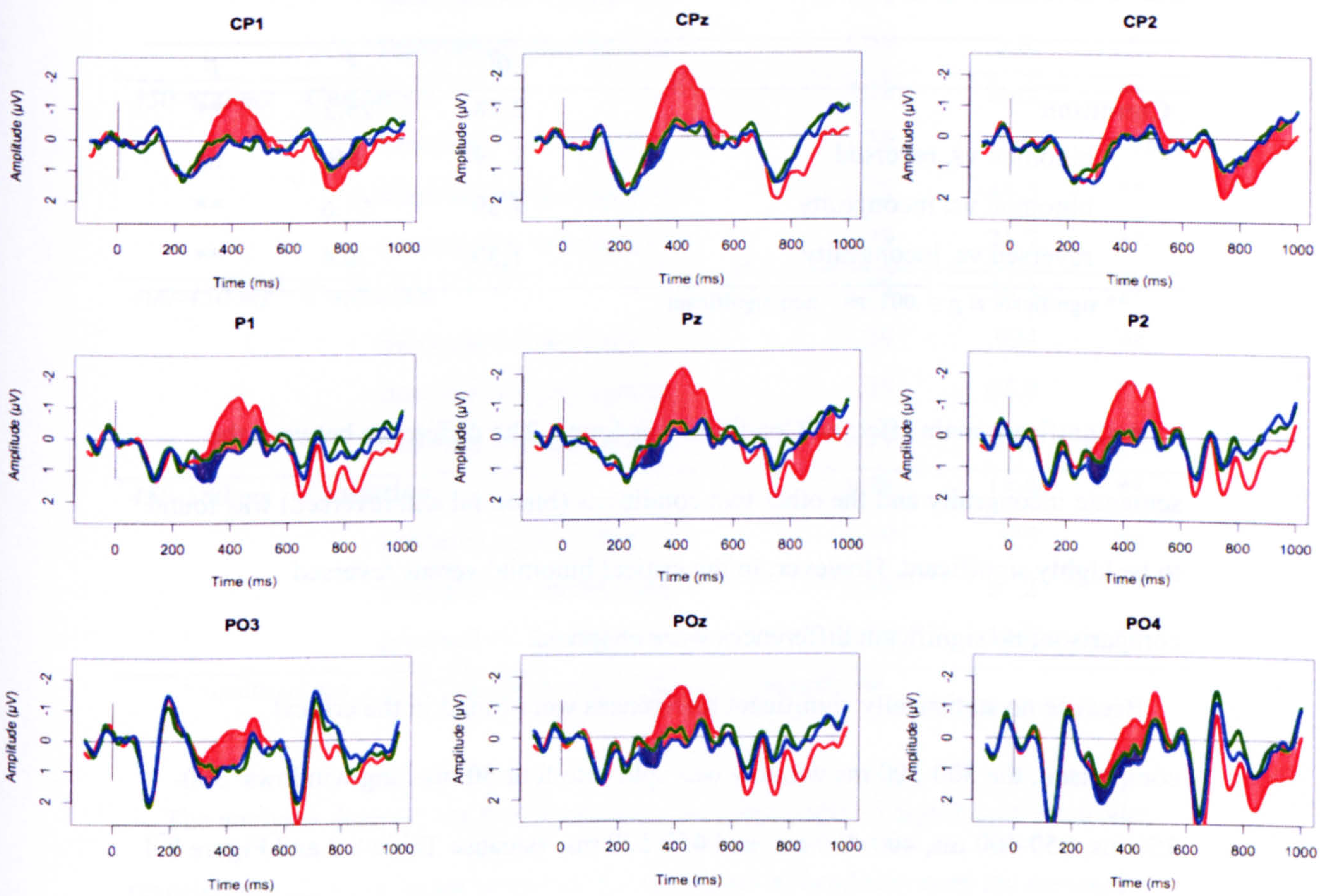


Figure 8.4. Waveforms for nine channels for Content Word 2 in the -100-1000 ms window. Blue lines indicate the binomial condition, green lines the reversed, and red the semantic incongruity. Red shaded areas represent significant differences between semantic incongruity and binomial/reversed conditions. Blue shaded areas represent significant differences in the critical comparison: binomial versus reversed.



To explore the differences between the three conditions in greater detail, I looked at the data from the 13 significant electrodes (Table 8.2) in the 300-500 ms window after the onset of the critical stimulus. The data were analysed using repeated measures ANOVAs treating participants as random variables (Table 8.3).

Table 8.3. Analysis of Variance in the 300-500 ms window for the binomial, reversed, and semantic incongruity.

	df	*F*	*p*
Condition	2,78	25.2	**
binomial vs. reversed	1,39	1.9	ns
binomial vs. incongruity	1,39	55.6	**
reversed vs. incongruity	1,39	20.4	**

** significant at $p \leq .001$, ns – non-significant

A significant main effect of Condition was found. The difference between the semantic incongruity and the other two conditions (binomial and reversed) was found to be highly significant. However, in the critical binomial versus reversed comparison, no significant differences were observed.

Because no statistically significant differences were found in the critical comparison, the 300-500 ms window was split into four 50 ms-long windows: 300-350 ms, 350-400 ms, 400-450 ms, and 450-500 ms. Because Table 8.2 and Figure 8.4 suggest that the difference between the binomial and reversed condition may appear before 300 ms after the onset of the critical word, an earlier 250-300 ms window was also added to the analysis. Table 8.4 shows ANOVAs for the five small windows for the data from the 13 significant centro-parietal electrodes.

Table 8.4. Binomials vs. reversed vs. semantic incongruity for five windows.

		df	*F*	*p*
250-300 ms	**Condition**	2,78	6.8	*
	binomial vs. reversed	1,39	10.2	*
	binomial vs. incongruity	1,39	16.8	**
	reversed vs. incongruity	1,39	.058	ns
300-350 ms	**Condition**	2,78	16.6	**
	binomial vs. reversed	1,39	14.8	**
	binomial vs. incongruity	1,39	33.5	**
	reversed vs. incongruity	1,39	5.8	*
350-400 ms	**Condition**	2,78	26.3	**
	binomial vs. reversed	1,39	3.3	=.08
	binomial vs. incongruity	1,39	46.5	**
	reversed vs. incongruity	1,39	21.7	**
400-450 ms	**Condition**	2,78	30.3	**
	binomial vs. reversed	1,39	.924	ns
	binomial vs. incongruity	1,39	63.0	**
	reversed vs. incongruity	1,39	27.2	**
450-500 ms	**Condition**	2,78	13.9	**
	binomial vs. reversed	1,39	.325	ns
	binomial vs. incongruity	1,39	31.2	**
	reversed vs. incongruity	1,39	14.0	**

* significant at $p \leq .01$, ** significant at $p \leq .001$, ns – non-significant

The analyses done on the five 50 ms windows across the 13 significant electrodes revealed significantly larger positivity for the binomial condition over the reversed one in two early windows: 250-300 ms and 300-350 ms. No statistically significant differences between the two critical conditions were observed in three late windows: 350-400 ms, 400-450 ms, and 450-500 ms (although the 350-400 ms window was found to be marginally significant). With regards to semantic incongruity, significantly larger negativity was observed for this condition when compared to the binomial and reversed conditions in all five windows (with the exception of the reversed vs. incongruity comparison in the 250-300 ms window).

Finally, I also looked at one parietal electrode separately – Pz (Figure 8.5). Table 8.5 shows ANOVAs for a 100 ms window (250-350 ms after stimulus onset) for this particular electrode. As expected (from Table 8.2), the analysis showed significantly larger positivity for the critical comparison, namely, the binomial condition versus the reversed one. Further, larger negativity was observed for the semantic incongruity condition when compared to the binomial and reversed conditions.

Figure 8.5. Waveforms for Pz for Content Word 2 in the -100-1000 ms window. Blue lines indicate the binomial condition, green lines the reversed, and red the semantic incongruity. Red shaded areas represent significant differences between semantic incongruity and binomial/reversed conditions. Blue shaded areas represent significant differences in the critical comparison: binomial versus reversed.

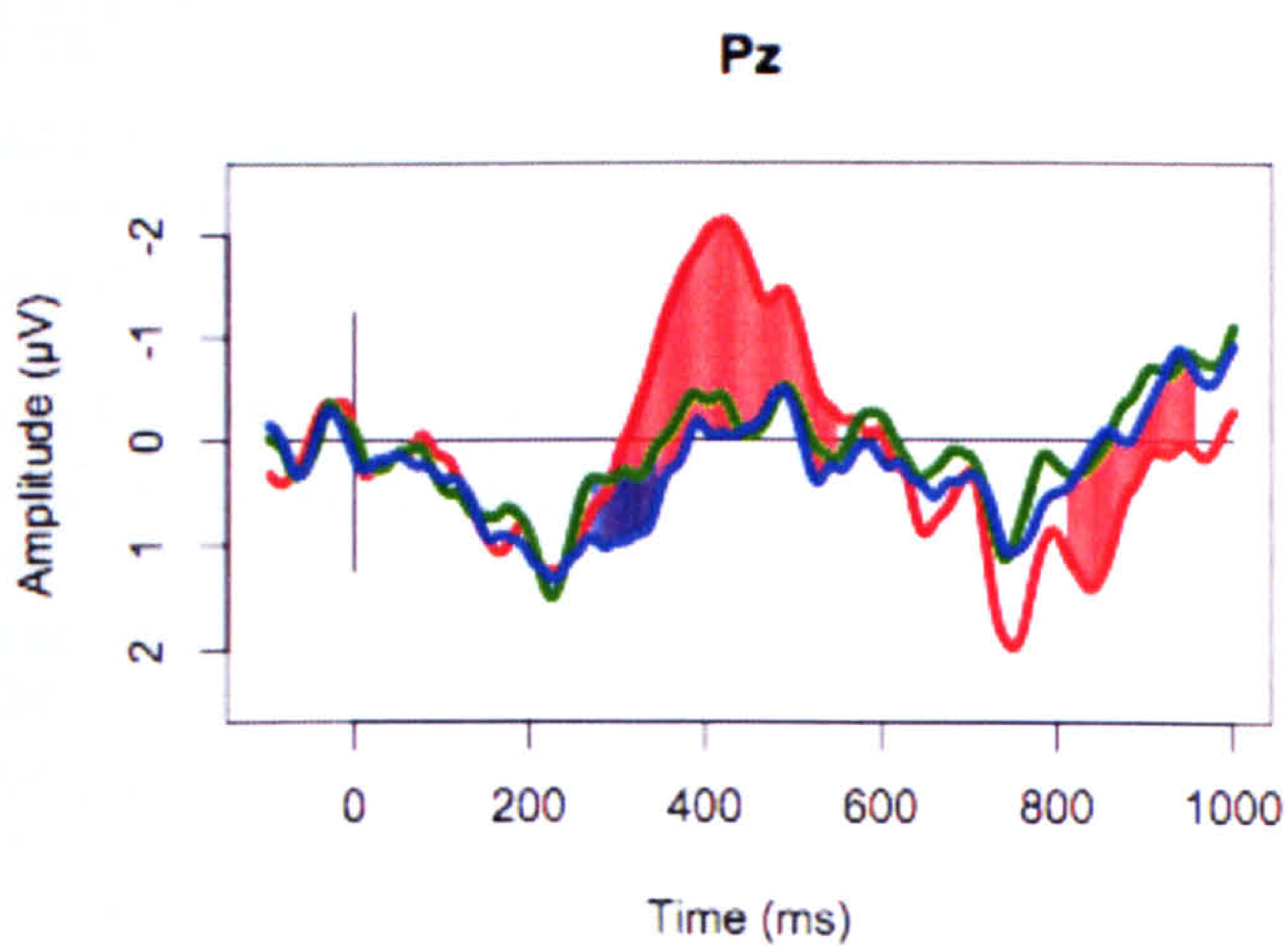


Table 8.5. Binomials vs. reversed vs. semantic incongruity for Pz in the 250-350 ms window.

	df	*F*	*p*
Condition	2,78	11.0	**
binomial vs. reversed	1,39	6.2	*
binomial vs. incongruity	1,39	26.4	**
reversed vs. incongruity	1,39	4.6	*

* significant at $p \leq .01$, ** significant at $p \leq .001$

Discussion

The above analysis revealed two findings. First, the semantic incongruity condition (e.g., *king and cloud*) elicited a large negative-going wave in the 300-500 ms window peaking around 400 ms after the onset of the critical word (e.g., *cloud*). This finding was expected and is consistent with previous research that has showed larger negative amplitudes, namely, the N400, elicited by semantically incongruous words. Of greater theoretical importance, however, is the finding of a positive deflection in the binomial condition (e.g., *king and queen*) in the early 250-350 ms window, peaking around 300 ms after the onset of the word *queen*. This deflection, although found to be significant only in a small group of parietal electrodes, is what I take to be the P300 effect, which is often associated with the phenomenon of "template matching" (e.g., Kok, 2001). This finding will be taken up in the general discussion.

The norming procedure in Experiment 1 showed that the words in the binomial and reversed conditions were not equally strongly associated. Namely, the word *king* was more strongly associated with the word *queen* than the other way round. A follow-up experiment was designed to address this issue. Experiment 2 further explores the processing of frequent binomial expressions versus infrequent novel word combinations. However, unlike Experiment 1, the constituent words in target phases are matched in association strength. In addition, stimuli are presented out of sentence context.

8.2.2. Experiment 2

The aim of Experiment 2 was to investigate the processing of frequent binomial expressions whose constituent words are strongly associated (e.g., *knife and fork*) versus infrequent novel phrases whose words are equally strongly associated (e.g., *spoon and fork)*. One of the aims of Experiment 2 was to investigate whether the P300 effect observed for binomials in Experiment 1 could have been due to association strength not being matched in the two critical conditions. If this was indeed the case, then there should be no processing advantage for binomials versus associates in Experiment 2 (since the association strength is matched). If there is something more to binomial expressions than just the two content words being strongly associated, then the processing pattern for binomials and associates should be different.

Materials

In Experiment 2, 120 matched triplets that contained items in the following three conditions (critical word underlined) were used:

(1) binomial condition – a binomial expression whose two content words are strongly associated (e.g., *knife and <u>fork</u>*).
(2) associate condition – a grammatically plausible but infrequent phrase whose two content words are as strongly associated as the two content words in the binomial condition (e.g., *spoon and <u>fork</u>*).

(3) semantic incongruity condition – a semantically anomalous phrase whose two content words are not associated at all (e.g., *theme and fork*).

Most of the binomial expressions used in Experiment 2 were borrowed from Experiment 1. Because Experiment 2 was conducted in the USA, a number of typically British binomials were substituted with those that would not be deemed unfamiliar or infrequent by American participants (all experimental items can be found in Appendix 7). It is noteworthy that a small number of binomials (e.g., *skin and bones* and *cream and sugar*) are of low frequency in the BNC, but were presumed to be familiar expressions to speakers of American English. To confirm this intuition, a Google search was done, which verified that they were indeed frequent phrases with a preferred word order.

Items in the associate condition were formed by means of substituting Content Word 1 of the binomial with an equally strong associate (e.g., *knife and fork* → *spoon and fork*). Items in the semantic incongruity condition were formed by means of substituting Content Word 1 in the binomial with a semantically unassociated word to create an anomalous word combination (e.g., *knife and fork* → *theme and fork*). Content Word 2 (e.g., *fork*) was always the same across the three conditions. The properties of the three experimental conditions can be found in (Table 8.6).

Table 8.6. Mean phrasal frequency (absolute frequency, i.e., per 100 mil words of the BNC corpus), phrasal length, and semantic association strength for binomials, associates, and semantically incongruous phrases.

	Binomial	Associate	Incongruity
Phrasal frequency	102	0.7	0
Phrasal length (characters)	13.6	14.6	14
Association strength	0.21	0.25	0

Association strength

The University of South Florida Free Association Norms Database [9] was used to match the constituents of binomials (e.g., *knife and fork)* and associate forms (e.g., *spoon and fork)* in semantic association strength. Table 8.6 illustrates the association strength for the forward association (Word 2 responses (*fork*) to Word 1 (*knife/spoon*)). Statistical analysis showed that the two content words in the binomial and associate condition were equally strongly associated ($t(119) = -1.6, p > .05$). It is noteworthy that some constituent words (e.g., *relaxation* as in *rest and relaxation*) were not provided as responses (in such cases, a zero value was used). The association database used in this experiment is different from the one used in Experiment 1, and hence the values provided are not directly comparable across the two experiments due to different measurements used obtaining and calculating association strength.

Participants

Forty-eight healthy native speakers of American English participated in the experiment. They were drawn from the undergraduate and graduate student population at the University of Florida, Gainesville, USA (23 females, 17 males; age

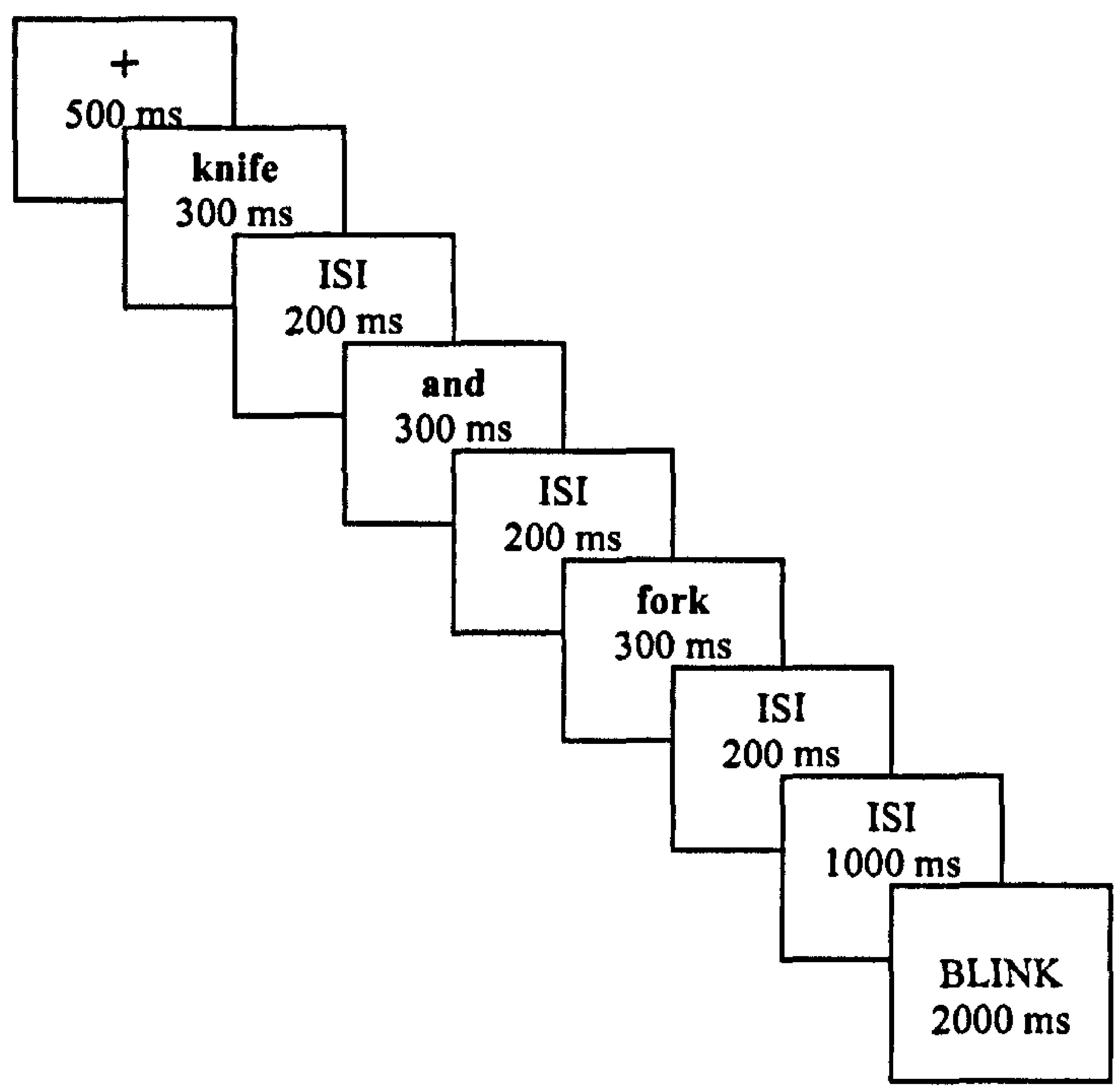
18–30; mean age 20.2). All participants were right handed as assessed by the Edinburgh handedness inventory (Oldfield, 1971), and none had suffered from any neurological diseases, or language-related impairments, as indicated by a self-report. All participants had normal or corrected-to-normal vision. All participants gave informed consent before the experiment and read an information sheet describing the EEG methodology, experimental procedure and instructions. After the experiment, they were paid $20 for their participation or were given course credit.

EEG procedure

Each participant first completed a practice session. In the experiment, the 120 triplets were pseudorandomised across three lists so that each participant read only one version of each triplet. Thus, each participant list contained 40 exemplars of the three conditions. The 120 target items (40 of each type) were intermixed with 40 fillers of the type '*Noun* and *Noun*', '*Adjective* and *Adjective*', or '*Verb* and *Verb*'. To encourage participants to read for comprehension, an animal categorisation task was performed. All filler items contained exactly one word denoting an animal. None of the experimental items contained 'animal' words. Fillers were designed in such a way that half of them had the 'animal' word in the first position (e.g., *lion and prey*), while the other half had the 'animal' word in the second position (e.g., *nest and eagle*). All stimuli were presented out of context.

Each trial started with a 500 ms fixation cross. Phrases were presented using RSVP, that is, one word at a time. The inter-stimulus interval (ISI) was 200 ms, while each word remained on the screen for 300 ms. There was a longer, 1000 ms, inter-stimulus interval after the last (critical) word so as to delay participants' blinks (Figure 8.6). After the last word of each trial, there was a 2000 ms blank screen with the word BLINK in the middle indicating that participants could blink and get ready for the next trial. Participants were urged not to start blinking until they saw the word BLINK and to stop blinking as soon as the word BLINK disappeared.

Figure 8.6. Experiment 2 presentation mode.



The words were presented in white lower case Courier New font against a black background in the center of a VGA computer screen. The viewing distance was approximately 100 cm. Participants were seated comfortably in a dimly lit, soundproof booth. Participants were instructed to read all words appearing on the screen and to press a designated button as soon as they saw a word denoting an animal. The experimental session lasted approximately twenty minutes and consisted of two blocks with a break in between the blocks during which an impedance check was performed.

EEG recording

EEG was recorded from 39 Ag/AgCl scalp electrodes, using a commercially available elastic cap with active shielding (Easy-Cap) combined with an ANT amplifier (ANT software B.V., Enschede, the Netherlands). Electrode positions were: midline - Fz, FCz, Cz, CPz, Pz; lateral left/right - FP1/2, F7/8, F5/6, F3/4, FT7/8, FC5/6, FC3/4, T7/8, C5/6, C3/4, TP7/8, CP5/6, CP3/4, P7/8, P5/6, P3/4, O1/2. Horizontal and vertical EOG was recorded from electrodes placed on the outer canthi, and below and above the right eye, respectively. Two additional electrodes were placed on the right (A2) and left (A1) mastoids. The signal was acquired using the mean of the electrodes as a common reference, but was arithmetically re-referenced off-line to the mean of the two mastoids. Electrode impedance was kept below 5 kΩ. The signal was sampled at a rate of 512 Hz, and was filtered off-line between 0.3 and 30 Hz.

Analysis and results

Data from eight participants were excluded from the analysis because of a large number of artifacts (e.g., blinks), participants' excessive body movement or lack of concentration. Thus, the data from 40 participants were included in the analysis. The EEG data were screened for eye movements, electrode drift, and other artifacts. Trials containing such artifacts were rejected (5.3% of the data). For each participant, average waveforms were computed across all remaining trials, using a 100 ms pre-critical-word baseline. The critical comparison was that of Content Word 2 across the three conditions. However, as in Experiment 1, I also looked at the amplitudes elicited by Content Word 1, as well as the conjunction 'and'.

Content Word 1

In the latency range (-100 to 1000 ms relative to the onset of the word), a series of running t-tests was performed in 24 ms bins, which shifted by 12 ms (e.g., 100-124 ms, 112-136 ms, etc.). The difference between any two waveforms was taken to be significant if five or more consecutive windows reached the significance level of $p < .05$. No significant differences for Content Word 1 were found across the three conditions (*Fs* $p > .05$). It can thus be argued that Content Word 1 in the binomial, associate, and semantic incongruity condition produced comparable amplitudes. This implies that even though Content Word 1 was not matched in length and lexical frequency across the conditions, these differences were relatively minor and did not affect participants' reading of this word.

The conjunction 'and'

As above, in the range of -100 to 1000 ms relative to the onset of the word, a series of running t-tests in 24 ms bins was performed. No significant differences for the middle word 'and' were found across the three conditions (*Fs* $p > .05$) in the -100-500 ms window.

Content Word 2

Content Word 2 was the critical word and where differences across the conditions were expected. Upon visual inspection, it became apparent that compared to the associate and incongruous condition, the binomial condition elicited a larger early positivity, peaking around 300 ms, as well as a reduced later negativity, peaking around 380 ms. The results of the running t-tests showed that the waveforms for the two critical conditions (binomial and associate) deviated significantly across a large number of electrodes (Table 8.7). Figure 8.7 shows the significant electrodes in the binomial versus associate comparison and their topography, while Figure 8.8 shows the waveforms in a selection of electrodes. The difference in the other two comparisons (binomials vs. incongruity and associate vs. incongruity) was observed across a large number of electrodes.

Table 8.7. Significant electrodes in the binomial versus associate comparison.

10-10 system	significant window (ms)	duration (ms)
F3	321 - 391	70
Fz	309 - 391	81
F4	309 - 391	81
F6	321 - 391	70
FC5	321 - 391	70
FC3	309 - 391	82
FCz	309 - 391	82
FC4	297 - 438	140
FC6	297 - 496	199
FT8	321 - 496	175
T7	321 - 438	117
C5	297 - 438	140
C3	297 - 426	128
Cz	297 - 426	128
C4	262 - 508	245
C6	297 - 508	210
T8	286 - 508	222
TP7	321 - 426	105
CP5	297 - 438	140
CP3	286 - 438	152
CPz	297 - 438	140
CP4	262 - 508	245
CP6	251 - 520	269

TP8	262 - 508	245
P7	309 - 438	128
P5	297 - 438	140
P3	297 - 426	128
Pz	297 - 426	128
P4	274 - 473	198
P6	262 - 508	245
P8	262 - 473	210
O1	297 - 426	128
O2	286 - 485	198

Figure 8.7. Binomial versus associate in the 300-500 ms window (red circles represent p < .05).

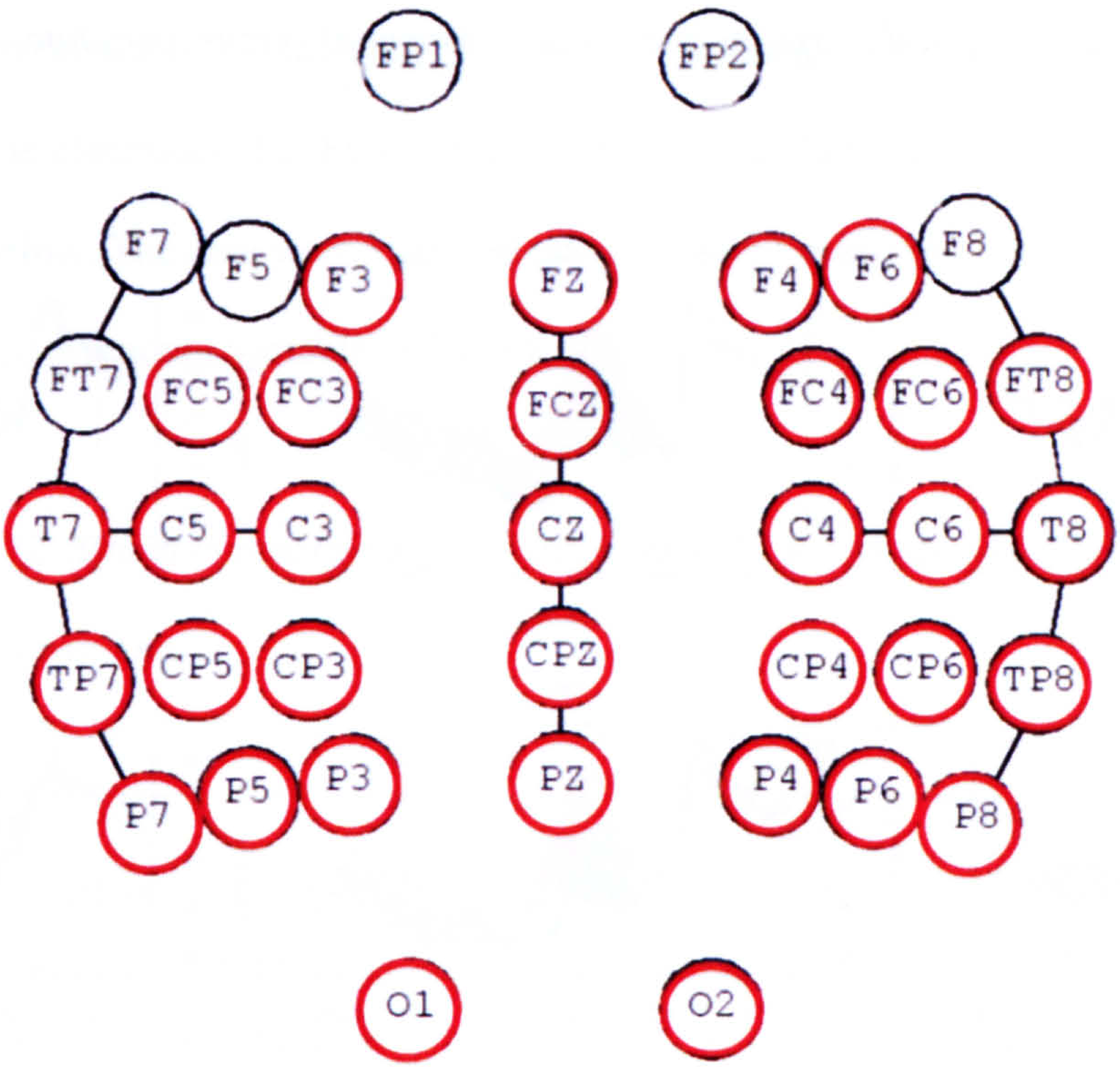
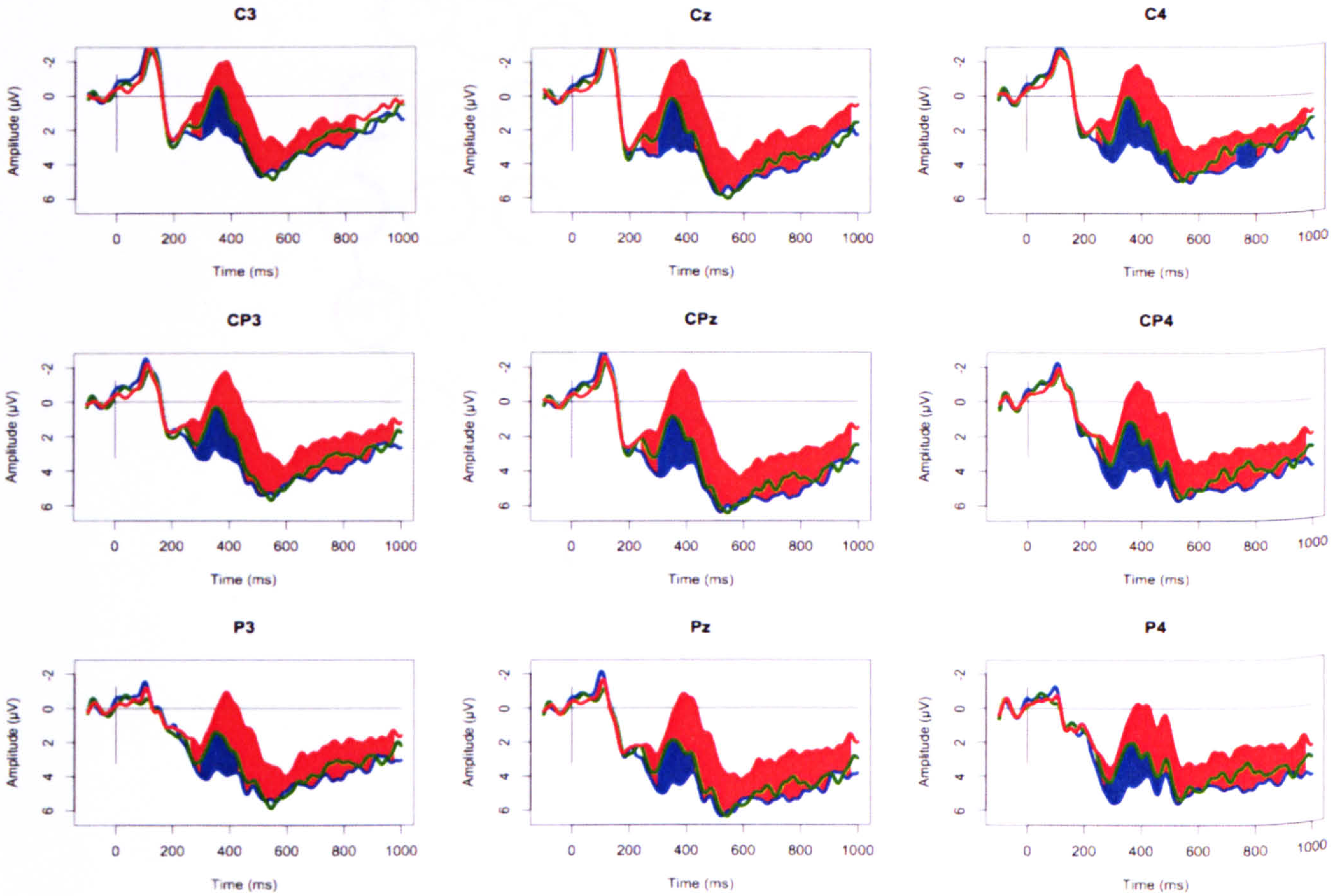


Figure 8.8. Waveforms for nine channels for Content Word 2 in the -100-1000 ms window. Blue lines indicate the binomial condition, green lines the associate, and red the semantic incongruity. Red shaded areas represent significant differences between semantic incongruity and binomial/associate conditions. Blue shaded areas represent significant differences in the critical comparison: binomial versus associate.



Upon visual inspection, the difference between the two critical conditions, binomials and associates, appeared to be distributed across almost the entire scalp (unlike Experiment 1, where the difference was significant only across 13 centro-parietal electrodes). Because so many electrodes were found to be significant, it was deemed unnecessary to include all of them in the statistical analysis. Hence, only the five Midline electrodes (Fz, FCz, Cz, CPz, Pz) were included in the statistical analysis reported below. The data were analysed using repeated measures ANOVAs treating participants as random variables (Table 8.8).

Table 8.8. Statistical comparisons for Content Word 2 for Midline electrodes in the 300-500 ms window.

	df	*F*	*p*
Condition	2,78	40.6	**
binomial vs. associate	1,39	8.6	*
binomial vs. incongruity	1,39	60.6	**
associate vs. incongruity	1,39	46.6	**

* significant at $p \leq .01$, ** significant at $p \leq .001$

A significant main effect of Condition was found. Planned comparisons revealed larger negative amplitudes in the 300-500 ms window for the semantic incongruity condition than for the binomial and associate conditions. In the critical binomial versus associate comparison, significantly deviant waveforms for the two conditions were found in the 300-500 ms window. Specifically, this difference manifested itself in a larger early positivity for the binomial condition peaking around 300 ms, and a larger later negativity for the associate condition, peaking around 380 ms (Figure 8.8).

Discussion

The above analysis revealed three findings. First, compared to the binomial and associate conditions, the semantic incongruity condition (e.g., *theme and fork*) elicited a larger N400 effect on the target word *fork*. This is consistent with Experiment 1 and with previous findings of a larger N400 on semantically incongruous words. Second, Experiment 2 results showed larger N400 amplitudes for the word *fork* in the associate condition (e.g., *spoon and fork)* than in the binomial condition (e.g., *knife and fork)*. Third, larger early positivity was observed in the binomial condition (e.g., *knife and fork*) relative to the other two conditions, peaking around 300 ms after the onset of the word *fork*. Similar to Experiment 1, this finding is interpreted as the P300 effect. It will be discussed in greater detail further on in the chapter.

It has been proposed in the literature that the N400 component serves as an index of semantic priming, a process wherein identification of a word is facilitated (as evidenced by reduced N400 amplitudes) if it is preceded by a related word (e.g., Steinschneider & Dunn, 2002). The two content words in the binomial and associate conditions in Experiment 2 were related (i.e., equally strongly associated). The two conditions, however, resulted in different waveforms. It is thus possible to hypothesise that if the processing advantage for frequent binomials (e.g., *knife and fork*) over less frequent novel phrases (e.g., *spoon and fork*) is due to their status of a conventional phrase represented in the mental lexicon, then removing the conjunction 'and' should eliminate this processing advantage, because the binomials will no longer be in the form of a fixed phrase (i.e., they will no longer be treated as a unit). In Experiment 3, this hypothesis is tested.

8.2.3. Experiment 3

Experiment 3 investigates the processing of individual constituents of binomial expressions (e.g., *knife fork*) and the constituents of novel phrases (e.g., *spoon fork).*

Materials

Materials used in Experiment 3 were identical to those used in Experiment 2, except that the conjunction 'and' was removed. Thus, the three conditions were as follows (critical word underlined):

(1) binomial condition – *knife <u>fork</u>*

(2) associate condition – *spoon <u>fork</u>*

(3) semantic incongruity – *theme <u>fork</u>*

Participants

The 48 participants who took part in Experiment 3 did so after completing Experiment 2.

EEG procedure

The EEG procedure was identical to that in Experiment 2 except for the conjunction 'and' that was absent in Experiment 3. The experimental session in Experiment 3 consisted of two blocks with a short break in-between, during which an additional impedance check was performed. Experiment 3 always followed Experiment 2.

As mentioned above, the same group of participants did both Experiment 2 and 3. To avoid repetition as much as possible, if a participant did List 1 in Experiment 2, s/he did List 2 or 3 in Experiment 3. Thus, no participant saw identical items in both experiments. However, some repetition was unavoidable. That is, if one participant saw *knife and fork* in Experiment 2, s/he would see *spoon fork* or *theme fork* in Experiment 3. Although Content Word 2 was read twice by each participant, this was the case across all conditions, and thus any repetition effect would have equally affected the three conditions.

EEG recording

EEG recording in Experiment 3 was identical to the one in Experiment 2.

Analysis and results

As in Experiment 2, the data from 40 people were included in the analysis. The EEG data were screened for eye movements, electrode drift, and other artifacts. Trials containing such artifacts were rejected (6.9% of the data). For each participant, average waveforms were computed across the remaining trials for each of the three conditions, using a 100 ms pre-critical-word baseline.

Content Word 1

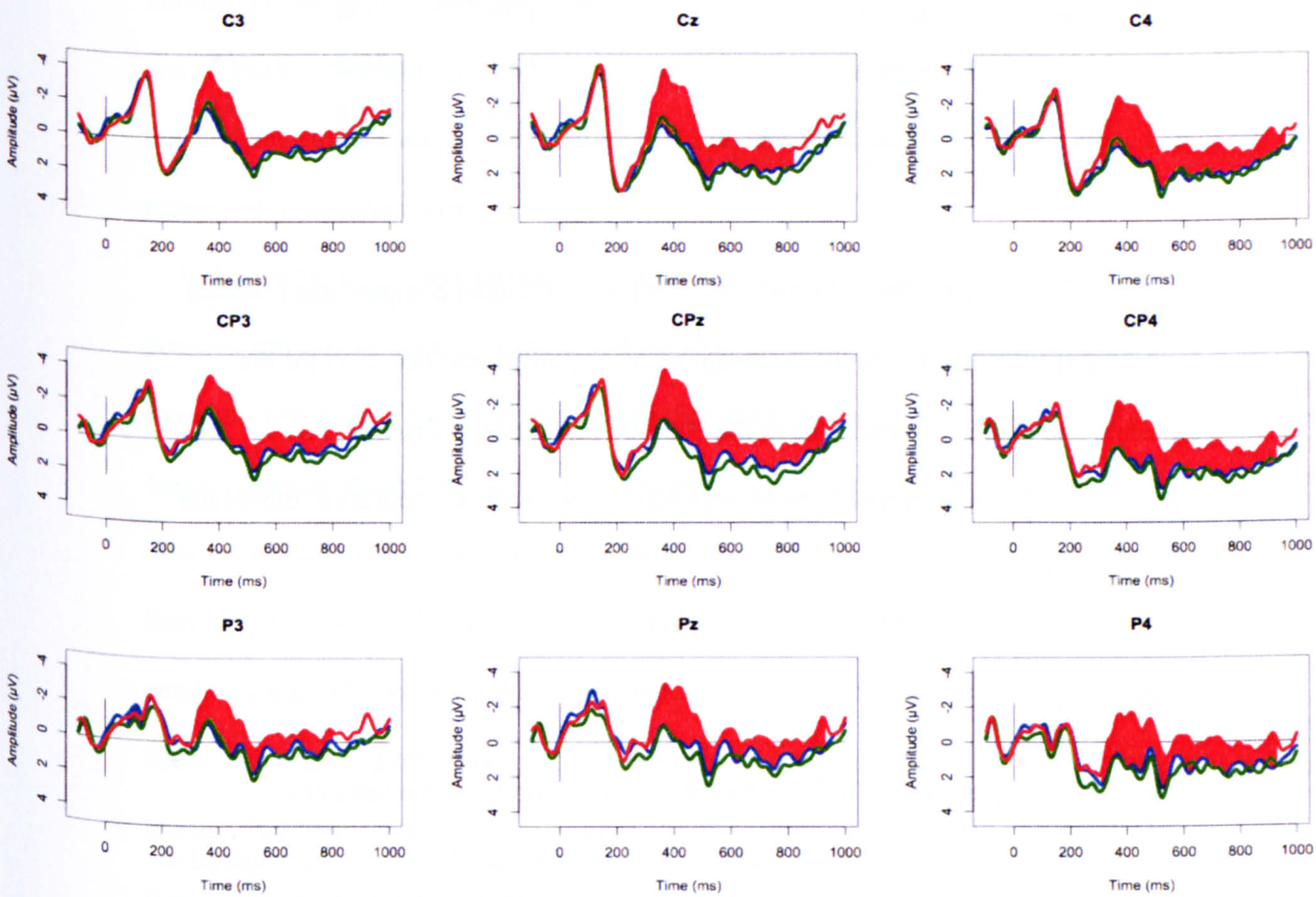
Following the same procedure as before, a series of running t-tests showed no significant differences for Content Word 1 across the three conditions (*Fs* $p > .05$).

Content Word 2

The results of the running t-tests showed that the waveforms for the two critical conditions (binomial and associate) did not differ significantly in the -100 and 1000 window (Figure 8.9). The difference in the other two comparisons (binomials vs.

incongruity and associate vs. incongruity) was found to be highly significant.

Figure 8.9. Waveforms for nine channels for Content Word 2 in the -100-1000 ms window. Blue lines indicate the binomial condition, green lines the associate, and red the semantic incongruity. Red shaded areas represent significant differences between semantic incongruity and binomial/associate condition.



The data were further analysed using repeated measures ANOVAs treating participants as random variables. As in Experiment 2, I looked at the data specific to the five Midline electrodes in the 300-500 ms window (Table 8.9).

Table 8.9. Statistical comparisons for Content Word 2 for Midline electrodes in the 300-500 ms window.

	df	*F*	*p*
Condition	2,78	32.2	**
binomial vs. associate	1,39	.244	ns
binomial vs. incongruity	1,39	38.5	**
associate vs. incongruity	1,39	59.8	**

** significant at $p \leq .001$, ns – non-significant

A significant main effect of Condition was found. Planned comparisons revealed larger negative-going waves for the semantic incongruity condition than for the binomial and associate conditions in the 300-500 ms window. Crucially, no significant differences were observed between the two critical conditions, binomials and associates.

Discussion

Similar to Experiment 2, in Experiment 3, it was found that the semantic incongruity condition (e.g., *theme fork*) elicited a significant N400 effect compared to the other two conditions. Importantly, however, in contrast to Experiment 2, it was established that when presented without the conjunction 'and', the associate (e.g., *spoon fork*) and binomial (e.g., *knife fork*) conditions exhibit identical waveforms. These results suggest that what drives the difference between the two critical conditions observed in Experiment 2 is, in fact, the phrasal (or unitary) status of binomial expressions, which is why this difference disappears completely in Experiment 3, where the same stimuli were used but where the conjunction 'and' was omitted, thus 'breaking' the unitary status.

8.3. General discussion

In a series of ERP experiments, I set out to investigate the processing of frequent familiar phrases (binomials) versus infrequent novel ones. Namely, the study had its aim to explore the neural correlates involved in phrasal processing. Crucially, one of the goals of Study 3 was to demonstrate that frequent multi-word sequences are processed in a *unitary way*.

In Experiment 1, where phrases were presented in a sentence context, larger positive amplitudes peaking around 300 ms after the onset of the critical word were found for binomials (e.g., *king and queen*) than their reversed forms (e.g., *queen and king*). Although the difference between these two conditions was found to be relatively small, it nevertheless reached significance for a group of parietal electrodes. Based on previous findings in literature, this early positivity observed in the binomial condition was interpreted as the P300 effect. In previous studies, the P300 effect has been linked (among other things) to the concept of "template matching" (e.g., Chao, Nielsen-Bohlman, & Knight, 1995; Ford, 1978; Squires, Hillyard, & Lindsay, 1973). It manifests itself in participants developing a neural representation, or a template, of a stimulus (given its preceding information). Kok (2007) argues that the closer the match between the incoming information and the template, the larger the P300 effect. As was mentioned earlier in the thesis, Roehm et al. (2007) investigated the processing of antonymous adjectives. The expected completion was shown to elicit a pronounced P300, because, as argued by Roehm and colleagues, there is a unique prediction that is either fulfilled or not.

Of greater relevance to Experiment 1 results is the study by Vespignani et al. (2009) who investigated the processing of idioms. The authors hypothesised that the expectations driven by the activation of a prefabricated chunk (e.g., idiom) should be different from those driven by general discourse-based constraints. In their study, the data were analysed before and after an idiom's recognition point. The crucial finding was that after the recognition point, idiomatic, but not literal, sentence completions resulted in the P300 effect. Because after the recognition point, only one idiom completion is viable (e.g., *take the bull by* can only be completed by *the horns*), Vespignani and colleagues proposed that the P300 effect observed in their study is the result of the categorical template matching, and that it "specifically operates for multi-word expressions ... when the compositional analysis must be integrated with the retrieval of prefabricated meaning from the semantic memory" (Vespignani et al., 2009, p. 15). The authors concluded that the electrophysiological correlates that underlie the processing of expected words in prefabricated phrases (where predictability is dependent on the knowledge of a specific phrasal configuration) differs from those that underlie the processing of expected words in sentences where predictability is down to general sentence-level information. Although, unlike idioms, binomials are used literally and are compositional, they are, nevertheless, also frequent, familiar, and highly predictable. Akin to idioms, they exhibit a canonical structure (i.e., word order), which mature language users have stored in their semantic memory. It is thus possible to take the results of Experiment 1 to support those reported in Vespignani et al. (2009). Namely, the processing of highly predictable and less so phrases differs not only in terms of reading times, as was shown in Studies 1 and 2 (Chapters 6 and 7), but also at an electrophysiological level. Crucially, the difference is observed in the case of idioms (highly predictable conventional phrases that can be used only figuratively, or both literally and figuratively), as well as completely compositional but fixed or semi-fixed phrases used literally, such as binomials.

Experiment 2 investigated the processing of frequent phrases whose words are strongly associated (e.g., *knife and fork*) versus infrequent phrases whose words are equally strongly associated (e.g., *spoon and fork*) out of sentence context. Unlike Experiment 1, where association strength was not matched in the binomial and reversed conditions, in Experiment 2, the two critical conditions were matched on association strength. The following findings were made. First, significantly larger N400 effects were observed in the associate condition (e.g., *spoon and fork)* than in the binomial condition (e.g., *knife and fork)*. Because the word *fork* in associates elicited a larger N400 than in binomials (even though association strength was matched), it is possible to suggest that the difference in the binomial and reversed conditions, which was found in Experiment 1, cannot be entirely due to the association strength not being matched. The smaller N400 waveforms observed in the binomial (compared to associate) condition further imply diminished processing costs for frequent linguistic patterns over less frequent ones. As such, this finding is in line with that of Strandburg et al. (1993) and Laurent et al. (2006) who observed reduced N400s on highly salient conventional phrases compared to matched novel ones. The result of this experiment, as well as the findings of Strandburg et al. and Laurent et al., demonstrates a processing advantage for multi-word speech.

Second, a larger early positivity, peaking around 300 ms after the onset of the critical word, was observed in the binomial condition relative to the associate one.

Similar to Experiment 1, this positivity is indicative of the P300 effect and, as such, supports the "categorical template matching" proposed by Vespignani and colleagues (2009) for multi-word expressions. Both Vespignani et al. (2009) and Roehm et al. (2007) observed the P300 on highly predictable words when presented in sentence contexts. However, Roehm and colleagues failed to replicate the P300 effect on the predictable upcoming word (e.g., *white* following *black*) when participants were presented with antonymous word pairs without a constraining sentence context. The results of Experiment 1 and 2 suggest that Content Word 2 of the binomial phrase is equally predictable in and out of sentence context and is likely to elicit the P300 in both cases. As Vespignani et al. (2009) propose, when a phrase becomes uniquely identifiable (for example, after the recognition point of an idiom), a categorical prediction mechanism operates resulting in the P300 effect.

Although the two content words in the binomial and associate conditions in Experiment 2 were equally strongly associated, they, nevertheless, exhibited deviant N400 waveforms. Because the N400 component serves (among other things) as an indicator of semantic priming (e.g., Steinschneider & Dunn, 2002), it was hypothesised that if the processing advantage for binomials over novel phrases observed in Experiment 2 was due to their status of a phrase (a prefabricated unit) in the mental lexicon, then this processing advantage should be eliminated in an experiment where the two content words are presented *without* the conjunction 'and', thus diminishing the phrasal (or unitary) status of a binomial. As a result, as proposed by Steinschneider and Dunn (2002), *knife* should prime *fork* in <u>exactly</u> the same way as *spoon* primes *fork*. However, the use of a priming paradigm should not diminish the processing advantage for *knife fork* or *spoon fork* over the anomalous *theme fork*, since in the latter, words are not associated. In Experiment 3, the above hypotheses were tested. It was found that when the conjunction 'and' was omitted, with only the two content words being presented to the participants, identical waveforms were elicited in the binomial and associate condition. The semantic incongruity condition, however, resulted in a significantly larger N400 (comparable to that in Experiment 2), suggesting that whether presented as a phrase or as a sequence of two content words, the semantic incongruity remains equally detectable. Because *no* differences were observed in the critical binomial versus associate comparison in Experiment 3, this leads to the conclusion that the processing differences between the frequent *knife and fork* and the infrequent *spoon and fork* reported in Experiment 2 are, indeed, down to the unitary (or phrasal) nature of binomial expressions.

Taken together, the results of the ERP study have added to the findings of the eye-tracking study (Study 2), providing a fuller picture of the processing of linguistic forms above the word level. While the eye-tracking findings showed that frequent linguistic patterns are read faster than infrequent ones, the results of the ERP experiments highlighted the very unitary nature of frequent phrases. Crucially, in Experiment 3, it was clearly shown that as soon as this unitary nature is distorted, the processing advantage for frequent phrases disappears and their processing starts to approximate that of infrequent phrases.

Overall, the results of the three ERP experiments suggest that, due to their frequency and predictability, binomials become encoded in the mental lexicon, and as a result, different neural correlates underlie their processing when compared to novel language. The above findings further support the view, according to which frequent multi-word sequences, such as binomial expressions, do not undergo the same semantic integration processes as instances of novel language, and that their processing is facilitated compared to that of novel language.

Chapter 9: Conclusion

9.1. General conclusions

The existence of recurrent patterns has long been acknowledged by linguists (e.g., Saussure, 1916/1966; Firth, 1957) and psychologists alike (e.g., Miller, 1956). While Saussure, proposed that two or more linguistic units can be fused into one, the "father" of collocation, John Firth, drew attention to the role of context-dependent nature of meaning (as he famously said: "You shall know a word by the company it keeps" (1957, p. 11)). The idea of 'chunking' has also been advanced in the field of psychology. In his paper on short-term memory limitations, Miller (1956) argues that chunking is an important strategy in linguistic processing. Miller proposed that, first, in order to be able to effectively process linguistic input, one has to operate with larger linguistic units – chunks; second, short-term memory has a capacity of seven plus or minus two chunks; and third, "the span of immediate memory seems to be almost independent of the number of bits per chunk" (pp. 92-93). If, indeed, language users operate with larger chunks, as well as single words, then it becomes apparent that the focus of linguistic enquiry should also be on multi-word units.

A more 'phrasal' perspective is gradually starting to gain ground. A view that has recently been gaining popularity is that language users are sensitive to frequency information at different levels (e.g., sublexical, lexical, phrasal, and clausal), and that this information affects the processing of different linguistic material (e.g., morphemes, words, multi-word phrases, and clauses), subsequently shaping our mental representations. Unsurprisingly, of all frequency effects, word frequency effects are the most well documented findings in psycholinguistic research. High frequency words are processed faster than low frequency ones in lexical decision (e.g., Balota et al., 2004; Whaley, 1978) and word naming tasks (e.g., Balota et al., 2004; Forster & Chambers, 1973), as well as in sentence comprehension (e.g., Rayner & Duffy, 1986). While frequency effects have been widely reported in word processing literature, a limited number of studies have investigated frequency effects in units larger than a word, such as two-word combinations (e.g., Bell et al., 2003), and larger syntactic structures (e.g., Frazier & Fodor, 1978; Reali & Christiansen, 2007).

With respect to longer sequences, the crucial questions to ask is: Do units above the word level, both fully compositional and less so, exhibit frequency effects in a comparable way to single words? It has recently been proposed that the frequency with which frequent multi-word sequences occur in language affects their representation and processing. For example, Sosa and MacFarlane (2002), Arnon and Snider (2010), Mondini et al. (2002), and Bannard and Mathews (2008) all report a processing advantage for frequent multi-word phrases of different kinds, such as collocations (e.g., *sort of*), two-word compounds (e.g., *red cross*), and regular compositional phrases (e.g., *don't have to worry*). These facilitative effects for frequent patterns over less frequent ones have been observed in studies with healthy adults (e.g., Arnon & Snider, 2010; Sosa & MacFarlane, 2002), speech-impaired patients (e.g., Mondini et al., 2002), and children (e.g., Bannard & Matthews, 2008). Electrophysiological studies have also suggested differences in processing patterns for frequent multi-word sequences (i.e., idioms and collocations) compared to novel phrases (e.g., Laurent et al., 2006; Strandburg et al.,1993; Vespignani et al., 2009). Furthermore, phrasal frequency effects have been observed not only in comprehension studies (as those mentioned above), but also in production studies (e.g., Bell et al., 2003; Bybee & Scheibman, 1999; Van Lancker, Canter, & Terbeek, 1981). Overall, it has been shown that frequent phrases are produced (i.e., articulated) faster and are phonetically reduced more than less frequent ones.

Although the above studies do suggest that frequent multi-word phrases are comprehended and produced differently from less frequent ones by native speakers (at the very least, they are processed faster), the evidence has nevertheless been rather scarce. With regards to nonnative speakers, the evidence is scarcer still. With this in mind, in the present thesis, I hoped to explore the issue of multi-word sequence processing and representation in native speakers, as well as to shed more light on the issue in relation to nonnative speakers.

Throughout the thesis, it is argued that multi-word sequences are processed "differently" from novel language. To be able to pin down what exactly "differently" is, in a series of studies, two techniques (eye-tracking and ERP), two participant groups (native and nonnative speakers of English), and two different types of multi-word sequences (idioms and binomial expressions) were employed.

Study 1, which uses an eye-tracking paradigm, investigated the comprehension of idioms used figuratively (*at the end of the day* – 'eventually'), literally (*at the end of the day* – 'in the evening'), as well as novel phrases (*at the end of the war*) in a first and second language. A number of findings were made. First, native speaker results suggested a robust processing advantage for idioms over novel phrases. This processing advantage for idioms suggests that they are not subject to computational processes in the same way that novel language is. Second, both full idiom and recognition point analysis indicated that in the presence of a preceding disambiguating story context, the higher frequency figurative idiom renderings were not read any faster than lower frequency literal ones. This finding highlights the important role of context. With regards to nonnative speakers, the results showed that, unlike natives, they process idioms with the same speed as novel phrases, which implies that idioms are comprehended in a more sequential fashion, akin to novel language. Further, it was found that figurative uses were processed more slowly than literal ones, even when the context biased the reader towards figurative interpretations. This suggests that figurative meanings are less strongly represented in the lexicon of nonnative speakers than in the lexicon of native speakers.

In Study 2, I explored on-line processing of another type of multi-word speech – binomial expressions. Despite being rather frequent, binomials have received no attention in psycholinguistic research. In Study 2, which made use of an eye-tracking paradigm, the following question was addressed: Are native and nonnative speakers sensitive to the frequency with which frequent and familiar but completely compositional phrases occur in language? In order to answer this question, I looked at the processing of frequent binomial expressions, such as *bride and groom,* and their infrequent reversed forms, such as *groom and bride.* As expected, native speaker reading times for binomials were faster than for their reversed forms suggesting that during their life-time, native speakers notice, register, and accumulate experiences with not just single words, but also regular compositional phrases. The nonnative data showed no overall advantage for binomials over their reversed forms. However, further analysis showed a significant processing advantage for the highest frequency binomials compared to their reversed forms, which suggests that with increased exposure to English, nonnative speakers' processing begins to approximate that of natives.

Finally, in three ERP experiments, which form Study 3, the processing of binomial expressions in a first language is further investigated. Overall, the results of the ERP experiments offered further support for the findings reported in the eye-tracking experiment. That is, the processing of binomials does not involve the same integration processes as their (novel) reversed forms. As a result, their processing is facilitated compared to novel language. This facilitation manifested itself in the presence of the P300 component for binomials (e.g., *king and queen*) compared to their reversed forms (e.g., *queen and king*); as well an increased P300 and a reduced N400 for binomials (e.g., *knife and fork*) over infrequent but equally strongly associated phrases (e.g., *spoon and fork*). Importantly, once the conjunction 'and' was removed from the stimuli and they were thus no longer in the form of a phrase, the differences between the ERP components elicited by the binomials and the strong associates completely disappeared. This indicates that the processing advantage for *knife and fork* over *spoon and fork* is likely to be due to the unitary (or phrasal) status of the former, and leads to the conclusion that different neural correlates underlie the processing of frequent familiar phrases as opposed to infrequent novel ones.

The above studies addressed the issue of processing and representation of familiar phrases from different perspectives. Namely, the use of two powerful techniques, eye-tracking and ERP, has allowed me to identify how the "different processing", that has been mentioned throughout, manifests itself with respect to idioms and binomial expressions versus novel phrases. The three studies point to the following:

- Frequent multi-word sequences are processed faster by native speakers. This is evidenced by shorter and fewer fixations made both on idioms and binomials when compared to novel strings.

- Frequent multi-word sequences are also processed faster by nonnative speakers, but *only* if these sequences are of a very high frequency. Nonnatives have been shown to have a *lexicon in transition*, that is, their processing begins to approximate that of natives only with respect to those items that they have encountered a sufficient number of times. This suggests that there may be a threshold for nonnative speakers when their processing becomes more native-like. Until this threshold has been reached, their processing of multi-word speech is more sequential.

- The processing advantage for binomials observed in the ERP studies suggests that "different processing" further presupposes processing differences at an electrophysiological level. The fact that different ERP components were shown to be involved in the processing of frequent and predictable, and infrequent and hence less predictable linguistic forms signifies that processing is not only quantitatively different (as was shown in the eye-tracking experiments), but is also qualitatively dissimilar.

- Native speakers process frequent multi-word sequences akin to a unit, or a chunk. This unitary nature is what distinguishes frequent and infrequent linguistic patterns at the level of representation and affects their on-line processing.

The fact that multi-word speech is processed qualitatively and quantitatively differently from novel speech has two major implications for linguistic theory. The first regards the nature of linguistic representation. It appears that the occurrence of a psychological event, a phrase in this case, leaves some sort of a trace in one's memory that facilitates its further reoccurrence. Through recurrence, even highly complex events can become routinised, and as a result, be executed effortlessly. This process is known as 'routinisation', or 'automatisation' (e.g., Segalowitz, 2003). The results outlined above suggest that due to their frequency of occurrence, multi-word sequences have become automatised to such an extent that they have become represented in the mental lexicon. What being represented presupposes is best put into Langacker's (1987) words: when a complex structure becomes a "pre-packaged" assembly, that is, it no longer requires conscious attention to its parts or their arrangement, then it is considered to have acquired a unitary status. This also implies that the unit becomes represented, or encoded, in the mental lexicon.

It is worth noting, that I am not arguing that the above findings entail that frequent multi-word expressions are stored and processed as *unanalysed wholes*. While this may be the case for the very frequent sequences, as has been proposed by some researchers (e.g., Bybee, 2002), it is not a claim I want to make, nor was it a question I set out to investigate at the beginning of my research. The results of the studies, I believe, have implications with regards to the way language is learnt, processed, and represented. However, they cannot be taken to indicate that idioms and binomial expressions are accessed as unanalysed wholes, because none of the studies presented in the thesis investigated whether the individual components of the multi-word sequences in question were activated or not. Furthermore, as was mentioned in Chapter 3, recent studies show that even the most idiosyncratic and arguably 'word-like' of all of the multi-word sequences – idioms – exhibit, at least to some degree, evidence of internal structure (e.g., Cutting & Bock, 1997; Konopka & Bock, 2009).

My proposition is that due to their frequency of occurrence, frequent multi-word sequences are represented in the lexicon of a native speaker, and, to a lesser degree, in the lexicon of a nonnative speaker. The issue of representation raised throughout the thesis does not, in my understanding, equal that of *holistic storage*. Whether instances of highly frequent linguistic patterns are stored and processed as wholes, that is without access to the their constituent parts, shall remain a topic of future research.

Second, the results of the studies have certain implications for the theories of language learning. Words have traditionally been viewed as primary units of language acquisition in first and second language learning. Chapter 4 reviewed a number of studies that focus on the acquisition and use of multi-word sequences in children and adults. Although these studies highlight the crucial role of multi-word sequences in language acquisition, many of the studies base their conclusions on rather naturalistic observations, providing little or no experimental evidence. The empirical finding that units above the word level may also serve as units of representation and processing in mature language users entails an interesting possibility with regards to the role that such units may play in language learning.

More generally, the results presented in the current thesis can be taken to support a number of usage-based (e.g., Bybee, 1998; Goldberg, 1995, 2006; Langacker, 1987; Tomasello, 2003) and exemplar-based theories (e.g, Bod, 1998, 2006; Pierrehumbert, 2001). At the core of these theories lies the idea that language learning and processing are affected by the amount of experience that language users have with linguistic exemplars. According to the proponents of these theories, all linguistic material is represented and processed in a comparable way, and frequency effects should be equally observable in smaller, as well as larger units: morphemes, complex words, regular compositional phrases, and more idiosyncratic ones. As Bybee (2006) argues, new experiences with linguistic exemplars play a pivotal role in shaping memory representations. Bod (2006) further notes that the allocation of representations to linguistic exemplars is accomplished purely on the basis of statistics. Thus, these usage-based and exemplar-based views predict faster processing for all frequent events over less frequent ones, be they words or phrases.

The results of the studies presented in the thesis are also in line with connectionist approach to language acquisition and processing (e.g., Christiansen & Chater, 1999; Elman, 1990; Rumelhart & McClelland, 1986). According to connectionist theory, linguistic units do not exist in isolation; rather, they form and exist in relationships (networks) with each other. Similar to the above two theories, connectionism puts an emphasis on statistical properties of the input in language learning and processing (e.g., Saffran, Aslin, & Newport, 1996) and argues that the same mechanisms operate for regular and irregular forms. Harris (1996) argues that a lexicon containing variable-sized units fits well into a connectionist framework, according to which the units of representation are not part of a 'fixed architecture', but appear via extracting regularities. The more strongly associated the structures are, the more likely they are to facilitate each other.

Usage-based, exemplar-based, and connectionist models differ in a number of important ways, for example in the use of symbolic and non-symbolic representations (e.g., Bybee & McClelland, 2005). However, what they do have in common is the idea that there is no obvious distinction between a stored and computed linguistic event, and, thus, all linguistic information, irrespective of its internal structure, is represented and processed in an analogous way, and hence should be similarly affected by experience (i.e., the frequency of occurrence).

To sum up, the amount of research and the wealth of findings that exist with respect to word recognition are rather impressive. What we know about the processing and representation of units larger than a traditional word, in both first and second language, is very little indeed. I believe the findings reported in the present thesis fill an empirical gap in the respect that they have demonstrated that the units that language users attend to during language comprehension are not limited to morphemes and words, but extend to multi-word sequences as well.

9.2. Limitations

The studies presented in the thesis are not without limitations. In Studies 1 and 2, I investigate the processing of idioms and binomial expressions by native and nonnative speakers of English. One of the drawbacks of the two experiments is that the nonnative participants came from various L1 backgrounds and thus, it was not possible to address the issue of L1 interference or facilitation.

Another limitation of Study 1 (where idioms were investigated) is that the relatively low number of items that were found to be familiar to nonnative speakers did not allow me to carry out a larger investigation, which could distinguish between the two types of idioms that were discussed in Chapter 3: decomposable and non-decomposable idioms. It has been argued that the two idiom types may be processed differently by native speakers, but little, if any, research has been done with nonnative speakers. Decomposable and non-decomposable idiom processing in a second language thus remains a largely under-researched, albeit interesting, area.

One of the limitations of the ERP methodology is that in order to obtain recordings with a good stimulus-to-noise ratio, an ERP experiment requires a large number of trials and items in each condition. Thus, unlike Study 2, where eye-tracking was used, Study 3 required a much larger number of items. Because their selection was already constrained by the need to match the items with controls in association strength and other factors, Study 3 also included a number of binomials that can be used both literally and figuratively (e.g., *bread and butter*), as well as binomials of the type *trial and error*, where the word order preference is marked in the sense that there is some chronological order to it. Although such items were included in all three ERP experiments, I, nonetheless, believe this factor was unlikely to influence the pattern of results obtained, as the number of such ambiguous items was very low.

Another limitation of the ERP methodology is that in order to avoid eye movements and to control the time-locking of the ERPs to critical stimuli, experimental stimuli need to be presented using RSVP (rapid serial visual presentation), that is, one word at a time. This way of stimulus presentation may create an additional load on working memory, which is absent in normal reading. Further, because it is necessary to control the time-locking of the ERPs to the critical stimulus, ERP reading experiments can never be self-paced. Finally, it is noteworthy that ERP interpretation can be obscured by overlapping components (e.g., Kutas & Van Petten, 1994). For example, the lexical decision task may result in the overlapping P300 and N400 because these components have a similar latency window. Thus, further, more fine-grained, analyses are needed in Study 3 (e.g., the independent component analysis (ICA)) in order to investigate the observed differences in greater detail, as well as to disentangle the P300 and N400 components.

9.3. Future directions

The present research has raised a number of important questions. With regards to idioms, their processing was investigated when a highly disambiguating context preceded the idiom used literally and figuratively. No statistically significant differences in the processing of the two renderings were found in the full idiom analysis, or in the recognition point analysis. This suggests that the processing of the two meanings was similar and highlights the important role of context in the processing of low frequency forms (i.e., the literal idiom meaning). However, it is unclear what processing patterns one might observe in the absence of a constraining context, or when the context follows the idioms, rather than precedes it (as was the case in my experiment). Because the figurative idiom uses are more frequent than literal ones, it may be that in this case, we might observe a speed-up for the frequent figurative meaning over the infrequent literal one (as the configuration hypothesis would predict (Caccairi & Tabossi, 1988)). This might be particularly the case after the recognition point.

Further, while researchers have looked at the processing of decomposable versus non-decomposable idioms (e.g., Gibbs & Nayak, 1989; Gibbs, Nayak, & Cutting, 1989; Titone & Connine, 1999), no study has investigated the processing of decomposable and non-decomposable idioms before and after the recognition point. The manipulations of the context (before and after the idiom, or completely neutral context), idiom decomposability (decomposable and non-decomposable), as well as language proficiency (native and nonnative speakers) are likely to allow for a more detailed and informative picture of idiom processing to immerge.

In recent years, there has been a strong interest in various aspects of bilingualism, for example, the nature of the *bilingual lexicon*. A number of studies have been conducted with bilinguals of various L1s and a number of models have been proposed with regards to *bilingual word activation* (e.g., Dijkstra & van Heuven, 2002; Green, 1998). Current research suggests that bilingual speakers access their two languages simultaneously (e.g., Dijkstra & van Heuven, 1998; Dijkstra, van Heuven, & Grainger, 1998). It has also been proposed that in the bilingual lexicon, translation equivalents in the two languages are connected via associative links (e.g., De Groot, 1992; Kroll & Stewart, 1994). That is, when bilinguals process words in one language, they necessarily activate translations in the other language (e.g., Hermans, Bongaerts, De Bot, & Schreuder, 1998). For example, De Groot (1992) found that translation equivalents are activated more for those words that share form than for words that do not. These studies, as well as the models proposed, have assumed a word to be a basic unit of language. However, throughout the thesis, it has been argued that highly frequent multi-word units, such as idioms and binomials, are also part of the lexicon of native and, to a lesser degree, nonnative speakers. If such recurrent multi-word expressions form part of the bilingual lexicon, then their processing should also be investigated along with that of singe words, and the models of bilingual processing should be able to account for lexical, as well as phrasal processing.

Further research is also necessary on nonnative phrasal processing using event-related brain potentials. Study 3, which looked at the processing of binomial expressions using ERP, dealt only with native speakers. However, it would be interesting to look at the pattern of activation of frequent phrases in proficient second language learners. In the eye-tracking experiment, it was found that, overall, nonnatives were not sensitive to frequent phrases compared to infrequent ones. However, when I looked at the very high frequency phrases, their processing diverged from that of infrequent phrases. Because the ERP technique taps into different processes compared to eye-tracking, it is conceivable that we may observe differences between binomials and their reversed forms for all binomials, both high and low frequency ones. It will also be elucidating to compare the interplay between the P300 and the N400 components with regards to nonnative speakers and to compare their processing to that of native speakers. So far, it has been argued that nonnatives process multi-word speech in a more sequential way and that their sensitivity to phrasal frequency is rather limited. However, future research using the ERP methodology, may be able to shed more light on phrasal processing in a second language.

Further, computational modelling may also be able to shed light on the representation and processing of frequent phrases. In recent years, there has been a surge in probabilistic modelling (for an overview, see Jurafsky, 2003). Jurafsky (2003) proposed that language is based on statistical mechanisms and that humans are, in fact, 'probabilistic reasoners'. This proposition is based on a wealth of psycholinguistic research that has shown that probabilities of various kinds play a crucial role in language comprehension, production, and learning. According to Jurafsky, et al. (2001), the language processor stores probabilistic relations between words. Given that the core properties of multi-word speech are that it is both frequent and has a high cloze probability, which have been the focus of models of single word processing, future models may try to extrapolate this to units larger than a single word.

The results presented above also have important pedagogical implications (e.g., how to better teach a foreign language in a classroom environment). It has been widely acknowledged that multi-word speech is ubiquitous and that it plays a fundamental role in both child naturalistic and adult classroom-based language learning. Its appropriate use has been recognised as a prerequisite for any second/foreign language learner who wants to achieve high proficiency and be accepted in an L2 community. However, it has also been documented that second language learners underuse native-like multi-word sequences and tend to use a large number of anomalous word-combinations that are grammatically correct but are simply 'not how native speakers say it' (e.g., see Chapter 4). This may be due to how languages are taught at schools and universities in learners' home countries, which has resulted from the view that has dominated linguistics in the past decades, namely, that the main unit of language acquisition and use is a word. Although in the thesis, I have not directly investigated pedagogical aspects of multi-word speech acquisition and use, it is apparent that future research does not only lie in the area of psycholinguistics, but also in the field of second language pedagogy. In fact, my interest in multi-word speech stems precisely from the research done in second language acquisition. In this field, it has long been proposed that multi-word speech differs radically from novel speech (for example, it is proposed that multi-word sequences are stored and retrieved holistically); however, little or no empirical evidence has been offered to support this claim. It is hoped that future research on multi-word sequences will not limit itself to corpus evidence and naturalistic observations with second language learners, but will strive to use experimental tools available in the field of psychology, such as, for example, reaction-time, eye-tracking, and ERP.

Last but not least, the results of the present studies are limited to idioms and binomial expressions. However, as was indicated in Chapter 2, the phenomenon of multi-word speech encompasses a large number of various multi-word sequences, including, but not limited to, phrasal and prepositional verbs, grammatical and lexical collocations, speech routines, and grammatical constructions. Future research should thus aim to incorporate a full array of multi-word sequences.

To conclude, it is probably fair to say that, with the exception of the research done on idioms, there is no established tradition in psycholinguistics that investigates how language users process units larger than a single word. The division of research into that pertinent to morphemes and words, and whole sentences, has been so deeply entrenched in psycholinguistics that few have attempted to address the issue of phrasal representation and processing. However, I hope, the studies reported in the present thesis have raised some important questions. I would like to finish my thesis as I started it, namely, by saying that an understanding of the use of familiar phrases is necessary to the understanding of the use of language as a whole (Becker, 1975).

Footnotes

[1] Not all idioms have both a figurative and literal interpretation. For example *shoot the breeze* meaning 'to talk without a purpose' can be used figuratively but not literally. Such idioms were <u>not</u> included in the current study.

[2] Because *war* was not provided as a completion for *at the end of* ___, it was decided that the novel phrases, such as *at the end of the war*, did not have a recognition point in the way that the idioms did. Thus, they were excluded from the recognition point analysis, as there was no point at which the completion *war* could be considered to be predictable from the previous portion of the phrase.

[3] The recognition point identification was done out of context, and was thus taken to be the same for literal and figurative meanings. It is possible that during the experiment, when the biasing context preceded the idiom, the recognition point shifted closer to the beginning of the idiom. However, because the context was designed to bias either a literal or figurative interpretation, if the recognition point did shift towards the beginning, it should have shifted for both meanings. Further, if we look at the target idioms (Appendix 2), it seems unlikely that the recognition point could have shifted closer to the beginning of the phrase in either context, as only one or two words would remain, leaving many likely completions (e.g., *at the* is unlikely to be predictive of *at the end of the day*).

[4] To establish how many times an idiom is used literally and figuratively, a search of the BNC was conducted. For idioms having more than 100 occurrences, I only looked at the first 100 of them (e.g., *on the other hand* appeared 5311 times), while for idioms having fewer than 100 occurrences I looked at every instance. The output was then rated as being either a figurative or literal use.

[5] The majority of binomials are regular expressions that are used literally. However, a few binomials can be used both literally and figuratively (i.e., *bread and butter*). Such binomials were not included in this study.

[6] This is not the case in some binomials. For example, in those expressions where the order of events plays an important role (with the 'V and V' structure in particular), the meaning does indeed change if the expression is reversed (e.g., *hit and run*). Such expressions were not included in this study.

[7] The Edinburgh Associative Thesaurus is used frequently for word association norms. More information about the norms, data collection and analysis can be found at http://www.eat.rl.ac.uk/.

[8] CELEX is a lexical database developed by Baayen, Piepenbrock, and van Rijn (1995). Available at http://celex.mpi.nl/.

[9] The University of South Florida Free Association Norms Database is frequently used for word association norms. More information about the norms, data collection and analysis can be found at http://w3.usf.edu/FreeAssociation/.

References

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, 23: 275–290.

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

Abkarian, G., Jones, A., & West, G. (1992). Young children's idiom comprehension: trying to get the picture. *Journal of Speech and Hearing Research*, 35, 580–587.

Ackerman, B. (1982). On comprehending idioms: Do children get the picture? *Journal of Experimental Child Psychology*, 33, 439–454.

Altarriba, J., Kroll, J., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed language sentences: Evidence from eye fixation and naming times. *Memory and Cognition, 24,* 477-492.

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

Baayen, H., Piepenbrock, R., & van Rijn, H. (1995). Available at http://celex.mpi.nl/.

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

Balota, D. (1994). Visual word recognition: The journey from features to meaning. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics*. Academic Press, 303-358.

Balota, D., & Chumbley, J. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance, 10*, 340-357.

Balota, D., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology,* 17, 364-390.

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, E., Bretherton, I. & Snyder, L. (1988) *From first words to grammar: individual differences and dissociable mechanisms,* Cambridge. New York, Cambridge University Press.

Becker, J. (1975). The phrasal lexicon. In R. Shank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing.* Cambridge, MA: Bolt Beranek and Newman, 60-63.

Bell, A., Brenier, J., 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.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, & M., Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America,* 113, 1001-1024.

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

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English.* Harlow: Longman.

Bihrle, A., Brownell, H., & Gardner, H. (1986). Comprehension of humorous and nonhumorous materials by left- and right-brain damaged patients. *Brain and Cognition,* 5, 399–411.

Biskup, D. (1992). L1 influence on learners' renderings of English collocations: A Polish/German empirical study. In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics*. Basingstoke, UK: Macmillan, 85–93.

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

Bod, R. (1998). *Beyond Grammar: An Experience-based Theory of Language*. Center for the Study of Language and Information, CA.

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

Bod, R. (2001). Sentence memory: storage vs. computation of frequent sentences. *Talk Presented at CUNY2001*, University of Pennsylvania, USA.

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

Bod, R., Hay, J., & Jannedy, S. (2003). Introduction. In R. Bod, J. Hay, and S. Jannedy (Eds). *Probabilistic Linguistics*. Cambridge: The MIT Press, 1-10.

Bolander, M. (1989). Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In Hyltenstam, K. & Obler, L. (Eds.) *Bilingualism Across the Lifespan: Aspects of Acquisition, Maturity and Loss*. Cambridge: Cambridge University Press.

Bolinger, D. (1962). Binomials and pitch accent. *Lingua*, 11, 34-44.

Bolinger, D. (1975). *Aspects of Language*. Harcourt, Brace, Jovanovich, New York.

Borensztajn, G., Zuidema, W., & Bod, R. (2009). Children's Grammars Grow More Abstract with Age-Evidence from an Automatic Procedure for Identifying the Productive Units of Language. *Topics in Cognitive Science*, 1, 175-188.

Bretherton, I., McNew, S., Snyder, L., & Bates, E. (1983). Individual differences at

20 months. *Journal of Child Language*, 10, 293–320.

Brinton, L., & Traugott, E. (2005). *Lexicalisation and Language Change.* Cambridge University Press.

Brown, R., & Hanlon, C. (1970) Derivational complexity and order of acquisition in child speech. In Hayes, J. R. (Ed.) *Cognition and the Development of Language.* London, John Wiley and Sons, Inc.

Brownell, H., & Stringfellow, A. (1999). Making requests: Illustrations of how right-hemisphere brain damage can affect discourse production. *Brain and Language,* 68, 442–465.

Brownell, H., Michel, D., Powelson, J., & Gardner, H. (1983). Surprise but not coherence: sensitivity to verbal humor in right-hemisphere patients. *Brain and Language,* 18, 20–27.

Brysbaert, M., & Vitu, E (1998). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception.* Oxford, England: Elsevier, 125-148.

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

Bybee , J. (2007). *Frequency of Use and the Organization of Language.* Oxford: Oxford University Press.

Bybee, J. (1985). *Morphology: A Study of Relation Between Meaning and Form.* Amsterdam: Benjamins.

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

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

Bybee, J. (2000). The phonology of the lexicon. In M. Barlow & S. Kemmer (Eds.), *Usage-based Models of Language*. Stanford, CA: CSLI Publications, 65-85.

Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, 24, 215-221.

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.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degree of constituency: The reduction of *don't* in American English. *Linguistics*, 37, 575–596.

Cacciari, C., & Glucksberg, S. (1991). Understanding idiomatic expressions: The contribution of word meanings. In G. Simpson (Ed.), *Understanding Word and Sentence*. Amsterdam: Elsevier, North-Holland, 251-273.

Cacciari, C., & Levorato, M. (1989). How children understand idioms in discourse. *Journal of Child Language*, 16, 387–405.

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

Carpenter, P., & Just, M. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye Movements in Reading: Perceptual and Language Processes*. New York: Academic Press, 275-307.

Chao, L., Nielsen-Bohlman, L., & Knight, R. (1995). Auditory event-related potentials dissociate early and late memory processes. *Electroencephalography and Clinical Neurophysiology*, *96*, 157–168.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Christiansen, M., & N. Chater. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23, 157-205.

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

Clark, R. (1974). Performing without competence. *Journal of Child Language*, 1, 1-10.

Colombo, L. (1993). The comprehension of ambiguous idioms in context. In C. Cacciari & P. Tabossi (Eds.), *Idioms: Processing, Structure, and Interpretation*. Hillsdale, NJ: Erlbaum, 163-200.

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.

Cooper, W., & Ross, J. (1975). World order. In R. Grossman, L. San., & T. Vance (Eds.), *Papers from the Parasession on Functionalism*. Chicago: Chicago Linguistic Society.

Coulson, S., King, J., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13, 21–58.

Cowie, A. (1998). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Croft, W., & Cruse, D. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Cruttenden, A. (1981). Item-learning and system-learning. *Journal of Psycholinguist Research*, 10(1), 79-88.

Curran, T., Tucker, D., Kutas, M., & Posner, M. (1993). Topography of the N400: Brain electrical activity reflecting semantic expectancy. *Electroencephalography and Clinical Neurophysiology*, 88, 188–209.

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.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: A case for contrastive analysis. *Studies in Second Language Acquisition,* 7(1), 73–79.

DeGroot, A. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 1001-1018.

Desmedt, J. (1980). P300 in serial tasks: An essential post-decision closure mechanism. In H. Kornhuber & L. Deecke (Eds.), *Motivation, Motor and Sensory Processes of the Brain: Electrical Potentials, Behavior and Clinical Use.* Amsterdam: Elsevier, 682– 686.

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

Dijkstra, A., van Heuven, W. J. B., & Grainger, J. (1998). Simulating competitor effects with the Bilingual Interactive Activation Model. *Psychologica Belgica*, 38, 177-196.

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

Donchin, E. (1981). Surprise! . . . Surprise? *Psychophysiology, 18,* 493–513.

Donchin, E., & Fabiani, M. (1991) The use of event-related brain potentials in the study of memory: Is P300 a measure of event distinctiveness? In J. R. Jennings & M. Coles (Eds.) *Handbook of Cognitive Psychophysiology: Central and Autonomic Nervous System Approaches*. New York: John Wiley, 471– 498.

Donchin, E., & Coles, M. (1988). Is the P300 component a manifestation of context-updating? *Behavioral and Brain Sciences*, 11, 355–372.

Drew, P., & Holt, E. (1988). Complainable matters: the use of idiomatic expressions in making complaints. *Social Problems,* 35, 398–417.

Drieghe, D., Brysbaert, M., & Desmet, T. (2005) Parafoveal-on-foveal effects on eye movements in text reading: Does an extra space make a difference? *Vision Research*, 45 (13), 1693-1706.

Duncan-Johnson, C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*, 456 – 467.

Duyck, W., van Assche, E., Drighe, D., & Hartsuiker, R. (2007). Visual word recognition by bilinguals in a sentence context: evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 663-679.

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

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

Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review.* 112(4), 777-813.

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

Espir, L., & Rose, F. (1970). The Basic Neurology of Speech. Oxford: Blackwell.

Federmeier, K., & Kutas, M. (1999). A rose by any other name: Long term memory structure and sentence processing. *Journal of Memory and Language*, 41, 469–495.

Fenk-Oczlon, G. (1989). Word frequency and word order in freezes. *Linguistics*, 27, 517-556.

Fillmore, C., Kay, P., & O'connor, M. (1988) Regularity and idiomaticity in grammatical constructions: the case of *let alone. Language,* 64, 500-38.

Finnigan S., Humphreys M., Dennis S., & Geffen G. (2002). ERP 'old/new' effects: memory strength and decisional factor(s). *Neuropsychologia*, 40, 2288–304.

Firth, J. (1957). A synopsis of linguistic theory, 1930-1955. In *Studies in Linguistic Analysis*. Oxford: Blackwell, 1-32.

Foldi, N. (1987). Appreciation of pragmatic interpretations of indirect commands: comparison of right- and left hemisphere brain-damaged patients. *Brain and Language,* 31, 88–108.

Ford, J. (1978). Does P300 reflect template match 0mismatch? In D. A. Otto (Ed.) *Multidisciplinary Perspectives in Event-related Brain Potential Research,* 181–183.

Forster, K. (1976). Accessing the mental lexicon. In R. Wales and E. Walter (Eds.). *New Approaches to Language Mechanisms*. Amsterdam: North Holland Publishing, 257-284.

Forster, K., & Chambers, S. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior,* 12, 627-635.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning. Teaching and Testing.* Harlow: Longman, 75-93.

Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech,* 31, 307-319.

Fraser, B. (1970). Idioms within a transformational grammar. *Foundation of Language,* 6, 22-42.

Frazier, L., & Fodor, J. (1978). The sausage machine: A new two-stage parsing model. *Cognition,* 66, 291-325.

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: a review of methodologies and experimental findings. *Second Language Research,* 21(2), 175-198.

Gardner, H., Winner, E., Bechhofer, R., & Wolf, D. (1978). The development of figurative language. In K., Nelson (Ed.), *Children's Language,* Vol. 2. New York: Gardner Press.

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

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

Gibbs, R. (1987). Linguistic factors in children's understanding of idioms. *Journal of Child Language,* 14, 569–586.

Gibbs, R. (1991). Semantic analyzability in children's understanding of idioms.

American Speech Language Hearing Association, 34, 613–620.

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

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

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.

Glucksberg, S. (1989). Metaphors in conversation: how are they understood? Why are they used? *Metaphor and Symbolic Activity*, 4, 125-43.

Goldberg, A. (1995). *Constructions*. Chicago: University of Chicago Press.

Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications,* Oxford: Oxford University Press, 79-100.

Green, D. (1998). Mental lexicon of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition,* 1, 67-81.

Gregory, M., Raymond, W., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society*, 35, 151-166.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438-441.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439–83.

Hahne, A., & Friederici, A. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11, 194–205.

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning, 24*(2), 287-297.

Hakuta, K. (1976). A case study of a Japanese child learning English. *Language Learning, 26*, 321-351.

Hanania, E., & Gradman, H. (1977). Acquisition of English structures: A case study of an adult native speaker of Arabic in an English-speaking environment. *Language Learning, 27*(1), 75-91.

Harris, C. (1996). Recognition of common word combinations: Towards a lexicon of variable sized units. *Research Report*, Psychology Department, Boston University, MA.

Harris, C. (1998). Psycholinguistic studies of entrenchment. In J.-P. Koenig (Ed.), *Discourse and Cognition: Bridging the Gap*, Stanford: CSI.I, 55–70.

Hauk, O., & Pulvermuller, F. (2004) Effects of word length and frequency on the human event-related potential, *Clinical Neurophysiology*, 115(5), 1090-1103.

Hebb, D. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.

Henderson, J., & Ferreira, E. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 417-429.

Heringer, J. (1976). Idioms and lexicalization in English. In J. P. Kimball (Ed.), *Syntax and Semantics*, Vol. 9: *The grammar of Causative Constructions.* New York Academic Press.

Hermans, D., Bongaerts, T., de Bot, K. & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1, 213-230.

Horgan, D. (1981). Rate of language acquisition and noun emphasis. *Journal of Psycholinguistic Research*, 10 (6), 629-640.

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

Huang, J., & Hatch, E. (1978). A Chinese child's acquisition of English. In E. M. Hatch (Ed.), *Second Language Acquisition: A Book of Readings.* Rowley, MA: Newbury House, 118-131.

Huebner, T. (1983). *A Longitudinal Analysis of the Acquisition of English.* Ann Arbor Karoma Publishers.

Hulstijn, J., & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition*, 11(3), 241–255.

Hyönä, J. (1993). Effects of thematic and lexical priming on readers' eye movements. *Scandinavian Journal of Psychology, 34,* 293-304.

Hyönä, J., & Bertram, R. (2004). Do frequency characteristics of non-fixated words influence the processing of non-fixated words during reading? *European Journal of Cognitive Psychology*, 16, 104-127.

Inhoff, A. (1989). Parafoveal processing of words and saccade computation during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance,* 15, 544-555.

Inhoff, A., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics,* 40, 431-439.

Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E. van der Linded, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and Psychological Perspectives.* Hillside, NJ: Erlbaum, 133-165.

Jackendoff, R. (2002). What's in the lexicon? In S. Nooteboom, F. Weerman, & F. Wijnen (Eds.). *Storage and Computation in the Language Faculty.* Dordrecht, Boston, London: Kluwer, Academic Publishers, 23-58.

Jackson, J. (1874). On the nature of the duality of the brain. In J. Taylor, (Ed.), Selected writings of John Hughlings Jackson. Vol 2. London: Hodder and Stoughton, 1932, 129–145.

Jescheniak, J., and Levelt, W. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 20: 824–843.

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

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

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

Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354.

Kaan, E. (2007). Event-Related Potentials and Language Processing: A Brief Overview. *Language and Linguistics Compass*, 1(6), 571-591.

Kaan, E., & Swaab, T. (2003). Electrophysiological evidence for serial sentence processing: a comparison between non-preferred and ungrammatical continuations. *Cognitive Brain Research,* 17, 621–635.

Kaan, E., Harris A., Gibson E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15, 159–201.

Kaplan, A., Brownell, H., Jacobs, R., & Gardner, H. (1990). The effects of right hemisphere damage on the pragmatic interpretation of conversational remarks. *Brain and Language*, 38, 315–333.

Katz, J. (1973) Compositionality, idiomaticity, and lexical substitution. In Anderson, S. & Kiparsky, P. (Eds.) *A festschrift for Morris Halle.* New York, Holt, Reinhart, & Winston, 357-376.

Katz, J., & Postal, P. (1963) Semantic interpretation of idioms and sentences containing them. *M.I.T. research laboratory of electronics quarterly progress report*, 275-82.

Kempler, D., Van Lancker, D., Marchman, V., & Bates, E. (1999). Idiom comprehension in children and adults with unilateral brain damage. *Developmental Neuropsychology*, 15(3), 327–349.

Kennedy, A. (1998). The influence of parafoveal words on foveal inspection time: Evidence for a processing trade-off. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception.* Oxford: Elsevier, 149-179.

Kenyeres, A. (1938). Comment une petite hongroise de sept ans apprend le francais. *Archives de Psychologie,* 26(104), 321-366.

Kleiman, G. (1980). Sentence frame contexts and lexical decisions: Sentences acceptability and word-relatedness effects. *Memory and Cognition, 8,* 336-344.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16,* 262–284.

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

Kok, A. (2001). On the utility of P300 amplitude as a measure of processing capacity. Psychophysiology, 38, 557–577.

Konopka, A., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology,* 58(1), 68-101.

Krashen, S., & Scarcella, R. (1978). On routines and patterns in language acquisition and performance. *Language Learning, 28*(2), 283-300.

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

Kutas, M., & Hillyard, S. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition,* 11, 539–550.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflectsemantic incongruity. *Science*, 207, 203–05.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics Electrified: Event-related potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics*, Academic Press, 83-143.

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences, *Electroencephalography and Clinical Neurophysiology*, 69, 218-233.

Lambrecht, K. (1984). Formulaicity, frame semantics, and pragmatics in German binomial expressions. *Language*, 60(4), 753-796.

Langacker, R. (1987). *Foundations of Cognitive Grammar.* Vol. 1. Stanford: Stanford University Press.

Laurent, J., Denhières, G., Passerieux, C., Iakimovac, G., & Hardy-Baylé, M. (2006). On understanding idiomatic language. *Brain Research*, 1068, 151–160.

Laval, V. (2003). Idiom comprehension and metapragmatic knowledge in French children. *Journal of Pragmatics, 35, 723-739.*

Levorato, M., & Cacciari, C. (1992). Children's comprehension and production of idioms: the role of context and familiarity. *Journal of Child Language,* 19, 415–433.

Levorato, M., & Cacciari, C. (1999). Idiom comprehension in children: are the effects of semantic analyzability and context separable? *European Journal of Cognitive Psychology*, 60, 261–283.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126-1177.

Lieven, E. (1987). Continuity and change in children's language development. In A. Ellis (Ed.), *Advances in the Psychology of Language*, Vol.3, Hillsdale, New Jersey: Lawrence Erlbaum Associates, 317-335.

Lieven, E., Behrens, H. Speares, J. & Tomasello, M. (2003) Early syntactic creativity: a usage-based approach. *Journal of Child Language*, 30, 333-370.

Lieven, E., Pine, J. & Barnes, H. (1992). Individual differences in early vocabulary development: redefining the referential-expressive distinction. *Journal of Child Language,* 19, 287-310.

Liversedge, S., Paterson, K., & Pickering, M. (1998). Eye movements and measures of reading time. In. G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception.* Elsevier Science, 55-75.

Locke, J. (1997). A theory of neurolinguistic development. *Brain and Language,* 58: 265-326.

Lodge, L., & Leach, E. (1975). Children's acquisition of idioms in the English language. *Journal of Speech and Hearing Research.* 18, 521-529.

Luck, S. (2005). *An Introduction to the Event-related Potential Technique.* Cambridge, MA: MIT Press.

MacNeilage, P. (1970) Motor control of serial ordering of speech. *Psychological Review,* 77, 182-196.

Malkiel, Y. (1959). Studies in irreversible binomials. *Lingua,* 8:113-160.

Manning, C. & Schutze, H. (1999). *Foundations of statistical natural language processing.* Cambridge, MA: MIT Press.

Martin, C., Vu, H., Kellas, G., & Metcalf, K. (1999). Strength of discourse context as a determinant of the subordinate bias effect. *The Quarterly Journal of Experimental Psychology,* 52A(4), 813-839.

Matlock, T., & Heredia, R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In R. Heredia, & J. Altarriba (Eds.), *Bilingual Sentence Processing.* Elsevier Press, 251-274.

McConkie, G., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception and Psychophysics,* 17, 578-586.

McConkie, G., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, 8, 365-368.

McDonald, J., Bock, K., & Kelly, M. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25, 188-230.

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

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

McDonald, S., Carpenter, R., & Shillcock, R. (2005). An anatomically constrained, stochastic model of eye movement control in reading. *Psychological Review*, 112, 814-840.

McFalls, E., & Schwanenflugel, P. (2002). The influence of contextual constraints on recall for words within sentences. *American Journal of Psychology*, 115(1), 67-88.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63, 81-97.

Mondini, S., Jarema, G., Luzzatti, C., Burani, C., & Semenza, C. (2002). Why is "red cross" different from "yellow cross"?: A neurophysiological study of non-adjective agreement within Italian compounds. *Brain and Language*, 81, 621-634.

Monsell, S., Doyle, M., & Haggard, P. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118*, 43-71.

Morris, R., Rayner, K., & Pollatsek, A. (1990). Eye movement guidance in reading: The role of parafoveal letter and space information. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 268-281.

Morrison, R. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance,* 10, 667-682.

Myles, F., Hooper, J. & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning,* 48, 323-64.

Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French L2. *Studies in Second Language Acquisition,* 21, 49-80.

Nattinger, J. (1980). A lexical phrase grammar for ESL. *TESOL Quarterly*, 14, 337–344.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, serial number 149, vol. 38, nos. 1-2.*

Nelson, K. (1981). Individual differences in language development: implications for development and language. *Developmental Psychology*, 17(2), 170-187.

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

Nesselhauf, N. (2004). *Collocations in a Learner Corpus.* Amsterdam: John Benjamins.

Nippold, M., & Martin, S. (1989). Idiom interpretation in isolation versus context: A developmental study with adolescents. *Journal of Speech and Language Research*, 32, 59–66.

Nippold, M., & Rudzinski, M. (1993). Familiarity and transparency in idiom explanation: a developmental study with children and adolescents. *Journal of Speech and Hearing Research,* 36, 728–737.

Nippold, M., & Taylor, C. (1995). Idiom understanding in youth: further examination of familiarity and transparency. *Journal of Speech and Hearing Research,* 38, 426–433.

Nippold, M., Taylor, C., & Baker, J. (1996). Idiom understanding in Australian youth: a cross-cultural comparison. *Journal of Speech and Hearing Research,* 39, 442–447.

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

Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia,* 9, 97–114.

Osterhout, L. & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language,* 31, 785–806.

Paterson, K., Liversedge, S., & Underwood, G. (1999). The influence of focus operators on syntactic processing of short relative clause sentences. *The Quarterly Journal of Experimental Psychology,* 52A, 717-737.

Paul, L., Van Lancker-Sidtis, D., Schieffeer, B., Dietrich, R., & Brown, W. (2003). Communicative deficits in agenesis of the corpus callosum: Nonliteral language and affective prosody. *Brain and Language,* 85, 313-324.

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

Pawley, A., and Syder, F. (2000). The One-Clause-at-a-Time Hypothesis. In H. Riggenbach (Ed.), *Perspectives on Fluency.* Ann Arbor: University of Michigan Press, 163-199.

Perera, N. (2001). The role of prefabricated language in young children's second language acquisition. *Bilingual Research Journal, 25*(3), 327-356.

Peters, A. (1977). Language-Learning Strategies: Does Whole Equal Sum of Parts. *Language,* 53, 560-73.

Peters, A. (1983). *The Units of Language*. Cambridge: Cambridge University Press.

Peterson, R., Burgess, C., Dell, G., & Eberhard, K., (2001). Disassociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(5), 1223-1237.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins, 137–157.

Pine, J. & Lieven, E. (1993) Reanalysing rote-learned phrases: individual differences in the transition to multi-word speech. *Journal of Child Language*, 20, 551-571.

Pinker, S. (1991). Rules of Language. *Science,* 253, 530-535.

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

Pinker, S. (1999). *Words and Rules: The Ingredients of Language.* New York: Harper Collins.

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

Plunkett, K. (1993). Lexical segmentation and vocabulary growth in early language acquisition. *Journal of Child Language,* 20(1), 43-60.

Pollatsek, A., Rayner, K., & Balota, D. A. (1986). Inferences about eye movement control from the perceptual span in reading. *Perception and Psychophysics, 40,* 123-130.

Pollio, H., Barlow, J., Fine, H. & Pollio, M. (1977). *Psychology and the Poetics of Growth: Figurative Language in Psychology, Psychotherapy, and Education.* Lawrence Erlbaum.

Pollio, M., & Pickens, J. (1980). The developmental structure of figurative competence. In R. Honeck & R. Hoffman (Eds.), *Cognition and Figurative Language*. Hillsdale, NJ: Erlbaum, 311-340.

Popiel, S., & McRae, K. (1988). The figurative and literal uses of idioms, or all idioms are not used equally. *Journal of Psycholinguistics Research*, 17(6), 475-487.

Prinz, P. (1983). The development of idiomatic meaning in children. *Language and Speech*, 26, 2263–272.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology,* 41, 211-236.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin,* 124, 372-422.

Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, 14, 191-201.

Rayner, K., & Duffy, S. (1988). On-line comprehension processes and eye movements in reading. In M. Daneman, G. MacKinnon, & T. Waller (Eds.), *Reading Research: Advances in Theory and Practice.* New York: Academic Press, 13-66.

Rayner, K., & Morris, R. (1992). Eye movement control in reading: Evidence against semantic preprocessing. *Journal of Experimental Psychology: Human Perception and Performance,* 18, 163-172.

Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading.* Englewood Cliffs, NJ: Prentice Hall.

Rayner, K., & Well, A. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin and Review,* 3(4), 504-509.

Rayner, K., Binder, K., & Duffy, S. (1999). Contextual Strength and the Subordinate Bias Effect: Comment on Martin, Vu, and Metcalf. *The Quarterly Journal of Experimental Psychology,* 52A(4), 841-852.

Rayner, K., Inhoff, A., Morrison, R., Slowiaczek, M., & Bertera, J. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance,* 7, 167-179.

Rayner, K., Sereno, S., Morris, R., Schmauder, A., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognition Processes,* 4, 21- 49.

Rayner, K., Sereno, S., Morris, R., Schmauder, A., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes,* 4, 21-50.

Rayner, K., Well, A., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception and Psychophysics, 27,* 537-544.

Reali, F., & Christiansen, M. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language,* 57, 1-23.

Reichle, E., Liversedge, S., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Science,* 13(3), 115-119.

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

Reichle, E., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences,* 26, 445-476.

Reynolds, R. & Ortony, A. (1980). Some issues in the measurement of children's comprehension of metaphorical language. *Child Development, 51,* 1110-1119.

Riess Jones, M. (1988). What does expectancy mean? *Behavioral and Brain Sciences,* 11, 387-388.

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience,* 19, 1259–1274.

Rugg, M., & Doyle, M. (1992). Event-related potentials dissociate repetition effects of high and low frequency words. *Memory and Cognition, 18,* 367–379.

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In: D. Rumelhart & J. McClelland, J. (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* MIT Press, 216-271.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science,* 274, 1926–1928.

Saussure, F. De (1916/1966). *Course in General Linguistics.* New York: McGraw-Hill.

Schmidt, R. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and Language Acquisition.* Rowley, MA: Newbury House, 137-174.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research,* 12(3), 329–363.

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

Schumann, J. (1978). *The Pidginization Process* Rowley, MA Newbury House.

Segalowitz, N. (2003). Automaticity and second languages. In M. Long & C. Doughty (Eds.), *The Handbook of Second Language Acquisition.* Malden, MA: Blackwell, 382-408.

Seidenberg, M., & MacDonald, M. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science,* 23(4), 569-588.

Sereno, S., Rayner, K., & Posner, M. (1998). Establishing the time-line of word recognition: evidence from eye movements and event-related potentials. *NeuroReport,* 9, 2195-2200.

Shallice, T. (1988). *From Neuropsychology to Mental Structure.* Cambridge: Cambridge University Press.

Shammi, P., & Stuss, D. (1999). Humour appreciation: a role of the right frontal lobe. *Brain,* 122, 657–666.

Shapira, R. (1978). The non-learning of English: Case study of an adult. In E. M. Hatch (Ed.), *Second Language Acquisition: A Book of Readings.* Rowley, MA: Newbury House, 246-255.

Sinclair, J. (1991). *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

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

Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word versus one-word verbs. *International Review of Applied Linguistics,* 45, 119–139.

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

Smadja, F. (1993). Retrieving Collocations from Text: Xtract. *Computational Lingustics,* 19(1), 143-177.

Smith, M., & Hallgren, E. (1989). Dissociation of recognition memory components following temporal lobe lesions. *Journal of Experimental Psychology. Learning, Memory and Cognition*, 15, 50–60.

Sosa, A., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of. *Brain and Language*, 83, 27-236.

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

Steinschneider, M., & Dunn, M. (2002). Electrophysiology in development neuropsychology. In: Segalowitz, S. & Rapin, I., (Eds.), *Handbook of Neuropsychology*, Vol. 7, Elsevier Press, 91-146.

Stemmer, B., Giroux, F., & Joanette, Y. (1994). Production and evaluation of requests by right hemisphere brain-damaged individuals. *Brain and Language*, 47, 1–31.

Strandburg, R., Marsh, J., Brown, W., Asarnow, R., Guthrie, D., & Higa, J. (1993). Event-related potentials in high-functioning adult autistics. *Neuropsychologia*, 31, 413–434.

Sutton S., Braren M., Zubin J., & John E. (1965) Evoked potentials correlates of stimulus uncertainty. *Science,* 150, 1187-1188.

Swiney, D., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behaviour*, 18, 23-534.

Tabossi, P., & Zardon, F. (1993). The activation of idiomatic meaning in spoken language comprehension. In C. Cacciari & P. Tabossi (Eds.), *Idioms: Processing, Structure, and Interpretation.* Lawrence Erlbaum Associates, Hillsdale, NJ, 145-161.

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory and Cognition,* 37 (4), 529-540.

Tabossi, P., & Zardon, F. (1995). The activation of idiomatic meaning. In M. Everaert, E. van der Linded, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and Psychological Perspectives.* Hillside, NJ: Erlbaum, 273-282.

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

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Cambridge, London: Harvard University Press.

Tomasello, M. (2006). Acquiring linguistic constructions. In W. Damon, R. Lerner, D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology* (6th ed.). Vol. 2. New York: Wiley, 255-298.

Tomasello, M., & Brooks, P. (1999) Early semantic development: a construction grammar approach. In Barret, M. (Ed.) *The Development of Language.* Cambridge: Cambridge University Press.

Tucker, D., Liotti, M., Potts, G., Russell, G., & Posner, M. (1994). Spatiotemporal analysis of brain electrical fields. *Human Brain Mapping,* 1, 134–152.

Ullman, M. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience,* 2, 717-726.

Ullman, M., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language,* 93, 185-238.

Underwood, G., Bloomfield, R., & Clews, S. (1988). Information influences the pattern of eye fixations during sentence comprehension. *Perception*, 17, 267-278.

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*. Amsterdam: John Benjamins, 153-172.

Van Lancker Sidtis, D., Postman, W., & Glosser, G. (2004). Feast or famine: Fixed expressions in the spontaneous speech of left hemisphere- and right hemisphere-damaged participants. *Brain and Language*, 91, 47-48.

Van Lancker, D. (1988). Nonpropositional speech: Neurolinguistic studies. In A. Ellis (Ed.), *Progress in the Psychology of Language*. Hillsdale, NJ: Lawrence Erlbaum, 49–118.

Van Lancker, D. (1990). The neurology of proverbs. *Behavioral Neurology, 3*, 169-187.

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.

Van Lancker, D., Canter, G., & Terbeek, D. (1981). Disambiguation of Ditropic Sentences: Acoustic and Phonetic Cues. *Journal of Speech and Hearing Research*, 24, 330-335.

Van Lancker-Sidtis, D. (2003). Auditory recognition of idioms by first and second speakers of English. *Applied Psycholinguistics*, 24, 45-57.

Van Patten, B., Williams, J., & Rott, S. (2004). Form-meaning connections in second language acquisition. In B. Van Patten, J. Williams, & S. Rott (Eds.). *Form-Meaning Connections in Second Language Acquisition*, Lawrence Erlbaum, 1-26.

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects and their temporal parameters. *Language and Cognitive Processes,* 8, 485–532.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition,* 18, 380–93.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context in open-and closed-class words. *Memory and Cognition,* 19, 95–112.

Verleger, R. (1988). Event-related potentials and cognition: A critique of the context-updating hypothesis and an alternative interpretation of the P300. *Behavioral and Brain Sciences*, 11, 343–427.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2009). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience.* Early on-line access at http://www.mitpressjournals.org/doi/abs/10.1162/jocn.2009.21293.

Vihman, M. (1982). Formulas in first and second language acquisition. In L. Obler & L. Menn (Eds.), *Exceptional Language and Linguistics*. New York: Academic Press, 261-284.

Vitu, F., Brysbaert, M., & Lancelin, D. (2004). A test of parafoveal-on-foveal effects with pairs of orthographically related words. *European Journal of Cognitive Psychology*, 16, 154–177.

Vu, H., Kellas, G., & Paul, S. (1998). Sources of sentence constraint on lexical ambiguity resolution. *Memory and Cognition*, 26, 979-1001.

Weinreich, U. (1969). Problems in the analysis of idioms. In J. Puhvel (Ed.), *Substance and Structure of Language*. Berkley: University of California Press, 23-81.

Whaley, C. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior,* 17, 143-154.

Wicha, N., Moreno, E., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience,* 16, 1272–1288.

Wong Fillmore, L. (1976). *The Second Time Around: Cognitive and Social Strategies in Second Language Acquisition.* Unpublished PhD thesis. Stanford University, CA.

Wong Fillmore, L. (1979). Individual differences in second language acquisition. In: Fillmore, C., Kempler, D., & Wang, W. (Eds.), *Individual Differences in Language Ability and Language Behavior.* Academic Press, London, 203–228.

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

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

Appendix 1: Norming Study 2 (for Study 1)

Norming study 2: Idiom decomposability.

Instructions:

Below, you will see a number of English idioms. Your task is to decide whether the individual components of each idiom make some contribution to the idiom's figurative meaning. For example, the idiom *miss the boat* means 'to miss an opportunity', hence the meaning of the individual component/s of this idiom contribute directly to its figurative meaning. In the idiom *spill the beans*, there is no obvious link between *the beans* and the secret; however, revealing a secret may be paralleled with spilling something. Such idioms are called **decomposable** idioms. On the other hand, the individual component/s of the idioms *kick the bucket* and *shoot the breeze* do not contribute to their figurative meaning. Such idioms are called **non-decomposable.**

Your task is to decide whether the idioms below are decomposable or non-decomposable idioms. Please put **D** for decomposable or **N** for non-decomposable.

Results:

	Idiom and its meaning	n. of responses D	N
1.	A breath of fresh air (something new and refreshing)	9	5
2.	A piece of cake (something very easy)	1	13
3.	Add fuel to the fire (to make the situation worse)	14	0
4.	As cold as ice (a cold person)	13	1
5.	At the end of the day (eventually)	9	5
6.	Kill two birds with one stone (to solve two problems at the same time)	10	4
7.	Left in the dark (be left in a bad situation not knowing what is going on)	9	5
8.	To cut a long story short (to get to the point)	11	3
9.	Not my cup of tea (something you do not like)	3	11
10.	On the other hand (alternatively)	9	5
11.	Pain in the neck (nuisance)	8	6
12.	Put your foot down (to be strict)	3	11
13.	Ring a bell (to remind)	6	8
14.	Sick and tired (be tired of something)	6	8
15.	The other side of the coin (another side of the situation)	9	5
16.	Tie the knot (to get married)	6	8
17.	Twist someone's arm (to make someone do something)	6	8
18.	Under your nose (if something happens when you don't expect it)	4	10
19.	You can't judge a book by its cover (you can't judge things by their looks)	11	3
20.	Leave a bad taste in your mouth (have bad memories of something)	9	5
21.	See which way the wind in blowing (see what the situation is like)	5	9

Appendix 2: Norming Study 3 (for Study 1)

Norming study 3: Idiom frequencies and their recognition point shown by "|" (n. = 21).

	Idioms and their recognition point	Frequency [a, b]	% [c]
1.	a breath | of fresh air	89	80
2.	a piece | of cake	70	70
3.	add fuel to | the fire	14	80
4.	as cold | as ice	24	90
5.	at the end | of the day	760	90
6.	kill two | birds with one stone	36	90
7.	leave a bad taste | in your mouth	13	90
8.	left in | the dark [d]	17	20
9.	make a long | story short	39	80
10.	not my cup | of tea	19	90
11.	on the other | hand	5311	100
12.	pain in | the neck [d]	36	60
13.	put your foot | down [d]	112	30
14.	ring a | bell [d]	75	50
15.	see which way the | wind is blowing [d]	23	60
16.	sick and | tired	58	90
17.	the other side of | the coin [d]	63	20
18.	tie the | knot	48	90
19.	twist someone's | arm	36	90
20.	under your | nose [d]	104	30
21.	you can't judge | a book by its cover	11	80

[a]Total frequencies were taken from the BNC (British National Corpus) and are given per 100 million words.

[b]The frequencies are given for the forms shown in the table above, as well as other idiomatically permissible variations, for example, 'tie the knot' + 'tied the knot', 'tying the knot', 'ties the knot', etc.

[c]The percentage of correct completions.

[d]Idioms that did not meet the 70% threshold

Appendix 3: Example of Short Stories (for Study 1)

An example of short stories that contained one of the three stimuli. In the experimental version, no target was underlined. All targets appeared in the middle of the line and never on the first or last two lines of the paragraph:

Figurative

I had my younger brother and my sister-in-law for dinner yesterday. They both have their degrees from Cambridge, whereas most of the people they work with have theirs from less well-known overseas and British universities. Personally, I think you can have the highest degree from the best university in the world, but <u>at the end of the day</u> it's your contribution to the society that matters, and not the name of the university you went to at all. Sadly, they didn't agree with me.

Literal

After my second year at university, I moved house. When I started packing, I realised that I had a lot more stuff than I had when I moved in as a first-year student. The house I was moving to was next door to the house I was moving from, which was very handy. However, I still had to carry most of my stuff in small boxes from my old room to the new one. I had to make at least 50 trips so <u>at the end of the day</u> I was absolutely exhausted. I'm hoping to stay at this house for at least another two years. I really don't want to move any more.

Novel

One of my granddads was an army officer for most of his life. Despite being an army guy, he's always been a very humane and kind person. He is also a very artistic and creative person. For example, one of his hobbies is writing poetry. He's a retired man now who served in Vietnam and who's been through many things in his life, so he's got plenty of things to write about. I know that at the end of the war he went on to teach students at the Military Academy. That was something he found particularly challenging but also rewarding in many respects.

Appendix 4: Binomial Word Order Constraints

(for Study 2)

Constraints underlying binomial word order (from Benor & Levy, 2006).

1. Semantic-pragmatic constraints:
   i. Formal markedness: a more general broader 'unmarked' word comes first, for example, 'pull and tug'.
   ii. Perception-based markedness: animate, singular, positive, concrete, masculine, more powerful concepts and entities come before inanimate, plural, negative, abstract, feminine, less powerful ones, for example, 'good and bad'.
   iii. Iconic/scalar sequencing, that is, chronological or incremental order where Word 1 notion precedes Word 2 notion, for example, 'months and years'.
2. Metrical constraint: Word 2 is longer than Word 1.
3. Frequency constraint: Word 1 is of higher frequency than Word 2, for example, 'pull and tug'.
4. Phonological constraints:
   i. Vowel length: Word 2 should have a longer main vowel than Word 1.
   ii. Vowel height: Word 2 should have a lower main vowel than of Word 1.

iii. Initial consonants: Word 2 should have more word initial (sound) consonants than Word 1 (when both Word 1 and Word 2 start with a consonant).

Appendix 5: Experimental Items in Sentence Context

(for Study 2)

Low frequency group:

Sentence:
His maternal grandfather is still **alive and well** despite his years.
John showed me pictures of the **bride and groom** both dressed in blue.
We were told that to all **intents and purposes** the case was won.
Despite the crisis the **king and queen** are still popular among the people.
She has such brilliant taste that most of her clothes **mix and match** easily.
My favourite special is **sweet and sour** but they didn't have it on the menu.
They bought some **stocks and shares** although they had no experience in this.
The truth is that in my **heart and soul** I've always believed in this.
Dan was relieved at the news that both **mother and child** were unhurt.
Jim was back home **safe and sound** despite his numerous adventures.
People are free to **buy and sell** their produce at the market.
The separation of **church and state** is important to many politicians.
The issues of **war and peace** are the central concerns in global politics.
It appeared in a number of **newspapers and magazines** in more detail.

High frequency group:

Sentence
Their inaccurate calculations of **profit and loss** were very misleading
Jane was given a clear understanding of **right and wrong** from early childhood.
In the majority of cases the **husband and wife** are both to blame.
You should find out the **name and address** of your nearest vet.
The amount of money spent on **research and development** is huge.
Ali never got along with his **brothers and sisters** even when he was a baby.
Such activities are beneficial for **mind and body** and are completely harmless
There should be a balance between **supply and demand** in the industry.
Events from the **past and present** will always affect the future
It is a free-trade zone linking **east and west** and providing work for people.
The different status of **men and women** is emphasised in his work.
They discussed this on **radio and television** but failed to come to an agreement.
Areas that are particularly rich in **flora and fauna** should be protected by law.
She could hardly **read and write** and was underdeveloped physically.

Appendix 6: Experimental Items

(for Study 3, Experiment 1)

Experimental items used in Study 3, Experiment 1. Due to space constraints, only three-word combinations are given, however, in the actual experiment, the items were presented in a sentence context.

	Type	Phrase	Phrasal frequency (BNC)
1	Binomial	here and abroad	16
	Reversed	abroad and here	0
	Incongruity	here and inside	0
2	Binomial	shirt and tie	23
	Reversed	tie and shirt	1
	Incongruity	shirt and fly	0
3	Binomial	sport and leisure	23
	Reversed	leisure and sport	1
	Incongruity	sport and despair	0
4	Binomial	milk and honey	27
	Reversed	honey and milk	3
	Incongruity	milk and snake	0
5	Binomial	horse and rider	33
	Reversed	rider and horse	4
	Incongruity	horse and torch	0
6	Binomial	fast and furious	35
	Reversed	furious and fast	0
	Incongruity	fast and liberal	0
7	Binomial	sweet and sour	36
	Reversed	sour and sweet	0
	Incongruity	sweet and slim	0
8	Binomial	winners and losers	39

	Reversed	losers and winners	0
	Incongruity	winners and boxers	0
9	Binomial	mix and match	43
	Reversed	match and mix	0
	Incongruity	mix and split	0
10	Binomial	pen and paper	60
	Reversed	paper and pen	5
	Incongruity	pen and story	0
11	Binomial	lock and key	50
	Reversed	key and lock	1
	Incongruity	lock and egg	0
12	Binomial	slowly and carefully	53
	Reversed	carefully and slowly	3
	Incongruity	slowly and generally	0
13	Binomial	questions and answers	56
	Reversed	answers and questions	0
	Incongruity	questions and efforts	0
14	Binomial	pick and choose	62
	Reversed	choose and pick	0
	Incongruity	pick and listen	0
15	Binomial	heaven and earth	66
	Reversed	earth and heaven	3
	Incongruity	heaven and plant	0
16	Binomial	pain and suffering	83
	Reversed	suffering and pain	4
	Incongruity	pain and admission	0
17	Binomial	king and queen	87
	Reversed	queen and king	0
	Incongruity	king and cloud	0
18	Binomial	look and see	87
	Reversed	see and look	1
	Incongruity	look and use	0
19	Binomial	direct and indirect	99
	Reversed	indirect and direct	4
	Incongruity	direct and advisory	0

20	Binomial	crime and punishment	109
	Reversed	punishment and crime	0
	Incongruity	crime and occupation	0
21	Binomial	alive and well	114
	Reversed	well and alive	0
	Incongruity	alive and away	0
22	Binomial	rich and poor	140
	Reversed	poor and rich	4
	Incongruity	rich and hard	0
23	Binomial	prince and princess	151
	Reversed	princess and prince	0
	Incongruity	prince and armchair	0
24	Binomial	trial and error	156
	Reversed	error and trial	0
	Incongruity	trial and angle	0
25	Binomial	aims and objectives	165
	Reversed	objectives and aims	1
	Incongruity	aims and exceptions	0
26	Binomial	top and bottom	195
	Reversed	bottom and top	1
	Incongruity	top and extent	0
27	Binomial	rules and regulations	204
	Reversed	regulations and rules	1
	Incongruity	rules and reputations	0
28	Binomial	food and drink	338
	Reversed	drink and food	4
	Incongruity	food and style	0
29	Binomial	bed and breakfast	492
	Reversed	breakfast and bed	1
	Incongruity	bed and democracy	0
30	Binomial	law and order	598
	Reversed	order and law	0
	Incongruity	law and sense	0
31	Binomial	rest and relaxation	17
	Reversed	relaxation and rest	1

	Incongruity	rest and deficiency	0
32	Binomial	nice and easy	21
	Reversed	easy and nice	0
	Incongruity	nice and open	0
33	Binomial	wit and wisdom	24
	Reversed	wisdom and wit	1
	Incongruity	wit and custom	0
34	Binomial	highs and lows	25
	Reversed	lows and highs	1
	Incongruity	highs and arks	0
35	Binomial	really and truly	35
	Reversed	truly and really	0
	Incongruity	really and aside	0
36	Binomial	thunder and lightning	36
	Reversed	lightning and thunder	1
	Incongruity	thunder and injustice	0
37	Binomial	thick and thin	38
	Reversed	thin and thick	2
	Incongruity	thick and vast	0
38	Binomial	weeks and months	45
	Reversed	months and weeks	1
	Incongruity	weeks and unions	0
39	Binomial	safe and sound	46
	Reversed	sound and safe	4
	Incongruity	safe and moist	0
40	Binomial	rights and responsibilities	49
	Reversed	responsibilities and rights	1
	Incongruity	rights and representatives	0
41	Binomial	today and tomorrow	55
	Reversed	tomorrow and today	0
	Incongruity	today and anywhere	0
42	Binomial	heart and soul	57
	Reversed	soul and heart	1
	Incongruity	heart and bone	0
43	Binomial	sick and tired	58

	Reversed	tired and sick	2
	Incongruity	sick and legal	0
44	Binomial	born and bred	71
	Reversed	bred and born	2
	Incongruity	born and rung	0
45	Binomial	stocks and shares	74
	Reversed	shares and stocks	0
	Incongruity	stocks and habits	0
46	Binomial	loud and clear	86
	Reversed	clear and loud	0
	Incongruity	loud and alone	0
47	Binomial	bride and groom	97
	Reversed	groom and bride	0
	Incongruity	bride and alien	0
48	Binomial	church and state	102
	Reversed	state and church	4
	Incongruity	church and light	0
49	Binomial	flesh and blood	109
	Reversed	blood and flesh	1
	Incongruity	flesh and glass	0
50	Binomial	true and fair	127
	Reversed	fair and true	1
	Incongruity	true and wild	0
51	Binomial	iron and steel	128
	Reversed	steel and iron	3
	Incongruity	iron and pride	0
52	Binomial	cause and effect	143
	Reversed	effect and cause	0
	Incongruity	cause and energy	0
53	Binomial	right and wrong	144
	Reversed	wrong and right	0
	Incongruity	right and happy	0
54	Binomial	good and bad	158
	Reversed	bad and good	4
	Incongruity	good and low	0

55	Binomial	pros and cons	167
	Reversed	cons and pros	0
	Incongruity	pros and tabs	0
56	Binomial	advantages and disadvantages	180
	Reversed	disadvantages and advantages	4
	Incongruity	advantages and professionals	0
57	Binomial	art and design	224
	Reversed	design and art	0
	Incongruity	art and cancer	0
58	Binomial	ladies and gentlemen	270
	Reversed	gentlemen and ladies	4
	Incongruity	ladies and employers	0
59	Binomial	terms and conditions	393
	Reversed	conditions and terms	2
	Incongruity	terms and traditions	0
60	Binomial	name and address	516
	Reversed	address and name	1
	Incongruity	name and traffic	0
61	Binomial	needle and thread	18
	Reversed	thread and needle	0
	Incongruity	needle and vacuum	0
62	Binomial	cheese and onion	23
	Reversed	onion and cheese	1
	Incongruity	cheese and lemon	0
63	Binomial	love and hate	23
	Reversed	hate and love	2
	Incongruity	love and beat	0
64	Binomial	hope and pray	26
	Reversed	pray and hope	3
	Incongruity	hope and rent	0
65	Binomial	drunk and disorderly	28
	Reversed	disorderly and drunk	0
	Incongruity	drunk and invariable	0
66	Binomial	marriage and divorce	30
	Reversed	divorce and marriage	4

	Incongruity	marriage and heating	0
67	Binomial	decline and fall	34
	Reversed	fall and decline	0
	Incongruity	decline and bath	0
68	Binomial	cat and mouse	37
	Reversed	mouse and cat	0
	Incongruity	cat and glove	0
69	Binomial	army and navy	42
	Reversed	navy and army	2
	Incongruity	army and kiss	0
70	Binomial	ready and willing	44
	Reversed	willing and ready	1
	Incongruity	ready and radical	0
71	Binomial	above and beyond	46
	Reversed	beyond and above	2
	Incongruity	above and beside	0
72	Binomial	pure and simple	46
	Reversed	simple and pure	0
	Incongruity	pure and united	0
73	Binomial	current and future	51
	Reversed	future and current	1
	Incongruity	current and silent	0
74	Binomial	start and finish	58
	Reversed	finish and start	0
	Incongruity	start and reduce	0
75	Binomial	early and late	64
	Reversed	late and early	0
	Incongruity	early and long	0
76	Binomial	aches and pains	69
	Reversed	pains and aches	0
	Incongruity	aches and noses	0
77	Binomial	beginning and end	77
	Reversed	end and beginning	1
	Incongruity	beginning and lot	0
78	Binomial	buy and sell	84

	Reversed	sell and buy	3
	Incongruity	buy and rise	0
79	Binomial	knife and fork	87
	Reversed	fork and knife	4
	Incongruity	knife and gulf	0
80	Binomial	intents and purposes	109
	Reversed	purposes and intents	0
	Incongruity	intents and pressure	0
81	Binomial	facts and figures	110
	Reversed	figures and facts	0
	Incongruity	facts and chances	0
82	Binomial	spring and summer	120
	Reversed	summer and spring	0
	Incongruity	spring and growth	0
83	Binomial	read and write	133
	Reversed	write and read	2
	Incongruity	read and bring	0
84	Binomial	landlord and tenant	152
	Reversed	tenant and landlord	3
	Incongruity	landlord and cattle	0
85	Binomial	hot and cold	168
	Reversed	cold and hot	4
	Incongruity	hot and dark	0
86	Binomial	bread and butter	204
	Reversed	butter and bread	0
	Incongruity	bread and toilet	0
87	Binomial	head and shoulders	218
	Reversed	shoulders and head	4
	Incongruity	head and ministers	0
88	Binomial	deaf and dumb	276
	Reversed	dumb and deaf	0
	Incongruity	deaf and lazy	0
89	Binomial	profit and loss	363
	Reversed	loss and profit	0
	Incongruity	profit and team	0

90	Binomial	research and development	739
	Reversed	development and research	4
	Incongruity	research and information	0
91	Binomial	shoes and socks	36
	Reversed	socks and shoes	17
	Incongruity	shoes and packs	0
92	Binomial	audio and video	41
	Reversed	video and audio	22
	Incongruity	audio and aroma	0
93	Binomial	major and minor	50
	Reversed	minor and major	10
	Incongruity	major and loose	0
94	Binomial	clean and tidy	56
	Reversed	tidy and clean	8
	Incongruity	clean and ripe	0
95	Binomial	see and hear	61
	Reversed	hear and see	16
	Incongruity	see and turn	0
96	Binomial	give and take	66
	Reversed	take and give	6
	Incongruity	give and find	0
97	Binomial	physical and emotional	77
	Reversed	emotional and physical	31
	Incongruity	physical and conscious	0
98	Binomial	vitamins and minerals	87
	Reversed	minerals and vitamins	13
	Incongruity	minerals and prophets	0
99	Binomial	health and welfare	111
	Reversed	welfare and health	7
	Incongruity	health and library	0
100	Binomial	formal and informal	131
	Reversed	informal and formal	18
	Incongruity	formal and ignorant	0
101	Binomial	trees and shrubs	133
	Reversed	shrubs and trees	33

	Incongruity	trees and chunks	0
102	Binomial	live and work	134
	Reversed	work and live	7
	Incongruity	live and talk	0
103	Binomial	age and sex	145
	Reversed	sex and age	32
	Incongruity	age and tax	0
104	Binomial	how and why	148
	Reversed	why and how	51
	Incongruity	how and who	0
105	Binomial	supply and demand	177
	Reversed	demand and supply	80
	Incongruity	supply and finger	0
106	Binomial	internal and external	191
	Reversed	external and internal	65
	Incongruity	internal and splendid	0
107	Binomial	schools and colleges	197
	Reversed	colleges and schools	13
	Incongruity	schools and kitchens	0
108	Binomial	social and cultural	202
	Reversed	cultural and social	54
	Incongruity	social and separate	0
109	Binomial	towns and cities	222
	Reversed	cities and towns	61
	Incongruity	towns and pounds	0
110	Binomial	fruit and vegetables	236
	Reversed	vegetables and fruit	38
	Incongruity	fruit and judgements	0
111	Binomial	radio and television	275
	Reversed	television and radio	151
	Incongruity	radio and department	0
112	Binomial	brothers and sisters	318
	Reversed	sisters and brothers	15
	Incongruity	brothers and bottles	0
113	Binomial	family and friends	331

	Reversed	friends and family	101
	Incongruity	family and streets	0
114	Binomial	boys and girls	339
	Reversed	girls and boys	85
	Incongruity	boys and sides	0
115	Binomial	east and west	380
	Reversed	west and east	63
	Incongruity	east and mile	0
116	Binomial	husband and wife	406
	Reversed	wife and husband	9
	Incongruity	husband and game	0
117	Binomial	education and training	544
	Reversed	training and education	82
	Incongruity	education and pleasure	0
118	Binomial	goods and services	643
	Reversed	services and goods	6
	Incongruity	goods and controls	0
119	Binomial	trade and industry	830
	Reversed	industry and trade	24
	Incongruity	trade and position	0
120	Binomial	black and white	1096
	Reversed	white and black	51
	Incongruity	black and young	0
121	Binomial	fish and chips	221
	Reversed	chips and fish	9
	Incongruity	fish and knots	0
122	Binomial	press and media	23
	Reversed	media and press	7
	Incongruity	press and giant	0
123	Binomial	safety and security	31
	Reversed	security and safety	11
	Incongruity	safety and instance	0
124	Binomial	lakes and rivers	44
	Reversed	rivers and lakes	25
	Incongruity	lakes and spaces	0

125	Binomial	singing and dancing	44
	Reversed	dancing and singing	16
	Incongruity	singing and reminding	0
126	Binomial	quickly and easily	64
	Reversed	easily and quickly	19
	Incongruity	quickly and hardly	0
127	Binomial	war and peace	72
	Reversed	peace and war	22
	Incongruity	war and dress	0
128	Binomial	domestic and foreign	77
	Reversed	foreign and domestic	31
	Incongruity	domestic and strange	0
129	Binomial	newspapers and magazines	96
	Reversed	magazines and newspapers	32
	Incongruity	newspapers and household	0
130	Binomial	wind and rain	96
	Reversed	rain and wind	15
	Incongruity	wind and gate	0
131	Binomial	front and back	113
	Reversed	back and front	31
	Incongruity	front and grey	0
132	Binomial	flora and fauna	132
	Reversed	fauna and flora	38
	Incongruity	flora and décor	0
133	Binomial	tea and coffee	134
	Reversed	coffee and tea	19
	Incongruity	tea and forest	0
134	Binomial	mind and body	139
	Reversed	body and mind	51
	Incongruity	mind and foot	0
135	Binomial	large and small	158
	Reversed	small and large	71
	Incongruity	large and whole	0
136	Binomial	costs and benefits	167
	Reversed	benefits and costs	18

	Incongruity	costs and machines	0
137	Binomial	sales and marketing	187
	Reversed	marketing and sales	39
	Incongruity	sales and amazement	0
138	Binomial	arms and legs	201
	Reversed	legs and arms	31
	Incongruity	arms and seas	0
139	Binomial	salt and pepper	202
	Reversed	pepper and salt	43
	Incongruity	salt and racket	0
140	Binomial	old and new	216
	Reversed	new and old	39
	Incongruity	old and big	0
141	Binomial	past and present	251
	Reversed	present and past	25
	Incongruity	past and concern	0
142	Binomial	hardware and software	268
	Reversed	software and hardware	45
	Incongruity	hardware and lipstick	0
143	Binomial	primary and secondary	286
	Reversed	secondary and primary	13
	Incongruity	primary and potential	0
144	Binomial	management and business	338
	Reversed	business and management	39
	Incongruity	management and movement	0
145	Binomial	first and second	362
	Reversed	second and first	6
	Incongruity	first and strong	0
146	Binomial	public and private	369
	Reversed	private and public	161
	Incongruity	public and natural	0
147	Binomial	north and south	439
	Reversed	south and north	11
	Incongruity	north and piece	0
148	Binomial	male and female	446

	Reversed	female and male	38
	Incongruity	male and narrow	0
149	Binomial	on and off	473
	Reversed	off and on	47
	Incongruity	on and far	0
150	Binomial	men and women	1956
	Reversed	women and men	251
	Incongruity	men and parts	0
151	Binomial	parks and gardens	25
	Reversed	gardens and parks	6
	Incongruity	parks and methods	0
152	Binomial	height and weight	38
	Reversed	weight and height	8
	Incongruity	height and object	0
153	Binomial	warm and dry	42
	Reversed	dry and warm	8
	Incongruity	warm and odd	0
154	Binomial	son and daughter	48
	Reversed	daughter and son	7
	Incongruity	son and hospital	0
155	Binomial	snow and ice	54
	Reversed	ice and snow	23
	Incongruity	snow and aid	0
156	Binomial	air and water	54
	Reversed	water and air	27
	Incongruity	air and group	0
157	Binomial	tables and chairs	57
	Reversed	chairs and tables	25
	Incongruity	tables and hotels	0
158	Binomial	inner and outer	76
	Reversed	outer and inner	23
	Incongruity	inner and cruel	0
159	Binomial	red and green	82
	Reversed	green and red	31
	Incongruity	red and final	0

160	Binomial	national and regional	93
	Reversed	regional and national	61
	Incongruity	national and detailed	0
161	Binomial	doctors and nurses	108
	Reversed	nurses and doctors	19
	Incongruity	doctors and shells	0
162	Binomial	income and expenditure	112
	Reversed	expenditure and income	17
	Incongruity	income and preparation	0
163	Binomial	help and advice	126
	Reversed	advice and help	42
	Incongruity	help and spirit	0
164	Binomial	shapes and sizes	130
	Reversed	sizes and shapes	13
	Incongruity	shapes and halls	0
165	Binomial	positive and negative	147
	Reversed	negative and positive	30
	Incongruity	positive and romantic	0
166	Binomial	upper and lower	156
	Reversed	lower and upper	33
	Incongruity	upper and blind	0
167	Binomial	gold and silver	173
	Reversed	silver and gold	52
	Incongruity	gold and vision	0
168	Binomial	central and local	184
	Reversed	local and central	52
	Incongruity	central and heavy	0
169	Binomial	parents and children	192
	Reversed	children and parents	47
	Incongruity	parents and examples	0
170	Binomial	theory and practice	211
	Reversed	practice and theory	10
	Incongruity	theory and standard	0
171	Binomial	backwards and forwards	223
	Reversed	forwards and backwards	60

	Incongruity	backwards and honestly	0
172	Binomial	life and death	242
	Reversed	death and life	9
	Incongruity	life and voice	0
173	Binomial	time and money	272
	Reversed	money and time	29
	Incongruity	time and party	0
174	Binomial	left and right	307
	Reversed	right and left	144
	Incongruity	left and short	0
175	Binomial	day and night	319
	Reversed	night and day	110
	Incongruity	day and point	0
176	Binomial	there and then	367
	Reversed	then and there	74
	Incongruity	there and even	0
177	Binomial	oil and gas	392
	Reversed	gas and oil	26
	Incongruity	oil and cup	0
178	Binomial	mum and dad	494
	Reversed	dad and mum	11
	Incongruity	mum and hut	0
179	Binomial	science and technology	616
	Reversed	technology and science	10
	Incongruity	science and revolution	0
180	Binomial	up and down	2118
	Reversed	down and up	17
	Incongruity	up and upon	0

Appendix 7: Experimental Items

(for Study 3, Experiment 2 and 3)

Experimental items used in Study 3, Experiment 2. The same items were used in Experiment 3, but without the conjunction 'and'.

	Type	Phrase	Phrasal frequency (BNC)
1	Binomial	aches and pains	69
	Associate	agony and pains	0
	Incongruity	tours and pains	0
2	Binomial	age and sex	145
	Associate	gender and sex	5
	Incongruity	tube and sex	0
3	Binomial	alive and well	114
	Associate	sick and well	0
	Incongruity	plenty and well	0
4	Binomial	angels and devils	2
	Associate	evil and devils	0
	Incongruity	necks and devils	0
5	Binomial	apples and oranges	10
	Associate	juice and oranges	0
	Incongruity	tablets and oranges	0
6	Binomial	army and navy	42
	Associate	sailor and navy	0
	Incongruity	vision and navy	0
7	Binomial	bacon and eggs	62
	Associate	omelet and eggs	0
	Incongruity	idiot and eggs	0
8	Binomial	bar and grill	9
	Associate	barbecue and grill	0
	Incongruity	guess and grill	0

9	Binomial	newspapers and magazines	96
	Associate	articles and magazines	0
	Incongruity	relations and magazines	0
10	Binomial	beginning and end	77
	Associate	conclude and end	0
	Incongruity	analysis and end	0
11	Binomial	black and white	1096
	Associate	pale and white	1
	Incongruity	busy and white	0
12	Binomial	boys and girls	339
	Associate	guys and girls	0
	Incongruity	pats and girls	0
13	Binomial	bread and butter	204
	Associate	margarine and butter	3
	Incongruity	angle and butter	0
14	Binomial	burgers and fries	0
	Associate	potatoes and fries	0
	Incongruity	violations and fries	0
15	Binomial	business and pleasure	17
	Associate	delight and pleasure	3
	Incongruity	surface and pleasure	0
16	Binomial	cap and gown	3
	Associate	robe and gown	0
	Incongruity	plug and gown	0
17	Binomial	car and truck	6
	Associate	van and truck	2
	Incongruity	bay and truck	0
18	Binomial	chapter and verse	36
	Associate	poem and verse	0
	Incongruity	depth and verse	0
19	Binomial	cops and robbers	13
	Associate	crooks and robbers	0
	Incongruity	dusks and robbers	0
20	Binomial	cream and sugar	6
	Associate	flour and sugar	2

	Incongruity	patch and sugar	0
21	Binomial	crime and punishment	109
	Associate	discipline and punishment	3
	Incongruity	owner and punishment	0
22	Binomial	deaf and dumb	276
	Associate	smart and dumb	0
	Incongruity	fluid and dumb	0
23	Binomial	decline and fall	34
	Associate	descent and fall	0
	Incongruity	fiber and fall	0
24	Binomial	early and late	64
	Associate	tardy and late	0
	Incongruity	fairy and late	0
25	Binomial	earth and sky	9
	Associate	stars and sky	0
	Incongruity	mouth and sky	0
26	Binomial	facts and figures	110
	Associate	forms and figures	0
	Incongruity	worth and figures	0
27	Binomial	fad and fashion	1
	Associate	trend and fashion	0
	Incongruity	spell and fashion	0
28	Binomial	family and friends	331
	Associate	fellows and friends	0
	Incongruity	sheets and friends	0
29	Binomial	far and away	56
	Associate	further and away	0
	Incongruity	down and away	0
30	Binomial	waiter and waitress	1
	Associate	hostess and waitress	0
	Incongruity	laundry and waitress	0
31	Binomial	seek and destroy	2
	Associate	make and destroy	1
	Incongruity	place and destroy	0
32	Binomial	fit and trim	1

	Associate	slim and trim	1
	Incongruity	final and trim	0
33	Binomial	flesh and blood	109
	Associate	vein and blood	0
	Incongruity	memory and blood	0
34	Binomial	floor and ceiling	17
	Associate	roof and ceiling	1
	Incongruity	bank and ceiling	0
35	Binomial	food and drink	338
	Associate	glass and drink	1
	Incongruity	green and drink	0
36	Binomial	forgive and forget	27
	Associate	remember and forget	0
	Incongruity	center and forget	0
37	Binomial	fruit and vegetables	236
	Associate	garden and vegetables	0
	Incongruity	limits and vegetables	0
38	Binomial	good and bad	158
	Associate	awful and bad	0
	Incongruity	done and bad	0
39	Binomial	hand and foot	53
	Associate	ankle and foot	1
	Incongruity	stop and foot	0
40	Binomial	heart and soul	57
	Associate	spirit and soul	2
	Incongruity	growth and soul	0
41	Binomial	heaven and earth	66
	Associate	ground and earth	0
	Incongruity	market and earth	0
42	Binomial	hopes and dreams	19
	Associate	wishes and dreams	0
	Incongruity	steps and dreams	0
43	Binomial	hot and cold	168
	Associate	shiver and cold	0
	Incongruity	aid and cold	0

44	Binomial	income and wealth	62
	Associate	success and wealth	0
	Incongruity	check and wealth	0
45	Binomial	intents and purposes	109
	Associate	functions and purposes	1
	Incongruity	editors and purposes	0
46	Binomial	iron and steel	128
	Associate	metal and steel	1
	Incongruity	chest and steel	0
47	Binomial	ketchup and mustard	0
	Associate	mayonnaise and mustard	0
	Incongruity	architect and mustard	0
48	Binomial	king and queen	87
	Associate	royalty and queen	0
	Incongruity	plastic and queen	0
49	Binomial	knife and fork	87
	Associate	spoon and fork	4
	Incongruity	theme and fork	0
50	Binomial	ladies and gentlemen	270
	Associate	officers and gentlemen	0
	Incongruity	periods and gentlemen	0
51	Binomial	lean and mean	3
	Associate	cruel and mean	0
	Incongruity	mint and mean	0
52	Binomial	life and death	242
	Associate	suicide and death	0
	Incongruity	room and death	0
53	Binomial	lock and key	50
	Associate	piano and key	0
	Incongruity	mare and key	0
54	Binomial	love and hate	23
	Associate	like and hate	0
	Incongruity	stand and hate	0
55	Binomial	marriage and divorce	30
	Associate	marry and divorce	2

	Incongruity	notion and divorce	0
56	Binomial	master and slave	11
	Associate	servant and slave	0
	Incongruity	symbol and slave	0
57	Binomial	milk and honey	27
	Associate	oats and honey	0
	Incongruity	dive and honey	0
58	Binomial	mix and match	43
	Associate	lighter and match	0
	Incongruity	beard and match	0
59	Binomial	mother and child	91
	Associate	doll and child	0
	Incongruity	piece and child	0
60	Binomial	name and address	516
	Associate	number and address	5
	Incongruity	hours and address	0
61	Binomial	neat and clean	13
	Associate	sweep and clean	2
	Incongruity	suck and clean	0
62	Binomial	needle and thread	18
	Associate	spool and thread	0
	Incongruity	update and thread	0
63	Binomial	nickel and dime	0
	Associate	quarter and dime	0
	Incongruity	puzzle and dime	0
64	Binomial	oil and vinegar	11
	Associate	wine and vinegar	1
	Incongruity	camp and vinegar	0
65	Binomial	old and new	216
	Associate	modern and new	0
	Incongruity	sacred and new	0
66	Binomial	pain and suffering	83
	Associate	hardship and suffering	2
	Incongruity	empire and suffering	0
67	Binomial	parents and children	192

	Associate	nursery and children	0
	Incongruity	models and children	0
68	Binomial	pass and fail	1
	Associate	succeed and fail	1
	Incongruity	spray and fail	0
69	Binomial	past and present	251
	Associate	absent and present	2
	Incongruity	edge and present	0
70	Binomial	peace and quiet	145
	Associate	passive and quiet	0
	Incongruity	medical and quiet	0
71	Binomial	pen and paper	60
	Associate	pad and paper	0
	Incongruity	tank and paper	0
72	Binomial	pick and choose	62
	Associate	decide and choose	0
	Incongruity	sleep and choose	0
73	Binomial	pins and needles	37
	Associate	syringes and needles	2
	Incongruity	syrups and needles	0
74	Binomial	plain and simple	26
	Associate	basic and simple	3
	Incongruity	rigid and simple	0
75	Binomial	scotch and water	4
	Associate	pool and water	1
	Incongruity	detail and water	0
76	Binomial	strawberries and cream	12
	Associate	ointment and cream	0
	Incongruity	students and cream	0
77	Binomial	pride and prejudice	33
	Associate	stereotype and prejudice	1
	Incongruity	brain and prejudice	0
78	Binomial	profit and loss	363
	Associate	gain and loss	4
	Incongruity	taste and loss	0

79	Binomial	public and private	369
	Associate	secluded and private	0
	Incongruity	pretty and private	0
80	Binomial	questions and answers	56
	Associate	responses and answers	0
	Incongruity	percents and answers	0
81	Binomial	radio and television	275
	Associate	cable and television	0
	Incongruity	drift and television	0
82	Binomial	far and wide	97
	Associate	vast and wide	0
	Incongruity	ahead and wide	0
83	Binomial	hope and pray	26
	Associate	kneel and pray	0
	Incongruity	trust and pray	0
84	Binomial	pride and joy	68
	Associate	happiness and joy	3
	Incongruity	builder and joy	0
85	Binomial	arms and legs	201
	Associate	thighs and legs	4
	Incongruity	birth and legs	0
86	Binomial	streets and roads	0
	Associate	highways and roads	0
	Incongruity	policies and roads	0
87	Binomial	read and write	133
	Associate	print and write	0
	Incongruity	shop and write	0
88	Binomial	ready and willing	44
	Associate	eager and willing	4
	Incongruity	worst and willing	0
89	Binomial	rest and relaxation	17
	Associate	comfort and relaxation	0
	Incongruity	shift and relaxation	0
90	Binomial	rich and poor	140
	Associate	homeless and poor	0

	Incongruity	phone and poor	0
91	Binomial	right and wrong	144
	Associate	immoral and wrong	0
	Incongruity	reason and wrong	0
92	Binomial	rise and shine	3
	Associate	polish and shine	0
	Incongruity	impact and shine	0
93	Binomial	safe and sound	46
	Associate	voice and sound	0
	Incongruity	risk and sound	0
94	Binomial	schools and colleges	197
	Associate	campuses and colleges	0
	Incongruity	killers and colleges	0
95	Binomial	science and technology	616
	Associate	computer and technology	0
	Incongruity	running and technology	0
96	Binomial	see and hear	61
	Associate	listen and hear	2
	Incongruity	peak and hear	0
97	Binomial	shapes and sizes	130
	Associate	measurements and sizes	0
	Incongruity	officials and sizes	0
98	Binomial	shirt and tie	23
	Associate	bow and tie	0
	Incongruity	sink and tie	0
99	Binomial	shoes and socks	36
	Associate	underwear and socks	0
	Incongruity	stairs and socks	0
100	Binomial	skin and bones	7
	Associate	joints and bones	0
	Incongruity	waves and bones	0
101	Binomial	slip and fall	6
	Associate	autumn and fall	0
	Incongruity	advise and fall	0
102	Binomial	snow and ice	54

	Associate	frost and ice	0
	Incongruity	boss and ice	0
103	Binomial	soap and water	43
	Associate	flood and water	0
	Incongruity	blank and water	0
104	Binomial	song and dance	68
	Associate	ballet and dance	0
	Incongruity	block and dance	0
105	Binomial	start and finish	58
	Associate	complete and finish	1
	Incongruity	direct and finish	0
106	Binomial	sticks and stones	25
	Associate	pebbles and stones	2
	Incongruity	blonds and stones	0
107	Binomial	straight and narrow	29
	Associate	broad and narrow	3
	Incongruity	best and narrow	0
108	Binomial	sun and moon	23
	Associate	crescent and moon	0
	Incongruity	battle and moon	0
109	Binomial	sweet and sour	36
	Associate	tart and sour	0
	Incongruity	count and sour	0
110	Binomial	tables and chairs	57
	Associate	stools and chairs	0
	Incongruity	ideas and chairs	0
111	Binomial	tea and coffee	134
	Associate	caffeine and coffee	0
	Incongruity	niece and coffee	0
112	Binomial	theory and practice	211
	Associate	method and practice	1
	Incongruity	summer and practice	0
113	Binomial	thick and thin	38
	Associate	skinny and thin	0
	Incongruity	notes and thin	0

114	Binomial	time and money	272
	Associate	taxes and money	0
	Incongruity	top and money	0
115	Binomial	towns and cities	222
	Associate	suburbs and cities	0
	Incongruity	finals and cities	0
116	Binomial	vitamins and minerals	87
	Associate	calcium and minerals	0
	Incongruity	murders and minerals	0
117	Binomial	wear and tear	153
	Associate	rip and tear	1
	Incongruity	curb and tear	0
118	Binomial	weights and measures	37
	Associate	scales and measures	1
	Incongruity	valleys and measures	0
119	Binomial	wild and crazy	7
	Associate	weird and crazy	0
	Incongruity	pink and crazy	0
120	Binomial	wind and rain	96
	Associate	storm and rain	1
	Incongruity	grace and rain	0