Found in Translation: a psycholinguistic investigation of idiom processing in native and non-native speakers

Gareth Carrol

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

December 2015

Abstract

Idioms, as highly familiar word combinations, are processed quickly by native speakers, but are problematic for non-native speakers even at high levels of proficiency. In this thesis I explore the representation of idioms in the monolingual and bilingual lexicons. In a series of studies I investigate how native and non-native speakers of English process English idioms and idioms translated from another language. In Study 1 I used a lexical decision task to test how much an expected word is primed following the first part of an idiom, e.g. *on the edge of your… seat*. English native speakers and Chinese-English bilinguals were tested using English idioms and translations of Chinese idioms (e.g. *draw a snake and add… feet*). In Study 2 I presented the same materials in short passages to allow for more natural presentation and used eye-tracking to investigate the reading patterns for all items. I also compared figurative and literal uses of the same items to see how easily non-native speakers were able to process non-compositional meaning in the L2. In Study 3 I used the same methodology (eye-tracking of idioms used in short sentence contexts) with a higher proficiency group (Swedish-English bilinguals), with much shorter, less predictable idioms (e.g. *break the ice/bryta isen*) and included a set of idioms that exist in both L1 and L2. All three studies point to the same conclusion: that even in an unfamiliar translated form, the expected lexical combination was facilitated (idioms showed faster processing than control phrases), but only the highest proficiency participants also showed evidence that they were able to process the figurative meanings without disruption. Congruent items show no additional advantage, hence it is clearly L1 knowledge of what words 'go together' that drives the effect in translation.

In Study 4 I extended this by contrasting idioms with other types of formulaic phrase: literal binomials (*king and queen*) and collocations (*abject poverty*). All types showed faster reading compared to equally plausible control phrases. I then used formulaic component words in separated contexts to see whether any lexical priming effects are observed when the formulaic frame is compromised. Only idioms showed evidence of a formulaic advantage in this condition, while binomials showed evidence of semantic priming and collocations showed evidence of disruption. Importantly, different factors relevant to each formulaic type show an effect on how they are processed, e.g. idioms were driven by predictability, while binomials were driven more by the strength of semantic association between component words.

The results overall provide a valuable new perspective on how formulaic units are represented in the mental lexicon. The fact that faster processing is seen for translated forms shows that idioms are not processed as unanalysed whole units, since L1 influence must be contingent on the individual words activating translation equivalent forms. This also shows that non-native speakers do not show fundamentally different processing in their L2 than native speakers, and ‘known’ word combinations are processed quickly regardless of the language of presentation. Compared to idioms, other formulaic types also show fast processing in canonical forms, but are more variable in whether or not the component words also show lexical priming in non-formulaic contexts. Formulaicity therefore exists at multiple levels of representation, encompassing lexical, structural and conceptual properties of word combinations.

Contents

List of Figures

Declaration

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

The methodology discussion in Chapter 3 was published in the *Journal of Eye Movement Research* (Carrol & Conklin, 2014a).

The study outlined in Chapter 4 was published in *Bilingualism: Language and Cognition* (Carrol & Conklin, 2014b).

The study in Chapter 5 has been published as part of a Special Issue of *Bilingualism: Language and Cognition* on Crosslinguistic Priming in Bilinguals (Carrol and Conklin, 2015).

Chapter 6 has been submitted and is currently under review (Carrol, Conklin & Gyllstad, under review).

Chapter 7 has been submitted and is currently under review (Carrol & Conklin, under review).

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

Published and submitted manuscripts

Carrol, G., & Conklin, K. (2014a). Eye-tracking multi-word units: some methodological questions. *Journal of Eye Movement Research*, 7(5), 5, 1-11.

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

Carrol, G., & Conklin, K. (2015). Cross language priming extends to formulaic units: Evidence from eye-tracking suggests that this idea 'has legs'. *Bilingualism Language and Cognition*, published online 20th April 2015.

Carrol, G., Conklin, K. & Gyllstad, H. (under review). Found in Translation: the influence of L1 on the reading of idioms in L2.

Carrol, G. & Conklin, K. (under review). No word is an island: Exploring the mechanisms of formulaic language processing.

Acknowledgements

My PhD research has been supported by a Sir Francis Hill Postgraduate Scholarship, awarded by the University of Nottingham. I have also been supported by research funding from the School of English and via the University of Nottingham Graduate School Travel Prize.

I would like to express my thanks to Kathy Conklin, my primary supervisor, for her invaluable guidance throughout my doctoral studies. I would also like to thank other members of the School of English at the University of Nottingham for their assistance in various ways throughout my time here, in particular my second supervisor, Norbert Schmitt, and the Head of School, Josephine Guy, as well as many others not named specifically here. My thanks also go to the members of the Bilingualism Research Group at the University of Nottingham, in particular Walter van Heuven.

I would also like to thank Henrik Gyllstad for his hospitality and assistance during my visit to Lund University, Debra Titone for her advice on a number of matters, Bodo Winter for email advice on statistical analysis, and Kurt Debono and Marcus Johnson at SR Research for their invaluable assistance on a number of eye-tracking projects.

Finally, my thanks go to my mother, Jill Carrol, and my grandmother, Rozel Chadderton, for their support, and most of all to Abigail Rhodes, for everything.

Terminology and typographical conventions

The terms formulaic language, formulaic units/sequences and multi-word units/sequences are used interchangeably throughout. I make no technical distinction between these terms.

Bilinguals in this thesis are considered to be speakers who have learned a second language to a degree that enables them to transact in that language. This definition is deliberately broad, since more or less rigorous standards of what constitutes a true bilingual can be found throughout the applied linguistics literature. More specific explanations of proficiency level are provided for the bilingual groups in the empirical chapters, as appropriate.

L1 refers to the native language for any speaker; L2 refers to a learned non-native language. Bilingual participants are therefore defined here according to their L1 and the L2 that they have learned, hence a Chinese-English bilingual is a person with L1 Chinese who has learned English as an L2.

Direct quotations are presented throughout in double quotation marks: "…". Terms used in a semi-technical sense are presented in single quotation marks: '…'

Examples of phrases are presented in italics, e.g. *spill the beans*, with meanings provided where required in double quotation marks, e.g. "reveal a secret". When discussing conceptual meaning, concepts as a distinct level of representation are presented in italic block capitals, e.g. *REVEAL A SECRET*.

Examples of unacceptable or ungrammatical phrases are presented with an asterisk, e.g. **spill the bean*.

“Then suddenly, he was struck by a powerful but simple little truth, and it was this: That English grammar is governed by rules that are almost mathematical in their strictness! Given the words, and given the sense of what is to be said, then there is only one correct order in which those words can be arranged.”

From *The Great Automatic Grammatizor* by Roald Dahl

“I know all those words, but that sentence makes no sense to me.”

Matt Groening

Chapter 1. Introduction: The Formulaic Nature of Language

Formulaic language is an umbrella term for the multitude of speech routines and multi-word expressions that are commonplace in natural language (Ellis, 2008; Pawley & Syder, 1983; Schmitt & Carter, 2004; Sinclair, 1991). Broadly, this refers to those word combinations that "appear to be processed without recourse to their lowest level of composition" (Wray, 2002, p.4) or which demonstrate a high degree of predictability, fixedness or conventionality, such as idioms (e.g. *kick the bucket*), phrasal verbs (e.g. *eat up*), binomials (e.g. *king and queen*), collocations (e.g. *abject poverty*), spaced compound nouns (e.g. *teddy bear*), routinised formulae (e.g. *how do you do?*) and frequently occurring lexical bundles or 'chunks' (e.g. *in the middle of*). Far from representing a marginal feature of language organisation, between a third and half of all naturally occurring language might be considered in some way formulaic (Erman & Warren, 2000; Foster, 2001).[1]

Formulaic or multi-word units are an important element of any study of the lexical semantic system, and have become of great interest to researchers in the fields of monolingual language processing and representation (e.g. Cacciari & Tabossi, 1988; Libben & Titone, 2008; Sprenger, Levelt & Kempen, 2006; Tabossi, Fanari & Wolf, 2009; Titone & Connine, 1994, 1999; Wray, 2002, 2008, 2012; Wray & Perkins, 2000), bilingualism and second language acquisition (e.g. Cieślicka, 2002, 2006, 2013; Conklin & Schmitt, 2008; Ellis, 2012; Ellis, Simpson-Vlach & Maynard, 2008;

[1] This figure is entirely dependent on the specific definition adopted, and measures of the proportion of formulaic language vary quite considerably in how they are calculated. For example, the widely-cited figure from Erman and Warren (2000) of around 50% of language being formulaic was based on a subjective analysis of 19 extracts of no more 800 words each, where the authors counted the proportion of items that they considered to be 'prefabs'. The authors state that such an analysis should certainly be interpreted with caution and be treated as an approximation rather than a detailed account.

Siyanova-Chanturia, Conklin & Schmitt, 2012; Siyanova-Chanturia, Conklin & van Heuven, 2011) and language impairment (e.g. Code, 2005; Sidtis, Canterucci & Katsnelson, 2009; Van Lancker Sidtis, 2004, 2012a, 2012b; Van Lancker Sidtis & Postman, 2006; Van Lancker Sidtis, Postman & Glosser, 2004; Wray, 2011). Research in all of these fields is extensive, and only a representative sample is listed here. From this has emerged an increasingly robust body of experimental evidence to elucidate the ways in which formulaic language is represented, processed and produced, allowing us to draw conclusions about the nature of the basic unit in language organisation, the schematic relationships between lexical entries, and the relationship between languages in bilingual speakers.

This thesis takes the idiom as its central concern. Idioms are often seen as prototypical examples of formulaic language (Cacciari, 2014; Siyanova-Chanturia, 2013; Titone, Columbus, Whitford, Mercier & Libben, 2015), and certainly are amongst the most studied of all formulaic types. Idiomaticity, as a broad concept, can refer to any patterns of usage that are specific to a speaker or speech community, but in a more narrow linguistic sense, idioms are non-compositional, lexically fixed combinations that denote a specific figurative meaning (e.g. *kick the bucket*). They are often seen as ‘frozen metaphors’ where an original denotation may have been lost, meaning that idioms can be arbitrary, gnomic, and difficult to fathom unless the meaning is known from prior experience.[2] In linguistic terms, semantically and syntactically idioms can

[2] *Kick the bucket*, which is generally used as an arch-example of a non-decomposable, non-transparent idiom, is a good example of this. Some sources (e.g. www.phrases.org.uk) suggest that in the 16th century *bucket* was a dialect word referring to a beam or yolk used to hang things, and specifically was used to refer to the roof beams of a slaughterhouse. Animals hung by their feet from this beam for slaughter therefore literally *kicked the bucket* during their death spasms. Knowing this, it is much easier to see the connection between the acts, much as *burying the hatchet* is a stereotypical act of making

be considered to represent single choices. That is, an idiom like *kick the bucket* has a single phrase level meaning (“die”), and is syntactically better analysed as a single intransitive verb than a combination of verb + object. This is clear if we compare the two examples below, which share a superficial structure but which differ in how they might best be analysed syntactically:

The old man	kicked	the ball	→	“The old man kicked the ball
[NP]	[VP]	[NP]		
[det-adj-noun]	[verb]	[det-noun]		

The old man	kicked the bucket	→	“The old man died”
[NP]	[VP]		
[det-adj-noun]	[verb]		

Idioms therefore represent a challenge to models of language that consider the ‘word’ to be central. Despite being much more syntactically flexible than is often assumed (Cacciari & Glucksberg, 1991; Konopka & Bock, 2009), idioms are nonetheless fixed in two key ways. They have a conventionalised meaning that is not altered by the context in which they appear (Cacciari, 2014), and they are lexically immutable, in the sense that substituting any of the component lexical items removes the figurative interpretation. For example, *booting the bucket* is not equivalent to *kicking the bucket*,

---

peace. Since this dialect meaning of *bucket* is now lost in modern English, no obvious, transparent relationship between the idiom and its components remains.

despite the near synonymy of *kick* and *boot*.[3] In some cases idioms can be deliberately changed to achieve specific stylistic effects, but generally speaking, unless for deliberate reasons of creativity, idioms are semantically and lexically fixed in a very predictable way. It is the fixedness of idioms, and their status as highly predictable 'known' units, that forms the basis of this thesis.

It is important to point out from the start that idioms represent only a drop in the ocean of formulaic language research, and processing is only one area of interest. Formulaic sequences range from very specific fixed units, such as idioms, to more general variable patterns, such as grammatical frames and schemata (e.g. Beckner et al., 2009; Van Lancker Sidtis, Cameron, Bridges & Sidtis, 2015). Wray (2012, p.237) suggests that by treating all of the many and varied subfields of formulaic language under the same banner, we risk "papering over cracks – even chasms – between distinct endeavors." This thesis therefore investigates specific questions about specific types of formulaic unit, namely those with a high degree of lexical fixedness, of which idioms are the clearest example. It is beyond the scope of this investigation to address issues such as the important social functions of formulaic language (Wray, 2002; Wray & Grace, 2007), although it should be noted that such functions are just as vital in a real-world sense as an understanding of how the brain deals with recurrent word combinations.

The studies reported here examine different aspects of how idioms are recognised, processed, and integrated into wider discourse contexts. They focus primarily on how native and non-native speakers process the form of idioms, which are generally highly

[3] There are possible exceptions to this in some rare cases. For example, an idiom like *hit the sack/hay* could be argued to be a case of lexical flexibility, but equally it could be argued that these are simply two different idioms that refer to the same action.

familiar and predictable, leading to the well documented 'idiom superiority effect' (Tabossi, Fanari & Wolf, 2009). Hence native speakers process the form of idioms more quickly than comparable 'novel' phrases, e.g. *break the ice* is recognised and processed more quickly than a non-formulaic phrase like *break the cup* (Swinney & Cutler, 1979).[4] Native speakers also show little difficulty in understanding the intended meaning of idioms, even when they are opaque or ambiguous. In contrast, most research into how non-native speakers process multi-word combinations has shown that even at high levels of proficiency, formulaic phrases continue to pose problems to learners (Barfield & Gyllstad, 2009; Boers & Lindstromberg, 2012; Kuiper, Columbus & Schmitt, 2009; Laufer & Waldman, 2011; Nesselhauf, 2005), both in terms of producing the accepted form, and in understanding the figurative meaning.

A key question is why such a difference should exist. An obvious starting point is to ask why it is that idioms are processed quickly by native speakers in the first place, and I will discuss this general question in the introductory literature review in the following chapter. Broadly speaking, however, idioms are recognised and processed quickly because they are known (Van Lancker Sidtis, 2012a), and native speakers have a remarkably consistent bank of idioms (and other multi-word combinations) with which they are highly familiar, as demonstrated by the high degree of agreement seen in many of the rating studies that form part of the research in this thesis. Logically, any bilingual speaker also has a store of 'known' word combinations in his

[4] 'Novel' language need not be entirely new in the sense of having never been heard before. Rather, in the context of formulaic language, 'novel' sequences are considered to be non-recurrent combinations that do not show any significant degree of cohesion or fixedness, while 'known' combinations are either highly frequent, highly cohesive, and/or have a single phrasal meaning.

or her first language, so an L1 French speaker will have a store of familiar word combinations that demonstrate the same properties as English idioms (faster recognition and processing, easy understanding of figurative meaning). Hence phrases such as *tomber dans les pommes* (*fall in the apples* – “to faint”) or *un grand cheval* (*a big horse* – “a disgraceful woman”) should show formulaic properties to French native speakers, but are meaningless to speakers of English with no knowledge of French. Importantly, these lexical combinations are also entirely unpredictable in English. The question underlying much of this research is this: what happens to these same word combinations when they are encountered in an L2? That is, do the component words that show such a high degree of cohesion in the L1 also ‘go together’ in the L2? If formulaicity is a result of having encountered such combinations enough times in the past for them to be registered as ‘known’, as argued by a primarily frequency-based approach to phraseology, translating idioms and other formulaic combinations should show no effect as this would effectively render them ‘novel’. As the results of the studies within this thesis show, such an assumption is not borne out, and the implications for models of a “Heteromorphic Distributed Lexicon” (Wray, 2002) in monolingual and bilingual speakers are extremely interesting. Three studies looking at idioms in translation from Chinese (Studies 1 and 2) and Swedish (Study 3) form the bulk of the empirical work in this thesis.

As well as investigating formulaic processing in bilingual speakers, in Study 4 I also present evidence on how monolingual speakers process formulaic language of different subtypes: idioms, but also literal/compositional sequences such as binomials and collocations, where the formulaicity is defined primarily by a conventionalised word combination rather than any degree of semantic idiomaticity. Baldwin and Nan

Kim (2010) call such items “statistical idioms”, and they are of great interest in understanding the complex patterns of co-occurrence that characterise natural language. They also add a rich dimension to models of how form and meaning are processed when formulaic language is encountered in natural linguistic contexts. Again, they fit the focus of this thesis since they are lexically fixed, predictable sequences, but in this case they differ from idioms by not being conceptually ‘single units’. This study therefore allows me to begin to explore this key question of how much idioms represent a special case because of their unique conceptual properties, or whether they represent the wider field of formulaic language as a whole.

Throughout the thesis I use experimental techniques (reaction times and eye-tracking) to investigate specific questions about the processing of formulaic language. The results of each study feed into the next, creating an overall series of investigations that add original observations to the wider literature on formulaic language. Importantly, the use of translation to investigate how words are linked within and between languages provides a novel perspective on formulaic language. In this way I use two mutually informing strands of applied linguistics research to help advance our understanding of both: what can bilingualism tell us about formulaic language, and what can formulaic language tell us about bilingualism?

Because of this experimental approach, the structure of the thesis is somewhat non-standard. There is a general literature review, and in this I discuss some overall questions relating to formulaic language and idioms, to provide an initial grounding for the studies to come. I also review some of the key work in bilingual word processing that is of relevance here. Each empirical chapter is then a self-contained study, and presents its own focused literature review to discuss issues that are specific to the particular investigation. Because of this, some repetition is inevitable in the general introduction to each study, but throughout I have tried to ensure that where information is repeated, this is because it helps to enrich the overall picture being created, for example by explicitly relating the discussion to key aspects that are relevant to each experiment. This also allows me to present each chapter as it was intended when written – as a stand-alone study, with minimal amendments from published versions where required.

1.1 Structure of the thesis

The structure of the thesis is as follows. Each empirical chapter is preceded by a short preface to situate it within the context of the wider discussion and link it to the preceding study. Experiments are numbered sequentially throughout the thesis, for ease of reference.

Chapter 2 provides a general literature review to summarise key issues in formulaic language and idiom research. This helps to provide a general grounding from which to develop the empirical work that follows. The focus is on formulaic language generally, then idioms and models of idiom processing within the literature. Bilingual lexical access is also discussed and related to the underlying research question of the thesis. Specific aspects of bilingual processing as it relates to formulaic language are introduced in the empirical chapters as appropriate.

Chapter 3 is a methodological chapter, introducing eye-tracking as the methodology used most often throughout the thesis. Although the first empirical chapter does not utilise this approach, all subsequent studies do, so it is important to discuss its use and also address some of the challenges inherent in applying it to multi-word analysis.

Eye-tracking is predominantly used to analyse individual words or sentences, hence multi-word units present a particular challenge in terms of the specific approach to analysis that needs to be adopted. I discuss this using evidence from the eye-tracking literature, and outline the method of analysing formulaic language that will be adopted in the relevant studies.

Chapter 4 is the first empirical chapter and presents a study of how non-native speakers (Chinese-English bilinguals) process translated forms of idioms. It uses reaction times in a lexical decision task to establish whether bilingual speakers process translations of known phrases more quickly than control phrases. So, in the same way as native speakers process an idiom like *on the edge of your seat* more quickly than a control phrase like *on the edge of your plate*, do Chinese-English bilinguals process a sequence like *draw a snake and add feet* (a translation of a common Chinese idiom) more quickly than a control phrase like *draw a snake and add hair*? The results of this study are discussed in terms of what they mean for the representation of ‘known’ word combinations in the bilingual lexicon.

Chapter 5 builds on the findings of Chapter 4 by conducting two eye-tracking studies with Chinese-English bilinguals. I again compare translations of Chinese idioms with control phrases to see firstly whether there is a processing advantage for familiar forms (evidenced by a more sensitive measurement than pure reaction times), and secondly whether there is any evidence that translations of familiar word combinations also activate the underlying figurative meanings. The results in this chapter enrich those of the first in a number of ways, and I discuss the implications for models of how formulaic language is represented in the bilingual lexicon.

Chapter 6 extends the previous investigations of translated idioms in several ways. Firstly, it uses participants of a much higher proficiency level: Swedish-English bilinguals, where the level of L2 proficiency is native-like or near native-like. Secondly, it includes only idioms of a short, compact form, where no ‘recognition point’ is available prior to the offset of the phrase. In long idioms like *flog a dead horse*, the final word may become predictable partway through, whereas in shorter idioms like *kick the bucket*, the final word is not unequivocally predictable until it has been seen. Hence this study presents minimally predictable idioms in context-neutral sentences to examine how they are processed. This study also introduces the dimension of congruency by including idioms that exist in both English and Swedish, to see whether these show different patterns of behaviour compared to either L2 only items or L1 translations. More specifically, this chapter investigates whether idioms that are common to both languages are privileged over those forms that only exist in L1 or L2 form. Results are discussed in terms of crosslinguistic influence at a formulaic level and in regards to the multi-word lexicon in first and second languages.

Chapter 7 presents a study of native speaker processing that extends this research to other types of formulaic language. In this study I consider idioms alongside binomials (sequences of noun-and-noun or verb-and-verb that are highly fixed in their order, such as *king and queen* or *salt and pepper*) and collocations (a broad definition of any frequently co-occurring adjective-noun or noun-noun combinations, such as *abject poverty* or *storm cloud*). The purpose of this study is to directly compare processing of different formulaic types (non-compositional idioms and literal/compositional sequences), and also to explore the relationships among the component words of formulaic sequences to test the model of ‘holistic’ storage that is widely assumed in the literature. To do this I analyse reading patterns for canonical structures, to directly compare formulaic subtypes, but also component words used in separate, non-formulaic contexts, to see whether there is any evidence of lexical priming between formulaic partners. I discuss the results in terms of usage-based and constraint-based models of language, and consider how the findings could be applicable to different formulaic units with varied properties and features.

Chapter 8 provides a general discussion of the findings of the studies as they relate to the wider literature on formulaic language. I discuss the implications for models of how formulaic language is stored and processed in native and non-native speakers, and propose some modifications to existing theories to account for the results seen here. I also provide some final conclusions to highlight some of the many areas for future research within the framework of this thesis.

Chapter 2. Putting Your Ducks in a Row

The purpose of this general introduction is to lay a broad foundation for the subsequent empirical chapters. As such, I review the literature on formulaic language generally, and look particularly at existing work on how idioms are processed. I also introduce a key aspect of bilingualism – the selective or non-selective nature of lexical access – since this is an important consideration in some of the studies that follow. It should be noted that the literature on bilingual language processing is extensive, and a complete review of it is unfeasible and not of primary relevance to the current research. Issues specific to each of the studies are discussed in the relevant chapters.

2.1 Formulaic language in linguistic theory

The study of formulaic language has grown into an important concern in modern linguistics. In general formulaic language refers to those sequences of words that are recurrent, cohesive, and highly familiar to native speakers. On a psychological level, it has been suggested that formulaic sequences exist to ease the burden on working memory by utilising the more abundant resource of long-term memory (Conklin & Schmitt, 2008; Tremblay & Baayen, 2010). Hence formulaic language supports speaker fluency (Pawley & Syder, 1983), as well as expressing a number of important ideational, referential, affective/attitudinal, social and discourse functions (Schmitt & Carter, 2004; Van Lancker Sidtis, 2012; Wray, 2002). A common view is that frequently encountered sequences are in some way instantiated as whole forms in the mental lexicon. This is exemplified in the famous “idiom principle” (Sinclair, 1991, p.110): “that a language user has available to him a large number of semipreconstructed phrases that constitute single choices, even though they might appear to be analysable into segments."

The basic tenet that commonly recurring sequences become represented in the lexicon in some way exemplifies the frequency-based approach to formulaic language, which has been empowered by the use of corpora to demonstrate just how predictable and repetitive natural language is. It is now abundantly clear that frequency effects are pervasive in language (Ellis, 2002; Bod, Hay & Jannedy, 2003) and high frequency may be one of the most important defining features of formulaic language (Wray, 2002). Frequency as it relates to multi-word units, however, is a much more complex metric than for single words, where a long-standing body of literature supports a clear facilitative effect of frequency in word recognition and processing (Forster & Chambers, 1973; Scarborough, Cortese, & Scarborough, 1977; Whaley, 1978). Frequency in multi-word units is generally much lower than amongst single words (Jurafsky, 2003) and may operate in a non-linear fashion (Columbus, 2010), but it is clear that on a broad level, statistical information about how often a particular sequence has been encountered is stored in some way and has an effect on subsequent processing. This has been demonstrated through numerous studies that show faster or more accurate processing for more frequently occurring word combinations compared to lower frequency controls (Arnon & Cohen-Priva, 2013; Arnon & Snider, 2010; Bannard & Matthews, 2008; Bod, 2001; Ellis, Simpson-Vlach & Maynard, 2008; Tremblay & Baayen, 2010; Tremblay, Derwing, Libben & Westbury, 2011). Such results, as well as many other studies into idioms and other formulaic subtypes, have led to detailed accounts of the distribution and processing of formulaic sequences, broadly distinguished as either compositional (lexical bundles, collocations) or non-compositional (idioms, phrasal verbs), according to the particular semantic properties of any given unit.[5] Whilst such a bipartite division is useful in an abstract way, a more representative picture is to present formulaic language as a continuum, with entirely literal, compositional utterances at one end, and entirely opaque, non-compositional sequences at the other. Van Lancker Sidtis (2012b) exemplifies such a scale, incorporating not just semantic properties but also attitudinal/affective qualities along a continuum from entirely novel to entirely reflexive/memorised.

A critical facet of any frequency-oriented account of language is that experience plays a vital role in its organisation. This is the central argument of approaches that come together under the broad heading of 'usage-based' (Beckner et al., 2009; Bybee, 2006, 2008; Bybee & McClelland, 2005; Tomasello, 2003). In such models, language representation is a dynamic network of linguistic experiences, where every encounter with a word or combination of words is registered and used to fine-tune the overall representation. In this way, grammar emerges from experience as abstractions of both specific and general patterns (Bybee & McClelland, 2005; Hopper, 1987). This means that language processing is affected by a vast and complex body of statistical knowledge relating to likelihood of lexical co-occurrence (McDonald & Shillcock, 2003a, b), which will be unique to any given speaker. The job of any language user is therefore to acquire a probabilistic map of how language is used within a language community, hence nativelike ability is not simply based on the nuts and bolts

[5] Such lines are not always particularly clear cut. For example, the definition of what constitutes a collocation as opposed to an idiom will often vary widely across studies. Bybee (2006) gives a number of examples (*break a habit*, *change hands*) that she describes as prefabs (collocations) that require at least some degree of figurative/metaphorical interpretation. The definitions in this thesis are applied broadly for the purposes of theoretical framing, and are more narrowly defined for experimental purposes as required.

knowledge of grammar and vocabulary, but requires the mastery of complex patterns of how words are combined. The importance of this to communication and interaction is exemplified by Wray (2009, p.194) who summarises that in this sense, formulaic sequences “ensure easy access to information, fluent delivery (which helps retain the turn), the effective conveying of messages, the meeting of physical and emotional needs, and self-preservation as a group member and as an individual”.

This lies in stark contrast to more traditional combinatorial approaches to linguistic description. In a traditional Chomskyan paradigm (e.g. Chomsky, 1957, 1965), formulaic sequences (especially idioms) represent an anomaly that is of marginal interest (Chafe, 1968). Rule-driven generative models of language therefore cannot account for the aberrant nature of idiomatic expressions, which often defy the normal rules of grammar in a variety of idiosyncratic ways. Further, language competence is distinct from real-world performance, which is seen as a poor reflection of the underlying organisation (reflected also in the langue/parole distinction made by Saussure, and later characterisations by Chomsky of I-language/E-language, e.g. Chomsky, 1995), hence factors like frequency and subjective experience have a limited role in underlying language structure. This in turn means that traditional approaches cannot account for the existence of frequency effects at multiple levels beyond the single word. More recent ‘words and rules’ models (Pinker, 1999; Pinker & Ullman, 2002) contrast with strictly rule-governed generative systems by positing that any sequence that cannot be generated by a series of rules must be memorised as a whole. Such models therefore accommodate idioms as effectively single entries in the lexicon (lexemes), stored in long term memory as a lexicalised form and phrase level meaning. The particular version of this model outlined in Pinker and Ullman

(2002) accounts primarily for irregular forms of inflected verbs, but by extension it does allow for the storage of "many kinds of verbal material". Pinker and Ullman (2002, p.458, emphasis in original), also state: "WR [words and rules] does not posit that regular forms are *never* stored, only that they do not *have to* be… Whether a regular form is stored, and whether stored regular forms are accessed, depends on word-, task-, and speaker-specific factors."

In theory, then, a words and rules approach could account for many of the patterns seen in formulaic language, at least in terms of non-compositional or (syntactically) non-standard forms. Two principle objections exist to suggest that such a model is not sufficient, however. Firstly, the plethora of evidence relating to formulaic patterns amongst entirely compositional, literal sequences (lexical bundles, etc.) demonstrates clear frequency effects even in situations where combination and the application of simple rules should be the default. A basic tenet of words and rules models is that frequency should only affect the representation of lexical items (Pinker & Ullman, 2002; Ullman, 2001), hence such effects for lexical bundles and other sentence fragments are problematic. Secondly, as will be discussed in the following sections, the evidence on idioms shows that they are far from being unanalysed wholes, at least on a lexical level, hence it is difficult to argue convincingly that they belong solely in the category of either lexicon or grammar.

Two more areas of importance should be highlighted, both of which support a usage-based account of how formulaic language is represented. The first relates to diachronic changes to language as a result of conventionalisation. Bybee (2006) outlines how higher frequency leads to a faster rate of various processes of language change, such as phonetic reduction. Hence rates of reduction in highly frequent sequences such as *I don't know* are higher than in less frequent sequences (Bybee & Scheibman, 1999), mirroring effects for single words (such as final consonant deletion). Lin (2010, 2012) discusses how the development of distinctive phonological patterns is also a feature of many types of formulaic language. In particular, idioms often show distinctive stress placement (Ashby, 2006) and intonation contours (Van Lancker, Canter & Terbeek, 1981) that contribute to them being interpreted appropriately by native speakers. Lin and Adolphs (2009) show that lexical bundles such as *I don't know why* operate as single intonation units more than half of the time they are used. Other researchers (e.g. Dahlmann, 2009; Wray, 2004) have also suggested that phonological coherence is evidence of holistic storage, at least at the level of articulatory sub-routines, since frequent use should lead to more fluent, more consistent production (Bybee, 2002).

A second point of interest relates to how formulaic language develops during first language acquisition. Language development is a vast area, so a detailed discussion is beyond the scope of the current review, but a key concept introduced by Wray (2002, 2008) is that of "needs only analysis". The principle underlying this is that in many instances, especially in the case of idioms and other set phrases, meaning may be assigned to the largest possible unit on first encounter with a phrase or chunk, and unless good reason to do so arises, word combinations might never be broken down into their constituent literal parts. Wray (2009) suggests that this can be the case in idioms and other semantically opaque phrases (e.g. *dog collar*), which have often lost their initial literal roots (as seen also in the now obsolete etymology behind *kick the bucket*). Attempting to break such items down is therefore less an important part of natural language processing and more a case of "post hoc linguistic game playing"

(Wray, 2009, p.193). As this applies to language development, it is clear that for children, segmenting individual words out of a continuous stream of speech is often difficult at first, hence it has been suggested that language learning is necessarily initially a holistic process (Bolinger, 1975; Lieven, 1987; Tomasello, 2003). Child language is therefore said to develop from a system of unanalysed holophrases to a more analytical method as grammatical and cognitive sophistication allows this to happen. This is of course not to suggest that children do not acquire single words at all, since much early development will be based on the acquisition of single words presented in isolation. However, Bannard and Matthews (2008) summarise that a broadly usage based account of development would allow for children to move from a restricted set of utterances of large grain sizes to a more productive, combinatory system by generalising from the input they do hear. Bannard and Lieven (2009) characterise this as the basis of formulaicity, since children effectively reuse and creatively recombine previously heard formulae as their developing linguistic and cognitive abilities allow.

The importance of such an argument for this thesis is that it proposes a fundamental principle of first language development, whereby chunks and phrases can first be acquired as wholes, and only later would these whole units broken down further into constituent parts. This will presumably be the case for the vast majority of early ‘chunks’ – commonly heard sequences such as *allgone* or *cupoftea*, which in the majority of cases will be broken down as cognitive and linguistic abilities develop. Needs only analysis posits that for longer, more semantically opaque phrases, this might not be the case, hence Wray (2002) further suggests that for native speakers, formulaic processing may represent the retention of previously established links between component words, rather than the binding of items into wholes. The importance of this will be revisited in the empirical chapters when I compare the way that second language learners approach formulaic items that they learn as older, more cognitively developed language users.

Formulaic language is therefore a vast and multifaceted topic, but clear evidence exists to suggest that it is at least partially represented at a level above the single word. At the very least, language users have a sophisticated and detailed record of how individual words are used together and the contexts in which they occur (Beckner et al., 2009), as borne out by extensive evidence of frequency effects for multi-word combinations of many different kinds. I next turn to idioms, which have been the most studied of all formulaic types, and which present a very particular set of challenges to our conception of the multi-word lexicon.

2.2 The case of idioms

Idioms are prototypical examples of formulaic language. They are lexically fixed, generally familiar expressions with a conventional figurative meaning. As demonstrated by the title of this chapter, this meaning can very often be entirely opaque and highly idiosyncratic (*putting your ducks in row* has a broad meaning of “getting things ready”). Titone et al. (2015) suggest that idioms vary along all of the dimensions relevant to formulaic language more generally: compositionality, literal plausibility, transparency, flexibility and frequency. The importance of idioms is exemplified both in their pervasiveness in natural language (Grant & Bauer, 2004; Grant & Nation, 2006; Jackendoff, 1995; Siyanova-Chanturia & Martinez, 2014), and in their contribution to our understanding of language processing in a general sense.

Despite this, idioms do not necessarily fit into a usage-based framework as well as certain other formulaic types. The main challenge is that individual idioms are simply not that frequent. Despite a clear tendency for people to speak idiomatically in general, idioms in their fixed, citation forms are surprisingly infrequent, at least in terms of corpus frequency (Moon, 1998). The paradox is that they are unquestionably familiar to native speakers, so a key question is how they acquire such a consistent formulaic status in the first place. Answering such a question is in itself beyond the purview of this thesis, but one factor of relevance is the role of salience. Bley-Vroman (2002) rightly points out that this term, or at least the mechanism underlying it, is rather mysterious, but generally it refers to a heightened level of noticing or attention for any given item or structure, which may be for a range of linguistic and non-linguistic reasons. This means that the most salient meanings for any given word or word combination are those that are most strongly encoded or consolidated in the lexicon (Giora, 2003). Huang (2009) differentiates global (context independent) and local (contextually determined) salience, and also considers factors such as personal preference and experience in how salience manifests during the interpretation of figurative or metaphorical sequences, i.e. which interpretation of a phrase will be most strongly activated in any given context. In the case of idioms, even though certain combinations are lower in frequency than comparable strings, it may be that they are instantiated better in memory by virtue of them receiving more attention at the time of first encounter because of their non-compositional nature (Wulff, 2008). Idioms also demonstrate rich pragmatic entries, which might also contribute to their representation and processing by allowing speakers to recall previously used routines (Vega Moreno, 2005). I will return to the notion of salience as it relates to idioms for native and non-native speakers in the empirical chapters, since it is argued that it plays a key role in the difference between native and non-native processing.

Much experimental evidence exists to show that idioms are processed more quickly than non-idioms. This is true both in terms of recognition of form (idioms vs. literal control phrases) and in terms of understanding the intended meaning. There is strong evidence that encountering the first part of an idiom generates a specific lexical expectancy, especially in the case of highly familiar, highly predictable idioms. Tabossi, Fanari and Wolf (2005) showed that this was the case in spoken idiom comprehension. They presented speakers with the initial fragments of idioms, which were then completed in a congruent but non-idiomatic way. Such items showed a significant processing cost compared to entirely literal phrases, suggesting that the first part of the idiom generated a specific expectation of what word *should* follow. In comparison, less predictable idioms showed no difference compared to literal phrases. Other studies concur that, especially in the case of predictable, well known idioms, the processing of form is facilitated compared to non-formulaic, literal control phrases (Gibbs, Bogdanovich, Sykes & Barr, 1997; Siyanova-Chanturia, Conklin & Schmitt, 2011; Swinney & Cutler, 1979).

Such results have been taken as evidence that idioms are privileged in the mental lexicon, but there is much disagreement in the extant literature about exactly how they are represented. Early models adhere to a largely lexical/non-compositional view, whereby idioms are aberrant constructions that do not conform to the processes of normal language. Consistent with the words and rules position (Pinker, 1999; Pinker & Ullman, 2002), the first psycholinguistic models suggested that idioms were effectively ‘big words’ in the lexicon, existing as single, unanalysed wholes (Bobrow

& Bell, 1973; Swinney & Cutler, 1979). The underlying mechanism is therefore that idioms are ‘retrieved’ whole, without the need for compositional analysis of the component words. Swinney and Cutler (1979) called this the Lexical Representation Hypothesis, and argued that effectively two mechanisms were at work simultaneously: a computation of the individual words, and a separate, direct retrieval of the whole idiom, which exists as a lexicalised form. Since retrieval is a faster process, this explains the speed advantage seen so consistently for idioms in comparison to literal phrases. Gibbs (1980) proposed a similar approach (the Direct Access Model), but suggested that consideration of an idiomatic meaning could bypass literal computation altogether. Hence anyone encountering an idiom would interpret it figuratively, and only additionally consider a literal meaning if there was cause to do so. Both of these models take the broad view that idiom recognition and retrieval is independent of literal computation, and that the form of idioms is to a greater or lesser extent fixed.

Much data has subsequently been presented to dispute this view that literal meaning plays no role in idiom processing. One important model proposing an alternative view is the Configuration Hypothesis (Cacciari & Tabossi, 1988). In this, idioms are distributed entries in the mental lexicon, but they are only accessible once enough of the idiom has been seen for it to be recognised – a point referred to as the “key” of the idiom. In this model, analysis proceeds as it would for any string of words, but once the key is reached, the idiom is retrieved directly, leading to activation of the figurative meaning (and termination of a literal interpretation of the component words). Evidence for this came from three cross-modal priming tasks where participants had to make a lexical decision to a word related either figuratively or literally to the idiom. When an idiomatic string was predictable, subjects were faster at judging figuratively related target words compared to control words. For idioms that were not recognisable until the whole combination had been seen, literal target words presented at offset were judged more quickly, but when target words were presented 300ms after the idiom, both figurative and literal target words were facilitated compared to control words. From this, it was suggested that literal computation is an obligatory precursor to idiomatic processing, since the sequence must be processed up until the point at which it is recognised as a known unit before the whole form can be accessed (see also Tabossi & Zardon, 1993; Titone & Connine, 1999). Subsequent studies have widely supported the literal activation of idiom components, but dispute the extent to which it is ‘switched off’ once the idiom is recognised (Cutting & Bock, 1997; Hillert & Swinney, 2001; Holsinger & Kaiser, 2013; Smolka, Rabanus & Rösler, 2007; Titone & Connine, 1994).

Evolution of the principles underpinning the Configuration Hypothesis led to a series of models that fall under the general heading of ‘hybrid’, since they consider idioms to be both distributed representations of individual words *and* single units at some level of representation. Cutting and Bock (1997) presented evidence via structural priming that idioms demonstrate internal syntax, arguing against their representation as unanalysed single units (see also Konopka & Bock, 2009; Peterson, Burgess, Dell & Eberhard, 2001). They found that following presentation of paired idioms, an elicitation task showed that blending errors were higher for pairs with the same syntactic structure (e.g. *kick the maker,* as an amalgamation of *kick the bucket* and *meet your maker*). From this, Cutting and Bock (1997) proposed that idioms are simultaneously complex syntactic phrasal frames, subject to compositional analysis like any other grammatical sequence, and word-like in that they are connected directly to an underlying phrase level meaning. Idioms are therefore represented as lexical-conceptual nodes, associated with the syntactic level to specify the structure for any particular item. This lexical-conceptual node links the individual lexical nodes (lemmas) and the overall phrasal concept. Syntactic priming and increased blending exist because of competing representations at equivalent levels (syntactic, leading to competition in idioms of the same structure, and conceptual for idioms with shared overall meanings).

Sprenger et al. (2006) presented perhaps the best known of the hybrid models of idiom representation. This model, like Cutting and Bock's (1997), was originally designed to account for idiom production, and proposed that idioms exist as individual word forms (lemmas) and an overall lexical-conceptual entry (a "superlemma"). This superlemma entry contains information about the phrase level meaning of the idiom, as well as defining syntactic properties, and is reciprocally linked to each of the component lemmas. Encountering the component words of an idiom therefore activates an underlying superlemma, which in turn activates the figurative meaning and the individual lemmas via spreading activation. Sprenger et al. (2006) showed that an idiom like *hit the road* was primed to a greater degree than a literal equivalent (*clean the road*) by the component word *road*, suggesting that both the individual component lemmas and a whole form entry was being activated. According to this and the extension of the model presented in Kuiper, van Egmond, Kempen and Sprenger (2007), idiomatic superlemmas could therefore compete with other individual lemmas at a lexical level during production. Despite some fundamental differences in productive and receptive language processes, the superlemma model is widely used in the literature to account for idiom processing and representation generally (e.g. Holsinger & Kaiser, 2013; Tabossi et al. 2009; Titone et al., 2015).

Equally important are those hybrid models that account for the multiple factors known to affect idiom recognition and processing. Hence the models described in Titone and Connine (1999) and Libben and Titone (2008) assume that comprehenders use all available information to aid in the recognition and correct interpretation of an idiom, including familiarity and predictability, various semantic factors like decomposability and literal plausibility, and higher level factors like discourse context and salience. Smolka et al. (2007) presented a similar model, showing that in idiom sentences, even once a combination was recognised as idiomatic, literal word meanings were still maintained. They argued for a unitary system for the processing of literal and figurative language, whereby literal word meanings (especially of verbs) make a vital contribution to processing in both cases. Such models can be grouped under a heading of ‘constraint-based’, and underline the fact that idioms are rarely encountered in isolation, but are treated like any other linguistic material in terms of the ways in which they are analysed and integrated into a wider context. Accordingly, context has been shown to affect the degree to which idioms are predictable (Titone & Connine, 1999), how fast they are processed (Gibbs, 1980; Mueller & Gibbs, 1987) and the extent to which a figurative or literal meaning is privileged (Cacciari, Padovani & Corradini, 2007; Colombo, 1993, 1998; Fanari, Cacciari & Tabossi, 2010).

The overarching conclusion that can be drawn from these models is that some description of dual route mechanism exists, whereby idioms are simultaneously analysed as compositional strings and retrieved as whole form representations. Such a dual system has been proposed for other linguistic categories, for example compound words (e.g. *strawberry*), which have been shown to be analysed both as constituent parts and whole units (Jarema, 2006; Libben, 1998; MacGregor & Shytrov, 2013). Much of the work on compounds focuses on the transparency of the component morphemes, and evidence converges on a view that multiple sources of information related to the constituent and the whole form are used during online processing (Kuperman, Bertram & Baayen, 2008).

In idioms this dual route means that individual words are analysed and their literal meanings are activated, but once the idiom is recognised then the underlying representation can be accessed directly. In this way the standard combinatorial mechanism of single word processing must be allied with the retrieval of an underlying lexicalised entry (Pesciarelli et al., 2014). According to the various models, this exists not only as a known combination of individual words, but also as an abstract structural representation of the syntactic properties and a conceptual entry for the figurative meaning of the whole phrase. The literature discussed so far on how idioms are processed therefore suggests that they are not unitary at a lexical level (i.e. they are not unanalysed lexical wholes), but at a higher level. Cutting and Bock (1997) and Sprenger et al. (2006) posited that this level was lexical-conceptual in nature. Arcara et al. (2010) proposed that for irreversible binomials (such as *kith and kin* in English) this was more likely to be at the level of the “input orthographic lexicon” (p.8), although they state that they cannot rule out the possibility that this is also linked to underlying semantic representations.

The importance of this dual route model to idioms will be discussed and exemplified further in Chapter 4.

2.3 Semantic models of idioms

Because of the discrepancy in how frequent idioms are compared to how well known they seem to be, alternative, more phraseological approaches to how they are processed have also been prominent in the literature. The work of Gibbs and colleagues exemplifies this tradition in idiom research, and is best represented in the Idiom Decomposition Hypothesis (Gibbs & Nayak, 1989; Gibbs, Nayak & Cutting, 1989). In this view, semantically decomposable idioms (those where the individual words make an identifiable contribution to the idiom as a whole, e.g. *pop the question*) are processed differently to non-decomposable idioms (e.g. *kick the bucket*). The predictions of this model are that decomposable idioms should be processed quickly, since the results of analysis and retrieval of the idiomatic meaning are consistent with each other, while for non-decomposable idioms the results of retrieval and analysis diverge, leading to slowed processing and impaired recognition due to competition between possible meanings and because of the inherent incongruity between interpretations. Cacciari and Glucksberg (1991) made a three-way distinction among analysable-opaque idioms (e.g. *kick the bucket*), analysable-transparent idioms (e.g. *break the ice*), where there is a clear metaphorical correspondence between components and the whole phrase, and quasi-metaphorical idioms (e.g. *bury the hatchet*), where the idiom is a prototypical exemplification of a particular act.

Many studies have since considered the contribution of semantic decomposability and transparency on idiom processing, with often conflicting results. Titone and Connine (1999) found that reading rates for non-decomposable idioms were significantly affected by a preceding biasing context, whereas decomposable idioms showed no such effect. They suggested that this was evidence that when the components contribute to both the figurative and literal meaning of the phrase, integration of meaning is easier than when this is not the case. Caillies and Butcher (2007) also found evidence supporting an advantage for decomposable idioms (see also Gibbs, 1991; Gibbs & Nayak, 1989). In their study, targets related to the figurative meaning of idioms were processed more quickly for decomposable idioms, where the meaning was available immediately, than for non-decomposable idioms, where effects were only seen around 500ms after the end of the idiom. In contrast, Tabossi et al. (2009) found no difference between non-decomposable and decomposable idioms and clichés (entirely literal formulaic sequences, such as *conquer the world*) on a semantic acceptability judgement task (judging whether a word sequence was a meaningful Italian phrase). All three conditions were faster than control items, and the authors suggested that this was evidence that familiarity rather than any inherent semantic properties, was responsible for the processing advantage. Cutting and Bock (1997), using a cued production task, and Tabossi, Fanari and Wolf (2008), using a series of semantic categorisation tasks, also found no evidence that decomposability played a prominent role in idiom processing.

Libben and Titone (2008) conducted a number of studies to show that the various dimensions of interest in idiom studies affect processing in different ways. For example, they showed that familiarity correlated highly with predictability, while semantic decomposability played a limited role in the earliest stages of recognition, becoming important only in tasks that explicitly required consideration of the semantic value of the phrase as whole. This is a vital and somewhat overlooked finding in the idiom literature, since it demonstrates that the initial recognition/retrieval of a lexical template is a separate process than comprehension of the overall meaning; the first is affected primarily by individual familiarity with the form of the phrase, while the latter is affected more by global measures of overall decomposability. This distinction is vital to the empirical studies in this thesis, since the key questions relate to how easily the form of idioms is processed and how easily the figurative meanings of specific lexical combinations are accessed. Overall, whilst it is clear that compositionality does therefore play some role in idiom processing, especially in terms of meaning, the particular task and the research question will undoubtedly be important in determining how it manifests.

One issue with the distinction between decomposable and non-decomposable sequences is that it is far from being neat and clear-cut, and multiple other factors (familiarity, underlying knowledge of the etymology) may well contribute to this in different speakers. However, what is clear is that idioms do not form a neat, homogenous class (Gibbs, 1995; Nunberg, Sag & Wasow, 1994), at least in terms of their semantic properties. For this reason, throughout the studies in this thesis, compositionality is considered to exist on a continuous scale, rather than as a binary distinction. This is broadly supported in a cognitive linguistic perspective (Wulff, 2013), where compositionality is seen as scalar. Rather than explicitly testing the contribution of this variable by comparing different categories of decomposable/non-decomposable idioms, I will include it as a continuous covariate in the analysis throughout the empirical studies as a way of accounting for its complex and multifaceted nature.

2.4 Neurolinguistic evidence in formulaic language studies

It is worth spending some time on more recent contributions to the literature on formulaic language, where neurolinguistic methodologies have been used to shed light on the processes underlying recognition and processing. Such techniques are of great value when it comes to understanding the mechanisms at work in language processing, since they tap directly into the brain's response rather than requiring the investigator to make inferences based on a behavioural task. Techniques like Electroencephalography (EEG) have a very high level of temporal resolution and a number of well-established components reflecting particular language processes, allowing for experimenters to test detailed predictions about how particular sequences are processed in real time. These components are known as Event Related Potentials (ERPs), which are specific patterns of electrical activity generated by the brain in response to certain stimuli. They are characterised as either positive or negative deflections in a waveform of voltage changes, and by the length of time they take to occur following the onset of a stimulus, hence a P600 would be a positive-going waveform occurring 600 milliseconds after a stimulus, while an N400 would be a negative deflection occurring 400 milliseconds post-stimulus. In formulaic language this literature is still in its infancy, since in general neurocognitive models of language have paid little attention to idiomatic and figurative language (Vespignani, Canal, Molinaro, Fonda & Cacciari, 2010), certainly in comparison to studies of literal language. However, some recent studies have begun to look at the brain's electrophysiological response to idioms using ERPs to compare idiomatic and non-idiomatic language.

One of the most robust ERP components is the N400 (Kutas & Hillyard, 1984). The N400 is elicited by all meaningful or potentially meaningful words (Kutas & Federmeier, 2000), and is well-known to reflect semantic integration. Its amplitude is inversely related to how predictable a word is in any given context (Federmeier, 2007;

van Berkum, Hagoort & Brown, 1999), i.e. less predictable or incongruent words produce larger N400s. Vespignani et al. (2010) presented evidence that the predictive mechanisms at work during idiom comprehension differ from those during literal language processing. They showed that idiomatic sentences produced a qualitatively different pattern to literal substitution sentences, e.g. the Italian idiom *have a hole in one's stomach* ("be hungry") produced a different response compared to the phrase *have a pain in one's stomach*. Specifically, the ERP component prior to its recognition point (its idiomatic key) was a comparable N400 for both types of phrase – suggesting that equivalent predictive mechanisms were at work – whereas after the idiom was recognised it displayed a P300, which was interpreted as an index of template matching. In other words, predictive processes for idioms (once recognised) are driven less by semantic expectancy and more by a categorical expectation of what the specific lexical combination is going to be. Behavioural results from the same study supported this mechanism, with idioms after the recognition point being read more quickly than literal sentences.

Rommers, Dijkstra and Bastiaansen (2013) conducted an ERP study of Dutch idioms and also found evidence that known, fixed lexical combinations behave differently to novel sequences. They tested the hypothesis that semantic integration processes were to some degree 'switched off' during idiom comprehension. In particular, they used an experimental paradigm first exemplified by Federmeier and Kutas (1999), whereby the amplitude of the N400 is 'graded' according to how closely related a target is to what is expected from the context. That is, following a sentence like: *They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of…*, Federmeier and Kutas (1999) found a smaller N400 to the highly expected word *palms* than to either a semantically related word (*pines*) or a more distantly related alternative (*tulips*). Crucially, despite the comparable cloze probability of both control words, the N400 to *pines* was significantly smaller than to tulips, reflecting its closer semantic relationship to the expected word. Rommers et al. (2013) conducted a similar study with idioms (e.g. the Dutch idiom *walk against the lamp*, meaning "get caught doing something illicit"), using comparable control phrases where the final word was either related (e.g. *walk against the candle*) or unrelated/incongruent (e.g. *walk against the fish*). The crucial manipulation was the sentence type, which either supported the idiom meaning or used the same terminal words in a literal context (e.g. *screw the bulb into the lamp/candle/fish*). For literal control sentences using this manipulation, the expected graded N400 was seen (N400 to the semantically related *candle* was smaller than to the incongruent *fish*). Importantly, in sentences supporting the idiomatic interpretation this was not the case. There was no N400 effect as a function of semantic relatedness, and instead the authors reported a P600 effect, which they state is generally seen in response to violations of agreement or orthography.[6] Hence they suggest that encountering an incorrect completion of an idiom represents an agreement error, which supports the lexical status of idioms at some level of representation. Although these results are compelling, it should be noted that a view whereby semantic integration does not take place at all in idioms does not necessarily accord with many of the studies discussed so far that show obligatory literal meaning activation during idiom processing.

---

[6] In general the P600 is taken to reflect syntactic violations (e.g. Friederici, 2002; Osterhout & Holcomb, 1993) although the exact nature of this is disputed in the literature. Broadly speaking, it reflects grammatical/syntactic integration, which in some studies has been shown to include errors in agreement (e.g. Hagoort, Brown & Groothusen, 1993).

A comparable pattern is reported by a number of studies showing similar effects for idioms in Chinese (Liu, Li, Shu, Zhang & Chen, 2010; Zhang, Yang, Gu & Ji, 2013; Zhou, Zhou & Chen, 2004). Collectively, these studies show evidence of early recognition of form/visual word-form mismatch, with later semantic involvement via an N400 which varied according to how compositional the idioms were. Zhang et al. (2013) demonstrated that even for high compositionality idioms, there was a significant difference in the magnitude of an early positivity (P250/300) compared to literal control phrases. This suggests that the nature of a known character sequence, rather than just the contribution of the semantics of the phrase, was triggering such effects. Liu et al. (2010) also showed evidence of P600 effects, and they suggested that this might be seen as a general index of the well-formedness of the phrases as complete syntactic units (in other words, a measure of agreement violation of an expected form).

Interestingly, other studies have shown that non-idiomatic formulaic language also shows comparable effects, which is in line with a view of idioms as prototypical examples of multi-word sequences rather than special entities in themselves. Molinaro and Carreiras (2010) compared literal and figurative collocations and found a similar P300 in both cases, relative to unexpected completions of clusters, which they interpreted as the matching of a word with a categorical expectation (similar to the results reported by Roehm, Bornkessel-Schlesewsky, Rösler and Schlesewsky, 2007, who found P300 effects for categorical predictions, such as in the case of antonyms). Figurative collocations showed additional later N400 effects which they took to reflect the additional semantic processing required to integrate an overall semantic meaning (in line with Coulson and Van Petten, 2002). Other evidence that the categorical P300 effect is not limited to idioms comes from an ERP investigation of binomials (Siyanova-Chanturia, 2010). In this study, binomials (e.g. *knife and fork*) were compared to their reversed forms (e.g. *fork and knife*) and to semantically associated word pairs that were matched with the binomials for association strength (*e.g. spoon and fork*). In both instances there was a significant P300 effect for binomials, with non-binomials showing N400 effects according to the degree of association between words. The same word pairs used without the connector (e.g. comparing *knife-fork* to *spoon-fork*) showed no difference in the waveforms for the binomial or associated pairs (no P300 in either case), and both showed a significantly smaller N400 than a semantically incongruent control pair (*theme-fork*). Siyanova-Chanturia (2010) concluded that once the phrasal nature of the binomials was compromised by removing the connector, their status as single units was removed and they were processed as individual words.

Finally, ERP evidence from Tremblay and Baayen (2010) suggests that lexical bundles (frequent combinations such as *in the middle of*) are processed in a ‘holistic’ manner. They found that whole string frequency for the combinations in their study showed an effect approximately 110-150 milliseconds after the stimulus onset. They argued that the speed with which this effect appeared prohibits the individual processing of four individual words, and supports a view whereby memory traces for the whole form are registered in some way. Combined with their behavioural results, where single word, trigram and whole sequence frequency all contributed to better recall of items, Tremblay and Baayen (2010) concluded that phrasal and non-phrasal combinations are stored and processed as both parts and wholes.

All formulaic units therefore show effects that seem to represent a 'unitary status'. The consistent presence of a P300 effect, taken to represent some form of template matching, suggests that formulaic sequences may be more lexically fixed than some of the idiom models would allow for. However, these positions are not necessarily incompatible, and in some ways the generation of a categorical lexical expectancy is entirely consistent with a view that formulaic sequences are highly cohesive distributed combinations of words. In other words, encountering part of a multi-word combination generates an overwhelming lexical expectancy that is based on conventionality and past experience rather than semantics. A key question throughout this thesis is how exactly this expectancy is generated, and how canonical the form has to be for this process to be triggered. This is not to remove the semantic dimension completely, and another key question, which will be pursued in Studies 2, 3 and 4 in particular, is how overall meaning for idioms is integrated into sentence contexts.

2.5 Bilingualism and lexical access

Bilingualism is used in this thesis as a tool to investigate formulaic language, rather than being the primary concern in itself. Because it is therefore not the major focus of this research, a complete review of the extensive literature in the area is not warranted. However, an important question for this thesis is how bilinguals access words in their two languages, and what implications this has for formulaic language, so the following section provides an overview of this key area.

A central concern in bilingual research has been the question of selective or non-selective lexical access. That is, do bilinguals store and access words from different languages in different ways, or is the lexicon effectively a single system, with all words coded along dimensions of form (orthography and phonology) and meaning? Dijkstra (2007) summarises that the majority of studies support a language non-selective view of bilingual organisation. Hence when processing words in one language, orthographically and phonologically related words in the other are shown to enter into competition (e.g. Dijkstra & van Heuven, 2002; Van Heuven, Dijkstra & Grainger, 1998; Spivey & Marian, 1999; Weber & Cutler, 2004). Other evidence comes from a number of methodologies that make use of interlingual homographs (word pairs with the same orthographic form but different meaning across languages, e.g. *coin*, where the same form means "piece of money" in English and "corner" in French), homophones (same phonological form, different meaning) and cognates (same form, same meaning, e.g. English-French *bus-bus*). For example, homograph pairs have been shown to cause interference in a number of tasks, including lexical decision for isolated words (e.g. de Groot, Delmaar & Lupker 2000; Dijkstra, Van Jaarsveld & Ten Brinke, 1998) and using a more natural sentence reading task (e.g. Conklin & Mauner, 2005). In all cases use of homographs leads to slowed response times since the same form obligatorily activates multiple competing meanings. Cognate facilitation is also shown to be a robust effect in bilinguals, demonstrated in faster processing of words that are cognate between languages (e.g. Allen & Conklin, 2013; Brenders, van Hell & Dijkstra, 2011; Costa, Santesteban & Caño, 2005; Rosselli, Ardila, Jurado & Salvatierra, 2012). This is often taken as evidence that such words share a common lexical semantic entry in the lexicon (Duñabeitia, Perea & Carreiras, 2010). Taken together the evidence supports a view that significant overlap of form between languages leads to co-activation, suggesting that lexical activation is language non-selective. This is accounted for via spreading activation processes, whereby shared formal features activate candidate words automatically, regardless of language membership (Dijkstra, Timmermans, & Schriefers, 2000; Dijkstra & Van Heuven, 2002).

Of more interest to the present study is whether the same is true of translation equivalent words, which share the same underlying meaning but which have a different orthographic and phonological form between languages (e.g. English-French *dog-chien*). Where no formal overlap exists to create interference, do we see lexical competition between languages in the same way? Some researchers suggest that bilinguals have multiple linguistic representations for shared underlying concepts (Bialystok, 2007; Kroll & de Groot, 1997), therefore competition is inevitable. As with cognates, evidence supports the automatic activation of translation equivalent words (e.g. Chen & Ng, 1989; de Groot & Nas, 1991; Dimitropoulou, Duñabeitia & Carreiras, 2011; Duñabeitia, Perea & Carreiras, 2007; Wang, 2007). Priming is generally stronger in the L2-L1 direction than the reverse (Jiang & Forster, 2001), but this effect evens out as proficiency increases (e.g. Kroll, Michael, Tokowicz & Dufour, 2002; Zhao, Li, Liu, Fang & Shu, 2011). Importantly, cross-language activation is shown both for shared-script languages and in language pairs with a different script: English-Japanese (Allen, Conklin & Van Heuven, in press; Hoshino & Kroll, 2008), English-Korean (Moon & Jiang, 2012), and in various studies showing cross-language activation in Chinese-English bilinguals (Thierry & Wu, 2007; Wu, Cristino, Leek & Thierry, 2013; Wu & Thierry, 2010; Zhang, van Heuven & Conklin, 2011). In such cases, where no orthographic overlap exists, it has been argued that any association must exist at a shared conceptual level (Forster & Jiang, 2001).

For the purposes of the studies in this thesis, the important findings can be summarised as follows. Bilingual speakers obligatorily activate lexical competitors in both languages during a range of tasks in auditory and visual lexical processing. In cases of form overlap (homophones, homographs) and complete overlap (cognates), it can be argued that this is due to multiple orthographic or phonological cues that activate lexical candidates in both languages. For translation equivalents, where meaning is shared but forms are different, conceptual mediation via a shared underlying entry may drive the effect. Broadly, this is reflected in the Revised Hierarchical Model (RHM) of bilingual organisation (Kroll & Stewart, 1994). In this model, originally designed to explain asymmetrical production effects for bilinguals (L2 to L1 activation/translation is faster than from L1 to L2), a shared underlying conceptual store underpins lexical representation in both L1 and L2, which are represented as separate lexicons. The lexicon for L1 is assumed to be larger than for L2, and links are stronger between L1 forms and concepts. Since the L2 forms are most often learned via L1-mediation, lexical links in this direction (L2-L1) are strong, and only become bi-directional over time. Links from L2 forms to concepts may also develop over time, as proficiency increases. The conception of separate L1/L2 lexicons runs contrary to the evidence discussed so far in support of an integrated lexicon, but for the purposes of the studies in this thesis, this aspect is relatively tangential.[7] What is important, given the clear evidence for the automatic activation of translation equivalent words, is whether this means that larger lexical entries (multi-

[7] Brysbaert and Duyck (2010) suggested that the overwhelming evidence in favour of language non-selection makes the RHM an outmoded model. In response, Kroll, van Hell, Tokowicz and Green (2010) took up this issue in their review and assessment of the RHM. They suggested that the evidence of language non-selection is not necessarily problematic to the RHM, and that a view of an integrated lexicon is not functionally any different to a view of separate lexicons with parallel access at multiple levels (lexical and sublexical).

word units, idiom superlemmas) also show evidence of cross-language activation. This is demonstrated in Figure 2.1, below, adapted from the RHM.

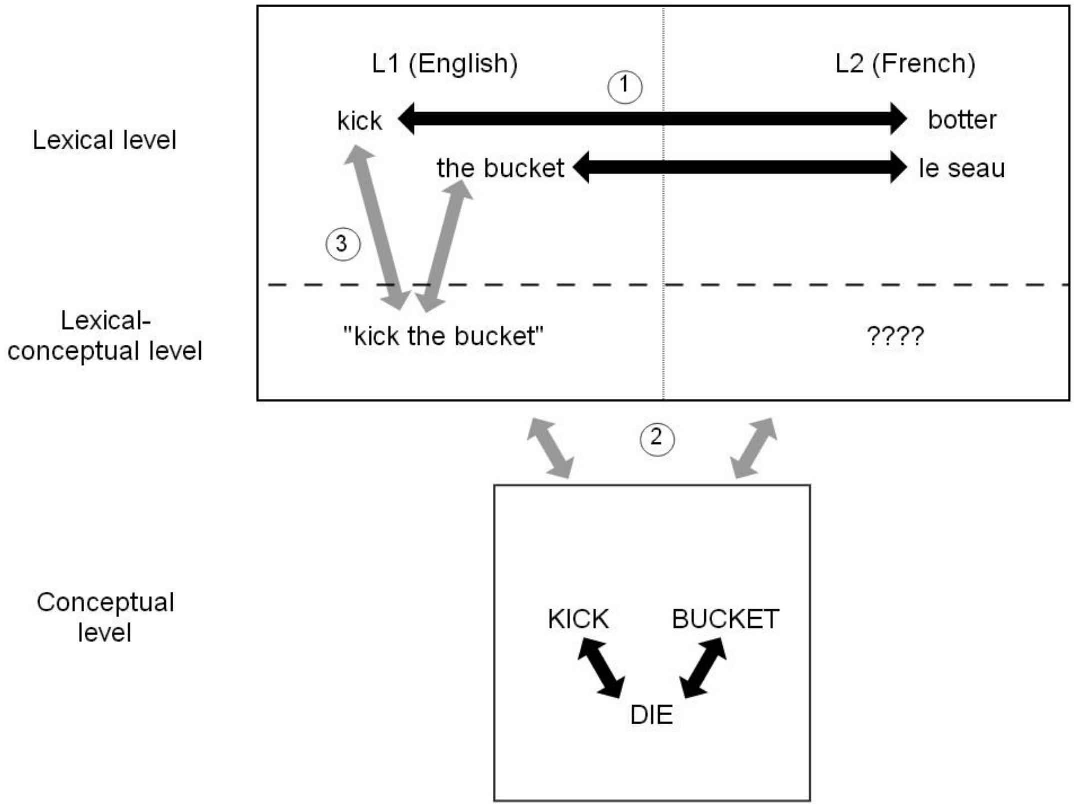

Figure 2.1. Adapted Revised Hierarchical Model (Kroll & Stewart, 1994), demonstrating the general research question of this thesis, i.e. to what degree is formulaicity a property of fixed, language specific word combinations? Black arrows represent links between items at the same level (lexical and conceptual); grey arrows represent links between lexical items and underlying concepts or between component words and 'superlemmas'. It is assumed that all links are reciprocal, but this is likely to be determined by proficiency (for L1-L2 links and L2-conceptual links) and familiarity (for links between individual words and superlemmas).

What Figure 2.1 demonstrates is one of the key questions underlying the present research: what would we expect for a translation of an idiom like *kick the bucket*, where the equivalent word forms in French (*botter le seau*) do not form a formulaic unit and do not carry the same phrasal meaning? It is assumed, based on the evidence discussed so far, that bilingual speakers have connected representations for lexical forms (1 in Figure 2.1), and that these forms are both linked to underlying concepts to varying degrees (2 in Figure 2.1); the reciprocity and strength of both lexical and conceptual links will be determined by proficiency level. Almost exclusively, models of bilingual activation operate at the level of the single word, so the question remains as to how this impacts fixed, formulaic lexical combinations. For an English native speaker, it is assumed (e.g. Sprenger et al., 2006; Smolka, Rabanus and Rösler, 2007) that encountering component words of idioms unlocks a lexical-conceptual representation of the whole form (3 in Figure 2.1). What then would we expect for translated forms, and what implications does this have for how idioms are modelled in monolingual and bilingual speakers? Are superlemmas, if this is an accurate way to characterise idiom entries, accessible via component words in either language? To my knowledge, no models of bilingual lexical access or formulaic language address this, so this is the first empirical study to extend existing bilingual models to multi-word units.

2.6 Summary

This brief overview has introduced several key concepts. Many of these will be explored in more detail in the subsequent chapters, but the following aspects should be highlighted.

Firstly, it is apparent that idioms are recognised quickly by native speakers. Overwhelming evidence for this comes from a range of experimental approaches and different tasks (lexical decision, semantic judgement, natural reading). The accepted view is that idioms are recognised quickly because they are fundamentally known to native speakers (Van Lancker Sidtis, 2012a). They appear to be highly predictable lexical sequences, but in many cases it is not obvious why this should be. In the example of *kick the bucket*, many more typical and likely objects of the verb *kick* are possible, so the question of why *bucket* is seemingly a highly predictable continuation is key. Models of idiom representation therefore converge on a view that they are whole units 'at some level of representation'. This means that they exist both as combinations of individual words, and have some whole-form entry in the lexicon, at least at a conceptual/semantic level.

Multiple factors appear to interact to affect how predictable idioms are, and how they are understood in context, and the task used to investigate this will be an important consideration in what conclusions can be drawn. For example, use of behavioural measures (reaction time studies) are useful in revealing the speed advantage for formulaic sequences but can tell us little else about the underlying mechanisms (Siyanova-Chanturia, 2013). In contrast, methods such as eye-tracking can scrutinise formulaic sequences in more detail, combining analysis of a whole unit with consideration of individual words, and utilising multiple measures to build up a clear picture of the early and late stages of processing. For idioms especially, this provides a valuable way to elucidate the processes of recognition and comprehension within the same study. For this reason, following the first study (Chapter 4) which uses reaction times, eye-tracking is used throughout this thesis as a primary method of investigation. However, before turning to the empirical work on idioms, the following chapter presents a discussion of the merits and challenges of the eye-tracking methodology as it relates to formulaic language.

In addition, although the studies in this thesis do not utilise ERPs as a way of investigating processing, the ERP evidence discussed above adds a valuable dimension to our understanding of how idioms and other formulaic units are processed. Integrating such findings with the behavioural results obtained in the wider literature is key to understanding the nature of formulaic language. In particular, the evidence of a P300 effect for all kinds of formulaic sequences (idioms, binomials, collocations) suggests that they are lexicalised to some extent, or at least that there is some expected, canonical lexical template that becomes active during processing. Combined with the phonological evidence reviewed previously, where formulaic sequences have been shown to develop compact, distinctive intonation patterns, there is some evidence that formulaic language is treated as unitary on a formal level, even if this is not actually a reflection of holistic/whole form storage in the sense of multi-word combinations being single, unanalysed units.

The question of bilingual processing has been dealt with briefly, but the importance to the present studies can be summed up quite simply. Idioms exist as known combinations of lexical items. If frequency and familiarity with the specific form is a key driver of formulaic processing, translating idioms should remove any advantage, as in the L2 they have zero frequency and are assumed to be completely unfamiliar. Given the often arbitrary nature of idioms, this seems like a logical prediction. In other words, translating idioms is effectively changing the lexical form, and as such should show comparable effects to substituting synonyms (e.g. *kick the bucket* → *boot* *the bucket*), since the canonical, expected lexical sequence is violated. Alternatively, since translation equivalent single words show consistent cross-language activation, it is possible that the specific lexical status of idioms is not the most important driver of their fast processing. Instead, language general conceptual properties might exist to unify idioms, leading to facilitation for underlying component lemmas in either language.

Chapter 3. An Eye for Detail: Applying Eye-tracking to Multi-word Units

This chapter introduces eye-tracking, which is the methodology used in the second, third and fourth studies (Chapters 5 to 7). Eye-tracking allows us to investigate normal reading processes and present stimuli in natural contexts. It also allows us to investigate several aspects of processing at once, through a combination of different measures and different foci of analysis. This technique has traditionally been applied to single words, hence its application in multi-word units is potentially highly fruitful but also not without its challenges. This chapter introduces eye-tracking and discusses how it might best be used to investigate idioms and other formulaic units. I present evidence from relevant previous studies that have used eye-tracking to explore language processing, as a way of expanding the literature review presented in the previous chapter, and I conclude with some proposals for how best to accommodate and analyse multi-word units within this framework. Much of the following discussion was published in *The Journal of Eye Movement Research* as "Eye-tracking multi-word units: some methodological questions" (Carrol & Conklin, 2014a).

3.1 When is a word not a word?

Eye-tracking has provided an invaluable tool in the armoury of the modern psycholinguist. For those concerned with the structure of the mental lexicon, it provides an online way to examine how words are recognised, processed and integrated into sentence structures, and to explore the various factors that affect these processes such as frequency, length, ambiguity and other variables. Eye-tracking has therefore been essential for models of single word processing, but, as pointed out by

Clifton, Staub and Rayner (2007), as the length of critical regions of interest increases, it becomes much harder to see precisely where an effect might occur within that region. For this reason, it is less straightforward to use eye-tracking for investigating formulaic language, where sequences of more than one word behave like “single choices” (Sinclair, 1991). This leads to the question of what the unit of analysis for formulaic language should be. The following discussion explores the notion of the ‘word’ and proposes different approaches that could be adopted in eye-tracking research based on those studies that have so far used this methodology to investigate multi-word units (MWUs).

The theoretical basis of eye-tracking as an approach to linguistic investigation is generally quite straightforward. As with other methods such as measurement of reaction times to a given stimulus, eye-tracking considers the amount of time spent on an item to be a reflection of the cognitive effort required to process it. Two assumptions are key to this: a principle of immediacy/incremental processing as each lexical item is encountered, and some degree of eye-mind equivalence, whereby it is assumed that what is being looked at is what is being processed (Pickering, Frisson, McElree & Traxler, 2004, but see also the discussion by these authors relating to how higher-level processes can call this assumption into question in certain contexts). Although different models of eye-movement control in reading vary in their predictions about specific features such as serial vs. parallel allocation of attention (see, for example, the predictions of the E-Z Reader (Reichle, Rayner & Pollatsek, 2003) and SWIFT (Engbert, Nuthmann, Richter & Kliegel, 2005) models), one common theme is that the analysis generally considers the ‘word’ as the primary unit of analysis. Fixations (or skipped fixations) are assigned to a single lexical item, and measurements have traditionally been separated into ‘early’ indicators – metrics like first fixation duration, first pass reading time/gaze duration and likelihood of skipping a given word – which are often taken to be a reflection of automatic processes, and ‘late’ measures – total reading time, total number of fixations and re-reading patterns – which can be seen as largely reflecting the more strategic, controlled processes involved in reading comprehension (Altarriba, Kroll, Scholl & Rayner, 1996; Inhoff, 1984; Paterson, Liversedge & Underwood, 1999; Staub & Rayner, 2007). This preference for treating each word as an individual unit of analysis is justified in Pickering et al. (2004), who suggest that long regions of interest are problematic for several reasons, not least that early effects such as first pass reading time become harder to interpret. The authors state that “our preference has always been to define one word critical regions where possible. Under such conditions, first-pass time, like first-fixation time, is spatially well-localized.” (Pickering et al., 2004, p.5).

Such an approach presupposes one key aspect: that the identification of a ‘word’ is a simple process. However, as argued by Reichle, Liversedge, Pollatsek and Rayner (2009), amongst others, this seemingly straightforward assumption can be deceptively hard to implement. They adopt a working definition of a word as “any sequence of letters that are separated by spaces and that have an accepted pronunciation and meaning in the language” (p.116), but take pains to point out the potential pitfalls for languages other than English where orthographic conventions might make it much harder to identify clear boundaries in this way. A further objection to this definition of the word is taken up by Cutter, Drieghe and Liversedge (2014), who propose that an approach based on this definition of the “basic lexical unit” is potentially vastly underspecified when we consider those items that are considered under the broad heading of formulaic language. This echoes a recent discussion by Wray (2014), who asks how we can even be sure that we know what a 'word' is, and who further argues that any vagueness in our definitions reflects the inherent problem that orthography imposes boundaries that do not always reflect any psychological validity.

Such calls for a rethink on how we might best describe a 'word' are in themselves reflections of the position taken by multiple researchers within the field of formulaic language, where strong evidence has been presented for the representation of (semi) fixed sequences as single entries that are retrieved directly from the mental lexicon (c.f. Arcara et al., 2012; Libben & Titone, 2008; Rommers et al., 2013; Sprenger et al., 2006; Titone & Connine, 1999; Titone et al. 2015). Given that idioms, and other forms of formulaic language, may therefore be represented at some level, an important question is how can we use eye-tracking to investigate the processing of such linguistic forms? I begin by reviewing the existing literature on the lexical and contextual factors that have been investigated to date in eye-tracking research to see what each might tell us about the processing of formulaic language.

3.2 What can single word processing tell us about formulaic language?

Cutter et al. (2014) suggest that (ease of) lexical processing is the main determinant of when the eyes move from one word to another. Or, as Clifton et al. (2007, p.348) put it, "how long readers look at a word is influenced by the ease or difficulty associated with accessing the meaning of the word", and this is an effect that emerges most clearly in early measures. Staub and Rayner (2007, p.330) outline the "intrinsic lexical factors" that affect the reading of individual words. Frequency is a primary determinant of fixation duration (Rayner & Duffy, 1986; Inhoff & Rayner, 1986) and likelihood of skipping (Rayner, Warren, Juhasz, & Liversedge, 2004), but in addition morphological structure (Andrews et al., 2004; Pollatsek, Hyona & Bertram, 2000; Juhasz, Starr, Inhoff & Placke, 2003) and meaning ambiguity leading to competition between lexical representations (Duffy, Morris & Rayner, 1988; Sereno, O'Donnell & Rayner, 2006) both show significant effects on single word reading patterns.

One of the main considerations here is the way in which formulaic language complicates many of these factors. Single word frequency is undoubtedly important, but for multi-word units we might also usefully consider whole phrase frequency and corpus-derived metrics such as mutual information (a measurement of observed unit frequency compared to the expected co-occurrence based on the individual word frequencies and the size of the sample they appear in) or transitional probability (the likelihood of seeing word B once word A has been encountered). It is clear that any given word can become significantly easier to process when it is used as part of a formulaic sequence, especially in the case of idioms (c.f. Conklin & Schmitt, 2008; Gibbs, 1980; Libben & Titone, 2008; Swinney & Cutler, 1979; Tabossi et al., 2009; Underwood, Schmitt & Galpin, 2004). This occurs despite the fact that idioms often use low frequency words (e.g. *bury the hatchet*), sometimes display non-standard morphology (e.g. *toing and froing*), can be inherently ambiguous (*drop the ball*), and often demonstrate highly context-specific meanings (e.g. *spill the beans*, where *beans* acquires a specific figurative meaning that is not assigned to it in any other context). When investigating formulaic language, other factors not relevant to individual words must also be taken into consideration. For example, previous studies on single word processing have generally shown unreliable *n + 2* preview effects (benefit derived from a parafoveal preview of the word two words further on from the point of fixation); when such effects exist they are generally limited to sequences where both *n* and *n + 1* are very short and highly frequent (Kligel, Risse & Laubrock, 2007; Radach, Inhoff, Glover, & Vorstius, 2013). However, a recent study by Cutter at al. (2014) investigating spaced compounds provided what they considered to be “one of the strongest pieces of evidence thus far in favour of MWUs [multi-word units] having unified lexical entries” (p.1784). They found an *n + 2* preview benefit, demonstrated in shorter fixation times for word *n + 1*, when *n + 1* and *n + 2* were constituents of a spaced compound (e.g. *teddy bear*), which they took as evidence that both words were being processed as part of a larger MWU. Crucially, *n + 2* effects were only seen when *n + 1* “licensed” the whole form, leading to an advantage that was not seen for any other combination (when either *n + 1*, *n + 2* or both were non-words). Cutter et al. (2014) argued that the increased length and lower frequency of the *n + 1* items in their study (compared to previous investigations) was evidence of this effect being driven by lexical rather than perceptual factors.

Juhasz, Pollatsek, Hyönä, Drieghe and Rayner (2009) also found *n + 2* preview effects for spaced compounds as well as for novel adjective + noun combinations; they suggested that for their stimuli the high syntactic predictability of the final noun was responsible for the effect in both spaced compounds and novel pairs. However, Cutter et al. (2014) argued that the predictability of word *n + 2* was not on its own a good explanation for their results: *n + 2* only became strongly predicted once *n + 1* had been seen, meaning that *n + 1* would have to be fully identified and integrated during fixations on word *n* if predictability was driving the effect. A similar finding emerged from a study by Siyanova-Chanturia, Conklin and van Heuven (2011), who looked at reading times for binomials (e.g. *bride and groom*). They found an advantage for binomials over their corresponding reversed forms (e.g. *groom and* *bride*) that was not solely attributable to predictability (as measured by a phrase completion test). They concluded that the processes involved in speeded reading of the binomials reflected something over and above simple predictability, and that the phrasal configuration itself played a crucial role.

Clearly predictability is a key component of the formulaic advantage. Previous research on predictability for single words has shown strong effects in terms of shorter first fixation durations and greater likelihood of skipping for more predictable words (Ashby, Rayner & Clifton, 2005; Rayner & Well, 1996), but formulaic language seems to show some level of 'added extra' advantage that goes beyond simple predictability. The question is therefore how eye-tracking might best be used to reveal the mechanism underlying this.

Cutter et al. (2014) do a good job of demonstrating how eye-tracking can usefully be applied to MWUs such as spaced compounds, but longer formulaic items would present considerably more of a challenge. Even for idioms of the common form V-det-N (*kick the bucket*, *spill the beans*, *chew the fat*), the presence of the determiner and the consequent extension of the unit to three words immediately raises the question of what we should be treating as the unit of analysis. The few studies that have used eye-tracking to look at idioms have broadly taken the same approach; that is, an idiom (e.g. *a pain in the neck*) is compared to a control phrase (e.g. *a pain in the back*) and the reading times are compared, either for the phrase as a whole or specifically for the final word (Siyanova-Chanturia, Conklin & Schmitt, 2011; Underwood, Schmitt & Galpin, 2004). This line of enquiry is an extension of other methodologies that have compared formulaic and novel phrases through, for example, phrase acceptability judgements (Swinney & Cutler, 1979; Tabossi, Fanari & Wolf,

2009) and self-paced reading studies (Conklin & Schmitt, 2008; Libben & Titone, 2008). The advantage offered by eye-tracking is that both phrase level and word level patterns can be examined in the same study. In this way Siyanova-Chanturia, Conklin and Schmitt (2011) were able to analyse idioms in terms of both whole phrase reading and sub-part reading (before and after the idiom recognition point or 'key'). They found an advantage for idioms (e.g. *at the end of the day*) vs. controls items (e.g. *at the end of the war*) for whole phrase reading times in late measures but not early measures, and found no effects for sub-part analysis for native or non-native speakers. Other studies (e.g. Underwood et al., 2004) have found facilitation at the single word level for the final word of sequences, which Columbus (2010) suggests is the locus of the formulaic advantage in most cases.

The discrepancy between Siyanova-Chanturia, Conklin and Schmitt (2011) finding effects only for the whole phrase and other studies finding effects for specific words underlines the need to adopt an approach that captures both the macro and micro features of formulaic units. An additional argument for such a dual approach is that it provides a way to accommodate skipping behaviour into analyses. Traditionally, duration measures on single words are only considered for those items that are not skipped entirely during first pass reading. For formulaic items, however, this means actively removing a substantial portion of the items that most clearly demonstrate the expected effect. For example, as will be seen in Chapter 5, native speakers show a tendency to skip the final words of idioms around 30% of the time (e.g. *seat* is often skipped in *on the edge of your seat*) compared to less than 10% for control phrases (e.g. *on the edge of your chair*). Removal of skipped items would therefore impact the idioms much more than other items, leading to an imbalance in the data for any subsequent analyses. Crucially, this would also mean that the clearest examples of the idiom advantage would be discounted from any further durational analysis. One solution, therefore, is to consider both word level measures (skipping rates, then duration measures for non-skipped words) and phrase level measures (duration measures for all items), thereby capturing the full range of behaviour. So for an example like *on the edge of your seat*, analysis of the word level measures may be limited (if *seat* is skipped then no further durational analysis is possible), but the overall phrase level reading times would still be informative across a range of measures, allowing for direct comparison with reading times for non-idiom control phrases. Of course, a notable practical consideration is the increased analysis time that such an approach necessitates, especially if multiple eye-tracking measures are used, but it seems that such a trade-off may well be worthwhile as a way of accounting for formulaic processing in as much detail as possible. Certainly skipping rates should form part of any word level-analysis, hence a method that allows for their consideration alongside other word and phrase level measures is essential.

The evidence discussed above is relatively clear in demonstrating formulaicity, i.e. there is a consistent advantage on a range of measures for idioms, and often the final words in particular, that can perhaps be best explained through their status as part of a formulaic unit. This is especially the case for short items (e.g. V-det-N idioms, binomials or simple two word combinations such as collocations or spaced compounds), where any unequivocal recognition point is not reached until all words have been seen. This is not to say that a whole unit/direct retrieval explanation is prohibitively implicated, and several alternative explanations are plausible (notably a lexical priming mechanism, similar to that proposed by Hoey, 2005). A key question is therefore how we might best utilise eye-tracking to differentiate potential mechanisms of formulaic processing. Clearly a fairly nuanced method of analysis is required if we are to distinguish whole form access from, for example, lexical priming or fast serial mapping of formulaic components (Wray, 2012).

An important conclusion is that those measurements that are typically used for single words (as delimited by orthographic considerations) may not necessarily scale up to formulaic units in a simple fashion. Additional variables that take into account the phrasal nature of such units (based on frequency and cohesion) might therefore be usefully included, as well as semantic considerations like transparency and decomposability. To this end it seems logical to consider phrasal variables in the design or analysis of any eye-tracking investigation of formulaic language as a way of capturing this specifically phrase level behaviour.

3.3 What can syntactic and global discourse context tell us about formulaic language?

The syntactic structure in which a word appears has also been widely investigated in the eye-tracking literature. A basic assumption is that when reading, the natural approach is to produce a word-by-word analysis of the syntactic structure as each word is encountered (the incremental processing assumption highlighted in Pickering et al., 2004). Syntactic ambiguity, therefore, has been the focus of much research, but Staub and Rayner (2007) summarise that very few, if any, studies have demonstrated that such structural competition leads to any cost in terms of reading times. Note that this stands in clear contrast to studies of meaning ambiguity, where lexical competition shows an unequivocal cost in terms of longer fixation durations (as summarised by Clifton et al., 2007). Overall then, it seems that the mechanisms that contribute to sentence level reading behaviour are not the same (or at least not as straightforward) as those that control single word reading. The importance of this to formulaic language is paramount, since often a word-by-word analysis is likely to provide an incorrect interpretation (e.g. for idioms such as *kick the bucket*). Arguably a word-by-word analysis of such items would present both a semantic and syntactic incongruity which would require re-assessment to resolve.

At a global discourse level, there seems to be an effect primarily in later measures of the coherence or otherwise of the overall discourse context, for example, resolution of anaphoric reference or completion of complex inferences within a multi-sentence text (Garrod, O'Brien, Morris & Rayner, 1990; Myers, Cook, Kambe, Mason & O'Brien, 2000; O'Brien, Shank, Myers & Rayner, 1988; Sturt, 2003). Some studies have looked at the global context more in terms of overall meaning, and the conclusion reached by, amongst others, Camblin, Gordon and Swaab (2007) is that global discourse context overrides any local, lexical effects when a rich enough context is provided. Thus, only when an absent or impoverished context is provided do lexical effects such as semantic relatedness emerge. In their study, Camblin et al. (2007) found that effects of disrupted global context were early to emerge and long lasting, as evidenced by significant effects in first pass reading time for a manipulation of the discourse context. When global discourse context was not influential (when it was impoverished or incongruous), low level semantic links showed an effect in terms of shorter reading times for semantically related words within a sentence.

One advantage of eye-tracking is that we can easily insert words into a variety of wider contexts to compare reading patterns. Semantic predictability of specific words as a result of preceding context has been shown to be a strong determinant of reading times (Ehrlich & Rayner, 1981; Frisson, Rayner & Pickering, 2005; Rayner et al., 2004), with words that are strongly predicted or highly constrained showing considerably shorter reading times as well as a higher likelihood of being skipped. Conversely, words that are semantically anomalous (and by definition therefore have low predictability) show inflated reading times (Murray & Rowan, 1998; Rayner et al., 2004). The predictability of formulaic units is, however, not entirely a function of the preceding discourse context: many studies of idioms presented in isolation have shown that the minimal lexical context provided leads to faster processing compared to a control phrase (e.g. Conklin & Schmitt, 2008; Swinney & Cutler, 1979). Underwood et al. (2004) showed that terminal words of formulaic sequences were read more quickly and with fewer fixations than the same words used in non-formulaic contexts, so it is clear that idioms (and specifically the highly predicted final words) are undoubtedly read more quickly and fixated less often than either control phrases or the same words used in non-formulaic contexts. Crucially, this is not driven by global discourse context in the way that semantic expectancy would be.

It seems that context, whether syntactically defined or whether it is provided by a more global discourse mechanism, shows effects that usually emerge in later eye-tracking measures. What is important when dealing with formulaic language is that we have to balance the local, lexical context provided by a very specific combination of words and the global discourse context that might lead a reader to expect a semantically congruent lexical item, whether this is a single word or a formulaic unit. In this sense, using the hybrid models of idiom representation as our guide might represent the best approach, where *the whole is greater than the sum of the parts*. Taking a holistic view of the phrase allows us to examine its behaviour as a whole, while analysis of the individual words (and in particular those that occur later in the sequence) might reveal more about precisely what is being activated; ‘hybrid’ is therefore an appropriate label for such analysis, since it actively combines the most useful elements of two different approaches. In some ways this echoes the overall conclusion reached by Staub and Rayner (2007) that models of naturalistic reading do a good job of accounting for the many lexical factors (length, frequency, predictability, etc.) that affect eye movements, but that higher level factors are to some degree under-explored. They suggest that the lexical factors should be considered as the “primary engine” (2007, p.336), and that higher level structural or discourse considerations will typically exert a later influence, for example in re-reading behaviour or total reading times when additional attention is required to make sense of a problematic text. (It is noteworthy, however, that results from Camblin et al. (2007) outlined previously argue in the opposite direction, suggesting that global features will very often override any lexical level effects.) Again, the conclusion is that using only single words as the base units of analysis in eye-tracking is likely to pose problems and will not necessarily tell us much about how formulaic language is parsed and processed in real time.

To summarise the issue thus far, eye-tracking as a way of investigating the form of idioms and other multi-word units is not necessarily a straightforward process. There is something of a paradox inherent in the analysis of ‘whole units’ through segmentation into component words, while to treat them only as single units is to eliminate the fine grained detail that eye-tracking can provide (and to ignore much of the evidence demonstrating the internal constituency of such units, e.g. Konopka & Bock, 2009; Sprenger et al., 2006). The multi-word space that idioms take up means that the traditional early measures become less reliable on a whole phrase level; at the same time, only utilising later measures would obscure the involvement of the automatic, intralexical processes that are also of interest.

3.4 Phrasal meaning and formulaic language

I have so far considered processing primarily in terms of form, but a second aspect of formulaic language particular to idioms is their meaning (e.g. “die” for *kick the bucket*). Thus, we can also ask to what degree a figurative meaning is activated (as opposed to incremental activation of the literal meanings of component words) and how might eye-tracking be used to explore this? In this regard it seems logical that later measures, broadly reflecting meaning integration, should be more important, i.e. the pattern of overall reading times alongside regression paths/refixation times should be most important in establishing how well any given sequence has been understood within a sentence. Especially in the case of idioms, which presumably always have their own semantic entry (Wray, 2012), a clear pattern should emerge for those items that are understood easily within a given context and those which are not (less transparent, less well known idioms). In this sense, effects should be comparable to those seen for single words. Results summarised by Clifton et al. (2007) regarding lexical ambiguity show that if disambiguating information encountered after an ambiguous word demonstrates that a subordinate meaning was intended, the result is significant disruption to reading (in the form of longer fixations and regressions) as a reflection of the reanalysis that is required. Similarly, Rayner et al. (2004) showed early effects for words that were semantically anomalous, but for words that were merely implausible the effects only emerged in later measures. If formulaic sequences are therefore treated as single units, similar patterns should emerge.

One study to look at this is Siyanova-Chanturia, Conklin and Schmitt (2011), who compared the reading times of figurative and literal uses of ditropic idioms (idioms that can plausibly have a literal and a figurative meaning, such as *at the end of the day*). They found that for native speakers there were no differences on any measures for the two meanings: both were read more quickly than a control phrase (*at the end of the war*) but neither was fixated fewer times or read more quickly in early or late measures than the other. Non-native speakers, on the other hand, showed a clear advantage for literal uses. Importantly, this was observed only in the later measures (total reading time and number of fixations), with first pass reading time showing no difference between a figurative use, a literal use or a control phrase. It seems clear here that the overall reading time, including the amount of time spent in revisiting material, is a fairly robust measure of how easily an idiom has been understood in the wider context, with more problematic (less compositional) material requiring greater consideration and cognitive effort.

Overall, these results seem to support the view of formulaic sequences as whole units (or at least as individual choices/meaning units), since the effects seen for both unknown idioms and implausible single words are comparable. The analysis of whole phrase reading in terms of meaning integration certainly seems to be more suited to late measures, and analysis of regions before and after the idiom might also be a useful way to approach this. For example, as well as the total time spent reading an idiom itself, how much do readers then need to return to the prior context in an attempt to integrate the meaning, or how much time is spent reading a following disambiguating region in the case of literally plausible items? Titone and Connine (1999) analysed idioms and the following disambiguating region and found that results differed according to whether the idiom was more or less decomposable: when literal and figurative analysis of the idiom overlap, meaning integration is facilitated, whereas when the results of literal and figurative analysis differ (for non-decomposable idioms) this process is more difficult, and costs are seen both in terms of idiom reading times and increased reading times for following regions. Cieślicka, Heredia and Olivares (2014) examined idiom processing in English-Spanish and Spanish-English bilinguals. Their results showed that idioms and post-idiom regions were affected by language dominance and contextual support. Total reading times for both idiom and post-idiom regions were shorter for English dominant participants and when context supported figurative meanings, and re-reading patterns for the idioms also demonstrated this effect. Overall, this study suggests that salience and context – key factors in allowing a reader to integrate the intended figurative meaning – are modulated by language dominance, and the effects were seen chiefly in late measures. This supports a view whereby formulaic sequences can be largely equated with single words, at least in terms of how they are understood in any given context. It therefore seems logical that, just as for single words, late measures like total reading time, total number of fixations and regression patterns should be the chief way of examining the dimension of meaning.

There is also a need to accommodate those idiom theories that posit automatic activation of the literal meanings of component words as an obligatory part of idiom comprehension (c.f. Cieślicka & Heredia, 2011; Holsinger & Kaiser, 2010; Smolka, Rabanus & Rösler, 2007; Sprenger et al., 2006; Titone & Connine, 1999). One clue to resolving this may come from the literature on compounds (both spaced, as in Cutter et al. (2014) discussed earlier, and non-spaced, e.g. *newspaper*). Ample evidence suggests that compounds are decomposed (Andrews, Miller & Rayner, 2004), and this is true whether they are semantically transparent or otherwise (Pollatsek & Hyönä, 2005; Juhasz, 2007). It is important, therefore, to also consider aspects such as compositionality and transparency (traditional metrics in idiom research) and their potential influence on eye movements when deciding on the best approach for the analysis of formulaic units. In this regard, it should also be noted that the discussion so far has focused largely on idioms, but it is equally important to consider how other types of formulaic language might best be analysed, especially items such as collocations (*abject poverty*) and binomials (*king and queen*) which are formulaic only by virtue of frequency and conventionality rather than because they represent a 'single meaning' in any way.[8] Again, a hybrid approach might represent the most flexible solution, but careful consideration of the many intralexical factors that have been identified in previous studies is equally important.

3.5 Conclusions

This discussion has aimed to highlight a gap in the application of eye-tracking to natural reading behaviour. The 'traditional' measures of eye-tracking relate broadly to single words, and more recently this has been applied to sentence-level syntactic processing and discourse-level understanding/integration, but formulaic sequences have become an important consideration in modern linguistics and must be accommodated in any theoretical approach to language and reading. The key issue is how we might distinguish between the determinants of processing for individual lexical items, such as predictability from context or single word frequency, and a

[8] The distinction between idioms and frequency-defined formulaic sequences will be revisited in detail in Chapter 7.

more complex representation of MWUs (which undoubtedly includes predictability but which may well reflect a more nuanced level of cohesion within the mental lexicon). In other words, how do we identify the ‘added extra’ advantage that formulaic sequences seem to have over matched, non-formulaic language, and how do we distinguish this from other language processing mechanisms that might be at play?

It is therefore an open question as to how we might best reconcile these lines of investigation. Eye-tracking has the considerable advantage of presenting the text all at once in a naturalistic way, so it is of great value to the investigation of formulaic language as it can be presented in highly natural contexts. Our methods of interpreting the data, however, must be refined if we want to say more about the nature of this important linguistic phenomenon. Clifton et al. (2007) make a clear distinction between those lexical factors that are best reflected in early measures and the higher level influences that may require a broader set of measurements. Given that formulaic sequences seem to fall to some extent between these two stools, it seems necessary to reconsider our approach to their analysis. A fruitful method might be to borrow the ‘hybrid’ model adopted in the idiom literature and consider formulaic sequences as simultaneously compositional strings and whole units, thereby gaining the maximum benefit of analysis of each word and an overall consideration of the phrase. Crucially, however, formulaic units are neither one thing nor the other: they are not simply combinations of individual words and they are not immutable, unanalysed wholes, so our analysis must bear this in mind and be tailored accordingly.

This discussion has shown that a traditional approach to eye-tracking that takes the single word as its basic unit of analysis is problematic when we consider the range of linguistic units that are inherently multi-word in their construction. The flexibility of eye-tracking and the range of measures available mean that the tools are already in place to tackle this issue, but clearly determining how to apply these measures represents one of the next challenges in the application of this methodology to the study of the 'word'.

Chapter 4. Getting Your Wires Crossed

The first empirical study is an initial exploration of how idioms are represented in the bilingual lexicon, and uses reaction times as a way to explore the question of how non-native speakers treat formulaic units. Specifically, it considers whether the formulaic advantage discussed so far is lost when idioms are translated into a second language. This allowed me to test predictions of idiom models as they relate to frequency of encounter (i.e. is fast processing contingent on an idiom being presented in a language-specific form?), and relate this to models of the bilingual lexicon. The study presented in this chapter was published in *Bilingualism: Language and Cognition* as “Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2” (Carrol & Conklin, 2014b).

4.1 Introduction

Formulaic language (idioms, speech formulae, clichés, etc.) is no *flash in the pan*. The definition of formulaic language used here is taken from Wray and Perkins (2000): “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.” Such sequences account for between a third and a half of spontaneous discourse (Erman & Warren, 2000; Foster, 2001). They contribute to speaker fluency (Pawley & Syder, 1983), facilitate real-time communication (Code, 1994) and reduce demands on working memory (Conklin & Schmitt, 2008). They present a particular challenge to non-native speakers as they are both an important part of native-like competence and one of the hardest aspects of a language to master. Cieślicka (2006) suggested that a better understanding of how non-native speakers acquire and utilise formulaic language should be a key goal of modern psycholinguistic and applied linguistic research.

As outlined in Chapter 2, a dual route model (Van Lancker Sidtis, 2012b; Wray, 2002; Wray & Perkins, 2000) can provide a means of describing formulaic language processing in native speakers. In this view, two possible approaches to processing are available to speakers: frequent, familiar phrases are stored in long-term memory and can be accessed or retrieved directly, while novel phrases are computed using a words-and-rules approach. It is important to note that whilst the advantage for formulaic language is often referred to as ‘retrieval’ or ‘holistic storage’, such terms are used here more as a convenient shorthand to describe attested processing differences between formulaic and novel language. The processing advantage for formulaic language could reflect the unitary storage of whole forms, but equally it could represent the simultaneous activation of the component parts of a phrase or the priming of multiple combinations via the base components (Wray, 2012, p.234). Throughout, ‘retrieval’ should therefore be taken to refer to access to the components and meaning of a familiar phrase in a way that is quicker than computing a comparable control phrase. This offers formulaic sequences an advantage over matched novel language as it is a qualitatively different and fundamentally faster process than computation (Tabossi, Fanari & Wolf, 2009).

Whilst a processing advantage is clear for native speakers (see Van Lancker Sidtis, 2012a or Wray, 2012 for reviews), formulaic language processing in non-native speakers remains comparatively unexplored, particularly in terms of how the bilingual lexicon might accommodate two distinct processing routes when more than one language is involved. I aim to address this question by investigating how sequences that would be formulaic in a first language (L1) are processed when they are encountered in a second language (L2). For example, if a French-English bilingual speaker encounters the English sequence *howl with the wolves,* will he or she recognise and retrieve the underlying French idiom *hurler avec les loups* (comparable to the English idiom *follow the crowd*)? If formulaic language represents the storage and association of frequently encountered forms, then it is logical to expect such units to be language specific, i.e. encountering a known sequence in an unfamiliar (L2) form should show no advantage relative to a matched control phrase. If an advantage is observed for unfamiliar translated forms, this would implicate some level of L1-L2 interaction in the processing of formulaic sequences, i.e. since the configuration *howl with the wolves* does not exist in English, any processing advantage cannot be located at a purely lexical level in the L2. Despite the wealth of research into formulaic language to date, no study has investigated this question directly.[9]

4.2 Evidence for a dual route model

Formulaic language is processed more quickly than matched novel language by native speakers. This has been consistently demonstrated for idioms (Gibbs, 1980; Swinney & Cutler, 1979), collocations (Durrant, 2008), corpus-derived multi-word units (Ellis, Simpson-Vlach & Maynard, 2008; Jiang & Nekrasova, 2007), binomials (Siyanova-Chanturia, Conklin, & van Heuven, 2011) and multi-word lexical verbs (Isobe, 2011). This formulaic/novel discrepancy is supported by widespread evidence of different

---

[9] Studies have examined the processing of congruent idioms (forms that exist in both languages), but to my knowledge, this is the first study to directly translate idioms that only exist in the L1. The only comparable study comes from Ueno (2009), which dealt with collocating word pairs, and which is discussed later in this chapter.

patterns in the brain’s electrophysiological response (ERP) to such stimuli (Siyanova, 2010; Tremblay & Baayen, 2010; Vespignani, Canal, Molinaro, Fonda & Cacciari, 2010), and by evidence of different patterns of performance for left hemisphere and right hemisphere brain damaged patients (Code, 2005; Van Lancker Sidtis & Postman, 2006). The wealth of psychological and neurological evidence strongly supports two distinct routes for language processing according to the nature of the material being processed (Van Lancker Sidtis, 2012). The retrieval route can be utilised for previously encountered phrases, and recognition of such a phrase will provide access to the underlying canonical form, its conventional meaning and its pragmatic conditions of usage. Subjective familiarity ultimately determines whether the direct route is available, and the dual route model can be seen as a race rather than an either/or choice: computation still takes place for known phrases, but direct access returns the same results more quickly (and in the case of figurative language is more likely to return the intended meaning than literal analysis), whereas for unfamiliar phrases only the computation route is available. Tabossi et al. (2009) showed that familiarity was the main driver of the processing advantage for both non-compositional formulaic sequences (idioms) and compositional units (clichés).

The present experiment uses idioms, which are “evidently formulaic” (Wray, 2008, p.28). Non-decomposable idioms, or what Grant and Bauer (2004, p.40) call “Core idioms”, present a particular problem because they are, at a surface level, incomprehensible, opaque and gnomic. Crucially, idioms are ubiquitous in discourse and their figurative meanings are processed without difficulty by native speakers. For the present purposes, it is their clear formulaicity that is important; this means that idioms have an attested L1 citation form that will be recognised and understood by a majority of native speakers, and the question is whether the advantage offered by direct access is based primarily on recognition of form. Given the importance of familiarity, it seems logical that presentation in a non-native language should impair recognition of the formulaic sequence. However, idioms are often much more flexible than people assume (Schmitt, 2005), and native speakers generally have little trouble dealing with non-standard and creative idioms provided they are not too far removed from the citation form (Omazic, 2008). Hence while early models (e.g. Bobrow & Bell, 1973; Gibbs, Nayak & Cutting, 1989; Swinney & Cutler, 1979) broadly described idioms as single entries in the lexicon, more recent hybrid accounts (e.g. Cacciari & Glucksberg, 1991; Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Sprenger, Levelt & Kempen, 2006) have attempted to incorporate the syntactic and lexical flexibility of idioms, as well as attempting to explain the finding that both literal meanings of individual words and the idiomatic meaning of the whole phrase seem to be available during idiom processing. Idioms may therefore be simultaneously compositional and non-compositional (Kuiper, van Egmond, Kempen & Sprenger, 2007), which argues against a view that they are represented as single, unanalysable units. They instead may represent configurations with distributed meanings in the lexicon, according to the Configuration Hypothesis proposed by Cacciari and Tabossi (1988), or they may represent separate lexical-conceptual entries – what Sprenger et al. (2006) call superlemmas – that are accessible via the component words. A dual route model therefore allows idioms to be directly accessed, which unlocks both their lexical components and the phrasal figurative meaning. Figure 4.1 shows a representation of a dual route model for the English idiom *flog a dead horse* (meaning “to persevere pointlessly with a task that will have no positive outcome”).

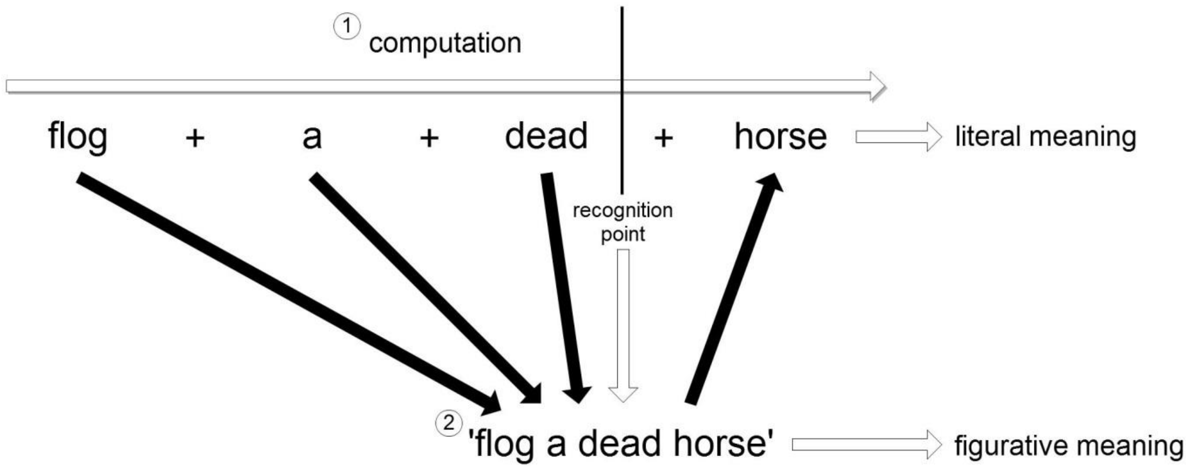

Figure 4.1. Dual route model for the English idiom *flog a dead horse*. The two routes represented are obligatory analysis and computation according to the individual words and grammar (1), and direct recognition and activation of the lexical-conceptual configuration of the idiom (2). Black arrows represent associative links between components and white arrows represent processes.

When *flog a* is encountered, obligatory analysis and computation begins (1 in Figure 4.1), until the recognition point is reached (the 'key' in Cacciari and Tabossi's (1998) terms). Logically this must be the default approach, because it is only by encountering enough of the component parts of a known phrase that it can be recognised and unlocked. For any sequence, therefore, the computation route is available, but previously encountered phrases, once the recognition point has been reached, may also be accessed directly (2 in Figure 4.1). Hence encountering the combination of *flog, a* and *dead* triggers enough associations to activate the known configuration of the idiom. As *horse* is a part of this configuration it is automatically activated before it is encountered as part of the computation. The retrieval route is therefore faster because, in the case of idioms with an early recognition point, the final components are activated before they are encountered via compositional analysis. Because *horse* has been activated as part of the full idiom, if this is the next word to be encountered then it will be processed more quickly, but if another word appears (e.g. in a control phrase like *flog a dead beast*), processing will continue compositionally.

For idioms with late recognition points (i.e. only after the final word, as in *kick the bucket*), the temporal advantage is perhaps not as clear. However, a processing advantage for such idioms could occur for two reasons. Firstly, encountering *kick the* should activate *bucket* to some extent, even though the idiom has not been fully recognised, especially if the context is supportive of the idiomatic usage. While unequivocal recognition might not occur until the final word has been seen, the idiom is likely to be already activated at least to some degree. This is congruent with Sprenger et al. (2006), who suggested that idiom recognition is contingent on reaching a threshold of activation based on encountering progressively more components of a phrase. This threshold may therefore represent confirmation of the idiom, but each component will contribute something toward idiom activation. Secondly, once the final word of an idiom has been encountered it will be activated both as part of the idiom and as part of a computational analysis. Hence *bucket* would be activated by both routes simultaneously, providing an advantage relative to a control phrase (e.g. *kick the packet*), which would only be activated via the computation route.

For any novel phrase, only the computational route is available. Until an idiom (or other formulaic sequence) has been encountered with enough frequency to form associative links between components and therefore create configurations, no direct access will be available, hence non-native speakers are unlikely to be able to use the direct access route until a certain level of proficiency has been reached. There is evidence that once they have encountered formulaic sequences in the L2 with enough regularity, non-native speakers demonstrate the same advantage as native speakers (Jiang & Nekrasova, 2007; Isobe, 2011). There should therefore be no fundamental difference in principle in how native speakers and non-native speakers process formulaic language, but there is likely to be a large discrepancy in the strength of associations available to trigger direct access. This means that non-native speakers are more likely to process L2 formulaic language compositionally and to encounter problems when this does not produce intelligible results (e.g. in the case of entirely opaque idioms).

4.3 Translated idioms and cross-language priming in bilinguals

The dual route model discussed previously makes clear predictions about why formulaic units should be processed quickly by native speakers. However, it is less clear in its predictions for translated idioms. There is widespread evidence to support priming effects in bilinguals for single words (Chen & Ng, 1989; de Groot & Nas, 1991). Translation equivalents in particular (e.g. *dog-chien*) consistently show cross language facilitation for bilingual speakers, which Wang (2007) suggests is a reflection of their shared conceptual representations. Therefore, there is clearly some level of interaction between single word representations in different languages. However, an important consideration is that such associative links are likely to be highly asymmetrical. Whilst a French-English bilingual is likely to have connected representations for the L1 and L2 forms of *hurler-howl*, *avec-with* and *les loups-the wolves*, the lexical associations between these items that unlock the underlying idiom should exist only in the L1 (French). Utilisation of the direct route across languages may therefore require mediation via a conceptual level, whereby the individual L2 forms activate their conceptual representations and the associations at this level trigger the concept underlying the idiom.

Some studies have shown cross language effects at a level above the single word, which would lend support to a conceptual basis for the dual route. For example, Japanese-English bilinguals responded more quickly to unconnected English word pairs that were translations of L1 Japanese collocations (e.g. *forgive marriage*) than to unrelated control pairs (Ueno, 2009). Wolter and Gyllstad (2011) found similar results for Swedish-English bilinguals, with facilitation for English word pairs that formed congruent collocations in English and Swedish (e.g. *give an answer*, which is the word for word translation equivalent of the Swedish phrase *ge ett svar*) relative to English only collocations (e.g. *pay a visit*, where the Swedish translation equivalent for *pay* cannot be used idiomatically in a phrase like **betala ett besök*). Both studies concluded that language non-selective conceptual associations can drive lexical effects in the L2. Given the evidence for cross language effects in single words and collocations, it seems logical that larger units (idioms) may demonstrate similar effects. The current experiment will explore that question by investigating whether Chinese-English bilinguals show any facilitation for Chinese idioms that have been translated into English relative to matched controls.

4.4 Chinese idioms

Chinese has a large set of homogenous idioms that are ideal for the purposes of the current investigation. Chengyu (“fixed expressions”) generally consist of four fixed characters, with no semantic substitution or syntactic flexibility possible without destroying the integrity of the idiom. Around 97% of all chengyu conform to the four character structure (Liu, Li, Shu, Zhang & Chen, 2010). They are generally semantically opaque, and many refer to a folk story or historical event. Understanding the intended meaning is therefore contingent on either knowing the underlying story or learning the arbitrary idiomatic meaning of the sequence.

Chengyu are formulaic units in Chinese (Simon, Zhang, Zang & Peng, 1989; Zhou, Zhou & Chen, 2004) and have been shown to hold the same processing advantage as English idioms. This has been demonstrated through faster reaction times to chengyu relative to matched control sequences (Liu et al., 2010; Zhang, Yang, Gu & Ji, 2013) and through ERP data showing different responses for idiomatic and matched non-idiomatic sequences (Zhou et al., 2004; Liu et al., 2010). Chung, Code and Ball (2004) described similar patterns of impairment in individuals with aphasia for Chinese and English speakers, i.e. differential performance on formulaic vs. novel language. This evidence supports a dual route mechanism for idiom processing in Chinese, just as in English.

For the current investigation, sets of English and Chinese idioms were prepared to explore the responses of Chinese-English bilinguals to formulaic language from the L2 (English idioms) and translated from the L1 (translated Chinese idioms). The responses of English native speakers were also collected for comparison. A lexical decision task was used to compare responses to idioms and matched controls for both languages. If the processing advantage for idioms is based on recognition and retrieval of known forms, there should be no advantage for translated idioms for the non-native speakers. It is also likely that any advantage for English idioms will be driven by proficiency. Native speakers should show an advantage for L1 (English) idioms vs. controls and no effect for translated idioms.

4.5 Experiment 1

4.5.1 Participants

Nineteen native speakers of English (with no experience of learning Mandarin) and 19 non-native speakers of English took part in the experiment for course credit. The non-native speakers all had Mandarin Chinese as their first language and were students undertaking a year of study abroad at the University of Nottingham. A summary of the non-native participants is shown in Table 4.1. All non-native participants were asked to complete a short language background questionnaire and a vocabulary test (modified from Nation & Beglar, 2007). The test presented a series of vocabulary items, each embedded in a short, context-neutral sentence (e.g. *Poor: we are poor*) and participants were asked to choose from five possible definitions: a correct response, three distractors and a *Don't know* option, added to minimise guessing (as per the suggestion in Zhang, 2013). The test included two items each from the first ten BNC word lists (the 10,000 most frequent word families in English) to give a total proficiency score out of 20 (see Appendix 1a for actual words used). This was augmented with any potentially unknown vocabulary items that appeared in the online experiment (e.g. in the Chinese idiom *a horse does not stop its hooves, hooves* might be an unfamiliar English word so was included in the test to verify whether it was known to the participants). Any words that appeared in the stimulus phrases (primes or targets) that were outside the 2000 most frequent word families in English were included in the test. If any participant failed to choose the correct response for a word from one of these idioms, the idiom containing that word was removed from the analysis for that participant. This meant that 33 words were included in the modified vocabulary test to give a total of 53 items. The language background questionnaire asked participants to provide information about the length of time they had been studying English and to estimate their English proficiency in reading, writing, listening and speaking (score out of five for each discipline). They were also asked to indicate how often they used English in their everyday lives (speaking to friends, attending lectures, reading in English for pleasure, watching TV, etc.). Each of these was scored on a five-point Likert scale and then aggregated into an overall usage rating (10 measures, each scored out of five to give an overall score out of 50). Both the vocabulary test and language background questionnaire were administered after the online experiment to eliminate any danger of repetition effects.

Table 4.1. Summary of non-native speakers' age, years of studying English, self-rating of English proficiency, estimate of usage and vocabulary test scores.

	Age	Years studying English	Reading	Writing	Speaking	Listening	Usage	Vocab
Mean	20.8	10.2	2.9	2.9	2.8	2.8	34.6	11.7
Range	19-22	5-15	2-4	2-4	1-5	2-4	26-46	7-16

Note: Reading, Writing, Speaking and Listening self-rated out of 5 (1 = Poor, 2 = Basic, 3 = Good, 4 = Very good, 5 = Excellent); Usage is an aggregated estimate of how frequently participants use English in their everyday lives (score out of 50 based on 10 measures such as reading for pleasure, watching TV, etc.); Vocab is a modified Vocabulary Levels Test with a total score out of 20.

4.5.2 Materials

The stimulus materials consisted of English idioms, English control phrases, translated Chinese idioms and translated Chinese control phrases. Control items were formed by replacing the final word of the corresponding idiom with an unrelated but logical alternative (e.g. *spill the beans* vs. *spill the chips*).

English idioms were selected from the Oxford Learner's Dictionary of English Idioms (Warren, 1994) and were chosen to have a monosyllabic final word that was either a noun (e.g. *jump the gun*) or in one case an adjective (*the coast is clear*). As recognition of familiar phrases was the main concern, no distinction was made between types of idioms, for example the core idioms, figuratives and ONCEs (one non-compositional element) classification developed by Grant and Bauer (2004). To ensure that the stimuli were generally well known, all English idioms were normed on a population of native speakers using a cloze test (i.e. to reveal a secret is to *spill the…*) and were correctly completed by at least 90% of respondents. Mean length of the final word of each idiom (the target) was 4.5 letters and mean occurrence in the British National Corpus (BYU-BNC, Davies, 2004) was 21 (per 100 million words). Control items were created by selecting an alternative final word that was matched with the original for part of speech, length and frequency. Independent samples *t*-tests showed no difference between the idioms and the control items for length (p = .69) or frequency (p = .43). All alternative phrases showed a phrase frequency of 0 in the BNC.

Chinese idioms were initially selected from the Dictionary of 1000 Chinese Idioms (Lin & Leonard, 2012). Only idioms where a literal translation provided a plausible English sequence with identical word order were considered, e.g. 畫蛇添足 – *draw-*

snake-add-feet = *draw a snake and add feet*, meaning "to ruin something by adding over-elaborate and unnecessary detail". The final character had to have a monosyllabic single word translation equivalent in English. The 20 that most closely matched the English idioms in length and frequency of the final word were retained. Four Chinese speakers confirmed that all were well known (all recognised by 4/4 speakers); this was not used as a strict norming test as all idioms were later assessed for subjective familiarity following the online experiment, but was intended simply to make sure that the idioms were likely to be recognised by the majority of participants. Translations were initially taken from the gloss provided by the Dictionary of 1000 Chinese Idioms. Because the intention was to recreate the form of each idiom as closely as possible, the translations were checked character by character using two different online translation engines (Google Translate and On-line Chinese Tools). In this way it was possible to get good agreement on the best literal translation of each character sequence. The translations were finally verified by three native speakers of Chinese who agreed that they were accurate representations of the Chinese originals. The mean length of the final word of each translated idiom was 4.7 letters and all translated Chinese idioms showed a phrase frequency of 0 in the BNC. Control items were created by replacing the final word of each translated idiom with a word matched for part of speech, length and frequency that formed a plausible sequence (e.g. *draw a snake and add hair*). Independent samples *t*-tests showed no difference between the idioms and the control items for length (p = .73), and a marginal difference for raw frequency (p = .09), although there was no difference for the frequency band of the items (p = .77). All alternative phrases showed a phrase frequency of 0 in the BNC.

A set of literal English phrases was constructed to act as filler material. All were literally plausible, grammatical English phrases (e.g. *carry the tray*) and each showed a phrase frequency of 0 in the BNC. Targets were monosyllabic and matched the idiom conditions for length (mean = 4.5 letters) and frequency. Non-word targets were created to make an equal number of word/non-word responses. All non-words were taken from the ARC non-word database (Rastle, Harrington & Coltheart, 2002), conformed to the phonotactic rules of English and were matched with the other conditions for length (mean = 5.0 letters). Primes for the non-words were a mix of unused items from the English idiom, Chinese idiom and English literal conditions.

All idioms were assessed for compositionality using a method adapted from Tabossi, Fanari and Wolf (2008). English native speakers (n = 16) were presented with the English and Chinese idioms and a literal paraphrase of each (e.g. to *spill the beans* means "to reveal a secret"). Participants were asked to judge on a seven-point Likert scale how easily they thought the meaning of the idiom could be mapped onto the literal paraphrase. The mean rating for English idioms was 4.6/7 and for Chinese idioms was 3.8/7. In addition, the Chinese idioms were presented in the original Chinese characters to a set of Chinese native speakers (who did not take part in the online task, n =12) who were asked to judge on a seven-point Likert scale how much they thought the individual characters contributed to the idiomatic meaning. The mean rating by Chinese native speakers was 5.5/7. There was no correlation between the two sets of compositionality judgements (r = .33; p = .16), and the discrepancy is itself a point of interest. In some ways the English speakers' ratings may represent a 'purer' measure of compositionality for the Chinese idioms, as they have no knowledge of the folk story or historical event that underpins the idiomatic meaning;

their judgements are therefore based entirely on how clearly the linguistic information contributes to the figurative meaning of the Chinese idiom. In contrast, the Chinese native speakers may see the idioms as more transparent as a result of knowing the underlying stories. The analysis will include both variables to see if either measure has an effect on response times.

The stimuli were divided into two counterbalanced lists with an idiom and its control appearing on opposite lists. Each participant saw 10 English idioms, 10 English controls, 10 translated Chinese idioms, 10 Chinese controls, 20 English filler items and 60 items with non-word targets (see Table 4.2). Independent samples *t*-tests showed no significant differences between the lists in target length (A = 4.55; B = 4.55; $p = 1.00$), target word frequency (A = 9860; B = 10101; $p = .95$) and phrase frequency (English idioms only: A = 20.8; B = 21.8; $p = .86$).

Table 4.2: Example of stimulus materials for each condition.

Condition	Prime	Target
English idiom	On the edge of your	seat
English control	On the edge of your	plate
Chinese idiom	Draw a snake and add	feet
Chinese control	Draw a snake and add	hair
Control phrase + real word	Put it in your	dish
Control phrase + non-word	Cut a long story	tealth

Care was also taken to ensure that the idioms on each list were balanced for compositionality, including both the scores by English native speakers (for both sets of idioms) and Chinese native speakers (for Chinese idioms only). The lists showed no significant differences for native speaker ratings of English idioms (A = 4.5; B = 4.7; $p = .52$), English native speaker ratings of translated Chinese idioms (List A = 3.3; List B = 4.3; $p = .17$) or Chinese native speaker ratings of Chinese idioms (List A = 5.3; List B = 5.7; $p = .43$). Stimulus materials from the experimental conditions are presented in Appendix 1b.

4.5.3 Procedure

The experiment was conducted in a quiet laboratory using E-Prime (v.1.4.1.1) to present participants with the prime phrases and the target words for the lexical decision task. Reading of the prime phrases was self-paced: participants were asked to read the phrase as quickly as possible, then to press a button to advance once they had finished reading. A self-paced protocol was adopted to allow for the variation in reading time between native and non-native speakers. Once the prime disappeared a line of asterisks appeared on screen. After 250ms this disappeared and the target was presented. Participants used a serial response button box to indicate whether the target was a real English word (YES/NO). Accuracy and response times (RTs) were recorded. The task was explained to each participant via on-screen instructions and two examples and six practice items were presented. The stimuli were then presented in random order until each participant had seen all 120 items.

Following this participants were asked to rate all idioms for how familiar they considered them to be. For native speakers all idioms were presented in English. Participants used a seven-point Likert scale to indicate familiarity with each phrase.

For non-native speakers the English idioms were presented in English and the Chinese idioms were presented in the original Chinese characters. Participants were again asked to rate how familiar they were with each phrase on a seven-point Likert scale.

4.6 Results and analysis

Two non-native speakers were removed from the analysis: both had a large number of extreme response times, suggesting that either they were not engaging in the task or that it was overly difficult for them. This left data from 17 non-native speakers and 19 native speakers. The non-word data and filler items were not included in the analysis. Incorrect responses were also removed, which constituted 2% of the data for both native and non-native speakers. Extreme values (RTs greater than 3000 milliseconds) were also removed, and for both native speakers and non-native speakers this represented less than 1% of the data.

The non-native speaker results were then adjusted to take into account any unknown vocabulary items, which removed 17% of the non-native speaker data. The distribution of unknown words was comparable for each of the conditions (Chinese idioms = 22 unknown words, Chinese controls = 21; English idioms = 36, English controls = 31).[10]

There were no significant differences in terms of errors for either native or non-native speakers (native speakers ANOVA by condition, p = .74; non-native speakers, p = .98). Only correct RTs were submitted to further analysis. Unsurprisingly, native

[10] The higher number of errors for English idioms probably reflects the number of words that are commonly used in English only in an idiomatic sense, therefore non-natives are unlikely to encounter such items in literal contexts or in isolation. For example, no non-native-speaker correctly identified the definition for *buck* as used in the English idiom *pass the buck*.

speakers had overall faster RTs than non-native speakers (NS mean = 576ms; NNS mean = 701ms), and independent samples *t*-tests showed that the difference was significant: $t_1(34) = -3.17$, $p < .01$; $t_2(45.6) = -8.23$, $p < .001$. Patterns of performance for each group were analysed separately with linear mixed effects models using R (R Development Core Team, 2009) and the lme4 (Bates & Maechler, 2009) and languageR packages (Baayen, 2009). Within the models MCMC sampling was used to calculate *p*-values of all factors.[11] RTs were log-transformed to reduce skewing as far as possible and LogRT was taken as the dependent variable. Distribution of response times for both participant groups is shown in Table 4.3.

Table 4.3. Mean response times in ms (SD in brackets) for native and non-native speakers in each of the four experimental conditions. Non-native speaker values are vocabulary-adjusted (any unknown items removed).

	English idioms	English controls	Chinese idioms	Chinese controls
Native Speakers				
Raw RT	539 (175)	610 (211)	579 (235)	577 (152)
Log RT	6.25 (0.27)	6.37 (0.28)	6.31 (0.28)	6.33 (0.23)
Non-native Speakers				
Raw RT	707 (243)	716 (312)	653 (212)	729 (259)
Log RT	6.52 (0.29)	6.51 (0.33)	6.44 (0.28)	6.54 (0.31)

[11] There is an ongoing debate about how to calculate degrees of freedom, and therefore significance values, in mixed effects modelling. In this chapter the degrees of freedom are calculated as ((number of observations) - (number of fixed effects)). The *p*-values of any effects are estimated through the MCMC sampling process.

4.6.1 Native speakers

A linear mixed effects model was fitted with the original language of each phrase (English vs. Chinese) and phrase type (idiom vs. control) as fixed effects. List, target length and log-transformed target frequency were also included as fixed effects, as were the experimental factors of trial order and response time to the preceding item. Subject and item were treated as crossed random effects. Non-significant effects were removed from the model (list, $p = .52$; target length, $p = .35$; log-transformed target frequency, $p = .21$).

There was a significant effect of language ($\beta = 0.0630$; $t(738) = 2.53$; $p < .01$) and phrase type ($\beta = 0.1166$; $t(738) = 4.69$; $p < .001$). The interaction between the two was also significant ($\beta = -0.1008$; $t(738) = -2.88$ $p < .01$). These effects were confirmed by fitting separate mixed effects models for the English and Chinese stimuli. For English idioms vs. controls phrase type was significant ($\beta = 0.1159$; $t(367) = 4.30$; $p < .0001$), while for Chinese idioms vs. controls it was not ($\beta = 0.0170$; $t(369) = 0.75$; $p = .44$).

4.6.2 Non-native speakers

The vocabulary adjusted values were used for analysis of the non-native speaker data.[12] A linear mixed effects model was fitted to assess the effects of original

[12] The non-vocabulary adjusted values were also analysed and a comparable pattern of results was found: Chinese idioms (mean = 668ms) were responded to significantly faster than Chinese controls (mean = 761ms). This was confirmed using a mixed effects model, where the interaction between language and phrase type was significant ($\beta = 0.0996$; $t(659) = 2.08$; $p < .05$); the effect of phrase type was significant for Chinese ($\beta = 0.0630$; $t(326) = 1.90$; $p < .05$) but not English ($\beta = -0.0520$; $t(329) = -1.77$; $p = .10$) stimuli.

language and phrase type. Fixed and random effects were the same as for the non-native speakers. Non-significant effects were removed (list, $p = .69$).

Neither language (β = -0.0440; $t(550) = -1.25$; $p = .16$) nor phrase type (β = -0.0271; $t(550) = -0.76$; $p = .40$) were significant on their own but the interaction did approach significance (β = 0.0796; $t(550) = 1.62$; $p = .07$). To further explore this, separate linear mixed effects models were fitted for Chinese phrases and English phrases. Phrase type (idiom vs. control) was significant for Chinese phrases (β = 0.0664; $t(285) = 1.93$; $p < .05$) but not English phrases (β = -0.0339; $t(261) = -1.04$; $p = .31$).

4.6.3 Familiarity, compositionality and proficiency

Because familiarity (Tabossi et al., 2009; Libben & Titone, 2008) and compositionality (Gibbs et al., 1989; Gibbs, 1991; Caillies & Butcher, 2007) have been suggested to influence idiom processing, these factors were explored further using linear mixed effects models.

All idioms were highly familiar to the respective native speaker groups and relatively unfamiliar to the opposite groups (on a seven-point Likert scale where 1 is completely unfamiliar and 7 is highly familiar, English idioms for native speakers, mean = 6.4 and for Chinese native speakers = 2.8; Chinese idioms for Chinese native speakers, mean = 6.5 and for English native speakers = 2.8). For native speakers relative familiarity was not a significant variable for English idioms (β = -0.0146; $t(182) = -1.22$; $p = .24$), but it was marginally significant for Chinese idioms (β = -0.0170; $t(185) = -1.83$; $p = .09$). This suggests that the English items, being at or near a ceiling of familiarity, showed very little variation in response times according to fine-grained differences. The Chinese items that are more predictable seem to have been judged as more familiar, for example, *doesn't know good from bad* was judged as familiar by English native speakers (mean = 6.4/7), even though it is not a common English phrase (0 occurrences in the BNC). Inclusion of association norms taken from the Edinburgh Associative Thesaurus (Kiss, Armstrong, Milroy & Piper, 1973) confirms this (i.e. using the score for the association between *good* and *bad* as an index of predictability): including this variable as a fixed effect was significant ($\beta = -0.1580$; $t(184) = -2.11$; $p < .05$), and this removed the effect of familiarity for Chinese idioms ($\beta = -0.0098$; $t(184) = -1.00$; $p = .34$). Non-native speakers showed no variation according to how relatively familiar the idioms were: familiarity was not significant for English idioms ($\beta = -0.0002$; $t(130) = -0.02$; $p = .98$) or Chinese idioms ($\beta = -0.0171$; $t(141) = -0.96$; $p = .42$). Taken together these results indicate that relative familiarity did not modulate response times for idiom completions, but it should be remembered that all items were deliberately chosen to be highly familiar so this lack of variation is perhaps unsurprising.

Compositionality was also included in the analysis to assess its contribution to response times. Two measures were used: compositionality ratings from English native speakers (judgement of English forms of both English and Chinese idioms) and an additional rating of the Chinese idioms in the original Chinese characters by a set of Chinese native speakers. I assumed that all control items are potentially just as compositional as their corresponding idioms, i.e. for native English speakers, the Chinese idiom *draw a snake and add feet* and the control *draw a snake and add hair* could both just as easily mean "ruin with unnecessary detail", hence they are equally compositional. In addition, because the prime phrases are the same (e.g. *draw a snake and add…*), the contribution of the compositionality of the prime phrase must be comparable across the idiomatic and control conditions. Table 4.4 summarises the results of analysis according to compositionality, showing analysis of all stimuli (idioms and controls) and of the idiom conditions separately.

Table 4.4. Contribution of compositionality to response times to English and Chinese stimuli.

	English stimuli	Chinese stimuli	
	English judgements	English judgements	Chinese judgements
Idioms + Controls			
Native Speakers	.95	.52	.88
Non-native Speakers	.61	.07	.82
Idioms only			
Native Speakers	.92	.44	.91
Non-native Speakers	.24	.04*	.46

Note: Data are p-values estimated from mixed effects models with compositionality as a fixed effect: $* \ p < .05$

For native English speakers none of the measures of compositionality demonstrated an influence on response times for either set of stimuli. For non-native speakers the English idioms were not affected by compositionality and for the Chinese items only the English native speaker judgements of the translated versions were significant. When idioms and controls were considered together there was a marginally significant effect of compositionality ($\beta = 0.0274$; $t(284) = 1.86$; $p = .07$) and a significant interaction with phrase type ($\beta = -0.0520$; $t(284) = -2.40$; $p < .05$). Analysis of the conditions separately showed a significant effect of compositionality for idioms (β = 0.0303; $t(141) = 2.19$; $p < .05$) but not controls (β = -0.0234; $t(139) = -1.14$; $p = .23$), so it is clear that compositionality did not drive the advantage for the idioms compared to controls for non-native speakers (when compositionality was included the difference between idioms and controls was still significant: β = 0.2608; $t(284) = 3.02$; $p < .01$), but the idioms themselves were affected by the degree of compositionality. This supports Caillies and Butcher (2007), who found an advantage for decomposable over non-decomposable idioms, however it should be noted that their study looked at meaning activation (lexical decision task on targets related to the figurative meaning), which was not required by participants in this study.

Proficiency level can also play a role in non-native idiom processing (Conklin & Schmitt, 2008; Ueno, 2009), so this was explored as a factor for the non-native speakers. Non-native speaker proficiency based on vocabulary score was non-significant for overall performance (β = -0.0107; $t(549) = 0.42$; $p = .66$) or as part of a three way interaction with language and phrase type (β = 0.0076; $t(549) = 0.29$; $p = .80$). All other direct measurements of proficiency (vocabulary score, self-ratings of speaking, reading, writing and listening skills and usage score) were shown to be non-significant (all *p*s > .05). The only significant indicator was the length of time studying English (β = 0.0313; $t(549) = -1.67$; $p < .05$), which may simply show that longer exposure leads to a better ability to recognise and judge English words (greater lexical knowledge, awareness of English forms, etc.). Importantly, analysis of the English and Chinese materials separately showed no interaction with phrase type for English items (β = -0.0086; $t(261) = -0.85$; $p = 0.39$) or Chinese items (β = 0.0044;

t(284) = 0.45; *p* = .68). Longer exposure to English therefore improved response times across the board, but did not affect the pattern of performance for any participant.

The lack of any direct effect of proficiency may be relatively unsurprising given the homogenous nature of the non-native participant group. All were from the same study abroad cohort and had broadly comparable proficiency and experience in English. In contrast, Ueno (2009) manipulated proficiency and found a significant difference between high and low proficiency groups. It is likely that in order to see an influence of proficiency, participants that have a wider range of proficiencies would need to be examined.

4.7 Discussion

A clear pattern of results for native (English) and non-native speakers (Chinese L1) was observed. Both native and non-native speakers responded most quickly to targets that formed idioms in their respective L1s, and the difference relative to matched control items was significant. The native speaker results are important as they support multiple previous studies of an advantage for idioms over matched novel language. They also show a clear pattern of performance according to overall familiarity: the English idioms showed an advantage over control items because they were known, whereas the Chinese idioms were not, and hence response times in the idiom and control conditions did not differ. Importantly, the English idioms showed no grading according to familiarity, so more familiar idioms were not significantly faster than less familiar ones. This may be simply be a reflection of the fact that stimuli were deliberately chosen to be common and familiar, hence any variation was likely to be extremely fine-grained (probably too fine-grained to significantly affect the RTs). Compositionality was not a significant factor either for fundamentally familiar

(English) or unfamiliar (Chinese) idioms. The native speaker results support the assertions of Tabossi et al. (2009) and Van Lancker Sidtis (2012a) that overall familiarity (whether an item was known or unknown) is the main driver of idiom recognition and therefore formulaicity.

A complementary pattern of results was observed for non-native speakers. Targets that formed English idioms were not reliably faster than controls, suggesting that these had not been encountered with enough regularity to form phrasal representations in English, which is contrary to evidence that advanced non-natives can show a formulaic advantage (Jiang & Nekrasova, 2007; Isobe, 2011; Underwood, Schmitt & Galpin, 2004). This is however in line with the general inconsistency of results, where sometimes non-native speakers show a processing advantage and other times they do not. Chinese idioms, despite being presented in an entirely unfamiliar form (English), did show an advantage over the control phrases. Relative familiarity within the idiom condition was not significant, suggesting that it was simply the status as known (idioms) or unknown (controls) that drove the advantage.

The finding that RTs to translations of L1 idioms by Chinese speakers is speeded poses an interesting problem for the dual route model. Van Lancker Sidtis (2012a) suggested that formulaic expressions differ from other utterances because they are not newly created. Importantly, in a purely formal/lexical sense, the translated Chinese idioms *were* novel, and the non-native participants are highly unlikely to have encountered the sequences in English (as evidenced by their 0 frequency in the BNC and the lack of familiarity for native speaker participants). Thus a canonical, learned configuration, stored as a result of multiple previous encounters and activated via associative lexical links, cannot explain the advantage observed for the translated idioms. What therefore accounts for the advantage for the translated Chinese idioms, and can this advantage be explained by the dual route model?

One possible explanation is that idioms can be activated at a conceptual level. Unlike some other forms of formulaic language, idioms have their own separate conceptual entry (i.e. *spill the beans* means *REVEAL A SECRET*); Wray (2012) suggested that it may be this aspect that offers them an advantage over non-idioms. One view of the bilingual lexicon is that there is an underlying shared conceptual system, hence learning L2 items involves the mapping of new forms onto existing concepts. Over time and as proficiency increases, direct links from L2 forms to concepts can be created, allowing bilinguals to bypass the L1 forms (Kroll & Stewart, 1994; Wang, 2007). If this is accurate, then idioms may exist as unitary concepts that are accessible via lexical forms in either language. Encountering the English prime (e.g. *draw a snake and add*…) therefore activates the underlying concepts of the component words (*DRAW, SNAKE, ADD*) in the shared bilingual conceptual store, and the associations of these at a conceptual level triggers the idiom entry (*RUIN WITH UNNECESSARY DETAIL*). This unitary concept activates the figurative meaning, but also activates the whole phrase and therefore the expected completion (*FEET*), making the lexical form of the target available either directly in the L2 if a strong enough link has been created (e.g. *feet*), or in the L1 (足). Because this L1 form is a translation equivalent of the presented target, facilitation for the English form *feet* is still observed in either case.

Such a view is broadly in accord with the conclusions reached by Ueno (2009) and Wolter and Gyllstad (2011). In their studies of collocations they proposed that lexical forms in the L2 (English) activated associative links in a language non-selective way, i.e. at a conceptual level, hence words that would form collocations in the L1 will be primed even when they are encountered in the L2. In particular, Ueno (2009) found that the effect increased with L2 proficiency: for her participants responses to both translated L1 collocations and L2 collocations became faster as proficiency increased, which she suggested was evidence of a strengthening of the separate links between the L1 and L2 lexical systems and the shared conceptual system. The results of Experiment 1 show no variation according to proficiency but do show faster response times as a result of increased number of years studying English. This may suggest that increased exposure can lead to more efficient access to L2 forms (or possibly just better ability to judge English words/non-words), but without a more rigidly defined set of high and low proficiency participants it is difficult to say any more about the development of direct conceptual access. If Ueno's hypothesis is correct, it is reasonable to expect that a higher proficiency group would show a more pronounced idiom superiority effect for the translated idioms, and probably also an effect for English idioms as increased exposure would be likely to generate idiom entries, at least for the most frequent English items.

A conceptual basis for cross-language priming beyond the single word level is therefore plausible, but the results do not provide unequivocal support for this. The task was designed to investigate whether the form of an idiom was the principle driver of recognition; participants therefore did not need to access any conceptual information in order to complete the task, i.e. a lexical decision could be made based solely on the form of the target word rather than on any associated semantic meanings (literal or figurative). A lexical translation-based process may therefore provide an alternative way to account for the results.

Zhang, van Heuven and Conklin (2011) demonstrated the process of fast automatic translation for Chinese-English bilinguals. They used English word pairs in a masked priming task with very short presentations (59ms) and found that the Chinese translation of the prime word was influential (i.e. when the prime-target showed a repeated morpheme in the Chinese translation there was facilitation compared to when the prime-target produced translations with unrelated morphemes). They concluded that the participants must be translating and decomposing the English primes quickly and automatically for the Chinese morphology to show an effect in the completely English task. The same process may be at work in the current study. Presentation of the prime phrases could be quickly and automatically translated and decomposed, hence the L1 characters are activated and their associations as part of an idiom are recognised at an L1 lexical level. This activates the overall Chinese idiom, which primes the final character; because this is a translation equivalent of the target in English, facilitation for the L2 form is observed.

In this explanation the configuration priming the idiom is language specific in that it is driven by associations at a lexical level in Chinese. Wang (2007) showed inter-language priming only for direct translation equivalents; in the current study, whilst the individual words are translation equivalents, the phrases are not (they do not exist in both languages), hence any associations at a lexical level must be driven by the L1 (Chinese). The study by Wang (2007) highlights another important factor, which is the influence of strategic processes. In the current study primes were not masked and were presented in a self-timed protocol, potentially giving participants ample time to read and translate them, make associations in the L1 and predict the final character, which would yield faster response times when the English target was a translation equivalent of the expected completion. Idioms present a particular challenge in this regard because they are generally unsuitable for masked priming because of their length, therefore alternative methods may be required in future to disambiguate automatic and strategic translation processes for such stimuli. However, whether translation was fast and automatic or strategic, an influence of the known L1 configurations was still observed in the L2. As with the conceptual explanation, increased proficiency might affect the process: as the lexical links between L1 and L2 are reinforced, activation of L1 forms via the L2 would become faster, hence the idiom advantage might also become stronger if the effect is driven by lexical/translation processes.

Proposing that faster processing for L1 configurations in an L2 could have a conceptual or lexical basis broadly reflects the position of Bley-Vroman (2002), who identified both a lexical frequency based and a meaning based motivation for formulaic language processing. Both explanations for the results can be incorporated into a dual route model, as shown in Figure 4.2

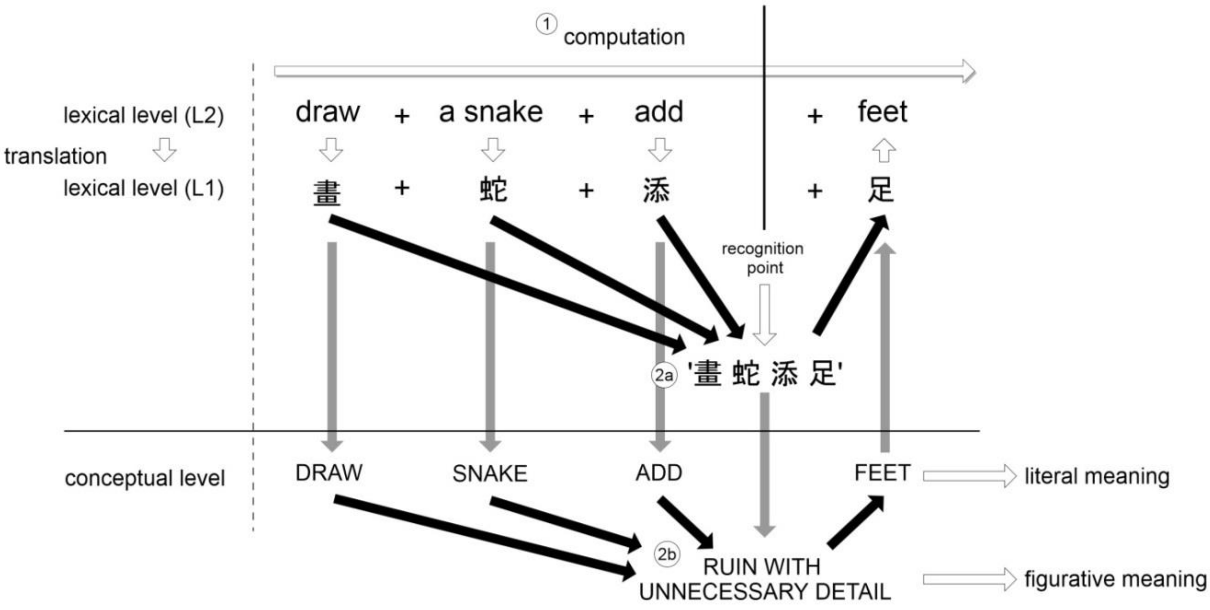

Figure 4.2. Modified dual route model for the translated idiom *draw a snake and add feet*. In this model two routes are available: analysis and computation of the phrase (1), and direct access either via a translation-based route at the lexical level (2a) or via a conceptual route (2b). In both of the direct routes a unitary entry is accessible, either as a lexical configuration (2a) or a distinct underlying concept (2b). Black arrows represent associative links between components, white arrows represent processes and grey arrows represent links between lexical items and their underlying concepts.

The modified dual route model allows bilinguals to access L1 idioms even when they are presented in the L2. An important consideration is how non-natives have been shown to process formulaic language in the L2. With idioms in particular, Cieślicka (2006) suggests that there is a fundamental difference in approach for native and non-native speakers: broadly speaking, native speakers tend to utilise a retrieval route wherever possible whereas non-natives are more likely to approach all material compositionally. In the present results no difference between English idioms and controls was observed for non-native speakers, which suggests that both sets of participants were processing the idioms compositionally. The difference between the groups is that for native speakers an additional configuration was recognised and unlocked, whereas for non-native speakers no such direct route was available. This indicates not necessarily a difference in approach, but rather a difference in available resources, i.e. non-native speakers are less likely to have formed associative links that can unlock the lexical configuration of an idiom and its underlying concept. Matlock and Heredia (2002) suggested that this leads to a situation where non-native speakers only recognise phrases as idioms once they have analysed them and found them to be incongruent.

For Chinese speakers encountering English idioms, even if they were recognised as non-compositional configurations and were potentially easy to 'spot' as idioms, no underlying lexical or conceptual configuration may be available. The Chinese idioms presented in English did show an effect of compositionality for the Chinese speakers if the compositionality ratings from English forms are used, but this did not negate the advantage they have over control phrases (assuming that the control phrases are just as compositional as their corresponding idioms). Similarly, taking the potentially more meaningful Chinese ratings of compositionality, all effects of this variable are non-significant for the Chinese idioms. This is consistent with the findings of Tabossi et al. (2009), who showed an overall advantage for familiar phrases but no variation for compositional items (clichés) compared with non-compositional items (idioms). Results from other studies in this respect have been mixed (for example, Gibbs et al., 1989; Gibbs, 1991; Caillies & Butcher, 2007), but a reasonable conclusion seems to be that the compositionality is strongly linked to meaningfulness and familiarity: the Chinese native speakers' ratings suggest that because the phrases (and their underlying stories) were known, the process of mapping idiomatic meaning onto the lexical items was facilitated. Including the Chinese speakers' ratings in the analysis, rather than the potentially 'purer' but less meaningful English ratings, demonstrated that the overall contribution of compositionality was not significant for the present task, i.e. form-based recognition.

For native speakers no effect of compositionality was observed for either set of idioms. For the English idioms, this is in line with some previous research (e.g. Tabossi et al., 2009). For translated Chinese idioms compositionality did not affect native speaker processing; this is unsurprising because no lexical or conceptual configurations would be available to aid English speakers' recognition for any of the Chinese idioms. Again, these results support Tabossi et al. (2009) rather than, for example, Caillies and Butcher (2007) in implicating overall idiom familiarity (known or unknown) as the key driver of the idiom superiority effect. Thus, English idioms, which are familiar and well known, show a processing benefit, while Chinese idioms, which are unfamiliar and unknown, are processed at the same speed as control items, but the degree of compositionality does not significantly affect either set of items.

In conclusion, non-native speakers were shown to respond more quickly to idioms translated from their L1 than to control phrases in a lexical decision task. This result mirrors native speaker performance for English idioms, suggesting that a dual route model can explain bilingual performance as well as monolingual access to formulaic language. Overall familiarity with the L1 form – i.e. recognising a 'known' phrase – was the main driver of the processing advantage for both native and non-native speakers. The ‘retrieval’ branch of the dual route model for bilinguals may represent a process at the lexical level, whereby English items were translated into their Chinese equivalents. This activated a known L1 lexical configuration, facilitating subsequent processing for translation equivalents in the L2. Alternatively, the same associations may exist at a language non-specific conceptual level, suggesting that it is the separate conceptual entry for idioms that drives their processing advantage. While the current results do not distinguish between these two explanations, some level of L1-L2 interaction is clearly indicated. This adds further support to the argument that idioms are not represented as single, unanalysable units in the lexicon, but instead represent a distributed meaning that is accessed via the component words.

Chapter 5. Data from Eye-tracking Suggests that Cross Language Formulaic Priming 'Has Legs'

The results in Chapter 4 provide suggestive evidence of crosslinguistic priming at the multi-word level. Specifically, based on the complementary pattern for English native speakers and Chinese-English bilinguals, I proposed two explanations for why idioms should show faster processing in their translated forms: a lexical-translation mechanism and a conceptually mediated mechanism. The aims of the next experiment were to replicate the findings of Chapter 4 using a more sensitive methodology (eye-tracking), and also to disambiguate the translation-based and conceptual-based arguments put forward to explain the crosslinguistic idiom effect. The two experiments described in this chapter have been published in *Bilingualism: Language and Cognition* as "Cross language lexical priming extends to formulaic units: Evidence from eye-tracking suggests that this idea 'has legs'" (Carrol & Conklin, 2015).

5.1 Introduction

Research into the bilingual lexicon has routinely looked at the relationship between single words in a first language (L1) and second language (L2) (Chen & Ng, 1989; de Groot & Nas, 1991; Wang, 2007), but there is a relative paucity of research into how translation equivalence might scale up to formulaic units. Some investigations of crosslinguistic influence have revealed an inherent reluctance to translate idioms (e.g. Kellerman, 1977, 1983, 1986), but other studies have shown effects of positive transfer, interference and avoidance in L2 idiom production (Irujo, 1986, 1993; Laufer, 2000) and comprehension (Liontas, 2001; Charteris-Black, 2002), generally finding facilitation for congruent items (those that exist in both languages). More recently, investigations into the online processing of such items have shown how congruence reduces the disruption caused during code switches in idiomatic and literal sentences (Titone, Columbus, Whitford, Mercier & Libben, 2015), and demonstrated the facilitatory effect of congruence in judging L2 collocations to be acceptable (Wolter & Gyllstad, 2011, 2013). I aim to add to this literature by exploring how translations of idioms are treated by intermediate proficiency Chinese–English bilinguals. Are ‘familiar’ sequences from the L1 treated as such even when they are encountered in an unfamiliar form? In other words, is the idiom priming effect that is evident when monolingual speakers read familiar phrases replicated when L1 idioms are encountered in the L2? The answer to this will have important implications for our understanding of how formulaic units are represented in the mental lexicon and will help to elucidate within-language relationships (how words are jointly represented) and between-language relationships (how different forms are represented across languages), both for single words and larger units. Translated idioms, therefore, provide a novel and potentially fruitful way to explore formulaic language in bilinguals.

5.2 Formulaic processing in L1 and L2

In native speakers the processing advantage for familiar phrases is well documented. Using a range of methodologies, it has been demonstrated that highly familiar idioms are processed more quickly than less familiar idioms or control phrases (Cacciari & Tabossi, 1988; Conklin & Schmitt, 2008; Libben & Titone, 2008; McGlone, Glucksberg & Cacciari, 1994; Rommers, Dijkstra & Bastiaansen, 2013; Schweigert, 1986, 1991; Schweigert & Moates, 1988; Siyanova-Chanturia, Conklin & Schmitt, 2011; Swinney & Cutler, 1979; Tabossi, Fanari & Wolf, 2009). This evidence supports hybrid models, whereby idioms exist in the mental lexicon both as individual words and whole units, variously described as ‘configurations’ (Cacciari & Tabossi, 1988), ‘superlemmas’ (Sprenger, Levelt & Kempen, 2006) or ‘formulemes’ (Van Lancker Sidtis, 2012). The view that frequently encountered combinations are lexicalised to instantiate their own unitary representations in the mental lexicon is consistent with usage based accounts of linguistic organisation (e.g. Bybee, 2006, 2008), and the processing of these lexicalised units and their component parts can be accounted for in different ways. Libben and Titone (2008; also Titone & Connine, 1999) describe a constraint-based view of idiom processing which utilises all possible information to help process any given combination of words appropriately; this helps to address the “paradox” of idioms seeming to be simultaneously unitary and compositional (Smolka, Rabanus & Rösler, 2007, p.228). Dual route explanations of the formulaic processing advantage (Van Lancker Sidtis, 2012; Wray, 2002; Wray & Perkins, 2000) propose that all linguistic material is analysed sequentially as it is encountered, but an additional (and quicker) direct route is also available for those sequences that have been encountered previously and registered as known combinations. Once an idiom or other formulaic sequence is triggered/recognised, it can therefore be accessed directly.

While this effect is robust in native speakers, second language learners rarely show the same level of formulaic advantage (Cieślicka, 2006, 2013; Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin & Schmitt, 2011; although see Isobe, 2011 and Jiang & Nekrasova, 2007 for alternative views). Second language learners may exhibit a fundamentally more compositional approach whereby sequential analysis is the default, meaning that literal meanings of words are likely to be more salient than figurative phrase level meanings (Cieślicka, Heredia & Olivares, 2014). The question is whether this is actually a difference in approach or simply in available resources: non-native speakers may not have encountered idioms in the L2 with enough regularity to allow for formation and direct retrieval of unitary entries. This is not to say that idioms cannot be understood in the L2, but the same direct processing route may not be available by default (or may be too slow to show any effect). The present investigation aims to explore this question by looking at combinations that are theoretically 'known' to non-native speakers, but which are encountered in an unfamiliar (translated) form. Given that congruence seems to facilitate L2 processing of formulaic language (Titone et al., 2015; Wolter & Gyllstad, 2011, 2013), it remains to be seen whether this is a direct effect of L1 knowledge. That is, are congruent forms facilitated because they have been encountered in both languages and are confirmed in the minds of bilinguals as transferrable, or is it the case that any lexical combinations that exist in the L1 will automatically show priming effects if the equivalent forms are encountered in an L2? For example, when a French–English bilingual speaker first encounters *bite the dust* (a word-for-word equivalent of the French *mordre la poussière*)¸ will this automatically be treated as an idiom because the forms are congruent, or would it only be accepted once the English version has been registered as the same as in the L1? In the present study I aim to investigate this for idioms that exist in the L1 but not the L2 (e.g. *call a cat a cat* – a non-idiom in English but a translation of the French *appeler un chat un chat*). Such items are therefore imbalanced in their relative L1–L2 frequency, hence any evidence of facilitation would be indicative of direct L1 influence.

The results discussed in Chapter 4 suggest that idioms may show formulaic priming effects in translation, but a recent similar study with Japanese collocations (Wolter &

Yamashita, 2014) found no advantage for acceptable L1 items presented in L2, so the extent of the effect remains unclear. As outlined in Chapter 4, two possible mechanisms could explain the underlying pattern of results for translated items: a lexical/translation route, whereby English words automatically activate Chinese equivalents, and a conceptual route, whereby English (L2) words directly triggered their underlying concepts. This conceptual priming mechanism fits the suggestion by Wray (2012) that the advantage for idioms may be a result of their distinct underlying concepts.

I present two experiments designed to further explore idiom priming in bilingual speakers, using eye-tracking as a way to tap into the automatic processes at play during reading. The aim of Experiment 2 was to investigate whether the local lexical context provided by an idiom is enough to facilitate lexical access to the final word. I compared reading times for idioms (e.g. *draw a snake and add feet*) and control items (e.g. *draw a snake and add hair*). Both variants were embedded in a short context that supported the idiomatic meaning, but neither would make sense in English without knowledge of the Chinese idiom. Shorter reading times for the final word in the idiom condition compared to the control would therefore be taken as evidence that bilingual speakers were utilising L1 knowledge to activate a known lexical combination and facilitate the expected completion.

The aim of Experiment 3 was to further explore the dimension of meaning in idiom processing. I specifically examined idioms that could also be used in a literal sense – what Van Lancker, Canter and Terbeek (1981) called “ditropic” idioms. Hybrid models suggest that literal meaning activation is obligatory (Cacciari & Tabossi, 1988; Cieślicka & Heredia, 2011; Holsinger & Kaiser, 2010; Sprenger et al., 2006;

but see Schweigert, 1991, on how relative familiarity and literal plausibility might moderate this). Siyanova-Chanturia, Conklin and Schmitt (2011) found that English native speakers showed comparable reading times for figurative and literal uses of highly familiar idioms: they read *at the end of the day* equally quickly in its idiomatic and literal senses, and both faster than a control phrase *like at the end of the war*. Non-native speakers read the literal uses significantly more quickly than the idiomatic uses, suggesting that the non-compositional nature of the figurative uses was problematic, or that the figurative meaning was simply not known. If L1 knowledge is being automatically activated when non-native speakers encounter translated forms, participants should have little difficulty interpreting idioms in figurative contexts, hence I would expect performance for Chinese native speakers on translated idioms to mirror that of English native speakers on English idioms, with no difference between figurative and literal uses for 'known' sequences. In both experiments I compare Chinese native speakers and monolingual English native speakers reading translated Chinese idioms/controls and English idioms/controls.

5.3 Experiment 2

In Experiment 2 I investigated whether 'known' sequences are facilitated in the L2: do native speakers of Chinese show facilitation for the final word of a translated idiom compared to a control word? As outlined in the previous chapter, Chinese is ideal for this kind of investigation because it has a large set of invariable idioms (chengyu) that are numerous in modern Chinese, and which have been shown to have the same formulaic properties as English idioms (Liu, Li, Shu, Zhang & Chen, 2010; Zhang, Yang, Gu & Ji, 2013; Zhou, Zhou & Chen, 2004).

5.3.1 Participants

Participants in Experiments 2 and 3 were taken from the same population, but were different in each study. All participants received course credit or £5 for participation. Chinese native speakers were students at the University of Nottingham (34 postgraduates, seven undergraduates; mean age = 24.8), hence had met minimum entry requirements to study at an English university (minimum IELTS score of 6.5), and had been in the UK for an average of 1.4 years. All had Mandarin Chinese as their L1.[13] Information regarding their English language background is shown in Table 5.1. English native speakers were undergraduate students at the University of Nottingham (mean age = 19.3), none of whom had any experience of learning Mandarin. Twenty English native speakers and 20 Chinese native speakers took part in Experiment 2. All norming described below used participants who did not take part in the main experiments and used a seven-point rating scale.

[13] Whilst the participants were all native speakers of Mandarin, it is possible that their time spent living in England could have led to slight language attrition that may have affected performance in these studies. The majority of students had spent a little over one year in England and when asked about their daily usage of Mandarin suggested that this was frequent since their social lives were largely constructed around other native speakers of Mandarin, although no data was collected to confirm this. I therefore assume that any effects of attrition would be minimal, but for future studies looking at L1 influence it might be useful to consider this in a more rigorous fashion.

Table 5.1. Summary of Chinese native speakers' language background for both experiments (all measures relate to proficiency in English)

	Age	Years studying English	Reading	Listening	Speaking	Writing	Usage	Vocab
Experiment 2								
Mean	26.2	14.1	3.4	3.2	2.8	2.8	35.3	10.7
Range	21-38	4-25	2-5	2-4	2-4	2-4	25-45	4-16
Experiment 3								
Mean	23.4	12.7	3.5	3.0	3.1	3.7	35.5	10.9
Range	21-30	7-16	2-5	1-4	1-4	2-5	27-43	7-16

N.B. Reading, Listening, Speaking and Writing are self-ratings of these skills out of 5 (1 = Poor, 2 = Basic, 3 = Good, 4 = Very good, 5 = Excellent); Usage is an aggregated estimate of how frequently participants use English in their everyday lives in a variety of contexts (total score out of 50); Vocab is a modified Vocabulary Size Test with a total score out of 20.

5.3.2 Materials

Chinese idioms were selected from the Dictionary of 1000 Chinese Idioms (Lin & Leonard, 2012). Only idioms where a literal translation provided a plausible English sequence with identical word order were considered, e.g. 畫蛇添足 – *draw-snake-add-feet* = *draw a snake and add feet*. For all items the final character had a single word translation equivalent in English. These idioms were judged to be highly familiar in the original Chinese form (mean = 6.5/7) by 27 native speakers of Mandarin. Translations were taken from the gloss provided by the Dictionary of 1000 Chinese Idioms then checked character by character using two different translation engines (Google Translate and On-line Chinese Tools) to ensure accurate transliterations into English. Control items were formed by replacing the final word of each idiom with an alternative, matched for part of speech, length and frequency (e.g. *draw a snake and add feet* vs. *draw a snake and add hair*). All Chinese idioms and control items showed a phrase frequency of 0 in the British National Corpus (BNC). Note that the intention was not necessarily to create a literally plausible control sentence in each case, but simply to replace the final word in such a way that I could compare speed of access based on the preceding sequence. Hence in the example of *draw a snake and add feet/hair*, neither is inherently more plausible in English unless the idiom is known, but if Chinese native speakers are activating the underlying L1 idiom then this should lead to facilitation for the expected word.

English idioms were selected from the Oxford Learner's Dictionary of English Idioms (Warren, 1994). Twenty-six idioms were judged to be highly familiar (mean = 6.6/7) by 19 English native speakers. Control items were formed by replacing the final word with an alternative matched for part of speech, length and frequency (e.g. *spill the beans* vs. *spill the chips*). As with the Chinese items, the intention was not to create literally plausible control items but rather to specifically test whether the 'correct' word was facilitated once an idiom had been encountered. All control items showed a phrase frequency of 0 in the BNC. The English and Chinese items used in both experiments are available in Appendix 2a.

All stimulus items were embedded in short sentence contexts supporting the figurative meaning, for example: "*My wife is terrible at keeping secrets. She loves any opportunity she gets to meet up with her friends and* <u>*spill the beans/chips*</u> *about anything they can think to gossip about*." All sentence contexts were of comparable length. Contexts for idioms and their corresponding controls were identical and all passages were presented over three lines with the idiom or control phrase appearing toward the middle of the second line. Forty filler items of comparable length were constructed, none of which contained idioms.

Compositionality ratings were gathered for all items, as this is often identified as an important factor in idiom processing (Caillies & Butcher, 2007; Gibbs, 1991; Gibbs, Nayak & Cutting, 1989). Sixteen English native speakers were presented with all English and Chinese idioms and asked how easily the meaning of the idiom could be matched to a literal equivalent (e.g. to *spill the beans* means “to reveal a secret”): English idioms: mean = 4.1/7; Chinese idioms: mean = 3.8/7. The Chinese idioms were also presented in the original Chinese characters to 12 Chinese native speakers who gave their own set of ratings (mean = 5.6/7).

Two counterbalanced stimulus lists were constructed so that each participant saw 13 English idioms, 13 English controls, 13 Chinese idioms, 13 Chinese controls and 40 filler items. Lists were matched for all lexical variables, for English idiom frequency and for the familiarity and compositionality of the idioms.

5.3.3 Procedure

The experiment was conducted using an Eyelink I (version 2.11) eye-tracker. Participants were seated in front of a monitor and fitted with a head-mounted camera to track pupil movements. Camera accuracy was verified using a nine-point calibration grid and recalibrations were performed throughout the experiment as required. Participants were asked to read the passages on screen for comprehension then press a button to advance once they had finished. Half of the items were followed by a yes/no comprehension question to encourage participants to pay attention and the rest were followed by a 'Ready?' prompt. After each trial a fixation dot appeared on the screen to allow for trial-by-trial drift correction. Each participant saw eight practice items, then the experiment began.

Afterwards, participants were asked to provide subjective familiarity ratings for all stimulus items. For English native speakers all items were presented in English (English items, mean = 6.4/7; Chinese items, mean = 2.1/7). For Chinese native speakers the English items were presented in the same way (mean = 3.5/7) but Chinese idioms were presented in the original Chinese characters (mean = 6.5/7).[14] Chinese native speakers were also asked to complete a short vocabulary test (modified from Nation and Beglar, 2007). This test was adapted to include a representative sample from the 10,000 most frequent word families in English, and was augmented with any low frequency vocabulary items that appeared in the stimulus items, as in Experiment 1 (Chapter 4). Any constituent words from the English or Chinese idioms that were outside the 3000 most frequent word families in English were added to the test, and incorrectly identified words were removed from the analysis on a per-participant basis. Finally, Chinese native speakers were asked to complete a language background questionnaire (see Table 5.1 for details).

5.4 Results and analysis

5.4.1 Word level analysis

One Chinese native speaker was removed from the analysis because of eye-tracker calibration problems. All data were cleaned according to the four stage procedure

[14] It is worth noting that such high levels of familiarity with these Chinese character sequences perhaps argue against any significant level of attrition for the Chinese native speakers.

within Eyelink Data Viewer software, meaning that fixations shorter than 100ms and fixations longer than 800ms were removed. Data were visually inspected and any trials where track loss occurred were removed, along with any trials containing words that were incorrectly identified on the vocabulary test (for non-native speakers only). Overall this accounted for 10.4% of raw data being removed from the analysis for Chinese native speakers.[15] No native speakers were removed from the analysis and 4.8% of the raw data was removed because of track loss. Participants generally had no difficulty answering the comprehension questions (English native speakers, mean = 93%; Chinese native speakers, mean = 89%), suggesting that the task of reading and understanding the passages was well within the capability of all participants.

I concentrated the analysis on the final word of each phrase with the rationale that if idioms are known and stored as whole units then reading the first few words should activate the underlying phrase. This in turn should facilitate the final word relative to any other completion, and this would be reflected in shorter reading times. For items that are unknown I would expect to see no difference in reading times for an idiom vs. a control since no expectation regarding the final word would be generated. Although there was some variability in how literally plausible the phrases were, if an item was unknown to any participant then there should be no expectation generated for either the correct or incorrect ending.

I utilised a range of early and late eye-tracking measures to examine the predictability of the final word. Broadly, early measures reflect automatic lexical access processes while late measures reflect post-lexical processes/integration of overall meaning into

[15] Despite this relatively high figure, the composition of the stimulus lists in terms of key balancing factors like word length, frequency and idiom familiarity was not differentially affected for idioms vs. control items.

wider context (c.f. Altarriba, Kroll, Sholl & Rayner, 1996; Inhoff, 1984; Paterson, Liversedge & Underwood, 1999; Staub & Rayner, 2007). The early measures are probability of skipping (how likely is it that a word is not fixated during first pass reading), first fixation duration (duration of the first fixation on the final word of the phrase) and first pass reading time (sum of all fixations before gaze exited either to the left or right). The late measures are total reading time (sum of all fixations on the target word throughout any given trial, including re-reading time) and total number of fixations (total number of times a target word was fixated during any given trial). Table 5.2 shows a summary of the word level reading patterns.

Table 5.2. Summary of reading patterns of final words of phrases for all measures for Chinese native speakers and English native speakers.

	Chinese phrases		English phrases	
	Idiom	Control	Idiom	Control
Chinese native speakers				
Likelihood of skipping	.03 (.16)	.00 (.07)	.04 (.20)	.03 (.18)
First fixation duration	272 123)	301 (118)	269 (116)	262 (119)
First pass reading time	344 (189)	380 (186)	307 (142)	315 (158)
Total reading time	484 (358)	538 (336)	440 319)	453 (310)
Total fixation count	1.8 (1.2)	1.9 (1.3)	1.7 (1.3)	1.7 (1.0)
English native speakers				
Likelihood of skipping	.07 (.23)	.09 (.28)	.31 (.46)	.09 (.28)
First fixation duration	199 (88)	201 (99)	134 (100)	183 (88)
First pass reading time	226 (121)	229 (136)	140 (109)	188 (93)
Total reading time	279 (176)	282 (194)	148 (122)	242 (197)
Total fixation count	1.3 (0.7)	1.3 (0.8)	0.8 (0.6)	1.2 (0.8)

Note: Data are mean values (SD in brackets) for likelihood of skipping expressed as a probability, raw values in ms for duration measures and raw values for fixation counts. Mean duration measures include a value of 0 for skipped items.

The data were analysed in an omnibus linear mixed effects model using the lme4 package (version 1.0-7, Bates, Maechler, Bolker, Walker, Christensen, Singmann & Dai, 2014) in R (version 3.1.2, R Core Team, 2014). Linear mixed effects models are able to incorporate random variation by subject and by item alongside fixed effects, thereby avoiding the “language as fixed effect fallacy” (Clark, 1973). I included the three treatment-coded main effects of group (Chinese native speakers vs. English native speakers), language (Chinese phrases vs. English phrases) and phrase type (idiom vs. control). Random intercepts for subject and item and by-subject random slopes for the effects of language and type were included (Barr, Levy, Scheepers & Tily, 2013). I included the covariates of idiom length in words, final word length in letters and log-transformed final word frequency in a stepwise fashion and compared the resulting models using likelihood ratio tests to see whether inclusion improved the fit; only covariates that significantly improved the model were retained. Separate models were fitted for each eye-tracking measurement. For the binary measure likelihood of skipping a logistic linear model was used (Jaeger, 2008). For subsequent analysis of durational measures any skipped items were removed from the dataset and all duration measures were log-transformed to reduce skewing. Fixation counts were analysed using a generalised linear model with poisson regression. The structure and output for all models is shown in Table 5.3.

Table 5.3. Omnibus linear mixed effects model output for final word, all eye-tracking measurements.

	Likelihood of skipping			First fixation duration			First pass reading time			Total reading time			Total fixation Count		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	z
Intercept	-0.85	0.75	-1.14	5.55	0.03	164.09***	5.85	0.01	46.88***	5.83	0.11	53.66***	0.24	0.13	1.88
Group: English	0.85	0.50	1.71	-0.24	0.05	-5.43***	-0.34	0.06	-6.06***	-0.44	0.08	-5.81***	-0.28	0.09	-2.88**
Language: English	-0.07	0.58	-0.12	0.02	0.04	0.57	-0.02	0.04	-0.61	-0.04	0.06	-0.70	-0.02	0.08	-0.25
Type: Control	-2.03	1.12	-1.81	0.08	0.03	2.46*	0.09	0.04	2.25*	0.12	0.05	2.31*	0.08	0.07	1.09
Group*Language	1.64	0.65	2.53*	-0.09	0.05	-1.83	-0.08	0.05	-1.45	-0.21	0.07	-3.17**	-0.49	0.12	-4.17***
Group*Type	2.24	1.17	1.91^{+}	-0.05	0.05	-1.10	-0.07	0.06	-1.16	-0.10	0.07	-1.39	-0.10	0.11	-0.93
Language*Type	1.18	1.23	0.96	-0.13	0.05	-2.90**	-0.09	0.05	-1.80	-0.10	0.06	-1.48	-0.08	0.10	-0.78
Group*Language*Type	-3.42	1.30	-2.63**	0.12	0.06	1.84	0.08	0.07	1.09	0.21	0.09	2.31*	0.53	0.16	3.33***
Control predictors:															
Word length (letters)	-0.58	0.13	-4.40***	n/a	n/a	n/a	0.02	0.01	1.77	0.03	0.02	1.88^{+}	0.06	0.02	2.90**
LogFrequency	n/a	n/a	n/a	n/a	n/a	n/a	-0.03	0.00	-2.86**	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.000			0.011			0.019			0.037			0.033		
Item	0.176			0.001			0.002			0.011			0.006		
Subject | Language	0.254			0.003			0.001			0.000			0.001		
Subject | Type	0.461			0.002			0.010			0.012			0.001		
Residual	n/a			0.106			0.141			0.217			n/a		

*Note: Significance values are estimated by the R package lmerTest (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): *** p<.001, ** p<.01, * p<.05, $^{+}$ p<.10. For likelihood of skipping a logistic linear mixed effects model was used and for fixation count a generalised linear model with poisson regression was used.*

In an initial model for skipping rates there was a significant three-way interaction of group, language and type ($z = -2.63$, $p < .01$). English native speakers showed a strong tendency to skip the final word of English idioms (31%) compared to control items (9%) but no effect for Chinese items. Chinese native speakers showed a small but non-significant tendency to skip the final words of translated idioms vs. controls and no difference for English items. The analysis of duration measures also supports a general pattern whereby L1 idioms are read more quickly than control words: native speakers of Chinese read the final word more quickly for translated idioms vs. controls but show no difference for English phrases, while English native speakers show an advantage for English idioms but not translations of Chinese phrases. This is seen in the three way interaction of group, language and type: for first fixation duration this is marginal ($t = 1.84$, $p = .07$) and is significant for total reading time ($t = 2.31$, $p < .05$) and fixation count ($t = 3.33$, $p < .001$). For first pass reading time this interaction is not significant, but this analysis has excluded all data for which the final word was skipped, which affected significantly more idioms than control phrases.[16]

Interactions were analysed further using the Phia package (version 0.1-5, De Rosario-Martinez, 2013) in R with separate models for the two speaker groups (available in Appendix 2b). Pairwise comparisons confirmed that Chinese native speakers showed an advantage for Chinese idioms vs. controls for first fixation duration ($\chi2$ (1, 841) = 5.39, $p < .05$), total reading time $\chi2$ (1, 841) = 4.81, $p = .05$) and marginally for first pass reading time ($\chi2$ (1, 841) = 4.12, $p = .08$), but not for likelihood of skipping or

[16] With this in mind, I also conducted a separate durational analysis where I retained all items but assigned all skipped words a single fixation duration of 100ms (the lower cut off in the dataset). This analysis revealed highly significant three way interactions for all measures (duration measures, all *t*s > 3, all *p*s < .01; fixation count, $t > 2$, $p < .05$). See discussion in Chapter 3 on this point, and also Appendix 2b, table 1 for the full output of this model.

fixation count. For English phrases no differences were significant. English native speakers showed significantly higher likelihood of skipping for English idioms vs. controls ($\chi2$ (1, 990) = 29.30, $p < .001$), significantly shorter total reading times ($\chi2$ (1, 990) = 5.78, $p < .05$) and significantly fewer fixations overall ($\chi2$ (1, 990) = 19.70, $p < .001$), but early duration measures were non-significant. Chinese phrases showed no difference on any measure.

5.4.2 Phrase level analysis

I also examined phrase level data to see whether the overall context could have contributed to the patterns described above. I considered first pass reading time, total reading time, and regression path duration for the phrase (once the phrase had been fixated, how much time was spent re-reading the context that preceded it) and specifically for the final word. These measures are summarised in Table 5.4.

Table 5.4. Phrase level reading patterns (all values in ms) for Chinese and English native speakers, all items.

	Chinese phrases		English phrases	
	Idiom	Control	Idiom	Control
Chinese native speakers				
First pass reading time	1397 (747)	1411 (777)	904 (527)	915 (573)
Total reading time	1959 (1055)	2030 (1179)	1348 (823)	1382 (787)
Regression duration (word)	748 (806)	850 (913)	674 (660)	703 (662)
Regression duration (phrase)	1213 (638)	1180 (660)	680 (454)	683 (404)
English native speakers				
First pass reading time	814 (460)	736 (422)	423 (211)	482 (242)
Total reading time	1244 (684)	1238 (652)	528 (269)	681 (437)
Regression duration (word)	513 (501)	495 (490)	199 (223)	354 (394)
Regression duration (phrase)	745 (395)	746 (398)	334 (170)	352 (198)

Table 5.5. Omnibus mixed effects model output for phrase level reading patterns

	First pass reading time			Total reading time			Regression duration (word)			Regression duration (phrase)		
Fixed effects:	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	6.62	0.17	39.37***	6.96	0.16	44.61***	6.27	0.10	65.36***	6.14	0.18	34.10***
Group: English	-0.55	0.08	-6.71***	-0.46	0.10	-4.46***	-0.28	0.13	-2.20*	-0.51	0.10	-4.96***
Language: English	-0.30	0.09	-3.53***	-0.28	0.07	-3.97***	-0.07	0.09	-0.82	-0.40	0.08	-4.95***
Type: Control	0.02	0.06	0.30	0.03	0.04	0.65	0.11	0.07	1.62	-0.05	0.04	-1.14
Group*Language	-0.17	0.08	-2.11*	-0.44	0.06	-7.85***	-0.45	0.11	-4.28***	-0.16	0.06	-2.64**
Group*Type	-0.13	0.08	-1.62	-0.02	0.06	-0.03	-0.13	0.09	-1.41	0.07	0.06	1.18
Language*Type	-0.07	0.08	-0.92	-0.01	0.05	-0.09	-0.07	0.09	-0.80	0.04	0.06	0.67
Group*Language*Type	0.31	0.11	2.91**	0.23	0.07	3.15**	0.29	0.13	2.28*	-0.02	0.08	-0.28
Control predictors:												
Word length (letters)	0.08	0.03	2.78**	0.09	0.03	3.66***	n/a	n/a	n/a	0.15	0.03	5.13***
Random effects:	Variance			Variance			Variance			Variance		
Subject	0.038			0.090			0.117			0.087		
Item	0.034			0.029			0.026			0.044		
Subject | Language	0.006			0.005			0.027			0.006		
Subject | Type	0.003			0.007			0.010			0.002		
Residual	0.329			0.154			0.426			0.167		

*Note: Significance values are estimated by the R package lmerTest (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): *** p<.001, ** p<.01, * p<.05*

The omnibus analysis (Table 5.5) shows significant interactions of group and language for all measures and a significant three way interaction for group, language and type (for all measures except phrase level regression durations). English native speakers had a tendency to read English idioms faster and to regress less. For control items, encountering an unexpected final word caused a regression to the immediate preceding context, but there was no difference in the amount of time spent re-reading the context prior to the phrase for idioms vs. controls. There was no difference between Chinese idioms and controls on any measure.

Chinese native speakers showed no difference on any of the phrase level measures for idioms compared to controls for either set of phrases (pairwise analysis by type, all *p*s > .05). Encountering the ‘incorrect’ completion of an idiom from either language did not lead to more time re-reading the phrase. Similarly, whole phrase reading times and overall regressions to the preceding context were comparable for both sets of idioms and controls. One way to interpret this is that the recognition of form evidenced in the analysis of the final words and integration of overall meaning may be exerting opposing forces. That is, Chinese native speakers may be reading the idiom and correctly predicting the final word, but they still need to spend time reading and re-reading the whole phrase and the prior context to attempt to resolve the meaning in both idiom and control conditions. This hints at a dissociation between recognition/prediction of the correct form and access to the overall phrase level meaning, which will be explored in more detail in Experiment 3.

5.4.3 Familiarity, compositionality and plausibility

I next analysed the data to assess the effect of subjective familiarity, relative compositionality and plausibility on each set of idioms. One possibility is that the difference in plausibility between idioms and controls might be exerting an effect: hence the advantage observed for idioms may in fact be a reflection of the disruption caused by implausible completions in the control items. To investigate this I collected plausibility ratings from 19 English native speakers to compare idioms and controls for both English and Chinese phrases. English phrases were considered more plausible than the controls (idioms: mean = 6.4; controls: mean = 4.0; $t(24) = 5.49$ $p < .001$), while Chinese phrases and controls were seen as equally plausible (idioms: mean = 3.5; controls: mean = 3.4; $t(24) = 1.49$, $p = .15$). This suggests that plausibility was not driving the effects for ‘unknown’ items. If plausibility was affecting Chinese native speakers reading English phrases, I would expect to see a significant slowdown for controls, rather than simply a null effect. Similarly, the Chinese items are equally plausible in their idiom or control forms to naïve readers (English native speakers), hence the only way a difference can emerge is if some underlying knowledge of the idioms is being utilised, as in the case of the Chinese native speakers. I further explore the effect of plausibility in the models below.

I fitted separate models to compare the effects of familiarity, compositionality and plausibility. All continuous predictor variables were centred. I considered Chinese native speaker and English native speaker participants separately. In each model language and type were fixed effects and the interaction with each variable of interest was considered individually. Random intercepts for subject and item and by-subject random slopes were included for each fixed effect. Models were fitted for all word and phrase level measures but only significant effects are described in detail here. (Full model outputs are provided in Appendix 2b, tables 4–10).

Familiarity

Subjective familiarity did not show significant effects for Chinese native speakers for Chinese idioms or English idioms. For English native speakers there was a marginal effect of familiarity on likelihood of skipping (β = 0.29, SE = 0.16, z = 1.87, p = .06). Closer inspection reveals that this reflects an interaction of familiarity and type for English idioms only (separate model for English phrases only, z = -1.86, p = .06). This pattern is repeated (although does not reach significance) for the later measures total dwell time and regression path duration. Hence for idioms, familiarity is facilitatory (more likely to skip, less likely to spend time re-reading the phrase or word). Conversely, controls of better known items are more likely to be read and re-read, presumably because the high familiarity generates a stronger expectation, the breaking of which is more problematic than for an idiom where the expected word is less strongly predicted. No significant effects were seen for Chinese items.

Compositionality

Compositionality showed no effects for Chinese native speakers for either set of phrases. This was also true of the compositionality ratings gathered from Chinese native speakers. English native speakers showed no effects of compositionality on any measure for English or Chinese items.

Plausibility

Plausibility showed no effect for Chinese native speakers reading English phrases, but was significant for Chinese phrases on early measures. For first fixation duration there was a significant interaction with phrase type (β = 0.08, SE = 0.04, t = 1.95, p = .05). This shows that more plausible phrases were read more quickly when the final word was correct, while for control phrases greater plausibility had an inhibitory effect. This trend was also seen in first pass reading time and total dwell time, although neither reaches significance. This means that for Chinese native speakers, who knew the ‘correct’ completion, there was a clear difference in the effect of plausibility between the two variants. Crucially, when reading Chinese phrases, English native speakers showed the same pattern for both idioms and controls: as they had no underlying knowledge of the ‘correct’ idiom, plausibility played an equal role for idioms and controls. In other words, *draw a snake and add…* could just as logically be completed with *hair* as it could *feet*, hence the effect was the same for either version. This shows that English native speakers did not consider the idioms or controls to be inherently more plausible (supporting the rating data). For English native speakers reading English phrases there was a significant interaction of plausibility and phrase type for skipping rate ($\beta = -0.56$, SE = 0.25, $z = -2.20$, $p < .05$). Hence greater plausibility increased the likelihood of skipping in idioms, whereas for other measures it had a generally facilitatory but non-significant effect on both idioms and control items.

Proficiency

A final set of models were fitted to assess the contribution of English proficiency level for Chinese native speakers, considered in terms of three variables: vocabulary test score, self-rated ability and estimated usage. Each proficiency measure was assessed in turn for its overall effect, then for its interaction with language and phrase type. No measure of proficiency had an effect for the final word or whole phrase, or on regression durations. This suggests that the Chinese native speakers were generally well-matched in their English proficiency, and this may explain why no effects were seen here: comparable studies that have found an effect of proficiency (e.g. Ueno, 2009) have done so with a deliberate high/low proficiency group manipulation.

5.5 Discussion

The results show complementary patterns for English native speaker and Chinese native speaker participants. Consistent with findings throughout the idiom literature (and in support of the results of Study 1), English native speakers showed significant facilitation for the final words of a known phrase compared to a control phrase. The fact that the effect was most clearly evidenced in the likelihood of skipping (31% for idioms) suggests that this was highly automatic behaviour. As a result of this relatively high skipping rate, the early reading measures did not show much difference, but total reading time also showed a significant advantage. Chinese native speakers showed no effect for English idioms, which is again consistent with the previous literature on non-native speakers processing formulaic sequences in the L2. The Chinese items were not processed differentially by English native speakers on any measure, and crucially there was no difference in the effect of plausibility for the idioms vs. controls – this demonstrates that there is fundamentally no reason to expect the correct completion (e.g. *feet*) over the control completion (e.g. *hair*) unless the idiom is known. There was a consistent difference across duration measures for the Chinese native speakers, suggesting that there was some degree of crosslinguistic influence that provided a boost to lexical access for the items that were known in the L1. The effect was most clearly seen in the early measure first fixation duration, suggesting a degree of bottom-up facilitation through something akin to an interactive-activation framework (as suggested by Cutter, Drieghe and Liversedge (2014) for their results on spaced compounds); it was also seen in total reading times, but not in phrase level reading times or regression path measures. This suggests that the lexical activation provided by the idiom is enough to facilitate the correct word, but not enough to overcome any inherent ambiguity in the non-compositional phrases. I will explore this dissociation further in Experiment 3.

One possible issue is that the idioms in the study were relatively long, and in particular the Chinese items were on average longer than the English items (Chinese items = 5.3 words; English items = 4.0 words, $t(50) = -4.55$, $p < .001$). However, in none of the analyses was the length of the prime a significant factor, i.e. a facilitative effect for the final word was seen whether the prime was relatively short (three words, e.g. *wine and meat* (*friends*)) or relatively long (six words, e.g. *beat the grass to scare the* (*snake*)). This suggests that the advantage seen for the Chinese native speakers was not necessarily strategic, although it is not possible to rule this out completely. Whether the result of strategic, active prediction or automatic lexical priming, I interpret the fact that there was an effect for Chinese native speakers as evidence of L1 influence, even though the phrases were entirely novel in terms of form.

5.6 Experiment 3

In Experiment 3 I wanted to examine how participants read figurative and literal uses of the same idioms. In Siyanova-Chanturia, Conklin and Schmitt (2011) native speakers showed no difference in reading times for literal or figurative uses of ditropic idioms, whereas for non-native speakers figurative uses were read more slowly than literal uses. This difficulty understanding non-compositional phrases in the L2 may indicate that either the figurative meanings of idioms are unknown to non-native speakers, hence there is no direct entry to access, or that if the idioms are known, they are not accessed directly in the same way as for native speakers, and consideration of the figurative meaning only occurs after the literal meaning has been rejected. For translated items, if the idiom advantage observed in Experiment 2 is the result of activation of the underlying L1 idiom entry, I would expect figurative and literal uses of the translated Chinese idioms to be read comparably by Chinese native speakers, since activating the idiom will presumably also make the semantic meaning of the phrase available. More specifically, they will be processed in the same way as English native speakers read English idioms. English native speakers should show a complementary pattern: difficulty reading the figurative uses of translated Chinese idioms compared to the entirely compositional literal uses.

5.6.1 Participants

Twenty-one English native speakers and 21 Chinese native speakers took part in Experiment 3, all from a similar population as Experiment 2.

5.6.2 Materials

The English idioms used in Experiment 2 were augmented with stimuli from Siyanova-Chanturia, Conklin and Schmitt (2011) to give an initial set of 35 items. Chinese idioms were selected with the same selection criteria as for Experiment 2 (literal translation gave a grammatical English phrase with congruent word order, final word was a single word translation equivalent), with the additional stipulation that all idioms had to be literally plausible. To confirm this all English and Chinese idioms were included in a norming study where 24 English native speakers judged on a seven-point Likert scale how acceptable each was in a literal context. The 20 English and 20 Chinese idioms that were judged most literally plausible were retained (all received mean scores of greater than 3.5/7).

The idioms were placed into short contexts to bias either the figurative or literal meaning. These were included in a further norming study to assess how acceptable each was as an English sentence: 36 English native speakers judged their acceptability on a seven-point Likert scale. English items were rated as very acceptable in both figurative and literal contexts (figurative, mean = 6.3/7; literal, mean = 5.7/7). Chinese idioms were rated as being very acceptable in the literal contexts (mean = 5.6/7) and less acceptable in their figurative contexts (mean = 3.8/7), which is not surprising given that the idioms are all unknown to English native speakers. Familiarity of all items was verified in a separate norming test with 10 English native speakers (to ensure that items were likely to be known, rather than as a strict norming test – by-participant familiarity ratings were gathered after the main experiment). All idioms were then included in further norming studies with English native speakers to assess compositionality (n = 20; Chinese idioms were also assessed by Chinese native speakers, n = 12, in the original Chinese, as in Experiment 2). Table 5.6 shows example stimuli used in figurative and literal contexts.

Table 5.6. Examples of ditropic English and translated Chinese idioms used in figurative and literal contexts.

English idiom	A piece of cake – "easy"
Figurative context	One of my hobbies is doing little jobs around the house. I find most things I try are <u>a piece of cake</u> if you make sure you have the right tools before you start.
Literal context	Yesterday I was in the canteen at work and I was very hungry. I really wanted to get <u>a piece of cake</u> for my lunch but I was good and just had a sandwich.
Chinese idiom	**Add oil and vinegar – "to embellish a story"**
Figurative context	I have a friend who always exaggerates whenever he tells stories. The problem is he tends to <u>add oil and vinegar</u> so it's hard to know whether or not to believe what he says.
Literal context	I read a really simple recipe for a salad dressing. You just chop up some garlic and then <u>add oil and vinegar</u> then you put it in the fridge until you need to use it.

Idioms were divided into two counterbalanced lists so that each participant saw 10 English idioms of each type (figurative/literal), 10 translated Chinese idioms of each type and 40 filler items. Within each list the idioms/controls were matched for number of words in the phrase, length and frequency of the final word, and literal plausibility of the idioms. The lexical coverage of all contexts was assessed using the Vocab Profile tool on the LexTutor website. All contexts had lexical coverage of greater than 96% at the K2 level (meaning that 96% of words were within the 2000 most frequent English word families) and greater than 99% coverage at the K5 level. In each item the idiom appeared toward the middle of the second line of a three-line block of text.

5.6.3 Procedure

All procedures were the same as in Experiment 2, however this time I took the whole phrase as the unit of analysis.[17] Because each analysis area was several words long, first fixation duration was discounted and first pass reading time was retained as the only early measure, with total reading time and total fixation count used as late measures. I also included regression path duration as an additional late measure to examine how participants used the preceding context to help understand each idiom.

Following the main experiment, participants were asked to provide subjective familiarity ratings for each idiom. English native speakers found English items highly familiar (mean = 6.4/7) and Chinese items unfamiliar (mean = 2.3/7). Chinese native speakers found Chinese items highly familiar (mean = 6.6/7) and English items less familiar (mean = 4.4/7). Chinese participants again completed a language background questionnaire and vocabulary test.

5.7 Results and analysis

5.7.1 Phrase level analysis

No participants were removed from the analysis and the same data cleaning procedure as in Experiment 2 was applied. All trials where track loss occurred were removed. For native speakers this accounted for 1.9% of the data. For non-native speakers, in addition to the removal of trials where track loss occurred, any items containing unknown vocabulary items were removed, accounting for 5.3% of the non-native

[17] Separate analysis to compare final word reading for figurative vs. literal contexts showed no significant differences on any measures for either set of stimuli for English native speakers or Chinese native speakers.

speaker data overall. English native speakers scored 92% on comprehension questions and non-native speakers scored 87%, suggesting that the task was again adequately completed by both groups. As with Experiment 2, duration measures were log-transformed to reduce skewing and for fixation count data a poisson regression was applied to the raw values. Table 5.7 shows a summary of results for all measures.

Table 5.7. Summary of reading patterns for whole phrases for all measures, Chinese native speakers and English native speakers.

	Chinese phrases		English phrases	
	Figurative	Literal	Figurative	Literal
Chinese native speakers				
First pass reading time	1350 (690)	1213 (701)	878 (517)	773 (434)
Total reading time	1985 (1082)	1807 (1022)	1242 (794)	1115 (757)
Total fixation count	7.8 (4.0)	7.3 (3.9)	4.9 (3.1)	4.5 (2.7)
Regression duration	1157 (568)	1033 (558)	644 (408)	576 (389)
English native speakers				
First pass reading time	739 (411)	681 (437)	394 (183)	400 (213)
Total reading time	1139 (601)	978 (490)	494 (244)	523 (343)
Total fixation count	5.4 (2.6)	4.7 (2.1)	2.6 (1.1)	2.6 (1.4)
Regression duration	644 (456)	585 (347)	308 (171)	316 (194)

Note: Data are mean values in ms for duration measures and raw values for fixation counts (SD in brackets).

An omnibus model was fitted in which fixed effects of group (Chinese native speakers vs. English native speakers), language (Chinese phrases vs. English phrases) and phrase type (figurative vs. literal) and their interactions were computed. By-subject and by-item random intercepts and by-subject slopes for language and phrase type were included in all models. The covariate idiom length (in words) was included in all models where log likelihood tests showed that this significantly improved the fit. Table 5.8 shows the omnibus results for all measures.

Table 5.8. Omnibus mixed effects model output for all eye-tracking measures

	First pass reading time			Total reading time			Total fixation count			Regression duration		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	t
Intercept	6.77	0.13	51.50***	7.28	0.14	53.70***	2.03	0.07	28.47***	6.63	0.15	44.40***
Group: English	-0.64	0.08	-8.21***	-0.57	0.09	-6.22***	-0.38	0.08	-4.73***	-0.69	0.10	-7.23***
Language: English	-0.38	0.07	-5.12***	-0.46	0.08	-6.07***	-0.46	0.08	-5.94***	-0.60	0.09	-6.82***
Type: Literal	-0.16	0.06	-2.78**	-0.12	0.04	-2.70**	-0.09	0.04	-2.33*	-0.15	0.05	-3.23***
Group*Language	-0.14	0.08	-1.70	-0.34	0.06	-5.49***	-0.27	0.07	-3.97***	-0.00	0.01	-0.08
Group*Type	0.06	0.08	0.71	-0.03	0.06	-0.55	-0.05	0.06	-0.91	-0.09	0.03	1.41
Language*Type	0.03	0.08	0.32	-0.02	0.06	-0.26	-0.02	-0.06	-0.31	-0.04	0.00	0.62
Group*Language*Type	0.09	0.11	0.81	0.18	0.08	2.24*	0.18	0.10	1.88	-0.03	0.00	0.33
Control predictors:												
Idiom length (words)	0.06	0.02	2.44*	0.04	0.02	1.86^{+}	n/a	n/a	n/a	0.07	0.02	2.70**
Random effects:	Variance			Variance			Variance			Variance		
Subject	0.029			0.069			0.050			0.076		
Item	0.019			0.037			0.045			0.053		
Subject | Language	0.001			0.004			0.002			0.011		
Subject | Type	0.002			0.001			0.000			0.003		
Residual	0.316			0.165			n/a			0.185		

*Note: Significance values are estimated by the R package lmerTest (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): *** p<.001, ** p<.01, * p<.05, $^{+}$ p < .10*

All measures showed significant main effects of group (English native speakers showed shorter reading times and fewer fixations, all *t*s > 4, all *p*s < .001), language (for all speakers English idioms were read more quickly than translated Chinese items, all *t*s > 2, all *p*s < .05), and importantly phrase type (literal phrases were read faster than figurative phrases, all *t*s > 2, all *p*s < .05). To further explore the data, separate models were fitted for Chinese native speakers and English native speakers (provided in Appendix 2b, tables 11–12).

Chinese native speakers show a significant main effect of type for all items (all *t*s > 2, all *p*s < .05) and no interactions between language and phrase type, suggesting that literal (compositional) uses were easier to understand than figurative uses for all phrases. In line with Siyanova-Chanturia, Conklin and Schmitt (2011) this was true for English (L2) idioms, but was also the case for translations of Chinese idioms. Therefore, despite the suggestion in Experiment 2 that known word combinations were being recognised/activated, this does not seem to translate into a straightforward understanding of the phrase level meaning. For English native speakers reading English idioms, the results differ according to whether or not the idiom is English or Chinese in origin. For measures where there is a main effect for phrase type, this interacts significantly with language, hence the Chinese but not the English items show longer reading times for figurative phrases. Specifically, pairwise comparisons show that English phrases are read comparably whether they are used figuratively or literally (all *p*s > .05), whereas there is a general slowdown for the figurative (non-compositional) uses of translated Chinese idioms. This is seen most clearly in the effect of type for Chinese phrases for total reading time ($\chi 2$ (1, 822) = 13.39, $p < .001$) and total fixation count ($\chi 2$ (1, 822) = 9.23, $p < .01$).

5.7.2 Familiarity, compositionality and plausibility

Separate models were fitted to assess the importance of these factors. Continuous predictor variables were centred. Chinese and English native speakers were considered separately, so models included language and phrase type as fixed effects and considered the interaction with each predictor variable in turn. Random intercepts for subject and item and by-subject random slopes for language and type were again included. Only significant results are discussed in detail below (all model outputs are provided in Appendix 2b, tables 13-19).

Familiarity

For Chinese native speakers familiarity was not a significant factor in how Chinese or English items were read. Similarly, English native speakers showed no significant effects of familiarity for either set of items on any measures. Although this might seem surprising, the fact that all items in the study were deliberately chosen to be highly familiar may explain this (especially for the English items). In other words, items were either well known and were facilitated or were unknown and were not, with no ‘sliding scale’ of facilitation.

Compositionality

For Chinese native speakers compositionality played a role only in later measures. There was a significant interaction with phrase type for total reading time ($\beta = 0.12$, SE = 0.03, $t = 3.63$, $p < .001$) and total number of fixations ($\beta = 0.11$, SE = 0.03, $t = 3.34$, $p < .001$) and a significant three way interaction with phrase type and language for total reading time ($\beta = -0.20$, SE = 0.08, $t = 2.39$, $p < .05$). In both cases, greater compositionality was facilitatory for figurative and inhibitory for literal Chinese items, whereas in English the effect was facilitatory for literal items and negligible for figurative uses. For Chinese native speakers I also considered the alternative measure of compositionality, as judged by Chinese natives for the idioms read in the original Chinese characters. When these ratings were considered, greater compositionality was facilitatory for figurative uses for total reading time (interaction with phrase type: β = 0.09, SE = 0.04, t = 2.24, p < .05) and fixation count (β = 0.10, SE = 0.04, t = 2.47, p < .05) and showed no effect for literal items.

English native speakers showed significant interactions between type and compositionality and language and compositionality across all measures. This meant that for all items (Chinese and English phrase), more compositional items were actually read slower in the control condition, whilst the effect for figurative uses was negligible.

Plausibility

Literal plausibility (how acceptable each idiom would be if used in a literal context) showed a clear main effect for both groups for the Chinese items (for all measures except first pass reading for Chinese speakers, all *t*s > 2, all *p*s < .05). In all cases both figurative and literal uses were significantly facilitated by being more literally plausible, but there was no interaction between literal plausibility and phrase type. For Chinese speakers reading English idioms, both literal and figurative uses were also significantly facilitated by increased literal plausibility; for English native speakers there was facilitation for literal English phrases across all measures but not figurative phrases.

Proficiency

Models were fitted to assess the effect of vocabulary test scores, self-rated ability and usage scores for the Chinese native speakers. Usage was not significant, but both vocabulary test score and self-rated ability had a significant effect on all late measures (total reading time, regression path duration and fixation count). There was no interaction with language or phrase type, so higher proficiency led to faster reading across the board (which is expected), but participants were overall well-matched in their knowledge of the idioms. Increased proficiency did not therefore cause participants to read idioms from either language in a different way (more like native speakers), at least within the relatively homogenous cohort investigated here.

5.8 Discussion

Native English speakers performed as predicted. Idioms were read equally quickly in figurative and literal contexts, suggesting that, at least for the highly familiar idioms used here, there is no difference between a compositional analysis of the literal meaning and retrieval of the figurative meaning: both are available at around the same time. Chinese idioms, being unfamiliar to English speakers, were read significantly slower in figurative contexts, suggesting that their non-compositionality and the lack of a known figurative concept made them difficult to understand.

Chinese native speakers displayed the same pattern for both English and Chinese idioms: the literal versions of phrases were read more quickly than the figurative equivalents. This suggests that the overall meaning of the literal phrases could be understood with little difficulty, whereas the non-compositional figurative uses were harder to integrate into the overall context. For the English stimuli this result is in line with comparable previous studies (e.g. Cieślicka, 2006; Siyanova-Chanturia, Conklin & Schmitt, 2011) that show a clear advantage for literal over figurative meaning for non-native speakers. The fact that this pattern seems to hold for the translated Chinese stimuli raises some interesting questions. Despite the apparent priming of known lexical combinations in Experiment 2, the figurative meanings of the translated Chinese phrases were still difficult to understand in context, leading to longer overall reading times, more re-reading and more fixations. Therefore, although some degree of lexical activation seems to occur for the translated items, it may not be the case that the underlying conceptual entries associated with the idioms are automatically activated.

5.9 General Discussion

The combined results of Experiments 2 and 3 provide novel data on a previously underexplored aspect of the bilingual lexicon. Experiment 2 suggested that the recognition of the component words of idioms is an automatic process, even when the idiom is encountered in an unfamiliar translated form. This was seen specifically in the early reading behaviour for Chinese native speakers, where recognition of the ‘correct’ word was significantly and consistently faster than an unexpected control word. L1 influence must be important, as this is the only factor that renders the Chinese idioms ‘known’ for non-native speakers and not for English native speakers. Experiment 3, however, suggests that this recognition of form does not automatically lead to the activation of meaning: Chinese native speakers showed some difficulty interpreting the figurative phrases that were English idioms (as expected) and showed the same pattern for Chinese idioms. This was most clearly shown in total reading times, which reflect how easily the phrase can be integrated into the overall discourse context. This was also hinted at in Experiment 2, where phrase level reading times and regression path durations were comparable for idioms and control phrases, suggesting that simply recognising the correct words did not prevent the Chinese native speakers from having to re-read the phrases to make any sense of them.

Based on these results, a conceptual route whereby idioms are represented in a language non-specific way seems unlikely; if this was the case, Chinese idioms should be understood relatively easily in translation. However, one important question relates to whether the figurative or literal meaning of an idiom is more salient, with the most salient in any given context being the one that is accessed first (Giora, 1997). As non-natives will almost always have encountered the component parts of idioms separately and in literal contexts more than in combination as an idiom, a literal, compositional reading is likely to be the default, and will remain the most salient interpretation until much higher levels of proficiency are reached (Cieślicka, 2006; Matlock & Heredia, 2002). For this reason a different set of results may emerge for participants of very advanced proficiency in terms of their reading of both English and Chinese phrases. As the participants in this study were all from the same cohort, this may explain the lack of any effect of proficiency level on the processing of the different phrase types. The following study (Chapter 6) will specifically focus on higher proficiency non-native speakers to begin to explore this question further.

Based on the advantage for the correct lexical forms seen in Experiment 2, a lexical-translation mechanism of idiom activation seems more plausible, but this is also not without its problems. If we assume that English stimuli are being quickly and automatically translated into Chinese and that this is triggering a known sequence, logically this should show some activation for the underlying concept. Thus, if a

Chinese–English bilingual reading *draw a snake and add…* is quickly activating the Chinese equivalents and priming the character sequence 畫蛇添足, the conceptual meaning of this stored L1 form should be available alongside the final character, so making sense of the phrase in a figurative context in English should not be problematic. One explanation for the pattern of results is provided by more recent developments in idiom models, which suggest that idioms actually exist as multiple entries in the mental lexicon (e.g. Holsinger, 2013). In other words, they exist as distributed representations of single words with strong associative links, but also as canonical structures with set meanings. Thus the priming effect that is seen in activating the form of an idiom may be the result of lexical facilitation among the individual parts, whereas the representation of a whole form structure and its associated figurative meaning is likely to be affected by familiarity and (language specific) frequency of encounter for any given speaker. For native speakers, strong intralexical links and strong whole form representations exist to allow easy activation of both the form and meaning of the idiom. For Chinese native speakers, representations of whole forms are likely to be much weaker, both for L2 idioms and translations of L1 idioms, neither of which will have been regularly encountered in English. L2 idioms therefore do not show any lexical priming effects (Experiment 2) and are more difficult to process when used figuratively than literally because links with underlying concepts, if they exist, are not strong (Experiment 3). For the translated idioms, fast automatic translation may be sufficient to trigger associations through simple lexical priming/spreading activation, thereby facilitating formal recognition (Experiment 2), but the less salient, non-canonical presentation may not be sufficient to also trigger the whole form structure/meaning units (Experiment 3), or the novelty of encountering this form in English may work against its recognition.

Overall, it seems likely that idioms do retain some level of cohesion in translation. I interpret the findings in Experiment 2 as evidence that ‘correct’ completions were being primed, even though the idioms themselves were unknown in their translated forms. As demonstrated in Experiment 3, this activation did not extend to the overall meaning of the idioms, suggesting that the processes underlying recognition of form and access to meaning may not be the same, or that the ‘compositional by default’ approach for non-native speakers may negate any possible idiom advantage in the L2 until much higher levels of proficiency are reached. This study adds to previous work on the facilitative effect of congruence in formulaic language, corroborates the findings of Experiment 1 (Chapter 4), and provides suggestive evidence of crosslinguistic interaction at the multi-word level, which adds a valuable new dimension to our understanding of the bilingual lexicon. The following chapter will expand this further by including a dimension of congruency in the investigation, and by exploring participants of higher proficiency in English to see how they process both translated and L2 forms.

Chapter 6. Found in Translation

The two studies presented so far strongly indicate that crosslinguistic priming does extend to the multi-word level. Using two separate methodologies, I have shown that non-native speakers process idioms translated from their L1 more quickly than unknown control phrases. Given that there was no inherent difference in the plausibility of the translated idioms and control items in Chapter 5, logically, L1 knowledge is the only factor that can explain the effects seen. The purpose of the next study was to extend these findings in a number of ways:

- By investigating another speaker group (Swedish native speakers), where the L1 is more closely related to English and where the level of proficiency is likely to be much higher.
- By investigating idioms of a much more compact (therefore less predictable) form, and using neutral contexts to eliminate any effect of overall discourse context on the prediction of upcoming words.
- To introduce the dimension of congruency, where idioms also exist in both languages with the same form and the same or very similar meaning, to see whether this provides any additional ‘boost’ to processing.
- To investigate formulaic word pairs used in non-formulaic combinations. The aim here is to see how important the canonical formulaic frame is to fast processing, and to see whether any evidence of lexical priming between component words exists when this underlying citation form is compromised. This will be explained in more detail in the methodology section, and is picked up again in Chapter 7.

This chapter is under review under the title “Found in Translation: the influence of L1 on the reading of idioms in L2” (Carrol, Conklin & Gyllstad, under review).

6.1 Introduction

One of the most vital facets of advanced proficiency in a non-native language is the acquisition of sufficient vocabulary to be able to communicate in a range of registers, contexts and social situations. But this on its own may not be sufficient, since ‘native-like’ proficiency in a language requires mastery of the vast array of word strings, conventionalised sequences and ‘turns of phrase’ that characterise native speaker interaction. This broad category of lexical knowledge is considered under the banner of formulaic language. It has been suggested that such multi-word combinations may be at least as numerous as the amount of single words in English (Jackendoff, 1995), and they present an ongoing challenge to non-native speakers, even up to advanced levels of proficiency (Barfield & Gyllstad, 2009; Boers & Lindstromberg, 2012; Kuiper, Columbus & Schmitt, 2009; Laufer & Waldman, 2011; Nesselhauf, 2005). Considerable attention has therefore been paid to the best ways to approach formulaic language in language learning, and to how knowledge of such combinations is represented in the mental lexicon for both native and non-native speakers.

Idioms are amongst the most studied of all formulaic language types, and arguably pose the greatest degree of difficulty for non-native speakers. For a native speaker of English, hearing that a person has *kicked the bucket*, *bought the farm* or *bitten the dust* would generally be cause for condolence, but for second language speakers there are no obvious clues that each of these phrases has the meaning “die”. Idioms, like certain other types of formulaic language, behave in many ways like single words, in that they perform a referential or ideational function (Boers & Lindstromberg, 2012), but their difficulty for language learners comes from the fact that they are often opaque and their meaning is difficult to infer without some prior knowledge. Idioms are also lexically frozen and otherwise fixed in highly idiosyncratic ways, such that minor changes can be enough to destroy the idiomaticity and render a phrase distinctly un-nativelike: *booting the bucket* or *kicking the pail*, for example, are lexically comparable (at least in a superficial way) but figuratively removed from the intended idiom, although there is no reason why either should be more or less plausible if the phrase were encountered for the first time.

With such idiosyncrasies in mind, it is easy to see why idioms continue to present such a challenge to language learners. Evidence is mixed as to how non-native speakers process, comprehend and produce idioms, and to what extent L1 knowledge is utilised to support accurate and appropriate deployment in communication. As such, there is still something of a research gap in terms of constructing an accurate and detailed model of idiom (and formulaic language more broadly) representation and processing in L2 speakers. To help address this, in the present study I investigate how non-native speakers process idioms that they encounter in their L2. Specifically, I present advanced learners of English with idioms in three categories: L2 only idioms, translations of L1 only idioms, and idioms that exist in both languages (same combination of words and same phrase level meaning), to see to what extent L1 knowledge is utilised and how this interacts with L2 formulaic competence. Before turning to the study, I will review two principle strands of previous research: the psycholinguistic literature relevant to the processing of idioms in L1 and L2, and the range of studies that have investigated idiom transfer from L1 in non-native speakers.

6.2 Formulaic processing in L1 and L2: different strokes for different folks

It is well-established that idioms and other types of formulaic language are processed more quickly than ‘novel’ language by native speakers. That is, when other factors like length and single word frequency are controlled for, ‘known’ phrases are processed in a qualitatively different and quantitatively faster manner than newly created sequences. Tabossi, Fanari and Wolf (2009) called this advantage the “idiom superiority effect”, and it has been long demonstrated, using a range of methodologies, that many types of formulaic combinations are processed differently to novel language. It is important to highlight that this ‘processing advantage’ for formulaic sequences can be considered in two ways: form activation and meaning activation. Hence form activation refers to the recognition of specific word combinations, leading to, for example, faster initial reading of formulaic sequences, or faster responses to tasks that require a judgement of lexical form (such as lexical decision tasks, where a participant must decide whether a letter string is a real word or not in the target language). Meaning activation refers to the ability to understand a word or word combination in terms of its intended semantic meaning, and to integrate this into any surrounding context. This would be apparent in, for example, overall reading times for sentences containing formulaic sequences, or for tasks requiring an explicit semantic judgement, such as phrase level judgements of meaningfulness (whether a given combination is a meaningful phrase in the target language).

Focusing on idioms, it is unclear what drives the privileged processing that is so robust amongst native speakers. Despite being considered to be highly ‘familiar’, they are relatively infrequent, at least based on traditional corpus data (Moon, 1998), hence it seems logical that their meaningfulness as unitary concepts contributes to their fast recognition. Jolsvai, McCauley and Christiansen (2013) demonstrated that this was the case, with idioms in their study being judged to be acceptable phrases more quickly than frequency matched compositional phrases and less meaningful fragments. Modern theories of idiom processing have converged on a view of idioms as being simultaneously compositional *and* non-compositional/unitary. That is, a non-compositional entry for the whole unit exists at some level of representation, and this is accessible via some combination of the component words, which are assumed to be compositional/analysable (Cacciari & Tabossi, 1988; Holsinger, 2013; Libben & Titone, 2008; Smolka, Rabanus & Rösler, 2007; Sprenger, Levelt & Kempen, 2006; Titone & Connine, 1999; Van Lancker Sidtis, 2012). Such hybrid, distributed or constraint-based models allow for idioms to be retrieved directly, leading to faster recognition than if the individual words have to be accessed in turn, whilst the existence of strong links to the individual words accounts for the internal syntactic constituency of idioms (Konopka & Bock, 2009) and the literal activation of the component words (Cieślicka & Heredia, 2011; Hillert & Swinney, 2001; Holsinger and Kaiser, 2010).

Despite this well-established evidence base in monolingual speakers, research into how non-native speakers process idioms in the L2 has been relatively sparse. Results are mixed, with some studies suggesting that the fast processing for idioms is absent in non-natives (Chapters 4 and 5; Siyanova-Chanturia, Conklin & Schmitt, 2011). Other studies have shown clear effects of non-native speakers being sensitive to L2 frequency for other formulaic sequences such as collocations (Durrant & Schmitt, 2010; Isobe, 2011; Jiang & Nekrasova, 2007; Wolter & Gyllstad, 2013) at least in some cases and especially at higher levels of proficiency. A logical assumption is that for collocational and idiomatic combinations in either L1 or L2, frequency of input is a key driver of how effectively patterns will be registered and abstracted, in line with a usage-based account of language organisation (Bybee, 2006, 2008; Tomasello, 2003; Wulff, 2008), hence language specific experience will be a strong predictor of performance in either language. However, Wray (2002) argues that the development of the lexicon in L1 and L2 may be fundamentally different: children acquiring an L1 proceed in a primarily holistic fashion, while L2 learners (referring to older, more cognitively developed learners) adopt a more analytical approach. From an early age, therefore, the L1 is characterised by the acquisition of 'chunks' and a process of "needs only analysis" (Wray, 2002, 2008), where children move from a system of coarse, unanalysed meaning units to one where chunks are broken down and analysed as cognitive abilities develop. In older second language learners, the default position is one of compositional analysis, and the basic unit is the individual word from the start (although this would not necessarily be the case for very young children acquiring more than one language from an early age). Various researchers (Cieślicka, 2006, 2012; Cieślicka et al., 2014; Cieślicka & Heredia, 2011; Kecskés, 2000) have suggested that the literal meanings of individual component words should therefore be more salient to non-native speakers, especially since learners are likely to have encountered such words used individually and literally earlier and more often in the language learning process.

Results discussed in Chapter 5 from Siyanova-Chanturia, Conklin and Schmitt (2011), support the idea that non-native speakers show a greater tendency to rely on literal meanings of individual words in the L2.[18] They also tend to fall back on the L1

---

[18] Recall that this study showed faster reading of both figurative and literal uses of idioms like *at the end of the day* compared to controls. Non-native speakers showed no advantage for either version

conceptual system (Kecskés, 2000), or to consider idioms to be more decomposable than native speakers would (Abel, 2003). In a related line of argument, Giora's Graded Salience Hypothesis (Giora, 1997, 1999, 2002, 2003) suggests that salience may interact with language dominance, meaning that compositional, literal meaning is likely to be more salient for non-native speakers, while figurative meaning is the more salient for native speakers. One study to demonstrate that this may change as proficiency increases, leading to a more 'native-like' representation of idioms, is Yeganehjoo and Thai (2012), who used a cross-modal priming task to show that advanced Iranian learners of English showed a greater degree of identity priming for idioms than literal phrases (e.g. *cake* primed the phrase *The test was a piece of cake* to a greater degree than *The test was to bake a cake*). This replicates the findings of Sprenger et al. (2006) for native speakers, and suggests that at high levels of proficiency and with sufficient exposure to idioms, non-natives may develop native-like entries that can be accessed directly.

Due to the combination of less exposure and a more analytical approach, it seems that in general non-native speakers do not show the same speeded processing of idioms in the L2 as demonstrated by native speakers. In other words, 'known' lexical combinations may not be as easily primed, and figurative meanings may not be available as early as literal meanings of words. This is not to say that idioms may not be understood, just that the mechanisms underlying their access are either qualitatively different than in the L1, or simply slower, although this may change as proficiency develops. However, an important question is how well learners are able to

---

compared to the control phrase, and figurative uses led to longer reading times. Hence neither overall form or meaning were facilitated for non-natives.

utilise existing L1 knowledge to aid understanding of L2 formulaic language, which is what I consider next.

6.3 Formulaic transfer from the L1: better the devil you know

It seems reasonable to assume that all languages contain idioms or formulaic patterns, so all language learners already have a store of pre-fabricated word combinations in their L1 to draw upon. Often idioms do cross the language barrier, probably because of the universality of the conceptual metaphors that underpin them (in some cases), but also due to linguistic, social and geographical proximity and interaction (e.g. German and Dutch are likely to share more idioms than either language would with Mandarin Chinese, for example, since the languages are more closely related and because the speakers are likely to have been in closer contact throughout history). Logically, learners should therefore already know certain idioms in the L2 if they are congruent (same form and meaning in both languages). However, Kellerman (1977, 1978, 1986, 1987) demonstrated that learners are often reluctant to transfer more idiomatic senses of words, believing them to be highly language specific. In his studies of Dutch learners of English, more peripheral (generally more figurative) uses of *breken* (*to break*) were rejected, even when verbatim translations of uses like *break a strike* would be acceptable in both languages. Proficiency was an important factor, and less proficient learners showed a greater willingness to accept such transfer, while more advanced learners were resistant, considering idioms to be too marked and language specific to be transferable. Similarly, Dagut and Laufer (1985) found that Hebrew learners of English showed a clear tendency to avoid using phrasal verbs in English, with idiomatic/figurative items being avoided the most often (although this is perhaps not surprising since phrasal verbs do not exist in Hebrew). Hulstijn and

Marchena (1989) built on this work with Dutch learners, hypothesising that they would be less likely to avoid phrasal verbs since these are common in both English and Dutch. Whilst their learners did not avoid phrasal verbs as a class, they did show a clear pattern of avoiding more idiomatic items that were perceived as too ‘Dutch-like’.

Contrary to Kellerman’s findings, subsequent studies have demonstrated that equivalence between languages can be facilitative, and often learners are willing to transfer idioms from the L1 to aid L2 production. Irujo (1986) showed that advanced learners (Spanish L1) were able to produce significantly more English idioms (via a recall and translation task) when they had congruent forms in Spanish, and that they could use L1 knowledge to generalise the meaning of idioms in the L2, even when there was some variation in form. Odlin (1991) demonstrated a high degree of direct translation of Irish idioms into English, although he accounted for this in sociological/environmental terms (the high proportion of bilingual speakers in the community) rather than it being a purely linguistic phenomenon. Laufer (2000) looked at how L1-L2 similarity affects avoidance in non-native speakers (L1 Hebrew). Her results showed that learners do not avoid idioms as a category, but that the degree of language overlap was a clear determining factor in which idioms were correctly used in a written translation test. Total language overlap led to greater likelihood of use, but conceptually equivalent idioms of entirely different forms were also used relatively freely. Partial overlap (similar meaning but different words, e.g. English *lip service* vs. Hebrew *lip tax*) and conceptual non-equivalence (idioms where only a literal version was available in the L1) were more likely to induce avoidance. Charteris-Black (2002) conducted a study with Malay learners and found that linguistic and conceptual equivalence with the L1 was a considerable help in determining L2 figurative proficiency. Of greater difficulty were those forms where there was linguistic overlap but a different conceptual meaning, or culture-specific expressions where no conceptual or linguistic equivalence exists in the L1.

Bulut and Çelik-Yazici (2004) showed that L2 learners favour a "heuristic model" (p.113) whereby they employ a range of strategies to understand idioms. In their study of Turkish L1 learners (all teachers of English, therefore all advanced English proficiency) they found that L1-L2 equivalent items were likely to be treated in the first instance as "false friends". Literal and figurative meanings (from the L1) were then considered, and the most widely applied strategy was to use context to work out the most likely of all possible meaning. Liontas (2000) found that Greek learners also tended to use multiple strategies (primarily consideration of the literal meanings of words and use of context) to identify and comprehend L2 idioms. For idioms presented with no supporting context, those items with matching L1 forms were understood and defined much better than non-matching items (albeit with a very small sample size of seven participants). Liontas (2000) concluded that the difficulty posed by non-matching items was that "they required additional processing effort beyond mere translation of the lexical units" (p.16). The addition of supporting context aided both matching and non-matching items, highlighting the use of both bottom up (L1 knowledge) and top down (contextual clues and more general inferencing ability) information.

In addition to those studies that have focused on the end result – comprehension or production – several more recent studies have focused on the online processing of idioms and other types of formulaic language. Titone et al. (2015) conducted a study to compare the effect of code-switching on sentences that contained English idioms and congruent English-French idioms. Their results suggested that code-switches during an idiom were more disruptive than during a literal sentence (supporting the idea that they are treated as single units), but that greater congruency between languages reduced the amount of disruption. That is, items that exist in the same form in both languages, or with partial overlap of form or meaning, caused less disruption when the final word was changed into French than items that only existed in English. The authors proposed that this is evidence for the representation of congruent idioms in both languages, hence both meaning and form could be directly retrieved in either intact or code-switched phrases.

Wolter and Gyllstad (2011, 2013) employed two different methodologies to explore how collocations are processed by advanced Swedish learners of English. In the first (Wolter & Gyllstad, 2011), they used a primed lexical decision task with verb + noun pairs. Participants saw the first word of each pair as a prime and then were asked to judge whether the second word was a real English word. Items were either congruent (acceptable in Swedish and English, e.g. *ge ett svar – give an answer*) or incongruent (acceptable in English only, e.g. **betala ett besök – pay a visit*, where *pay* cannot be used with the same meaning in Swedish). They found a consistent advantage for congruent items in online (faster lexical decisions for collocation pairs) and offline measures (higher scores on a test of receptive collocation knowledge). In a second study (Wolter & Gyllstad, 2013), a phrase level judgement task was used with adjective + noun pairs. Again, congruent items (e.g. *high profile*) were judged to be acceptable more quickly and with fewer errors than incongruent (English only) collocations (e.g. *false teeth*). Yamashita and Jiang (2010) found a similar result for

Japanese-English learners, with congruent collocations judged more quickly and more accurately than incongruent ones, although this varied as a function of proficiency (higher level learners showed a difference in error rates but not response times, suggesting that L1 influenced the accuracy but not the speed of access for these learners). The authors interpreted their results as evidence that L2 exposure and L1 congruency combine to affect acquisition of formulaic patterns in non-native speakers.

Other studies have specifically considered items where there is a significant imbalance in the L1-L2 frequency, that is, L1 formulaic items that do not exist in the L2. As seen in Chapters 4 and 5, Chinese-English bilinguals showed faster processing for the form of translated items (faster recognition of expected words vs. controls), but this did not automatically lead to easier understanding of the figurative meanings of the phrases: known combinations were judged/read more quickly than control phrases, but overall more time was spent reading figurative phrases than literal ones, suggesting that these were more difficult to integrate into the overall discourse context (in line with results from Siyanova-Chanturia, Conklin and Schmitt, 2011, and the literal first view of non-native processing discussed previously). Conversely, Wolter and Yamashita (2014) conducted a similar test with Japanese learners, investigating whether patterns that would be acceptable in the L1 were facilitated in the L2 (e.g. *forgive marriage*, which would be an acceptable collocation in Japanese but which is not in English). They found no advantage relative to baseline items in a phrase level decision task either for adjective-noun (*bitter win*) or verb-noun (*drink tears*) combinations.

Ueno (2009) also investigated Japanese collocations presented in English using a primed lexical decision task (similar to Wolter & Gyllstad, 2011) and did find evidence of facilitation for such combinations, but only for very advanced learners. She suggested that this was evidence that as proficiency develops, rich semantic networks are formed that encompass both L1 and L2 in a non-selective manner. Wolter and Yamashita (2014), despite their null result, agreed with this theoretical standpoint. They invoked the three-stage model outlined by Jiang (2000; itself built on models first proposed by Levelt, 1989) to explain how L1 knowledge might be activated by L2 forms. In Jiang's model, all lexical items consist of a lemma level and a lexeme level; the lemma contains information about semantics and syntax, while the lexeme level relates to formal properties like morphology, orthography and phonology.[19] Jiang (2000) argued that due to the "practical constraints imposed on L2 learning" (p.47), many words fossilise at lemma mediation stage, meaning that lemma information from the L1 is copied to the newly acquired L2 lexeme. Wolter and Yamashita (2014) argue that this L1 lemma information may include aspects such as collocational links, hence encountering an L2 form will activate known associates and connections from the L1. This raises the possibility that, far from being single units, idioms and other formulaic units are linked and co-activated through something more akin to a spreading activation/lexical priming mechanism (Hoey, 2005; Pace Sigge, 2013). In line with broader views of how the bilingual lexicon is organised, language membership seems to be largely non-selective, that is, information about which language a lexical item belongs to seems to be a fairly late feature of processing (Dijkstra, 2007). In this regard, there is no reason why semantic or associative links

[19] Broadly we can relate this to the conceptual and lexical levels of the Revised Hierarchical Model and the model depicted in Chapter 2, Figure 2.1.

between words should not hold across languages, i.e. connections that exist in the L1 should be triggered even if words are encountered in the L2. Arguing against this, Williams and Cheung (2011; see also de Groot & Nas, 1991; Williams, 1994) showed that more central aspects of semantics but not associative relations showed cross-language priming. They found significant cross-language priming for translation equivalents (e.g. *squirrel-écureuil*) and semantically similar words (e.g. *sofa-chaise (chair)*), but not for semantic associates (e.g. *desk-chaise*).[20] They argued that associate relationships were established more through experience, hence they highlight "the importance of individual learning episodes in providing the meanings with which they are associated" (p.93). If this view is accurate, information such as how a word combines with other words to create formulaic configurations may not form part of the core lemma level knowledge that is assigned to the L2 form. Knowledge of idioms and other formulaic combinations (e.g. collocations) would therefore be dependent on language-specific frequency of encounter.

In summary, there is clear evidence that formulaic language holds a privileged processing status for native speakers, but this is not necessarily the case for non-natives. L1 knowledge and L2 proficiency/exposure are both important factors in how formulaic language is processed in the L2, especially in receptive tasks where learners can use multiple sources of information to reach a decision about the likely meaning of idioms and other phrases. It seems clear that congruency between languages can

[20] Williams and Cheung tested semantic priming from L3 (French) to L1 (Chinese) via English (L2), which was the language of instruction. They used French prime words (e.g. *chaise*) and Chinese target words (e.g. 書桌(desk)), on the assumption that since English had been the language of instruction, no episodic memory connections could exist between the French and Chinese forms, hence any priming should be the effect of direct semantic connections. For the sake of simplicity, I have presented only the English-French forms to demonstrate the priming effects that were/were not observed.

have a facilitative effect when learners encounter L2 formulaic language, but that the extent of this will be determined by many factors (including the nature of the task, the perceived transferability of the item in question, and learner specific factors like proficiency, creativity, etc.).

The present study aims to add to the literature on non-native processing of formulaic language in a number of ways. Primarily, I intend to explore the importance of L1 knowledge in the on-line processing of idioms presented in the L2. Given the variability in results of previous studies (e.g. Chapters 4 and 5; Ueno, 2009; Wolter & Yamashita, 2014), this study allows me to further test whether translations of formulaic phrases show privileged processing by non-native speakers. A key question also relates to the effect of congruency between languages, and whether this shows any additional facilitatory effects compared to items that only exist in the L1. In other words, is cross-language facilitation the result of L1 knowledge, or is additional awareness/experience of the same combinations in the L2 an added benefit?

In this study I will explore these questions with a different speaker group (L1 Swedes), enabling me to test participants that are likely to be of very advanced proficiency and also to test a language that is more closely related to English than in the studies mentioned above (which compare Chinese and Japanese to English). This more advanced proficiency group also allows me to test whether advanced non-natives show any evidence of formulaic processing for L2 only idioms, i.e. do very high levels of proficiency lead to more native-like processing for formulaic items that must be learned in the L2? This was generally not observed for the Chinese participants in the previous two studies, and in general is inconsistent in previous studies in the literature.

An additional question is whether shorter, less predictable idioms of the form X-the-X (e.g. *kick the bucket*, *spill the beans*, etc.), where the idiom is not uniquely identifiable until the whole phrase has been seen, also subject to fast recognition/processing? Monolingual and bilingual idiom studies generally use idioms of variable length (e.g. Chapters 4 and 5; Rommers et al., 2013; Titone et al., 2015), meaning that often predictability can be a potentially confounding factor (e.g. native speakers will be likely to guess the completion to *flog a dead*… (*horse*), hence it is hard to determine whether the advantage seen is the result of automatic, lexical processes or strategic, predictive processes). Related to this, context can be an important factor in the processing of different kinds of idioms (Cieślicka, 2012; Titone & Connine, 1999), with a biasing context greatly increasing predictability. I will therefore examine idioms in neutral contexts where no clues are provided to aid prediction of the upcoming words.

Finally, I aim to investigate whether there is evidence that the idiom advantage is the result of lexical priming/spreading activation, rather than whole form access. In other words, is there a connection between component words such that seeing one component of an idiom primes the other components, regardless of whether the words are used as part of a formulaic unit? For example, in the sentence *I saw him kick it and the bucket went flying across the room*, would we expect to see any priming for *bucket* once *kick* has been encountered? Camblin, Gordon and Swaab (2007) and Carroll and Slowiaczek (1986) demonstrated that semantically associated words showed priming in this way, so evidence of comparable priming for idiom components might indicate an intrinsic link between components rather than a separate, whole-form representation.

6.4 Experiment 4

In the present study participants read idioms embedded in short, context-neutral sentences. All materials were presented in English, and I recorded the reading patterns for the whole idiom and its final word. I also split idioms into their component parts and recorded the reading times for the second word when it appeared later in a sentence. For example:

Idiom condition

1) Idiom sentence: It was hard for him to *break the ice* when he was at the party last week.
2) Control sentence: It was hard for him to *crack the ice* when his locks froze last week.

Component words condition:

3) Idiom sentence: It was hard to *break* at first but the *ice* finally gave way after a few minutes.
4) Control sentence: It was hard to *crack* at first but the *ice* finally gave way after a few minutes.

English native speakers and non-native (L1 Swedish) participants were tested on a set of English idioms, translated Swedish idioms and congruent idioms. I used eye-tracking to measure the number and duration of word fixations during natural reading. As discussed in Chapter 3, in this study I adopt a ‘hybrid’ method of analysis, whereby features of both the whole phrase and the final word are examined, and I consider a range of early and late measures for each unit (see Table 6.1). Generally speaking, early measures are taken to be a reflection of lexical access and other automatic processes, while late measures are seen as reflecting post-lexical strategic effects (Altarriba, Kroll, Sholl & Rayner, 1996; Inhoff, 1984; Paterson, Liversedge & Underwood, 1999; Staub & Rayner, 2007). This can also be related to the distinction between form and meaning activation: early measures can be taken as an index of form activation, since they reflect how easily the 'correct' lexical combinations are activated, while later measures show how easily the overall meaning is activated and integrated into the wider sentence. Based on previous research of eye-tracking formulaic units, I expect idioms to show shorter overall reading times, and the final words to be skipped more often or be read more quickly than in control phrases, reflecting their status as 'known' units (Chapter 5; Siyanova-Chanturia, Conklin & Schmitt, 2011; Underwood, Schmitt & Galpin, 2004).

Table 6.1. Eye-tracking measures used in the experiment along with descriptions and stage of processing

Stage of processing	Type of measure	Description
	Phrase level	
Early	First pass reading time	The sum duration of all fixations on the phrase the first time it is encountered in the sentence
Late	Total reading time	The sum duration of all fixations on the phrase during the trial (including re-reading)
	Total fixation count	The total number of fixations on the phrase during the trial
	Word level	
Early	Likelihood of skipping	The likelihood that a word is skipped (not fixated at all) during first pass reading
	First fixation duration	The duration of the first fixation on the word
	First pass reading time	The sum duration of all fixations on the word the first time it is encountered in the sentence
Late	Total reading time	The sum duration of all fixations on the word during the trial (including re-reading)
	Regression path duration	The duration of all regressions to material preceding the word once it has been fixated for the first time (including the prior context and the start of the phrase)

6.4.1 Participants

Twenty-four English native-speakers and 24 Swedish native speakers took part in the study and received a fee for their participation. Native English speakers were all undergraduates at the University of Nottingham with L1 English and no experience of learning Swedish. Non-native speakers were all students at Lund University, which is an English language university in Sweden. Most of the participants were undergraduates (one postgraduate) and were studying in the Department of English Language and Literature. All had Swedish as their L1. Following the main experiment demographic and language background data were collected, including self-rating of proficiency in English and an estimate of usage in various contexts (e.g. at university, at home with friends and family, reading for pleasure, etc.). A short vocabulary test was also administered, consisting of a shortened version of the Vocabulary Size Test (Nation & Beglar, 2007). In this test items are presented in a short, neutral context (e.g. Shoe: This is a shoe) and participants select the correct definition from four alternatives, with an additional don't know option to minimise guessing. The original test sampled 10 items from each of the first 14 BNC word levels (level one represents the 1000 most frequent word families in English, level two the next 1000, and so on). I randomly selected two items from the first ten bands to give a total of 20 items, so a score of 20/20 would correspond to a vocabulary size of around 10,000 words. The mean score on this test was high (16.2/20, corresponding to around 8000 word families), which was in keeping with the assumption that the non-native participants in this study were advanced level learners of English. A summary of the non-native participants is provided in Table 6.2.

Table 6.2. Summary of non-native speakers (all L1 Swedish).

	Age	Years of English	Reading	Listening	Speaking	Writing	Usage	Vocab
Mean	23.7	11.5	7.4	8.1	7.0	7.0	39.5	16.2
Range	19-45	9-19	4-10	5-10	4-9	4-10	29-49	11-20

Note: Years of English is based on the years of formal instruction each participant had undergone at the time of the study; Reading, Listening, Speaking and Writing are all self-rated proficiency measures out of 10; Usage is an aggregated estimate of how often participants use English in their everyday lives (10 measures, each estimated out of 5 to give a total score out of 50); Vocab is the score out of 20 on the modified vocabulary size test.

6.4.2 Materials

Three categories of item were created: English only idioms, Swedish only idioms and congruent idioms (those which exist with the same or very similar form and meaning in both languages). In all cases items were selected to conform as closely as possible to the structure X-det-N, where X was normally a verb (e.g. *kick the bucket*) or in some cases a noun (*neck over head*) or preposition (*under the ice*). The determiner was sometimes a personal pronoun (e.g. *pull your weight*), was sometimes replaced by a preposition (*fall from grace*) and was sometimes omitted (*tread water*). The key criterion was that each item must contain two main lexical items: some flexibility was permitted to ensure that sufficient numbers of items could be found in each of the three categories. All experimental items are available in Appendix 3a.

English idioms were first selected from a variety of sources, including items from previous published studies and from various idiom dictionaries (principally Warren, 1994). An initial pool of around 100 items was prepared, all of which were considered to be common in modern British English. This list was examined by a Swedish native speaker (a lecturer in the Department of English Language and Literature at Lund University, who therefore had nativelike proficiency in English), who identified all items that have a corresponding version with identical or near-identical form in Swedish, for example *break the ice*, which has a direct equivalent *bryta isen*.[21] In all cases the main lexical items had single word translation equivalents and appeared in the same order in both languages, although because Swedish definite articles are attached to the end of the noun they modify, some variation in form was inevitable

[21] This judgement was primarily based on the Swedish native speaker's personal experience, but was also checked by him using a variety of Swedish idiom dictionaries and lists (principally Hübinette & Odenstedt, 1988; Hargevik & Ljung, 1989).

(e.g. *ice* = *is*, *the ice* = *isen*). All items were also considered to be common in modern Swedish. Final sets of 40 items were created for each condition (congruent and English only). Items that were removed were done so to avoid use of very low frequency or obscure vocabulary items and to avoid too much repetition of the same lexical items. The phrase frequency of each item was checked in the British National Corpus (BYU-BNC, Davies, 2004). English only items showed a mean of 36 occurrences (per 100 million words) and congruent items showed a mean of 62.

A final list of Swedish only items was prepared with the assistance of the same Swedish native speaker. These consisted of items of the same general form as the other conditions: two main lexical items, mostly V-det-N but also in some cases N-Prep-N (*a cow on the ice*) and det-Adj-N (*the red thread*). The majority of items in this condition (around 80%) conformed to the V-det-N structure. All items were chosen from various Swedish idiom dictionaries and word lists, as before. The list was reviewed to ensure that none of the items existed in English. These items were then transliterated into English as closely as possible, with the core meaning of each word taken as the basis for translation by a Swedish native speaker. These translations were checked using Google Translate, and then submitted to a translation norming test using three more Swedish native speakers who were advanced learners of English (all employed in the Department of English Language and Literature at Lund University as either lecturers in English or in one case a post-doctoral researcher, so their proficiency was native-like or near-native like). They were asked to assess the English translations for their accuracy using a five point scale and where appropriate suggest any improvements. Overall ratings were high (mean=4.7/5) and any items that received scores below 4/5 were amended as per the suggestions given by the raters.

These suggestions were generally very minor (e.g. *neck* instead of *throat* for the item *hals över huvud*). Mean phrase frequency of translated items taken from the BNC was 24 (per 100 million words), although this was hugely inflated by the high occurrence of one item, the Swedish *gå bort* (*walk away*, meaning "to die"), which occurs in its literal form in English 834 times. Without this item, the mean phrase frequency for the translated items was three.

All items were presented in a short norming study to assess how well known they were to native speakers of English. Participants were asked to indicate familiarity with each phrase on a seven-point Likert scale (1 = completely unfamiliar, 7 = very familiar). English only items and congruent items were generally very well known (6.2/7 and 6/7, respectively), while translations of Swedish items were generally unknown (1.6/7). Subjective familiarity ratings for the test items were collected following the main experiment on a by-subject basis.

For all idioms a control item was created by changing the first content word for a logical alternative matched for part of speech and, where possible, length and frequency, e.g. *break the ice* became *crack the ice*. All control phrases in all three conditions therefore formed logical, acceptable, but non-idiomatic phrases in English. For each of the stimulus items four variants were created for use in the main experiment. As shown previously in the example sentences, all items were used in a formulaic condition – idiom (1) vs. control phrase (2) – and a component parts condition – idiom (3) vs. control (4). In each case the context was created to be neutral, i.e. it did not bias a figurative or literal meaning of the idiom, hence encountering the first word (e.g. *kick* in *kick the bucket*) would not lead participants to expect an idiom completion any more than they might expect a literal completion. The context was created so that both the idiom and control phrase sounded natural, and it was only after the idiom/control phrase that the context differed so that overall each item was an acceptable, logical sentence in English. In all cases the material preceding and immediately following the phrase was the same for the idiom and the corresponding control. Idioms/controls were therefore automatically matched for number of preceding words (mean = 4) and were comparable for the number of words following the phrase (idioms = 11.2, controls = 11.8). By creating controls where the first word of each phrase was changed rather than the terminal word, I could directly compare reading times for the same word in different contexts, rather than comparing different words as has often been the case in previous idiom studies (e.g. *break the* *__ice__* vs. *break the* *__cup__*).

For the component parts condition the same idioms/controls were used but the two lexical items were separated and used in a non-formulaic configuration. In this condition the number of words preceding (mean = 4.6) and separating (mean = 4.2) the two lexical items was constant for idioms/controls, and the following context was more often than not identical as well. Although this did differ for some items, the average number of words following the second lexical item ended up being identical (mean = 7.1 for idiom and control conditions).

Four counterbalanced lists were created to ensure that no participant would see more than one variant of the same item. Lists were matched internally (across conditions) and externally (relative to each other) for phrase frequency, and for length and frequency of the individual words. Because it was considered very unlikely that participants would spot the component parts sentences and realise that they were composed of words that formed idioms, no additional filler items were added. This meant that each list contained only 25% idioms, with the rest of the items being literal sentences of some form (overall there were 30 idioms, 30 controls, 30 idiom component parts and 30 component part controls per list).

6.4.3 Procedure

The study was conducted using an Eyelink 1000+ system for the native speakers and an Eyelink 1000 system for the non-native speakers. Recording was performed with a desk-mounted eye-tracking camera and was monocular at a sampling rate of 500Hz. For all participants the right eye was tracked unless setup proved problematic, in which case the left eye was used. Participants were seated in front of a 1280x1960 resolution widescreen monitor with a refresh rate of 144Hz. Head position was stabilised with a desk-mounted chin rest. Participants read an initial instruction screen, following which camera setup and calibration were performed.

Participants were randomly assigned to one of the four stimulus lists. The presentation sequence was the same throughout. At the start of each trial a fixation cross appeared toward the centre-top of the screen. Once the experimenter had verified and accepted the participant gaze position, the cross disappeared and a sentence appeared across the middle of the screen. All sentences appeared on one line. Participants were asked to read each sentence as naturally as possible for comprehension. They were to try to read each sentence only once, but rereading was allowed if required. As soon as they had finished reading they were asked to press the spacebar. One third of all items were followed by a simple yes/no comprehension question (included to ensure that participants were actually reading for comprehension rather than just skimming the sentences), and the remainder were followed by a *Ready?* prompt. Participants saw the stimulus items in two blocks of 60 sentences. Each block was balanced across conditions and within each block the trial order was randomised for each participant. Participants were shown five practice trials, following which the main trial began. Following block one participants were given a short, self-timed break, after which the camera was re-calibrated and validated. Trial by trial drift correction was monitored throughout and re-calibration performed as required. The eye-tracking part of the study took around 30 minutes for non-native participants to complete and around 20 minutes for native speakers.

All participants were then asked to complete a rating questionnaire to indicate subjective familiarity with the items used in the main study. They were asked to indicate familiarity with each item (whether they had seen the item before and whether they knew the figurative meaning) on a seven-point Likert scale. For native speakers all 120 items were presented in English in random order. For non-native speakers two versions of the task were used. One presented the English only items and the congruent items in English, and the second presented the Swedish only items and the congruent items in Swedish. In both cases the order of presentation was randomised, and to minimise repetition effects for the congruent items (which appeared on both lists but in different languages) half of the participants saw the English list first and half saw the Swedish list first. Participants were specifically asked to indicate their familiarity with the items in the language in which they appeared.

Finally all participants were asked to provide some background information. For native speakers this consisted of basic information such as age and study status. For non-natives this included a more detailed background questionnaire and vocabulary test, as described earlier and as summarised in Table 6.1.

6.5 Results and analysis

Results of the familiarity ratings were first calculated to ensure that all items had been generally well known to the participants. For non-native speakers, Swedish only items were well known in Swedish (5.1/7). English only items were also well known (4.9/7), and this again demonstrates that the learners in this study were advanced and had a relatively high level of exposure to English. The congruent items were familiar in both languages: 5.7/7 in Swedish and 5.4/7 in English. For native speakers English only idioms (5.6/7) and congruent idioms (5.5/7) were well known, while Swedish idioms were not (2.8/7).

Prior to analysis of the eye-tracking all trials were visually checked for missing or unusable data. Any trials where track loss occurred were removed, although this accounted for a very small fraction of all data (less than 0.01%). Data were cleaned according to the four stage process within the Eyelink Data Viewer software, so all fixations shorter than 100ms or longer than 800ms were removed. Fixation data were extracted for all trials for the whole phrase and final word (for items used in their formulaic configurations) and for the second word (for the component parts condition). Results for the formulaic configurations and component parts conditions were analysed separately.

6.5.1 Formulaic configurations

Results were analysed using an omnibus linear mixed effects model using the lme4 package (version 1.0-7, Bates, Maechler, Bolker, Walker, Christensen, Singmann & Dai, 2014) in R (version 3.1.2, R Core Team, 2014). Three treatment coded main effects of group (English L1 vs. Swedish L1), type (literal phrase vs. idiom) and condition (Congruent vs. English vs. Swedish) were included, as were random intercepts for subject and item and by-subject random slopes for the effects of phrase type and condition (Barr, Levy, Scheepers & Tily, 2013). A summary of the raw results is shown in Table 6.3.

Table 6.3. Results for all speakers for formulaic configurations, split by participant group and by phrase/word level measures.

	Swedish only		Congruent		English only	
	Idioms	Controls	Idioms	Controls	Idioms	Controls
Swedish native speakers						
Whole phrase						
First pass reading time	625 (352)	670 (432)	597 (299)	596 (338)	564 (274)	609 (291)
Total reading time	1176 (683)	1309 (808)	997 (654)	1062 (637)	977 (590)	1021 (590)
Fixation count	5.0 (2.9)	5.4 (3.4)	4.2 (2.6)	4.6 (2.5)	4.2 (2.6)	4.4 (2.4)
Final word						
Likelihood of skipping	.08 (.26)	.02 (.13)	.13 (.34)	.04 (.19)	.13 (.33)	.13 (.34)
First fixation duration	237 (116)	256 (108)	211 (116)	229 (104)	215 (126)	207 (111)
First pass reading time	282 (155)	299 (160)	237 (138)	250 (126)	235 (147)	247 (147)
Total reading time	455 (318)	535 (376)	349 (318)	378 (275)	329 (247)	348 (271)
Regression path duration	739 (595)	867 (737)	524 (580)	617 (581)	507 (507)	531 (535)
English native speakers						
Whole phrase						
First pass reading time	450 (228)	463 (281)	361 (145)	415 (194)	367 (178)	430 (191)
Total reading time	832 (536)	652 (407)	475 (246)	561 (333)	466 (268)	582 (390)
Fixation count	3.9 (2.4)	3.0 (1.6)	2.4 (1.1)	2.7 (1.4)	2.3 (1.6)	2.8 (1.1)
Final word						
Likelihood of skipping	.10 (.31)	.11 (.32)	.29 (.45)	.25 (.43)	.33 (.47)	.23 (.42)
First fixation duration	202 (103)	197 (102)	149 (103)	161 (113)	135 (105)	166 (104)
First pass reading time	223 (123)	208 (115)	150 (104)	166 (118)	140 (111)	173 (114)
Total reading time	337 (267)	248 (162)	179 (157)	213 (195)	159 (144)	216 (212)
Regression path duration	541 (489)	360 (313)	211 (228)	278 (303)	199 (233)	291 (364)

Note: Mean values (SD in brackets): for duration measures reading times in ms are reported; fixation count is a raw value; likelihood of skipping is reported as a probability.

Table 6.4. Omnibus mixed effects model estimates for all phrase level eye-tracking measures. For condition, Congruent is taken to be the baseline.

	First pass reading time			Total reading time			Fixation count		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z
Intercept	6.13	0.10	63.30	6.18	0.07	91.33	0.96	0.07	14.16
Group: Swedish	0.29	0.07	3.96***	0.64	0.09	7.58***	0.52	0.08	6.30***
Type: Idiom	-0.12	0.05	-2.61**	-0.13	0.04	-3.13***	-0.11	0.06	-1.98*
Condition: English	0.05	0.06	0.87	0.04	0.06	0.71	0.03	0.07	0.37
Condition: Swedish	0.07	0.06	1.19	0.15	0.06	2.52**	0.11	0.07	1.67
Group*Type	0.16	0.07	2.38*	0.04	0.06	0.68	0.03	0.07	0.39
Group*Condition: English	0.02	0.07	0.27	-0.08	0.06	-1.44	-0.09	0.07	-1.26
Group*Condition: Swedish	0.01	0.07	0.14	0.03	0.06	0.52	0.02	0.07	0.27
Type*Condition: English	-0.04	0.07	-0.60	-0.07	0.05	-1.24	-0.07	0.08	-0.81
Type*Condition: Swedish	0.10	0.07	1.60	0.34	0.05	6.28***	0.36	0.08	4.72***
Group*Type*Condition: English	-0.08	0.09	-0.85	0.11	0.08	1.47	0.11	0.10	1.07
Group*Type*Condition: Swedish	-0.18	0.09	-1.92*	-0.35	0.08	-4.58***	-0.35	0.10	-3.63***
Control predictors:									
Word 1 Length	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Word 1 Frequency (log)	-0.02	0.01	-2.62**	n/a	n/a	n/a	n/a	n/a	n/a
Word 2 Length	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Word 2 Frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance		
Item	0.022			0.038			0.031		
Subject	0.037			0.069			0.053		
Subject | Type	0.003			0.004			0.001		
Subject | Condition: English	0.002			0.001			0.002		
Subject | Condition: Swedish	0.006			0.004			0.004		
Residual	0.258			0.178			n/a		

*Note: Table displays coefficients, standard error (SE) and t-values (z-values for fixation count), with significance values estimated by the lmerTest package in R (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): *** p < .001, ** p < .01, * p< .05*

Table 6.5. Omnibus mixed effects model estimates for all word level eye-tracking measures. For condition, Congruent is taken to be the baseline.

	Likelihood of skipping			First fixation duration			First pass reading time			Total reading time			Regression path duration		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-0.06	0.39	-0.15	5.31	0.03	155.99	5.42	0.10	55.32	5.49	0.14	38.53	5.65	0.09	62.99
Group: Swedish	-2.29	0.40	-5.65***	0.10	0.05	2.17*	0.15	0.05	2.89**	0.32	0.08	4.22***	0.51	0.12	4.39***
Type: Idiom	0.22	0.22	0.99	-0.05	0.03	-1.34	-0.06	0.04	-1.63	-0.11	0.05	-2.11*	-0.18	0.07	-2.64**
Condition: English	-0.24	0.27	-0.89	-0.01	0.04	-0.36	0.01	0.04	-0.16	-0.03	0.06	-0.47	-0.01	0.07	-0.12
Condition: Swedish	-1.01	0.30	-3.38***	0.04	0.04	1.12	0.03	0.04	0.62	0.02	0.06	0.27	0.14	0.07	1.81
Group*Type	1.18	0.44	2.66**	0.06	0.05	1.27	0.09	0.05	1.63	0.06	0.07	0.93	0.06	0.09	0.68
Group*Condition: English	1.55	0.45	3.41***	0.03	0.05	0.61	0.06	0.05	1.22	0.03	0.07	0.48	-0.05	0.09	-0.61
Group*Condition: Swedish	0.18	0.65	0.28	0.04	0.05	0.97	0.08	0.05	1.49	0.24	0.07	3.59***	0.18	0.09	2.00*
Type*Condition: English	0.42	0.31	1.36	-0.01	0.05	-0.21	0.00	0.05	-0.01	-0.01	0.07	-0.17	-0.01	0.09	-0.01
Type*Condition: Swedish	-0.31	0.37	-0.85	0.06	0.05	1.24	0.11	0.05	2.13*	0.34	0.07	4.95***	0.52	0.09	6.03***
Group*Type*Condition: English	-1.89	0.57	-3.31***	0.01	0.06	0.16	0.07	0.07	-0.97	0.00	0.10	0.02	0.08	0.12	0.63
Group*Type*Condition: Swedish	0.54	0.77	0.71	-0.08	0.06	-1.27	-0.14	0.07	-2.03*	-0.40	0.09	-4.27***	-0.47	0.12	-4.02***
Control predictors:															
Word 2 Length	-0.25	0.07	-3.68**	n/a	n/a	n/a	0.02	0.01	2.64**	0.04	0.01	3.01**	n/a	n/a	n/a
Word 2 Frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	-0.02	0.01	-2.96**	-0.02	0.01	-1.89*	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance		
Item	0.345			0.003			0.007			0.017			0.030		
Subject	0.295			0.012			0.016			0.043			0.119		
Subject | Type	0.000			0.000			0.003			0.002			0.007		
Subject | Condition: English	0.030			0.002			0.001			0.004			0.004		
Subject | Condition: Swedish	0.000			0.002			0.004			0.005			0.008		
Residual	n/a			0.099			0.122			0.220			0.347		

*Note: Table displays coefficients, standard error (SE) and t-values (z-values for likelihood of skipping), with significance values estimated by the lmerTest package in R (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): *** p < .001, ** p < .01, * p< .05*

Tables 6.4 (phrase level) and 6.5 (word level) show the omnibus mixed effects analysis for all eye-tracking measurements. For word level analysis, likelihood of skipping was analysed with a logistic mixed effects model and skipped items were removed from the analysis for subsequent durational measures. All duration measures are log-transformed to reduce skewing. Fixation counts were analysed using a generalised linear model with a poisson regression. The omnibus analysis shows clear effects of group (English native speakers were faster readers than Swedish native speakers), which is not surprising. There was also an overall effect of type for most measures, and this was qualified by an interaction between phrase type and condition (congruent vs. Swedish) for all measures except the word level measures likelihood of skipping and first fixation duration. This suggests that while Swedish native speakers treated both congruent and Swedish items as ‘known’, English native speakers showed a significant difference in how they read these items. To further explore the data, separate models were fitted for the Swedish native speaker and the English native speaker data (for both groups the L1 only condition was taken as the baseline). Interactions were explored using the Phia package (version 0.1-5, De Rosario-Martinez, 2013) in R to conduct pairwise comparisons as appropriate. Significant results are described here and full model outputs are provided in Appendix 3b (tables 1, 2 and 3).

Swedish native speakers showed a pattern of overall facilitation for idioms compared to controls in all three conditions. At the phrase level, there were no effects for first pass reading time but there was a significant overall effect of type for total reading time (participants spent less time reading idioms than controls, $t = -2.65$, $p < .01$) and total number of fixations (participants fixated fewer times overall on idioms than controls, $z = -1.98$, $p < .05$). For word level analysis, likelihood of skipping was significantly higher for idioms (overall effect of type, $z = 2.96$, $p < .01$) and interacted with condition for the Swedish vs. English contrast ($z = -2.71$, $p < .01$). Pairwise comparisons confirmed that the final words of idioms were skipped more often than controls in the Swedish only condition (χ^2 (1, 1434) = 8.78, $p < .01$) and congruent condition (χ^2 (1, 1434) = 12.49, $p < .01$) but not the English only condition (χ^2 (1, 1434) = 0.04, $p = .84$). Other early measures (first fixation duration and first pass reading time) showed no significant effects. Total reading time showed an overall effect, so idioms in all conditions were read more quickly than controls ($t = -2.27$, $p < .05$). Regression path duration showed no effects. Importantly, on no measure was there any interaction between phrase type and condition for Swedish only and congruent idioms, suggesting that there were no differences between these two conditions (congruency did not provide any advantage relative to L1 only phrases).

English native speakers showed a clear pattern across all measures except for first fixation duration and first pass reading time at the word level (although it should be remembered that these are strongly affected by the removal of any skipped items). For all other measures there was a main effect of phrase type and an interaction between type and condition (English vs. Swedish). As expected, there was no interaction with phrase type for English vs. Congruent items, demonstrating that to English native speakers there was no difference between these conditions and all items were treated as known phrases. Pairwise comparisons confirmed that for all phrase level measures and late word level measures (total reading time and regression path duration), native speakers spent less time on English and congruent items compared to controls (all *p*s $<.05$). For Swedish idioms significantly longer was spent during initial reading and subsequent re-reading for all late measures (all *p*s < .01), while for early measures (phrase level first pass reading and likelihood of skipping the final word) the difference was not significant. This suggests that English native speakers had relatively little difficulty when the Swedish idioms were first encountered, or at least they read them as easily as control phrases, but post-lexical integration of the overall meaning was severely disrupted. For English only and congruent items, even though the literal control items were all perfectly plausible, there was a consistent advantage for idioms on all measures, as predicted by the previous literature. Figure 6.1 demonstrates the different patterns for English native speakers and Swedish native speakers on phrase level reading time, likelihood of skipping the final word and regression path duration from the final word.

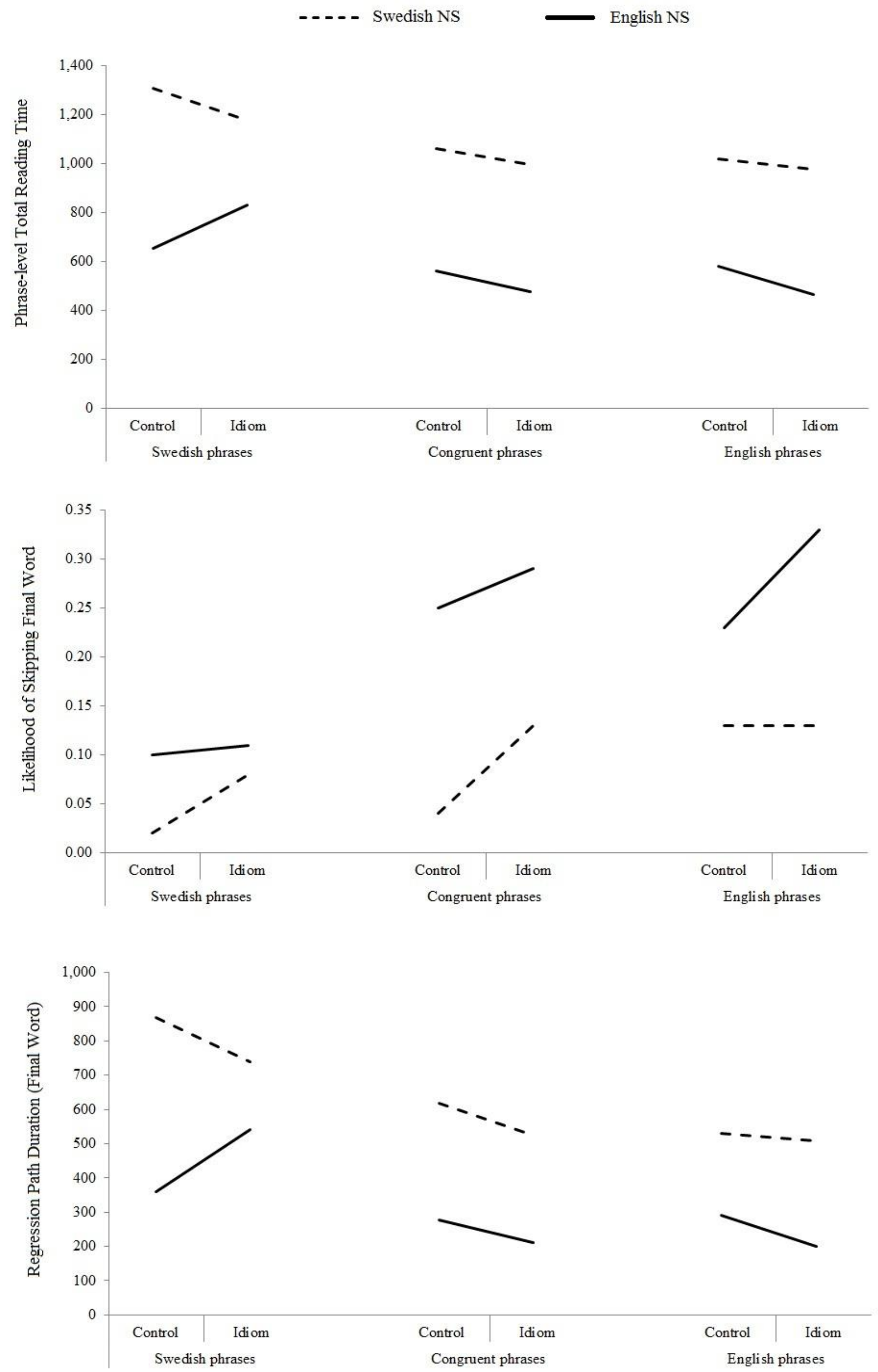

Figure 6.1. Interaction plots for Experiment 4: phrase level reading time in ms (top row), probability of skipping the final word (middle row) and final word regression duration in ms (bottom row).

The top row of Figure 6.1 (whole phrase total reading time) demonstrates that Swedish native speakers spent less time overall reading idioms than control phrases in all three conditions. English native speakers showed a clear advantage for idioms in congruent and English only conditions, but show significant disruption when reading the unfamiliar Swedish idioms. The second row demonstrates that the idiom advantage, as indexed by the automatic activation of component words leading to higher skipping rates in idioms than controls, may not be as well established for L2 idioms as those that exist in the L1 for Swedish native speakers. That is, both Swedish idioms and congruent items show higher likelihood of skipping the final word than controls, but for English only phrases there is no difference between the two types. English native speakers show a clear advantage for idioms in the congruent and English conditions but no difference for Swedish only items, hence there was nothing in the literal control phrases or idioms that made the final word more or less predictable for Swedish phrases. Despite this, regressions to the prior context and the start of the idiom once the final word had been fixated (bottom row) were consistently shorter for idioms in all conditions for non-native speakers (although this effect was not significant), while for English native speakers the Swedish item condition again shows a clear effect of disruption for the idioms compared to controls (more time spent overall on regressions for idioms compared to controls).

6.5.2 Familiarity

A final set of models was fitted to assess the effect of subjective familiarity. This was analysed separately as different rating sets were used for the English and Swedish native speakers (detailed below), therefore there is no straightforward way to explore this in an omnibus model. Separate models were created for English native speakers and Swedish native speakers, with the interaction between familiarity rating and type (idiom vs. literal phrase) computed for each measure.

For English native speakers the English and congruent categories were collapsed into one, and Swedish idioms were discounted on the grounds that they were all fundamentally unknown. English native speakers showed significant interactions between familiarity and phrase type for phrase level total reading time ($t = -3.32$, $p < .01$) and word level regression path duration ($t = -2.53$, $p < .05$); in both cases greater familiarity led to shorter reading and re-reading times for English idioms. No early measures showed any effect of familiarity.

For Swedish native speakers the effects on each condition were considered separately; for congruent items both Swedish ratings and English ratings of familiarity were considered. Swedish only items showed no effects for early measures but there was a significant interaction between phrase type and familiarity for phrase level total reading time ($t = -1.97$, $p < .05$), a marginal interaction with word level total reading time ($t = -1.74$, $p = .08$), and a significant interaction with regression path duration ($t = -2.10$, $p < .05$). Familiarity with the L1 idiom, therefore, leads to less time being spent on the English translation for late measures, suggesting that the meaning could be more easily understood the better the idiom was known in the L1.

For congruent items there were no effects of English familiarity rating on any measure, however for the Swedish familiarity ratings there were marginal or significant interactions with phrase type for phrase level total reading time ($t = -1.86$, $p = .06$), word level total reading time ($t = -1.99$, $p < .05$) and regression path duration ($t = -1.89$, $p = .06$). Congruent items were therefore affected positively by L1 familiarity for late measures (increased familiarity was facilitatory), just like Swedish only items, but showed no evidence that specific L2 familiarity was important. No effect on early measures for either set of ratings was demonstrated.

For English only items there were no effects of familiarity on early measures, however there were significant interactions between phrase type and familiarity for phrase level total reading time (t = -3.58, p < .001), likelihood of skipping the final word (z = 2.57, p < .01), word level total reading time (t = -3.23, p < .001) and regression path duration (t = -3.98, p < .001). For items that only exist in the L2, therefore, specific L2 familiarity is a strong predictor of how easily the idiom will be understood, and also whether the final word is predictable enough to be skipped (whether the form of the idiom is known).

Overall, familiarity showed significant effects in late measures only (with the exception of skipping rates for Swedish native speakers reading English only idioms). For native speakers this suggests that better known idioms were more easily understood, but this was not reflected in the automatic activation of known lexical combinations (no effect for early measures). For non-native speakers, L1 familiarity is used whenever possible (Swedish only and congruent phrases) to aid understanding of the phrase level meaning. When no L1 knowledge is available, specific L2 familiarity can also be utilised. This seems logical, given that knowledge of the correct form of an idiom and knowledge of the overall meaning may often not be equivalent. That is, knowing the meaning of an idiom is multi-faceted, and a speaker may have some idea about the general meaning without knowing how the idiom might be specifically deployed in context, for example how to use it in a pragmatically appropriate way, therefore greater familiarity should impact understanding in an incremental fashion. In many ways this mirrors single word knowledge, where a distinction can be made between form learning aspects and meaning aspects (Ellis, 1994; Nation, 2013). As Nation (2013, p.73) outlines, learners can have knowledge of a word form – they can recognise it as a legal word – without necessarily connecting it to an underlying concept. Laufer and Goldstein (2004) also demonstrated a hierarchy of vocabulary knowledge, from passive recognition to active recall, hence learners can demonstrate a high degree of form recognition without necessarily having a strong link to an underlying meaning. Certainly the range of knowledge that contributes to the meaning of a word (polysemy, connotations, collocations, constraints, grammatical functions) is in most cases much broader than the form knowledge required, which is largely restricted to orthography and phonology, and this can be seen as a much shallower level of word knowledge than that required to fully grasp the meaning. If idioms are ‘meaning units’ that behave like single choices, it seems logical that the same should apply to them.

6.5.3 Component parts

I next analysed the results to compare component words of idioms used separately in sentences with control items. I conducted a word level analysis of the second word in each item. Results are summarised in Table 6.6 and omnibus analysis is provided in Appendix 3b (Table 4).

Table 6.6. Results for all speakers for component parts, split by participant group.

	Swedish only		Congruent		English only	
	Idioms	Controls	Idioms	Controls	Idioms	Controls
Swedish native speakers						
Likelihood of skipping	.08 (.27)	.05 (.23)	.10 (.31)	.08 (.28)	.03 (.16)	.06 (.24)
First fixation duration	219 (94)	220 (83)	215 (108)	230 (108)	230 (88)	223 (97)
First pass reading time	253 (130)	253 (125)	235 (135)	253 (122)	253 (114)	249 (124)
Total reading time	385 (259)	385 (268)	343 (247)	397 (279)	381 (233)	389 (313)
Fixation count	1.7 (1.1)	1.7 (1.1)	1.5 (1.0)	1.7 (1.2)	1.7 (1.0)	1.6 (1.2)
Regression path duration	511 (397)	498 (393)	469 (413)	544 (460)	501 (362)	514 (464)
English native speakers						
Likelihood of skipping	.21 (.41)	.20 (.40)	.22 (.42)	.26 (.44)	.26 (.44)	.20 (.40)
First fixation duration	165 (100)	167 (101)	159 (99)	145 (98)	154 (113)	164 (99)
First pass reading time	174 (111)	173 (114)	160 (101)	150 (105)	160 (121)	169 (107)
Total reading time	214 (162)	222 (204)	197 (166)	177 (146)	194 (171)	211 (158)
Fixation count	1.1 (0.8)	1.1 (0.9)	1.0 (0.7)	0.9 (0.7)	0.9 (0.7)	1.0 (0.7)
Regression path duration	255 (225)	255 (270)	228 (227)	203 (179)	237 (240)	247 (210)

Note: For duration measures reading times in ms are reported; fixation count is a raw value; likelihood of skipping is reported as a probability (SDs in brackets)

Results for the component parts are inconclusive. For non-native speakers, there was no real indication of any facilitation for idiom words vs. controls words for any condition. The only measures that differed were total reading time and regression path duration, both of which were shorter for congruent idiom words than control words. Comparison with the native speakers showed the unexpected pattern that English only but not congruent idiom words were facilitated. Since there is no principled reason to distinguish these two conditions for English native speakers, it is more logical to consider them as a single class, in which case no differences were apparent. Overall, the paucity of significant results from the omnibus analysis suggests that there are no notable patterns according to whether the component words were part of an idiom or not.

6.6 Discussion

The non-native speakers in this study showed a consistent advantage when reading idioms compared to novel, literal control phrases. This was true for L2 only idioms, idioms that exist in both L1 and L2, and L1 only idioms, which by definition should not be familiar in their translated forms. In all conditions, late measures (phrase level total reading time and regression path durations) confirm that non-native speakers had no difficulty understanding the meaning of these phrases and in general spent less time on the idioms than the literal phrases. This was also partially supported in early measures for the final words (likelihood of skipping), where Swedish and congruent items but not English items showed an advantage. I interpret this as evidence that these 'known' combinations were being automatically triggered in such a way that lexical access for the final word was significantly quicker. For English only idioms, despite the relative ease with which they were understood, no such automatic boost was observed, suggesting that the lexical combinations were perhaps not as well entrenched in the mental lexicon, even though the figurative meanings were accessible/understandable. In contrast, English native speakers performed exactly as predicted on English idioms, showing facilitation for the form (through early measures) and meaning (through late measures) of idioms compared to comparably plausible literal phrases. However, when faced with unfamiliar idioms (translated Swedish forms) they showed considerable disruption in all late measures, suggesting that they had to spend more time reading and re-reading the idioms in an attempt to work out the meaning.

The implications for bilingual processing of formulaic language are extremely interesting. The non-native participants in this study were all at a very high level of proficiency; they rated the English idioms as familiar (average of around 5/7, considering English only and congruent items together), and this seems to have been borne out in their performance during the eye-tracking study, at least in terms of how well they understood the figurative meanings. Clearly, then, there is nothing fundamentally stopping non-native speakers from instantiating idioms in the mental lexicon in a way that enables them to process them in a comparable way to native speakers. Equally clear, however, is that the exposure and level of proficiency necessary for this to happen is high: even for the advanced speakers seen in this study the advantage was modest, and was not really evident in the most automatic lexical access measures (skipping rates and early measures for the final words) for the English only idioms. Although the effects for English only idioms were not as clear cut as for the English native speakers processing familiar phrases, direct comparison between the 'non-native' conditions for the two groups is telling: English native speakers reading 'L2' idioms (translated from Swedish) showed clear disruption in later measures, whereas Swedish native speakers reading English idioms showed none of the same difficulty in processing L2 non-compositional sequences, demonstrating that they were perfectly capable of understanding the figurative meaning of these in an online fashion during natural reading.

Of potentially greater interest is the clear finding that non-native speakers did treat L1 idioms like formulaic units when these were encountered in the L2. This was true for congruent items, which conceivably could have been encountered in English as well as Swedish, but also for the Swedish only items where this cannot be the case: the only source of knowledge about these configurations is that the same words go together in the L1, and it is highly unlikely that any of these combinations would ever have been encountered by the non-native speakers in English. Importantly, despite the unfamiliar form of these translations, there is a clear advantage for idioms vs. literal controls, especially in terms of the ease with which these were understood in the overall context of the sentence. Non-native speakers therefore had no difficulty in understanding the phrase level meaning of these items (as shown via the late measures), and show some evidence that the expected word was being predicted, even in the 'wrong' language (higher skipping rates for idiom final words in the Swedish and congruent conditions). Importantly, this was the case despite the fact that no biasing context was provided, and despite the fact that all idioms were short, hence there was no unequivocal recognition point until the final word had been read. This strongly implicates a language non-selective view of bilingual processing (Dijkstra, 2007), whereby at the earliest stages lexical activation is not specific to either language, but a known combination of lexical items can be triggered and accessed by encountering the forms in either L1 or L2.

There is also no clear evidence that congruency has any additional facilitative effect over and above those items that exist only in the L1. Titone et al. (2015) suggested that their results (less disruption during code-switching of idioms when the items were congruent) provide evidence for the representation of idioms in both languages. The present results dispute this, since there is no evidence that congruent items were treated any differently to Swedish only items, but both were facilitated relative to incongruent items (English only idioms). L1 knowledge appears to be the main driver of this effect, irrespective of whether the item is also 'known' in the L2. The effect of relative familiarity is important here. For both Swedish only and congruent items, increased familiarity with the Swedish version of the phrase showed a facilitatory effect for idioms in late measures. Thus, items that were better known in the L1 were more easily understood when encountered in the L2. Crucially, the congruent items showed no evidence that familiarity with the English form had any effect, which strongly implicates L1 knowledge over and above direct experience in the L2 in how these items were treated. In other words, whether or not these items were also known in the L2, it was the specific familiarity with the L1 version that determined how easily they were understood. In the case of English only idioms, where no L1 knowledge exists to aid with either the form or meaning of the idioms, experience directly in the L2 can be utilised and shows a facilitatory effect. This again strongly suggests that non-native speakers can develop ‘native-like’ formulaic performance in the L2, in line with various studies that have shown this to be the case at high levels of proficiency (Gyllstad & Wolter, in press; Isobe, 2010; Yamashita & Jiang, 2007; Yeganehjoo and Thai 2012), but whenever possible this is superseded by existing L1 knowledge.

On the question of why L1 knowledge should show such a strong influence, an increasing body of evidence suggests that when bilinguals process language in their L2, they demonstrate ballistic activation (Phillips, Segalowitz, O’Brien & Yamasaki, 2004). That is, they obligatorily activate the equivalent words in their native language (Thierry & Wu, 2007; Wu, Cristino, Leek & Thierry, 2013; Wu & Thierry, 2010; Zhang, van Heuven & Conklin, 2011). Assuming that this is the case, it is easy to see why both congruent and translated forms might show privileged processing in the same way as we see for native speakers encountering L1 forms. Reading the first word of an idiom will automatically trigger the L1 equivalent (e.g. *break* → *bryta*). If the L1 idiom is known and has its own idiom entry, this entry will be activated, leading to language non-selective activation of the whole unit (*bryta isen*), which in turn will provide a boost in lexical access to the expected word (*isen/ice*). For the control phrase, encountering *crack* will also trigger the L1 equivalent (*knäcka*), but since *knäcka isen* is not an idiom in either language, no ‘whole form’ entry or association between the two words exists, therefore, there is no reason for *isen/ice* to be activated over and above any other plausible continuation. Under this view, both Swedish only and congruent items should activate L1 equivalents, leading to facilitation. English only idioms have no L1 equivalents, therefore no strong and well-established idiom entries will be triggered, but experience in the L2 will have developed entries for at least some idioms (presumably the most frequent/common ones), leading to the modest level of facilitation seen in the results, and the clear effect of specific L2 familiarity in this condition.

An alternative lemma-based view is conceptually very similar, but does not rely so heavily on the idea that idiom entries exist as unitary concepts. In line with the view put forward by Jiang (2000), Ueno (2009) and Wolter and Yamashita (2014), lemma level information may be copied over from the L1 when a word is learned in the L2, hence connections between words in the L1 (semantic links but also associative and experientially derived connections – contrary to the results of Williams and Cheung, 2011) may be automatically created between equivalent forms in the L2. Alternatively, lemma-level information may be language non-specific, with information such as semantic and associative networks being tied to the conceptual values of words rather than a language specific form. This may also explain why for congruent items there is an effect of the well-established L1 familiarity over and above any effect of specific L2 familiarity, as this is likely to be much more strongly represented and more strongly linked to the underlying concept/lemma. One way to test this might be to perform this study in reverse by translating the English items into Swedish to see how the non-native speakers process them. If lemmas are language non-specific then there should be some level of facilitation for translated English items in Swedish, while Swedish native speakers with no knowledge of English should show the same pattern as the native English speakers in the present study: considerable disruption reading non-compositional idioms compared to literal control phrases.

One argument against a lemma-based view is the lack of any clear results for the component parts of idioms. If the idiom priming effect, seen so consistently in the literature, is the result of intralexical priming between lemmas, then we would expect to see some degree of activation for idiom components, just as Camblin et al. (2007) and Carroll and Slowiaczek (1986) did for semantically related words. The fact that there was no such pattern is not conclusive proof against this view, and the strong effects of familiarity for late measures but not early measures in the formulaic configurations might lend some support to the idea that idioms are best conceptualised as distributed representations of single words (Holsinger, 2013). For both speaker groups there was a sliding scale of familiarity for late measures (which I take to reflect meaning access) but not early measures (reflecting formal/lexical recognition), suggesting that these two aspects of idiom representation might not necessarily be equivalent. A more nuanced way of exploring the nature of idioms in the mental lexicon may be required to shed further light on this.

In summary, these results show clear support for L1 influence on the processing of idioms by non-native speakers. Importantly, as well as being evident in offline tasks as shown in previous research (Bulut and Çelik-Yazici, 2004; Charteris-Black, 2002; Irujo, 1986; Laufer, 2000; Liontas, 2000), this study suggests that this knowledge is used in an online fashion, facilitating lexical access and semantic integration for known combinations from the very earliest stages of processing. The fact that this is true whether or not the combination also exists in the L2 is crucial since it prioritises L1 knowledge directly, rather than fitting a 'confirmatory' account whereby L2 idioms have been encountered and mentally registered as transferrable in the minds of individual learners, or where congruent idioms are dually represented in both the L1 and L2 lexicons.

Chapter 7. No Word is an Island

The three empirical chapters so far have concentrated on translation as a way of investigating how idioms are recognised and processed. The evidence is clear that the specific form of an idiom is not necessarily the only driver of fast processing. That is, even when the form is changed via translation, ‘known’ sequences are facilitated. Whilst it is possible that this process is entirely mediated by automatic translation and ballistic L1 activation, there is also the strong possibility that the conceptual unity of idioms is at least partly responsible for their faster processing. Hence language-independent concepts allow for language-specific word forms to be activated in specific configurations.

One way to test this is to extend the discussion to other forms of formulaic language that do not have the same degree of underlying conceptual unity, and this is the intention in Chapter 7. Although idioms are ‘prototypical’ in formulaic terms, it is possible that their specific properties (non-compositionality, conceptual unity) mean that their behaviour is not representative of all formulaic types. The following study therefore includes compositional, frequency-defined formulaic units in order to compare patterns of behaviour, to see whether faster processing is a feature of all formulaic types. It also extends the idea raised in Chapter 6 that formulaicity is a property of inter-lexical priming and spreading activation, rather than reflecting dedicated ‘whole form’ storage. That is, in the same way that I have so far tried to remove any formulaic advantage by altering the form using translation, I now investigate what happens to schematically related formulaic ‘partners’ when they are used in proximal but non-formulaic contexts, i.e. when the component words are used outside of their canonical formulaic frames. This study is currently under review as

"No Word is an Island: Exploring the mechanisms of formulaic language processing" (Carrol & Conklin, under review).

7.1 Introduction

There is mounting evidence that formulaic language is stored 'holistically' in the mental lexicon and that it has a privileged processing status, but further research is required to explore how the individual component words of different types of formulaic language are wired together in such a cohesive manner. To address this I conducted two studies. In Experiment 5, I compare different types of formulaic sequence (idioms, binomials and collocations) to establish whether such items demonstrate a similar processing advantage (relative to novel control phrases) during natural reading. This will allow me to explicitly test how different types of formulaic phrase are processed and assess the factors that contribute to this, which is important since very little experimental evidence currently exists to compare different types of formulaic language in this way. In Experiment 6, I explore the individual component words of the same formulaic units to see whether they show cohesion in non-formulaic configurations. This will provide valuable insight into the relationships between the component words in different types of formulaic language and assess whether they are fundamentally different in nature. Thus, this research addresses a fundamental question about whether the processing advantage found for formulaic language is indicative of 'holistic' storage, or simply a reflection of multiple schematic relationships among frequently co-encountered words.

7.2 Formulaic language and the mental lexicon: evidence for 'holistic' storage

Formulaic language varies along multiple continua, including frequency, semantic transparency, compositionality, fixedness/flexibility and literal plausibility. Idioms are seen as prototypical examples within the broader class of formulaic language (Cacciari, 2014; Titone et al., 2015), since they vary along all of the dimensions listed above. Studies using a variety of methodologies have shown that highly familiar idioms are processed more quickly than non-idiomatic control phrases or less familiar idioms (Chapters 4, 5 and 6; Conklin & Schmitt, 2008; Libben & Titone, 2008; Rommers, Dijkstra & Bastiaansen, 2013; Schweigert, 1986, 1991; Schweigert & Moates, 1988; Siyanova-Chanturia, Conklin & Schmitt, 2011; Swinney & Cutler, 1979; Tabossi, Fanari & Wolf, 2009). Familiarity with an idiom therefore allows for the "initial retrieval of idiomatic configurations as lexicalised units" (Libben & Titone, 2008, p.1117), leading to facilitation for the expected form and the overall figurative meaning. Many factors influence the degree of literal/figurative activation and the predictability of an idiom, including preceding context (Cieślicka, 2012; Colombo, 1993, 1998; Titone & Connine, 1999), ambiguity (Mueller & Gibbs, 1987) and literal plausibility (Cronk & Schweigert, 1992; Schweigert, 1991).

There is disagreement over whether this privileged processing is restricted to idioms, or whether it is a feature of all subtypes of formulaic language. We can broadly distinguish formulaic language that conceptually seems to represent 'single choices' from those that are defined by a high degree of frequency and co-occurrence rather than any unitary conceptual properties or semantic idiomaticity. In the first class, alongside idioms we can consider phrasal verbs (e.g. *get into* (an argument)) and spaced compounds (e.g. *teddy bear*). Phrasal verbs display properties of single words and longer syntactic units (Blais & Gonnerman, 2013; Cappelle, Shtyrov & Pulvermüller, 2010; Konopka & Bock, 2009), leading to faster processing for conventionalised figurative forms (Matlock & Heredia, 2002; Paulmann, Ghareeb-Ali & Felser, 2015). Similarly, spaced compounds demonstrate aspects of unitary semantics and phrasal syntax (De Cat, Klepousniotou & Baayen, 2015), and Cutter, Drieghe and Liversedge (2014) presented compelling eye-tracking evidence to suggest that spaced compounds are processed as part of a single lexical unit (see Chapter 3 for a more detailed discussion of this study). For any such conceptually 'whole' items, the semantic unity might plausibly explain their faster processing.

In contrast, 'statistical idioms' (Baldwin & Nan Kim, 2010) are not semantic wholes. This class includes lexical bundles, chunks, clichés, non-idiomatic collocations and literal binomials. Binomials (noun-and-noun sequences with a highly conventional order, e.g. *king and queen*) in particular have been shown to have an advantage compared to novel combinations or reversed forms. This is true for both irreversible (idiomatic) binomials (Arcara et al., 2012) and literal binomials (Siyanova-Chanturia, Conklin & van Heuven, 2011), both of which seem to be treated more like lexical items than compositional sequences. Highly frequent sentence fragments or lexical 'chunks' also show faster processing than less frequent combinations (Arnon & Cohen-Priva, 2013; Arnon & Snider, 2010; Bannard & Matthews, 2008; Ellis, Simpson-Vlach & Maynard, 2008; Tremblay & Baayen, 2010; Tremblay et al., 2011), e.g. a sequence like *don't have to worry* shows an overall processing advantage relative to the substring matched but overall less frequent *don't have to wait* (Arnon & Snider, 2010). Fast processing is therefore not restricted to semantically or syntactically complete phrases, but can be a property of any recurrent sequence of words.

Despite the evidence relating to specific formulaic classes, relatively little evidence exists to directly compare different subtypes in terms of their key features. Jolsvai, McCauley and Christiansen (2013) showed that idioms were processed more quickly than equally frequent compositional phrases or sentence fragments, but other studies have shown that it is simply the familiar form of known sequences that leads to fast recognition (Tabossi et al., 2009), regardless of semantic properties. Columbus (2010) compared reading time for idioms, lexical bundles and restricted collocations; all three types were read more quickly than non-formulaic controls, and idioms were processed the most quickly overall. She concluded that these differences may not be the result of the different subtypes per se, but that different variables relevant to each type produce different effects. The present study also seeks to address this, and explore different variables in formulaic language processing. Gyllstad and Wolter (in press) showed that degree of semantic transparency and phrase frequency contributed to speed of processing, with entirely literal items (‘free combinations’) showing shorter response times than partially transparent word pairs. Therefore, semantic and statistical properties appear to contribute to how formulaic units are processed, but there remains a significant lack of direct comparisons of distinct formulaic subtypes in the literature.

Other compelling evidence for the ‘unitary’ nature of formulaic language comes from ERP studies. Compared to literal sentences, idioms show reduced N400 effects alongside evidence of a P300 (which is often taken as reflection of matching a stimulus with a stored template) or other evidence of early form recognition/mismatch

(Liu, Li, Shu, Zhang & Chen, 2010; Vespignani, Canal, Molinaro, Fonda & Cacciari, 2010; Rommers et al., 2013; Zhang, Yang, Gu & Ji, 2013; Zhou, Zhou & Chen, 2004). Similar results have been found for literal collocations (Molinaro & Carreiras, 2010) and binomials (Siyanova-Chanturia, 2010), as well as other types of categorical prediction, such as antonyms (Roehm, Bornkessel-Schlesewsky, Rösler & Schlesewsky, 2007). Molinaro and Carreiras (2010) also found evidence of increased N400 effects for figurative collocations, which they suggested reflect the integration of an overall semantic meaning (in line with Coulson and Van Petten, 2002). Similar processes of recognition are therefore reported for all 'known' combinations, but aspects of later semantic processing vary according to how conceptually 'whole' the units are considered to be.

7.3 Summary

There is a clear body of evidence that frequently occurring or familiar word combinations enjoy faster and even qualitatively different processing than infrequent, non-formulaic combinations. Importantly, formulaic language exists as a continuum, with a number of broad taxonomic distinctions that may or may not also exist as discrete psychological categories. While it has been said to be processed quickly because it is 'known' (Van Lancker Sidtis, 2012) and therefore highly predictable (Cacciari, 2014; Vespignani et al. 2010), questions remain about whether the evidence discussed so far (in this chapter and in the preceding studies) is an indication of 'holistic' processing and storage, or some other mechanism. Wray (2012) suggests that several explanations for fast formulaic processing are plausible within a broadly usage-based framework (Bybee, 2006; Bybee & McClelland, 2005; Tomasello, 2003). It could entail the simultaneous activation of multiple component words, it could represent the mapping of overall conceptual meaning to larger base components, or it could, in the case of those items that are truly frozen, entail a bypassing of the componential route altogether (Wray, 2012).

A key aim of this chapter is to explore precisely what drives this cohesion in different phrase types, i.e. those that differ in terms of their semantic and statistical properties. I also aim to test whether the fast processing of formulaic items is contingent on components appearing in their canonical form. That is, do the individual words of formulaic units show any evidence of cohesion when they are encountered in non-formulaic configurations, and is this consistent for different formulaic types?

7.4 Experiment 5

Experiment 5 investigates the processing of formulaic phrases where there is no ‘recognition point’ until the whole phrase has been seen. The formulaic phrases are presented in neutral context sentences, which minimises any discourse level or strategic influence on the predictability of each phrase. Any advantage must therefore be a direct result of the ‘formulaic’ nature of the phrase. Crucially, the nature of this formulaicity is different for idioms and compositional units, so Experiment 5 provides a way to directly compare subtypes with different characteristics. I investigated the processing of formulaic sequences in their canonical forms, compared to equally plausible, matched control phrases. I compared idioms (*spill the beans*) and two types of literal/compositional units: binomials (*king and queen*) and collocations (*abject poverty*).

7.4.1 Participants

Twenty-four undergraduate students (all native speakers of English) participated in Experiment 5 for course credit.

7.4.2 Materials

Items of each subtype were of the same form. Each had two main lexical components and no unequivocal recognition point until the end of the phrase. Idioms conformed to the structure verb-X-noun, where X was a determiner or personal pronoun, or preposition-det-noun. Items were chosen from the Oxford Learner's Dictionary of Idioms (Warren, 1994) and various published idiom studies. Items were normed for familiarity by native English speakers (n = 21) using a seven-point Likert scale and a final list of 45 items was selected (mean familiarity = 5.9/7). These items were further normed by native speakers (n = 19) for compositionality using a literal paraphrase (e.g. If you make peace with someone you *bury the hatchet*), and judged on a seven-point Likert scale (mean = 4.1/7).

Binomials were all of the form noun-and-noun or verb-and-verb. Items were collected from online lists and previous published studies. Items were considered as binomials if the ratio of forward to backward occurrence was greater than 4:1, and if phrase frequency in the British National Corpus (BNC) was greater than 20 (per 100 million words). To ensure that only literal/compositional binomials were used, all items were normed on a population of native speakers for idiomaticity/literalness (n = 25) and reversibility (n = 23). There was a high correlation between reversibility and literalness ($r = .64$, $p < .01$), demonstrating that literal items are more reversible, while idiomatic items tend to be irreversible (Lohmann, 2012; Mollin, 2012). A final set of 45 items was created (mean ratio of forward/reversed forms = 9:1; mean literalness = 2.7/3; mean reversibility = 4.6/7).

Collocations were defined as non-idiomatic frequently co-occurring word pairs (either noun-noun or adjective-noun). A list of the most frequent two-word combinations was extracted from the BNC and from this I selected candidate items that I considered to be common collocations and obtained phrase frequency and mutual information (MI) scores from the BNC.[22] I adopted a minimum threshold of 10 occurrences in the BNC for phrase frequency and 2.9 for MI score.

The collocations were further classified as being semantic associates or non-associates, according to the same rationale as Durrant and Doherty (2010), who found automatic priming only for collocating word pairs that were also semantically associated.[23] I classified the collocations using scores obtained from the Edinburgh Associative Thesaurus (EAT: Kiss, Armstrong, Milroy & Piper, 1973), an online database of free association norms. A pair was considered to be unassociated if the first word of the collocation returned an association score of 0 for the second word, and associated if the score was greater than 0. Final lists of 45 associated and 45 unassociated collocations were selected based on these criteria.

[22] Mutual information is a measure of collocation strength. It compares the observed number of occurrences of a word pair with the expected number of co-occurrences based on the individual word frequencies and the size of the corpus. In this study the formula for MI was: log(observed/expected).

[23] Durrant and Doherty (2010) used a primed lexical decision task to test whether seeing the first word of a collocating word pair speeded up responses to the second word. In their first study, an unmasked prime of 600ms produced a facilitative effect for high frequency collocations, regardless of any semantic link between words. In their second study, when a masked prime of 60ms duration was used, only those items that were high frequency collocations *and* semantic associates produced facilitation.

For all items I created two control phrases, matched with the formulaic phrases for length and single word frequency. For control type 1 phrases, the first word was changed (e.g. *spill the beans* became *drop the beans*) and for control type 2 the second word was changed (e.g. *spill the beans* became *spill the chips*). I collected association scores between component words for all items and calculated a measurement of phrase completion likelihood, based on the percentage likelihood that seeing the first part of the phrase would lead to the formulaic completion based on BNC counts (all frequencies per 100 million words, e.g. for *spill the beans*, overall phrase frequency = 39, frequency of *spill the* = 93, therefore 39/93*100 = 42%).

For all items a short context sentence was created (see Table 7.1). Contexts immediately preceding the unit were as neutral as possible and were created so that all three variants of any item would form a logical continuation. The immediate post-context was also the same across conditions, then the final part of the sentence was tailored so that each version was completed in a coherent manner. Mean number of words preceding the phrase was 3.8 and following the phrase was 11.6. All context sentences were normed by native speakers (n = 25) to ensure that they were equally plausible as English sentences. There were no differences between formulaic units and either control type (formulaic units, mean = 4.5; control type 1, mean = 4.4; control type 2, mean = 4.5; one way ANOVA by condition: $F = 0.91$, $p = .40$).

Table 7.1. Example context sentences for idioms and control phrases, Experiment 5

	Pre-context	Phrase	Post-context
Formulaic	It was hard not to	spill the beans	when I heard such a juicy piece of gossip.
Control type 1	It was hard not to	drop the beans	when I burned my hand on the hot pan.
Control type 2	It was hard not to	spill the chips	when I stumbled on my way out of the kitchen.

Finally, for all items cloze probability scores were collected via a fill in the blank task asking participants to provide the most likely continuation for each phrase. Native speakers (n = 69, spread across four lists so that no-one saw any item in more than one condition) were presented with the first part of the context sentence for formulaic units and control items (e.g. *It was hard not to spill the*…) and asked to provide the first word that came to mind that could plausibly continue the sentence. It was stressed that these were sentence fragments, and that the word did not have to complete the sentence, simply to continue it in a reasonable way. Cloze probability was calculated as the percentage of participants who provided the intended completion in each case. Overall, for each formulaic unit type there were 45 items, counterbalanced over three lists so that no participant saw the same item in more than one condition. A summary of item characteristics for the stimuli is shown in Table 7.2 and all stimulus items are provided in Appendix 4a.

Table 7.2. Example phrases and item characteristics for all stimuli, Experiment 5.

	Phrase	Phrase Fr	%	Ass	Cloze	MI
Idiom	*Spill the beans*	54	10.2	0.00	35%	n/a
Control type 1	*Drop the beans*	12	0.5	0.00	3%	n/a
Control type 2	*Spill the chips*	11	0.9	0.00	4%	n/a
Binomial	*King and queen*	253	28.2	0.30	68%	n/a
Control type 1	*Prince and queen*	5	1.0	0.04	13%	n/a
Control type 2	*King and prince*	8	0.9	0.01	4%	n/a
Collocations						
Associated	*Abject poverty*	258	3.7	0.15	35%	7.9
Control type 1	*Awful poverty*	9	0.1	0.00	3%	2.2
Control type 2	*Abject misery*	5	0.1	0.00	1%	2.3
Unassociated	*Ancient history*	118	3.5	0.00	6%	6.8
Control type 1	*Recent history*	10	0.1	0.00	2%	1.8
Control type 2	*Ancient stories*	6	0.2	0.00	2%	1.8

Note: Phrase frequency is a raw value from the BNC (per 100 million words); % is the phrase continuation likelihood; Ass is the strength of association based on EAT scores; Cloze is the mean cloze probability; MI is the mutual information score for collocations.

7.4.3 Procedure

The experiment was administered on an Eyelink 1000+ eye-tracking system from SR Research. Participants were seated at a comfortable height approximately 60 cm away from a widescreen computer monitor (resolution 1920 x 1080, refresh rate 60hz). Their heads were stabilised using a table-mounted chin rest. Eye movements were recorded using a desk-mounted camera (sample rate 500hz). Recording was monocular and for all participants the left eye was tracked. Following initial setup a nine-point calibration and validation procedure was used to verify accuracy, repeated at regular intervals throughout the experiment.

Participants were asked to read each sentence for comprehension and to press the spacebar when they had finished. Each sentence was preceded by a fixation cross to allow for trial by trial drift correction. One third of the sentences were followed by a simple yes/no comprehension question to ensure that participants paid attention throughout. Participants saw a total of 180 sentences, with a short break and recalibration after every 60 items.

7.5 Results and analysis

All eye-tracking data was cleaned according to the standard procedure within the Eyelink Data Viewer program, so fixations below 100ms and above 800ms were removed. Data were also visually inspected and any trials where data was unusable or track loss had occurred, or where the whole phrase was skipped, were discounted. In total 4.6% of data were excluded.

I again applied a ‘hybrid’ analysis where I considered the units as a whole, but also specifically examined the reading patterns for the final word, which is assumed to be the locus of the formulaic advantage (Columbus, 2010). I considered the same early and late eye-tracking measures as in Chapter 6 (reproduced in Table 7.3).

Table 7.3. Eye-tracking measurements used in the analysis, with description and stage of processing

Stage of processing	Type of measure	Description
	Phrase level	
Early	First pass reading time	The sum duration of all fixations on the phrase the first time it is encountered in the sentence
Late	Total reading time	The sum duration of all fixations on the phrase during the trial (including re-reading)
	Total fixation count	The total number of fixations on the phrase during the trial
	Word level	
Early	Likelihood of skipping	The likelihood that a word is skipped (not fixated at all) during first pass reading
	First fixation duration	The duration of the first fixation on the word
	First pass reading time	The sum duration of all fixations on the word the first time it is encountered in the sentence
Late	Total reading time	The sum duration of all fixations on the word during the trial (including re-reading)
	Regression path duration	The duration of all regressions to material preceding the word once it has been fixated for the first time (including the prior context and the start of the phrase)

I compiled linear mixed effects models using the lme4 package (version 1.0-7, Bates, Maechler, Bolker, Walker, Christensen, Singmann & Dai, 2014) in R (version 3.1.2, R Development Core Team, 2014). Each eye-tracking measure was considered in its own model. The fixed effect of phrase type (formulaic/control) was treatment-coded so that formulaic was considered to be the baseline.[24] I included random intercepts for subject and item and by-subject random slopes for the effect of phrase type and condition (Barr, Levy, Scheepers & Tily, 2013). The covariates of word length, word frequency and phrase frequency (both log-transformed), phrase continuation

[24] Since there are two types of control item, formulaic items are considered to be the baseline to avoid making an arbitrary decision as to which control type should be used for comparison. This also means that each control type can be compared to the formulaic units in the same model.

likelihood and cloze probability were included where this significantly improved the fit of any model. Type-specific predictors were included in the models for each formulaic unit type, detailed below. Duration measures (in milliseconds) were log-transformed to reduce skewing. For the analysis of fixation count I used a generalized linear model with poisson regression, and for the binary variable likelihood of skipping I used a logistic linear model. Skipped items were discounted from any subsequent word level analysis. Table 7.4 shows a summary of results for all measures.

Table 7.4. Summary of all data, phrase level and word level eye-tracking measures.

	Phrase level			Word level				
	First pass reading time	Total reading time	Fixation count	Likelihood of skipping	First fixation duration	First pass reading time	Total reading time	Regression path duration
All types	339 (173)	538 (340)	2.7 (1.4)	.22 (.41)	152 (96)	157 (108)	202 (174)	290 (291)
Control1	343 (167)	554 (342)	2.8 (1.5)	.15 (.35)	174 (91)	182 (102)	235 (174)	363 (337)
Control2	331 (164)	538 (329)	2.7 (1.4)	.13 (.33)	176 (89)	185 (102)	250 (185)	380 (343)
Idioms	333 (134)	487 (266)	2.6 (1.2)	.31 (.46)	133 (101)	134 (104)	151 (125)	232 (250)
Control 1	388 (178)	596 (374)	3.0 (1.6)	.15 (.36)	173 (96)	177 (101)	206 (150)	346 (355)
Control 2	372 (164)	552 (322)	2.8 (1.4)	.13 (.34)	173 (89)	176 (94)	204 (124)	330 (313)
Binomials	326 (170)	479 (276)	2.4 (1.2)	.18 (.39)	160 (99)	163 (106)	196 (156)	281 (275)
Control 1	355 (186)	544 (314)	2.7 (1.4)	.14 (.35)	187 (94)	193 (100)	239 (166)	362 (295)
Control 2	354 (165)	552 (332)	2.8 (1.4)	.11 (.31)	184 (90)	195 (104)	243 (167)	369 (304)
Collocations:								
Associated	296 (142)	475 (322)	2.4 (1.4)	.20 (.40)	157 (94)	164 (108)	213 (183)	306 (305)
Control 1	316 (154)	502 (274)	2.6 (1.2)	.14 (.34)	168 (87)	180 (106)	234 (170)	330 (374)
Control 2	325 (159)	546 (314)	2.7 (1.4)	.14 (.35)	170 (85)	177 (94)	254 (200)	367 (315)
Unassociated	316 (168)	537 (338)	2.7 (1.5)	.17 (.37)	157 (89)	167 (109)	247 (209)	343 (321)
Control 1	327 (180)	610 (415)	3.0 (1.8)	.15 (.36)	169 (87)	178 (100)	261 (203)	416 (403)
Control 2	346 (190)	642 (418)	3.1 (1.7)	.12 .(33)	178 (92)	192 (114)	300 (223)	454 (417)

Note: All figures are mean values (standard deviation in brackets). Duration measures are expressed in milliseconds, fixation count is a raw figure and likelihood of skipping is expressed as a probability.

An omnibus model was fitted to explore the overall effect of formulaicity (Tables 7.5 and 7.6). Idioms were treatment-coded to be the baseline, since this is the subtype that is most often seen as ‘prototypically’ formulaic. Collocations (semantically related/unrelated) were combined into one group, with association strength included as a covariate in all models where it was significant.

Phrase level analysis showed no effects of phrase frequency or phrase continuation likelihood on any measure. For first pass reading time there was a significant effect of semantic association (t = -1.94, p =.05) and for total reading time there was an effect of cloze probability (t = -2.23, p <.05). For all measures there was a significant interaction between control type 2 phrases and the collocations, and for total reading time and total fixation count there were also interactions between collocations and control type 1 phrases.

Table 7.5. Omnibus mixed effects model estimates for all items, phrase level measures.

	First pass reading time			Total reading time			Fixation count		
	β	SE	t	β	SE	t	β	SE	z
Fixed effects:									
Intercept	5.64	0.06	88.19	5.96	0.09	64.55	0.80	0.09	8.90
Subtype : Binomial	-0.11	0.04	-3.02***	-0.10	0.06	-1.61	-0.10	0.06	-1.63
Subtype : Collocation	-0.28	0.04	-7.32***	-0.22	0.06	-3.74***	-0.20	0.06	-3.49***
Condition: Control 1	-0.03	0.03	-0.99	-0.05	0.04	-1.47	-0.04	0.04	-0.90
Condition: Control 2	-0.12	0.03	-3.63***	-0.14	0.04	-3.23***	-0.14	0.05	-2.98**
Binomial*Control 1	0.06	0.05	1.33	0.09	0.05	1.70^{+}	0.06	0.06	1.00
Binomial* Control 2	0.09	0.05	1.75	0.09	0.05	1.77^{+}	0.05	0.07	0.72
Collocation*Control 1	0.07	0.04	1.61	0.13	0.04	2.90**	0.11	0.06	2.01*
Collocation*Control 2	0.20	0.04	4.86***	0.28	0.04	5.53***	0.26	0.06	4.63***
Control predictors:									
Total length	0.02	0.00	5.08***	0.03	0.01	3.60***	0.03	0.01	3.40***
Phrase Fr (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Phrase %	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Cloze	n/a	n/a	n/a	-0.01	0.00	-2.23*	n/a	n/a	n/a
Ass forward	-0.15	0.08	-1.94*	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance		
Subject	0.044			0.035			0.021		
Item	0.006			0.059			0.040		
Subject | Binomial	0.000			0.000			0.000		
Subject | Collocation	0.007			0.005			0.000		
Subject | Control 1	0.000			0.001			0.000		
Subject | Control 2	0.002			0.005			0.000		
Residual	0.185			0.231			n/a		

*Note: Table displays coefficients (β) standard error (SE) and t-values (z-values for generalised linear models), with significance values estimated by the lmerTest package in R (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): *** p < .001, ** p < .01, * p< .05.*

Table 7.6. Omnibus mixed effects model estimates for all items, word level measures.

	Skipping rate			First fixation duration			First pass reading time			Total reading time			Regression path duration		
	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Fixed effects:															
Intercept	-0.59	0.32	-1.83	5.28	0.04	144.37	5.30	0.04	132.83	5.26	0.06	83.54	5.65	0.10	56.55
Type: Binomial	-0.82	0.24	-3.37***	0.02	0.03	0.69	0.04	0.03	1.13	0.01	0.05	0.26	-0.07	0.09	-0.83
Type: Collocation	-0.75	0.24	-3.17**	-0.01	0.03	-0.40	0.02	0.03	0.63	0.11	0.05	2.35*	0.06	0.07	0.82
Condition: Control 1	-0.76	0.23	-3.37***	0.02	0.03	0.66	0.02	0.03	0.55	0.06	0.04	1.47	0.04	0.05	0.74
Condition: Control 2	-1.02	0.26	-4.00***	-0.01	0.03	-0.26	-0.01	0.03	-0.18	0.05	0.04	1.18	-0.01	0.05	-0.14
Binomial*Control 1	0.83	0.31	2.70**	-0.01	0.03	-0.32	-0.02	0.04	-0.44	0.05	0.05	0.89	0.07	0.07	0.92
Binomial* Control 2	0.53	0.32	1.66	0.01	0.00	0.19	0.01	0.04	0.38	0.08	0.05	1.48	0.16	0.07	1.99*
Collocation*Control 1	0.72	0.26	2.79**	-0.04	0.03	-1.33	-0.04	0.03	-1.22	-0.01	0.05	-0.30	-0.04	0.06	-0.61
Collocation*Control2	0.83	0.27	3.06**	-0.01	0.03	-0.35	-0.01	0.03	-0.43	0.07	0.05	1.64	0.08	0.06	1.31
Control predictors:															
W2 Length	-0.16	0.05	-3.47***	n/a	n/a	n/a	n/a	n/a	n/a	0.02	0.01	1.70^{+}	0.03	0.01	1.93*
W2 Freq (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Phrase Freq (log)	0.11	0.05	2.43*	-0.02	0.00	-3.24**	-0.02	0.01	-3.74***	n/a	n/a	n/a	-0.03	0.01	-3.04**
Phrase %	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Cloze	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Ass forward	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.22	0.12	1.85^{+}
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.310			0.015			0.017			0.014			0.023		
Item	0.282			0.004			0.005			0.028			0.092		
Subject | Binomial	0.082			0.003			0.003			0.005			0.006		
Subject | Collocation	0.404			0.003			0.004			0.004			0.008		
Subject | Control 1	0.055			0.001			0.002			0.000			0.002		
Subject | Control 2	0.267			0.001			0.001			0.001			0.002		
Residual	n/a			0.073			0.091			0.185			0.290		

*Note: Table displays coefficients (β) standard error (SE) and t-values (z-values for generalised linear models), with significance values estimated by the lmerTest package in R (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): *** p < .001, ** p < .01, * p< .05.*

At the word level, likelihood of skipping in idioms was significantly more likely than in all other subtypes and conditions. Phrase frequency was also a significant predictor of skipping rates ($t = 2.43$, $p < .05$). Other early measures showed significant effects of phrase frequency but not of subtype or condition. Total reading time showed no effects of phrase frequency and no interactions between type and condition. Regression path duration was significantly affected by phrase frequency and association strength.

In sum, idioms showed a markedly higher likelihood of the final word being skipped, which suggests that they are somehow more predictable than any of the other types. Collocations behaved differently to idioms and binomials, although it remains to be seen whether the associated and unassociated collocations demonstrate different patterns since association strength showed no consistent effects in the omnibus models. The results suggest that there may indeed be a reason to distinguish formulaic subtypes, so I constructed separate models to analyse each subtype in more detail. For the following analysis I summarise significant results, and individual model outputs are provided in Appendix 4b.

7.5.1 Idioms

For idioms I examined the effects of condition (idiom vs. control 1 vs. control 2), the covariates listed above, and the idiom specific metrics of subjective familiarity and compositionality (both centred).

At the phrase level, phrase frequency was a significant predictor of first pass reading ($t = -5.11$, $p < .001$) and total reading times ($t = -3.92$, $p < .001$). Fixation count showed no effects of phrasal frequency or any other predictor variables, but there were significant differences between idioms and both control type 1 phrases ($t = 3.06$, $p < .01$) and control type 2 phrases ($t = 2.10$, $p < .05$).

At the word level, cloze probability ($z = 1.81$, $p = .07$) was a marginally significant predictor of skipping rate, and there were significant effects of condition for control type 1 ($z = -2.12$, $p < .05$) and control type 2 phrases ($z = -2.98$, $p < .01$). Phrase frequency was not a significant predictor of skipping rates. Early durational measures for the non-skipped items showed effects of phrase frequency, cloze probability and also individual familiarity; in each case inclusion of these in the model meant that any between group differences were not significant. For later duration measures, total reading time showed an effect only of compositionality ($t = -2.74$, $p < .01$), and regression path duration showed an effect only of phrase frequency ($t = -1.93$, $p < .05$).

7.5.2 Binomials

For binomials I compared effects of condition (binomial vs. control 1 vs. control 2) and the covariates listed above. I also included the reversibility ratio of each item (as a measure of how fixed the binomial was) and association strength; neither of these was a significant predictor for any measures.

At the phrase level binomials showed significant effects of cloze probability for first pass reading time ($t = -2.32$, $p < .05$) and a significant effect of phrase frequency on total reading time ($t = -2.10$, $p < .05$). Fixation count showed significant differences according to condition (control type1: $z = 2.20$, $p < .05$; control type 2: $z = 2.71$, $p < .01$).

At the word level, skipping rate was marginally higher for binomials than control type 1 phrases ($z = -1.74$, $p = .08$) and significantly higher than control type 2 phrases ($z = -3.01$, $p < .01$), but showed no effects of cloze probability or phrase frequency. Early duration measures showed significant effects of cloze probability (first fixation duration: $t = -2.30$, $p < .05$; first pass reading time: $t = -2.40$, $p < .05$). Total reading time showed a main effect of condition (control type 1: $t = 2.41$, $p < .05$; control type 2: $t = 3.32$, $p < .01$) and regression path duration showed an effect of phrase frequency ($t = 2.00$, $p < .05$).

7.5.3 Collocations

I considered the collocations as a whole and included whether or not the items were also associates as a fixed effect in the models. As well as the covariates listed above, I included MI score and association strength. Neither phrase continuation likelihood nor cloze probability showed significant effects on any measure.

At the phrase level there were main effects of type (associated collocations were consistently read more quickly than unassociated items), but there was no interaction between type and condition. Collocations overall showed shorter first pass reading time ($t = 2.50$, $p < .05$), total reading time ($t = 2.30$, $p < .05$) and fewer fixations ($t = 2.52$, $p < .05$), compared to control type 2 phrases. Compared to control type 1 phrases, collocations only showed shorter first pass reading times ($t = 2.40$, $p < .05$) and a marginal difference for total reading time ($t = 1.71$, $p = .09$), although both phrase frequency ($t = -1.99$, $p < .05$) and association strength ($t = 2.24$, $p < .05$) were significant predictors.

At the word level, skipping rates were significantly higher for collocations compared to control type 1 phrases (z = -2.06, p < .05) and control type 2 phrases (z = -2.15, p < .05). There was no effect of association type, and no interactions between type and condition. For subsequent duration measures there was a distinction between early and late measures. Phrase frequency was a significant predictor for first fixation duration (t =-2.53, p < .05) and first pass reading time (t = -2.24, p < .05), while MI score was significant for total reading time (t = -2.14, p < .05) and regression path duration (t = -2.51, p < .05).

7.6 Discussion

The results show an overall pattern of facilitation for all formulaic subtypes such that the whole phrase was read more quickly when it was a formulaic unit than a control item. The locus of the advantage in all cases was the final word. This is most evident in idioms, where the final word was skipped almost one third of the time, despite there being no overall context to support an idiom continuation rather than a literal sentence, and despite the idioms having no unequivocal recognition point until the end of the phrase. Cloze probability was a significant predictor of skipping, but even when this was included in the models, group differences still existed. This suggests that predictability alone does not explain why idioms are processed in this way. It is also noteworthy that compositionality – often seen as a key component of how idioms are processed – was only significant in the total reading time of the final word, hence it is only when participants were required to integrate the idiomatic meaning of this into the sentence as a whole that effects were seen, which is in line with previous research (e.g. Libben & Titone, 2008, who showed that predictability correlated with familiarity, and variables like compositionality only came into play when a task specifically required consideration of the meaning of the idiom as a whole).

Binomials showed between-group differences and effects of cloze probability (in early measures, but notably not in terms of likelihood of skipping the final word) and phrase frequency, but showed no effects of semantic association. This is consistent with the findings of Siyanova-Chanturia (2010), who found reduced N400 effects for binomials in their canonical configurations, and with other studies of formulaic language (particularly idioms) that suggest that normal semantic integration processes are to some extent ‘switched off’ once a known formulaic template has been matched (Rommers et al., 2013; Vespignani et al., 2010). Importantly, the present results support the inclusion of literal binomials (rather than just irreversible/idiomatic binomials) in the broad class of formulaic language.

Collocations showed no difference according to whether or not they were semantic associates. Durrant and Doherty (2010) found a difference between the two types only in the most automatic processes, so it may be that in natural reading, where multiple sources of information are available, such differences are minimised. Instead, there was a consistent effect of frequency, whereby collocations were read more quickly than less frequent control phrases. In later measures it was the MI score – in some ways a measure of the strength of the link between the words – rather than simply phrase frequency that was important. Later processes may therefore reflect the coherence of the unit as whole as it fits into the broader sentence context.

Overall, Experiment 5 has demonstrated a clear formulaic processing advantage for all of the subtypes considered here. Importantly, specific variables affect different types of unit: cloze probability/predictability in idioms, phrase frequency in binomials, and MI in later processing of collocations. These results validate the claim of Columbus (2010) that distinct features underpin the processing of different subtypes of formulaic language. The final words of idioms showed a much higher tendency to be skipped than other subtypes, despite having lower phrase frequency and cloze probability than binomials. This suggests that their status as single conceptual units may contribute to ‘holistic’ processing, whereas the advantage for compositional units is driven by experience/frequency based processes. In Experiment 6 I set out to test this by investigating whether the cohesion of the same items is retained when the underlying formulaic frames are compromised. If the processing advantage for idioms and compositional units is driven by different mechanisms, I expect different patterns to emerge once the formulaic frames are compromised.

7.7 Experiment 6

In Experiment 6 I explore the relationships between the component words of formulaic units. If formulaic units are stored as whole forms, then we might expect that encountering the component parts in non-formulaic configurations should remove the formulaicity and negate any formulaic priming effects. Few studies have looked at the processing of non-standard variants of formulaic sequences in this way, despite the widely acknowledged fact that idioms are more flexible than many researchers assume (Konopka & Bock, 2009; Vietri, 2014). Where formulaic variants have been investigated, results are mixed as to whether creative forms show the same speeded processing as canonical configurations. McGlone, Glucksberg & Cacciari (1994) compared idioms (*spill the beans*) with variants (*spill a single bean*) and literal paraphrases (*not say a word*) in a self-paced line by line reading study. The idiom was processed more quickly than both other phrase types, which did not differ in their processing times. As discussed previously, Siyanova-Chanturia (2010) found that removing the *and* from binomials made them behave in the same way as comparably related word pairs, but differently from formulaic language: the previously observed P300 effect was removed and replaced with a more expected N400. The only study to show some evidence of formulaic priming in non-standard configurations is Bonk and Healy (2005), who used a primed naming task with collocating word pairs that had no semantic relationship (e.g. *bend-rules*) and presented items in forward and backward configurations. They found an effect of the prime only when the collocates were presented in a reversed form (*rules-bend*). They suggested that this was evidence that lexical networks make conventional word pairs active during language processing, leading to priming. Molinaro, Canal, Vespignani, Pesciarelli and Cacciari (2013) showed that inserting additional elements into collocational complex prepositions (e.g. *in the hands of*) did not disrupt processing of the overall string, and N400 effects on the noun were actually smaller when it was preceded by an inserted adjective. They concluded that such items were not processed by a simple 'lexical look up' procedure but were open to regular decompositional analysis and transformational procedures that could enrich the overall meaning. Overall, given how little research has been done on formulaic variants, the nature of precisely how formulaic word pairs are connected is very much open to investigation.

I again use a natural sentence reading task to investigate how participants treat formulaic word pairs encountered in non-formulaic sentence contexts. Carroll and Slowiaczek (1986; see also Camblin, Gordon & Swaab, 2007 and Traxler, Foss, Seely, Kaup & Morris, 2000) investigated within-sentence priming for semantically associated words and found priming for word pairs that had a close semantic relationship over unrelated or neutral pairs. For example, in a sentence like *Although the AUTHOR / ECONOMIST / GUY had wanted to finish the BOOK about the economic crisis, speaking engagements took all of his time*, the processing of the target word (*BOOK*) was facilitated when it followed the closely related word (*AUTHOR*) rather than either the non-associate or neutral word. Carroll and Slowiaczek (1986) found evidence of this effect for category-exemplar pairs (e.g. *Science-Chemistry*) and primary associates (e.g. *Author-Book*). I adopt a similar method to test whether any associative relationships between formulaic word pairs are evident when the formulaic frame has been removed. This allows me to further explore the 'holistic' storage of formulaic language and determine whether it is of a similar nature for the formulaic subtypes from Experiment 5. To my knowledge, this is the first study to examine formulaic components in context in this way.

7.7.1 Participants

Forty undergraduate students (all native speakers of English) took part in Experiment 6 for course credit. None had taken part in Experiment 5 or any of the norming procedures for either experiment.

7.7.2 Materials

Stimulus items were the same as in Experiment 5, although some control pairs were altered from those used in Experiment 5 to ensure that the new sentence contexts sounded natural. In such cases the alternative was matched for length, frequency and association strength. Formulaic word pairs were inserted into sentence contexts, separated by a minimum of two and a maximum of six words, in forward and backward combinations. I added a category of semantic associates for comparison;

this allowed me to see whether lexical-semantic priming could be induced within sentences, and provided a non-formulaic baseline category for analysis. Semantic associations were based on EAT scores. Association strength in both directions was obtained and the direction with the highest score was deemed the forward configuration.

Examples of items and their characteristics are shown in Table 7.7. I expected to see an effect of association strength for semantically related pairs vs. controls, with no distinction between ‘forward’ and ‘backward’ pairs. For formulaic units, forward priming but not backward priming may indicate ‘holistic’ storage. For example, encountering *spill* may prime *beans*, whereas *beans* may not similarly prime *spill*, since encountering the start of a formulaic unit should trigger the underlying whole phrase. However, if priming occurred in both directions this would implicate a ‘lexical networks’ view of formulaic word pairs, where words are linked by association (Bonk & Healy, 2005). Lack of any priming could also support a lexical network view, since the links may be too weak to retain cohesion in non-contiguous contexts. Whatever the results, different patterns for idioms compared to literal subtypes would indicate that different mechanisms underlie the formulaic advantage seen in Experiment 5.

Table 7.7. Examples of stimulus pairs for Experiment 6, phrase frequency and association strength.

Type	Forward pair	Phrase Fr	Ass	Backward pair	Phrase Fr	Ass
Semantic	Bread-baker	0	.18	Baker- bread	0	.05
Control	Fruit-baker	0	.00	Grocer-bread	0	.00
Idiom	Spill-beans	54	.00	Beans-spill	54	.00
Control	Drop-beans	8	.00	Chips-Spill	5	.00
Binomial	King-queen	251	.29	Queen-king	251	.25
Control	Prince-queen	5	.04	Prince-king	8	.03
Collocations						
Associated	Ancient-history	228	.15	History-ancient	228	.05
Control	Distant-history	6	.00	Stories-ancient	5	.00
Unassociated	Abject-poverty	108	.00	Poverty-abject	108	.00
Control	Total-poverty	9	.00	Agony-abject	5	.00

Note: For all reversed pairs the phrase frequency was considered to be the frequency of the underlying formulaic unit, i.e. spill-beans and beans-spill are both based on the frequency of spill the beans.

Context sentences for all of the items were created. As with Experiment 5, sentences were created so that the same context made sense when either the formulaic or control pair was inserted. Importantly, the contexts preceding the first word, in-between words and following the second word were identical for formulaic and control sentences. All sentences were normed by native speakers (n = 36) for naturalness. No differences existed between formulaic types and controls in either direction (one way ANOVA by condition: $F = 0.69$, $p = .56$). Examples are presented in Table 7.8.

Table 7.8. Example sentences for all conditions, Experiment 6.

	Pre-context	Word 1	Inter-context	Word 2	Post-context
Forward					
Formulaic	I tried not to	spill	them but the	beans	still ended up on the floor
Control	I tried not to	drop	them but the	beans	still ended up on the floor
Backward					
Formulaic	He grabbed for the	beans	but then they	spilled	all over the kitchen floor
Control	He grabbed for the	chips	but then they	spilled	all over the kitchen floor

7.8 Results and analysis

All data cleaning procedures were the same as in Experiment 5 and 3.7% of trials were removed due to track loss. Since formulaic units were presented as separated words, I only considered reading patterns for the second word of each pair. I compiled an omnibus mixed effects model in R with treatment-coded fixed effects of subtype (semantically related pairs as baseline), condition (related vs. control) and direction (forward vs. backward configuration). Random intercepts were included for subject and item as well as by-subject random slopes for the effect of word condition and direction. The covariates of word length, word frequency and phrase frequency (both log-transformed), association strength, and the length of the region separating the prime and target words were included. Table 7.9 shows a summary of results for all eye-tracking measures and omnibus analysis is presented in Table 7.10.

Table 7.9. Summary of eye-tracking measures for all items, second word of each pair.

Word pair	Direction	Likelihood of skipping	First fixation duration	First pass reading	Total reading	Regression path duration
Semantic	Forward	.21 (.41)	152 (94)	156 (102)	202 (213)	286 (327)
Control	Forward	.19 (.39)	161 (95)	167 107)	218 (188)	296 (295)
Semantic	Backward	.22 (.42)	156 (101)	159 (107)	192 (160)	284 (275)
Control	Backward	.19 (.40)	162 (99)	169 (109)	210 (166)	321 (335)
Idiom	Forward	.29 (.45)	143 (104)	148 (109)	168 (137)	221 (237)
Control	Forward	.20 (.40)	157 (95)	165 (105)	193 (148)	247 (224)
Idiom	Backward	.16 (.37)	171 (97)	175 (101)	214 (159)	277 (240)
Control	Backward	.18 (.38)	166 (100)	171 (107)	212 (159)	283 (259)
Binomial	Forward	.24 (.43)	146 (96)	151 (107)	173 (137)	238 (242)
Control	Forward	.18 (.39)	160 (93)	165 (105)	197 (146)	246 (227)
Binomial	Backward	.24 (.43)	149 (102)	153 (110)	176 (140)	239 (239)
Control	Backward	.23 (.42)	158 (104)	163 (111)	182 (136)	239 (210)
Collocations						
Associated	Forward	.21 (.41)	155 (95)	161 (108)	186 (160)	267 (295)
Control	Forward	.20 (.40)	159 (98)	164 (105)	196 (156)	282 (289)
Associated	Backward	.16 (.37)	169 (97)	177 (106)	221 (171)	301 (261)
Control	Backward	.16 (.37)	169 (95)	178 (108)	224 (170)	328 (309)
Unassociated	Forward	.18 (.38)	165 (96)	172 (103)	210 (153)	311 (297)
Control	Forward	.16 (.37)	164 (92)	169 (98)	202 (142)	311 (273)
Unassociated	Backward	.14 (.35)	175 (93)	188 (111)	250 (211)	352 (339)
Control	Backward	.11 (.31)	184 (94)	198 (116)	263 (198)	401 (403)

Note: All figures are mean values (standard deviation in brackets). Likelihood of skipping is expressed as a probability and duration measures are in ms.

Table 7.10. Omnibus mixed effects models for second word reading, all measures.

Omnibus	Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-1.40	0.33	-4.27	5.32	0.04	140.85	5.37	0.04	143.04	5.54	0.07	83.08	5.90	0.08	70.97
Subtype: Idiom	0.42	0.18	2.30*	0.04	0.02	1.75$^+$	0.05	0.03	1.89*	0.01	0.04	0.23	-0.05	0.06	-0.84
Subtype : Binomial	0.11	0.18	0.60	0.00	0.02	0.12	0.01	0.03	0.22	-0.04	0.04	-1.13	-0.09	0.06	-1.44
Subtype : Ass-Coll	0.06	0.18	0.33	0.02	0.02	0.86	0.03	0.03	1.06	-0.03	0.04	-0.79	-0.03	0.06	-0.43
Subtype : Non-Coll	-0.06	0.20	-0.30	0.04	0.02	1.94*	0.05	0.03	2.10*	0.06	0.04	1.54	0.13	0.06	2.14*
Condition: Control	-0.04	0.19	-0.23	0.02	0.02	1.00	0.02	0.02	0.94	0.05	0.03	1.65	0.07	0.04	1.53
Direction: Backward	0.21	0.18	1.18	0.02	0.02	0.99	0.01	0.02	0.59	-0.02	0.03	-0.52	0.05	0.04	1.09
Idiom*Condition	-0.44	0.25	-1.76$^+$	-0.04	0.03	-1.47	-0.03	0.03	-1.00	-0.04	0.05	-0.79	-0.04	0.06	-0.73
Binomial*Condition	-0.18	0.25	-0.72	-0.01	0.03	-0.36	-0.01	0.03	-0.40	-0.01	0.05	-0.16	-0.05	0.06	-0.89
Ass-Coll*Condition	0.07	0.25	0.27	-0.01	0.03	-0.24	-0.02	0.03	-0.48	-0.01	0.05	-0.23	0.00	0.06	0.02
Non-Coll*Condition	-0.08	0.26	-0.32	-0.04	0.03	-1.47	-0.05	0.03	-1.63	-0.10	0.04	-2.19*	-0.08	0.06	-1.37
Idiom*Direction	-0.98	0.25	-3.92***	-0.01	0.03	-0.50	-0.01	0.03	-0.27	0.07	0.05	1.42	0.03	0.06	0.54
Binomial*Direction	-0.33	0.25	-1.37	-0.02	0.03	-0.57	-0.01	0.03	-0.33	0.04	0.05	1.42	-0.03	0.06	-0.54
Ass-Col*Direction	-0.28	0.25	-1.10	-0.00	0.03	-0.14	-0.00	0.00	-0.01	0.08	0.05	1.79$^+$	0.04	0.06	0.64
Non-Coll*Direction	-0.40	0.26	-1.53	-0.01	0.03	-0.47	0.01	0.03	0.22	0.07	0.05	1.53	-0.02	0.06	-0.28
Condition*Direction	-0.15	0.25	-0.58	-0.01	0.03	-0.47	0.00	0.03	0.03	0.03	0.05	0.01	0.01	0.06	0.25
Idiom*Condition*Direction	0.73	0.36	2.04*	0.03	0.04	0.63	0.00	0.05	0.06	-0.01	0.06	-0.19	-0.03	0.08	-0.16
Binomial*Condition*Direction	0.40	0.35	1.14	0.06	0.04	1.35	0.05	0.05	1.04	-0.01	0.06	-0.11	0.04	0.08	0.43
Ass-Coll*Condition*Direction	0.10	0.36	0.28	0.00	0.04	0.06	-0.00	0.05	-0.02	-0.03	0.06	-0.42	-0.01	0.08	-0.12
Non-Coll*Condition*Direction	-0.04	0.38	-0.10	0.05	0.04	1.28	0.05	0.04	1.06	0.08	0.06	1.35	0.08	0.08	0.95
Control predictors:															
Length	-0.17	0.03	-6.46***	-0.01	0.00	-1.88$^+$	n/a	n/a	n/a	0.01	0.01	2.19*	n/a	n/a	n/a
Frequency (log)	0.08	0.03	3.09**	-0.01	0.01	-2.48*	-0.01	0.00	-4.19***	-0.02	0.01	-3.16**	-0.03	0.01	-4.45***
Association	0.63	0.29	2.21*	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.17	0.08	2.03*
Separating region length (words)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	-0.02	0.01	-2.94**	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.304			0.008			0.008			0.012			0.029		
Item	0.065			0.001			0.003			0.007			0.043		
Subject | Condition	0.004			0.000			0.000			0.000			0.000		
Subject | Direction	0.000			0.000			0.000			0.000			0.000		
Residual	n/a			0.070			0.086			0.170			0.281		

*Note: Table displays coefficients (β) standard error (SE) and t-values (z-values for generalised linear models), with significance values estimated by the lmerTest package in R (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): *** p < .001, ** p < .01, * p< .05.*

The omnibus analysis shows a significant three-way interaction for the idiom subset, condition and direction ($z = 2.04$; $p < .05$). Idiom word pairs in their forward configuration were significantly more likely to induce skipping of the second word than any other subtype or configuration. There was also an overall effect of semantic association strength ($z = 2.21$; $p < .05$). Neither cloze probability nor phrase frequency showed significant effects for any measure. To more fully explore the behavior of the different formulaic subtypes I fitted separate models for each. As before, significant results are summarised below, with full model outputs available in Appendix 4b (table 4-10).

7.8.1 Semantic pairs

Semantically related word pairs showed minimal effects. Analysis with condition and direction as fixed effects showed no significant results, so I reanalyzed the set of items as a whole without splitting them into forward and backward configurations. The effect of condition (semantic vs. unrelated) was only significant for total reading time ($t = 2.16$, $p < .05$) and marginally for regression path duration ($t = 1.73$, $p = .08$). I also ran a model with no categorical fixed effects and just association strength as a predictor. This was marginally significant for first pass reading time ($t= -1.75$, $p = .08$) and regression path duration ($t = -1.91$, $p = .06$). Overall, there was a small effect of semantic relatedness, but no consistent evidence that semantically related word pairs are primed within sentences, at least in the non-contiguous contexts used here.

7.8.2 Idioms

For idioms I included cloze probability, familiarity and compositionality (both centred) as covariates if they significantly improved the models. Familiarity was not significant for any measure. Formulaic pairs showed a greater likelihood of skipping the second word in the forward but not backward configuration. This was clearly seen in the interaction of condition and direction ($z = 2.26$, $p < .05$), and cloze probability was also a significant predictor of skipping rates ($z = 2.15$, $p < .05$). In the duration measures there were no significant effects of condition or direction.

I fitted a final model to further explore just the forward configuration for formulaic/control pairs, since this configuration of idiom word pairs showed the clearest effect. There were significant effects of cloze probability for likelihood of skipping ($z = 2.98$, $p < .01$), total reading time ($t = -2.15$, $p < .05$) and regression path duration ($t = -2.73$, $p < .01$) and marginally for first fixation duration ($t = -1.76$, $p = .08$). In all cases higher cloze probability for the underlying idiom meant the second word of each pair was more likely to be skipped or read more quickly.

7.8.3 Binomials

Binomials showed a marginal effect of association strength ($z = 0.67$, $p = .06$) on likelihood of skipping and significant effects on first fixation duration ($t = -2.10$, $p < .05$) and first pass reading time ($t = -2.34$, $p < .05$). In later measures the effects of association strength were non-significant, but for total reading time there was a significant effect of phrase frequency ($t = -2.14$, $p < .05$). Regression path duration showed no significant effects.

7.8.4 Collocations

As in Experiment 5, collocations were combined and associated/unassociated was included as a fixed effect. MI score was included in all models but showed no significant effects.

There were no effects of any variable on likelihood of skipping, but the duration measures showed an interaction of subset and type and a three-way interaction of subset, type and direction for total reading time ($t = 2.13$, $p < .05$). Because of these interactions I fitted separate models for the associated and unassociated collocations to further explore each set.

Associated collocations showed overall effects of type for all duration measures; in all cases control words were read more slowly than collocating words. In addition there were significant effects of association strength and cloze probability on all measures, but this was in the opposite direction to what might be expected. In both cases higher scores (stronger associations, higher cloze probabilities) increased reading times.

Unassociated collocations showed no effects for early measures. For total reading time there was a significant interaction of type and direction ($t = 1.99$, $p < .05$), and both total reading time and regression duration showed significant effects of cloze probability. Again, this was not in the expected direction, with higher cloze probabilities leading to longer reading and re-reading times.

Phrase frequency showed no significant effects for either type of collocation on any of the eye-tracking measures.

7.9 Discussion

The results from Experiment 6 yielded some interesting patterns. For semantically related pairs there was some evidence of within-sentence priming but this was limited. Previous studies have found variable effects depending on the type of semantic relationship. Traxler, Foss, Seely, Kaup and Morris (2000) found evidence of identity priming in early reading patterns but no effect for synonyms or associates. Camblin, Gordon and Swaab (2007) found early priming effects but only for impoverished or incongruent discourse contexts. It may be that in Experiment 6 the designation of ‘semantically associated’ was too broad to show consistent effects. So alongside close associates (e.g. *bread-baker*, *plane-pilot*) there were a number of more distant, schematically related pairs (e.g. *kettle-steam*, *water-bridge*). This broadness was required to differentiate semantically related pairs from those which also formed parts of formulaic units, and it is noteworthy that at least some of the stimuli used in Carroll and Slowiaczek (1986) and Camblin et al. (2007), both of whom found semantic priming, could be considered to be binomial pairs (e.g. *father-mother*, *arms-legs*). It may be that the formulaic nature of these pairs contributed in some way to the priming effects seen in these studies, hence when this is excluded entirely, as in our semantically related pairs, effects are minimised.

In comparison, the binomial pairs are predominantly strongly related primary semantic associates (*king-queen*, *son-daughter*, *north-south*) and formulaically related items, which may lead to a greater degree of priming than for either connection alone: effects were much less apparent for semantic pairs, which have no formulaic underpinning, and collocations, which have weaker underlying semantic links. For the collocations, links also seem to be of a fundamentally different nature, being mostly associative (e.g. *express train*, *modern art*) rather than truly ‘semantic’. Since association strength was based on EAT scores, which is a measure of free association, many different types of association are reflected in our categories. One type of association, associative links, is largely schematic and learned through experience (Williams & Cheung, 2011). Associative links may be subject to different processes than more central ‘core’ connections, e.g. episodic matching, rather than more semantically driven processes like spreading activation or semantic matching (Jones & Estes, 2012). Overall, it seems likely that a combination of stronger/more central semantic relations and the added formulaic association of binomials led to priming for this type compared to the word pairs that were semantically but not formulaically related.

Collocations showed an unexpected pattern of results. There was an effect of association strength, but this meant that more closely associated words were read more slowly. This might at first sight seem odd, but it is again perhaps an indication that the types of associations between collocates are fundamentally different from the types of primary semantic relations that exist between binomial partners. Both association strength and cloze probability (significant for all collocations) reflect this 'learned' relationship, and encountering the second word of a collocating pair several words downstream might therefore be unexpected, since the schematic relationship between words means that they are almost always encountered as contiguous pairs. This might in turn lead to increased processing time as the now unexpected item is reintegrated into the sentence. Hence the more expected the item is in its canonical form (when both words appear together), the more disruptive it is to have the other word presented several words downstream.

The clearest finding is that idiomatically related words show a distinctly different pattern compared to the other stimuli. The likelihood of skipping the second word was much higher for idiom pairs than any other types of formulaic pair, and I interpret likelihood of skipping as a clear indication of increased predictability (Rayner &

Well, 1996; Staub & Rayner, 2007).[25] Even when skipped items were removed from the analysis there was a consistent effect of cloze probability on all duration measures. The predictability of the underlying idiom is therefore an important factor in whether or not the same word pairs show priming in non-formulaic configurations. Importantly, all idiom pairs had EAT scores of 0, so no semantic or associative links can be driving this effect. The fact that facilitation is only seen in the forward configuration is also crucial, as it clearly implicates some aspect of the overall structure in how such a unit is activated. In comparison, binomial word pairs showed an overall effect of association strength and no effect of direction. This suggests that an intrinsic and bidirectional semantic link, rather than an effect of an underlying configuration, might better explain the priming effect in binomials. Alternatively, it may simply be the case that the coordinated nature of binomials makes them fundamentally reversible, whereas idioms and collocations are not. The order of components may not be as important as in other constructions, where grammatical constraints will place additional restrictions on ordering of components (although results from Siyanova-Chanturia (2010) and Siyanova-Chanturia, Conklin and Van Heuven (2011) argue against this, since both found a significant advantage for binomials compared to reversed forms).

One final result worthy of mention is that the distance separating words showed no consistent effects. This was true in the omnibus model for total reading time, but there was also no effect on early measures, either in the omnibus model or the individual models for the different subtypes. In terms of the automatic priming of words,

[25] Although Staub and Rayner (2007) also summarise research showing that perceptual and visual information are a primary determinant of skipping behaviour, the fact that in this study formulaic and control sentences were identical except for the prime word shows that this could not explain the results.

indexed by likelihood of skipping and early duration measures, the distance separating prime and target therefore did not appear to be important.

7.10 General discussion

In two experiments I have examined the processing of formulaic units and formulaic component words in natural reading. Experiment 5 demonstrated a consistent advantage for all formulaic subtypes, with idioms showing significant differences compared to binomials and collocations. Columbus (2010) found that idioms were read faster than sentences containing other formulaic units, and the results in this chapter extend this by providing a direct comparison of formulaic units and controls across a number of distinct formulaic subtypes. In all cases it seems that predictability drives the formulaic effect, but the underlying reason for this predictability varies. Whilst phrasal frequency appears to account for the predictability of binomials and collocations (consistent with a usage based-view that these items have simply been encountered more often), cloze probability was a more important factor for idioms. Importantly, even when the effects of cloze probability were accounted for in the analysis, significant differences were apparent between idioms and control items (which was not the case for binomials and collocations, where including phrase frequency in the models removed differences between conditions). Idioms are therefore processed quickly for reasons other than simply frequency of previous encounter, or because seeing the first word makes the second word highly predictable.

A more telling difference between idioms and other types of formulaic unit is evident in the results of Experiment 6, where component words of idioms show a clear advantage even in non-formulaic configurations. Crucially, effects of this are only seen in the ‘correct’ order, in other words, when idiom components are presented in the forward configuration. Also important is the fact that the component words were used in their individual literal senses, rather than with the particular "phrase induced polysemy" (Glucksberg, 1993) that is ascribed to them in specific idiomatic configurations. This is consistent with the idea that idioms represent a case of categorical template matching (as per the ERP results of Vespignani et al., 2010 and Rommers et al., 2013), whereby the lexical expectation that is generated is a binary value of expected/unexpected word. Whilst this explains why canonical idioms show speeded processing, what is less clear is why this should lead to facilitation for idiom words used separately from each other.

Existing models of how idioms are represented propose that if a whole form entry exists, this should be accessible via the base components. This whole form has been referred to as a superlemma (Sprenger et al., 2006), formuleme (Van Lancker Sidtis, 2012), configuration (Cacciari & Tabossi, 1988) and lexicalised unit (Libben & Titone, 2008). In all of these models there exists some unified representation, allowing the entries to specify complex structural information, as well as the lexical make-up and figurative meaning. Sprenger at al. (2006) proposed that when any of the components of an idiom is encountered, this results in "spreading activation from the element to all the remaining elements via a common idiom representation, resulting in faster availability of these elements" (p.167). It is conceivable that this increased activation for individual lemmas may still exert an influence several words downstream, hence encountering *spill* means that the superlemma *spill the beans* and in turn the individual lemma *beans* will be primed, even if this word is not encountered immediately. This would allow for flexibility in how idioms are deployed, which may be driven by the functional relationships and mapping between the underlying concept of an idiom and its component parts (Cacciari & Glucksberg, 1991; Konopka & Bock, 2009).

However, Sprenger et al. (2006) suggest that idiom entries should be accessible via any of the base components. Their study showed a larger advantage for idioms (e.g. *hit the road*) compared to literal sentences (*clean the road*) when the final word (*road*) was used as the prime in a cued recall task. Experiment 6 failed to show this kind of backward priming, so it may be that word order is a vital aspect of the whole form entry. Although some idioms can be passivised or otherwise transformed in creative ways, the canonical, default forms are much more frequent, and the links between individual lemmas and whole form superlemmas should reflect this. In other words, encountering a word that starts an idiom should trigger automatic consideration of the idiom meaning by activating the whole form, whereas encountering words that appear later in the default form may either not activate the idiom at all or may show much weaker activation. Alternatively, the relative contribution of the component words to the idiom as a whole may be crucial. Hamblin and Gibbs (1999) demonstrated that intuitions about the overall phrasal meaning of idioms are directly related to the meaning of the main verb. Often the motivation of an idiom comes from the verb, hence this ‘licenses’ the idiomatic meaning of the noun, which otherwise retains only its core, literal meaning (i.e. *beans* only acquires the specific figurative meaning of “a secret” when used in conjunction with *spill*). It is also the case that verbs can more often be used to denote the idiomatic meaning on their own. For example, the idiom *spill the beans* can easily be shortened to simply *spill* (e.g. *I knew he was hiding something and I was just waiting for him to spill*), whereas the same cannot be said of *beans* (e.g. **I knew he was hiding something and I really wanted to find out the beans*).

Even if we accept this view of idioms as whole units, accessible by the more salient component words, we still need to account for the faster processing of other formulaic types in Experiment 5 and in previous research. Here, the lexical priming view proposed by Hoey (2005) and elaborated by Pace-Sigge (2013) seems of relevance. In this view, individual words become associated through experience, hence collocations and other multi-word combinations (literal binomials, lexical bundles) develop links in the mental lexicon based on previous encounter/co-occurrence. The ERP results for all formulaic types support this: a consistent finding of a P300 type effect, reflecting categorical lexical expectation based on an existing template (Siyanova-Chanturia, 2010; Vespignani et al., 2010; Molinaro & Carreiras, 2010). Thus, encountering any of the formulaic units in Experiment 5 generates a lexical expectancy based on previous encounters with the phrase. This expectancy is categorical, hence the 'template' is either matched or not, and encountering any other continuation does not fulfil the expectation. In Experiment 6, where this categorical expectation was broken, only those items where an underlying conceptual relationship exists to bind lexical items together showed any effect. This manifests in different ways: for idioms, a 'whole form' concept underpins the effect, allowing the unseen component word to remain active later in the sentence; for binomials, a strong, central semantic relationship (which may be additionally strengthened by the formulaic link between components) exists to allow bidirectional priming.

Such an argument is broadly in line with the constraint-based accounts proposed for idioms, metaphors and language processing more generally (c.f. Gibbs, 2006; Libben

& Titone, 2008; Macdonald, Pearlmutter & Seidenberg, 1994), all of which assume that multiple sources of information are utilised to aid processing in any given situation. Importantly, idioms are familiar word combinations *and* unique conceptual wholes; the individual words are processed like other configurations in the lexicon, but the meaning is retrieved directly (Tabossi, Fanari & Wolf, 2008). They may also require the integration of more distant semantic relations by default (Cacciari, 2014), hence maintaining possible lexical partners for longer may simply be a logical by-product of this.

Interestingly, the effects seen for collocations were in the opposite direction to idioms and binomials. That is, once the initial word had been encountered and the predicted ‘template’ had not been matched, reading the collocate several words downstream was actually disruptive, with higher cloze probability and association strength leading to longer reading and re-reading times. This suggests that reintegration of a schematic expectation is difficult, since the categorical lexical expectation has already been broken. A similar argument should hold for binomials, but two additional factors may mitigate this. Firstly, as discussed, the relationship between words is much stronger, being a core semantic link rather than a schematic association; secondly, binomials are fundamentally coordinated in nature, hence may be more amenable to separation/reversal than the noun-noun or adjective-noun collocations. Together, this means that binomials are underpinned by more central, semantically driven associations and are grammatically more amenable to separation, without this causing the kind of disruption seen in collocations.

The picture this paints of the mental lexicon is intriguing. Certain types of formulaic units – lexical bundles, non-compound collocations, unassociated binomials – may only represent statistical patterns of co-occurrence, which will be dynamic, speaker-specific and transient in the sense that as soon as the lexical 'template' is violated, consideration of the expected sequence ceases. A view of 'holistic' storage of whole units may be seen as a useful metaphor, but does not easily explain the effects seen for such sequences. Connections are schematic, acquired through experience, and this forms associative links of varying strengths between words, comparable to those that exist to link words that share semantic or phonological properties. Crucially, the types of links reflected are much shallower, being based purely on patterns of previous encounter rather than any core linguistic features, which may exist as deeper, conceptual associations. In some instances, like the binomials in this study, such associations at a conceptual level may also exist, but these exist as intrinsic links rather than as part of a shared whole form. This is consistent with a number of studies showing differences in the degree and nature of priming engendered by semantic and associative relationships (e.g. Estes & Jones, 2009; Rhodes & Donaldson, 2008).

Idioms, on the other hand, are conventional lexical sequences *and* single choices in a conceptual sense, hence the cohesion that they display is fundamentally different. Like other formulaic items, lexical priming exists to activate likely continuations based on previous experience, but only idioms also show the flexibility and complexity to explain the delayed priming effect seen here. That is, component words of idioms demonstrate (potential) separate figurative meanings that are maintained for longer because of their underlying conceptual whole, leading to lexical facilitation even when the expected partner is not encountered immediately. Figure 7.1 demonstrates the different representations of the formulaic units considered here.

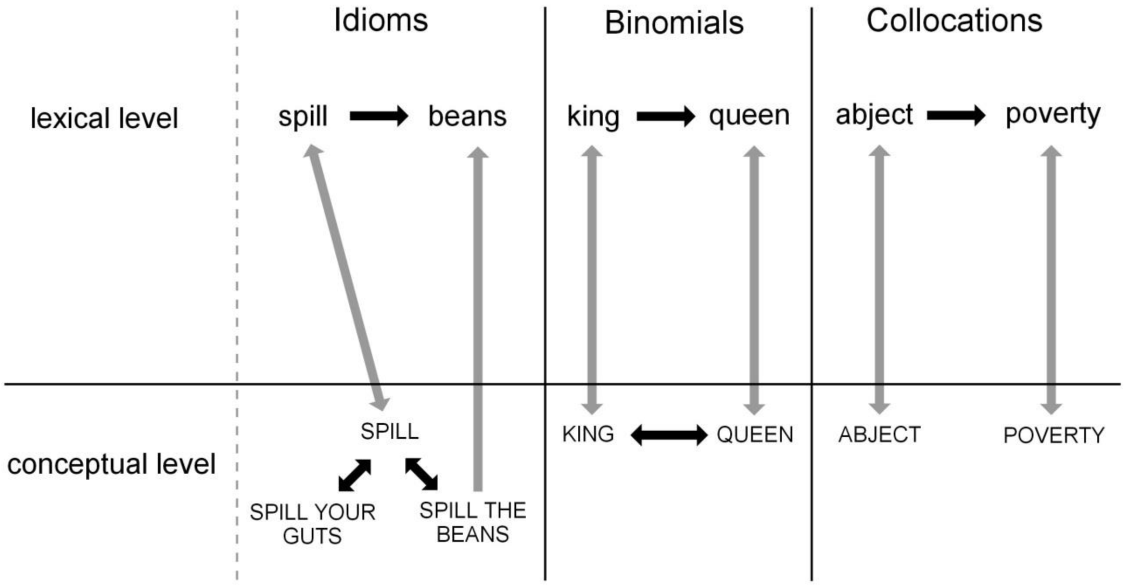

Figure 7.1. Schematic diagram of formulaic links for different subtypes. Black arrows show links between lexical items or between concepts; grey arrows are links between lexical items and underlying concepts. All show unidirectional priming at a lexical level, driven by frequency of previous encounter. At a conceptual level, only idioms have unique conceptual entries. Hence encountering *spill* activates the lemma *SPILL*, as well as whole form entries for any idioms of which it is a part (*spill the beans*, *spill your guts*, etc.). The unidirectional arrow from *SPILL THE BEANS* to *beans* reflects the forward only priming seen in the data. Binomials show lexical priming and bidirectional semantic priming at a conceptual level. The relationship between *abject* and *poverty* is schematic and learned and there is no underlying semantic relationship, hence priming exists only at a lexical level and is disrupted if the canonical sequence is not presented.

In summary, I have presented clear evidence that formulaic units – both those defined by semantic idiomaticity and purely statistical combinations – are recognised/processed more quickly by native speakers in their canonical forms in context-neutral situations. Given the structure of all stimuli, where no recognition point was available until the end of the phrase, it seems clear that some form of lexical priming exists to bias an expected continuation, based primarily on previous encounter and probabilistic expectation. I also found evidence that idioms retain their predictability in a delayed way, such that when the components are presented several words apart we still see a processing advantage. I interpret this as support for the special status of idioms, in that they represent complex structural combinations which are nonetheless lexically immutable. The underlying conceptual unity binds together specific lexical items, allowing for them to remain active during sentence processing in a way that other statistically defined formulaic combinations do not. A view of 'holistic' representation is therefore plausible at a conceptual level, but does not adequately explain the lexical priming of literal/compositional formulaic sequences, although binomial pairs show a high degree of semantic association that may be in part a reflection of their 'special' formulaic status.

Chapter 8. Summing the Parts

This chapter provides a general discussion of the results of the four empirical chapters. Each study has provided data that enriches our understanding of formulaic language in the monolingual and bilingual lexicons. In this chapter I discuss how this contributes to existing models of formulaic language and idioms in particular. I also present some general conclusions that highlight the key contributions of this thesis, as well as some of its limitations and some directions for further research.

8.1 General discussion

It is now well established that language shows effects of frequency at multiple levels of granularity. As outlined throughout this thesis, large bodies of evidence exist to show that more frequent combinations of words are processed more quickly than less frequent ones. Usage based models contend that the lexicon is made up of constructions of various sizes and degrees of abstraction, with a construction simply referring to any form-meaning pairing. For idioms, this is obfuscated by the presence of multiple semantic properties that mean they are diverse and difficult to consider as a single homogenous class. Still, the underlying common ground for idioms is that they are lexically fixed and highly predictable to native speakers, despite their relative infrequency (in multi-word terms), according to corpus data.

This fixedness is a key part of their representation. The few studies that have looked at idiom variation have found that fast processing is a property only of the canonical, citation configurations. McGlone, Glucksberg and Cacciari (1994) found no advantage for creative forms like *spill a single bean*. Holsinger and Kaiser (2010, also Holsinger, 2013) showed that inserting a sentential boundary into an idiom (e.g.

…kicked. The bucket…) was disruptive to early processing of the idiomatic meaning (but did not prevent later consideration, which they described as “a post processing recognition of the idiomatic string”, p.81). This suggests that the canonical form of an idiom is a vital component of how it will be recognised, which is consistent with usage-based models, which allow for both general and specific patterns to emerge from linguistic experiences. Although creativity and flexibility are not uncommon in idioms (Cacciari, 2014; Omazic, 2008; Schmitt, 2005; Vietri, 2014), they will be encountered predominantly in their citation forms, hence there is a frequency effect in that ‘correct’ versions will be recognised quickly in a way that variants may not.

It could therefore be argued that idioms in their citation forms do show some evidence of lexicalisation (McGlone et al., 1994; Tabossi et al., 2005). This is also supported by the phonological evidence discussed in Chapter 2, where frequent word strings develop consistent and predictable intonation contours (Lin, 2010; Lin & Adolphs, 2009; Van Lancker, Canter & Terbeek, 1981), and by the ERP evidence discussed throughout, where predictable lexical sequences (idioms, but also binomials and literal collocations) show evidence of template matching and categorical expectation via a P300 effect (Liu et al., 2010; Molinaro & Carreiras, 2010; Rommers et al, 2013; Vespignani et al., 2010; Zhou et al., 2004). While this suggests that at the level of articulation/formal representation idioms and other formulaic sequences do become lexically ‘fixed’ in some way, this is a long way from the lexical representation or ‘big words’ arguments of early idiom models, and simply reflects a natural process of grammaticisation that is seen in many forms of frequently occurring linguistic material (Bybee, 2002, 2006; Beckner et al., 2009). At least in semantic terms, idioms do behave as whole entries, and underpinning a canonical lexical form, hybrid models posit the existence of an underlying lexical/conceptual frame for idioms (Cutting & Bock, 1997; Sprenger et al., 2006), which specifies a distinct lexical combination and links to an underlying phrasal meaning.

The questions raised by the results of the empirical studies in this thesis are therefore intriguing. If idioms are lexically fixed, or even lexically specified by an underlying superlemma, why did translated forms show such a consistent formulaic advantage? Logically, if the expected citation form is what contributes to fast recognition, translating idioms should change the form enough to render them unformulaic and therefore negate any advantage. Instead, as seen consistently throughout Chapters 4, 5 and 6, Chinese-English and Swedish-English bilinguals of intermediate to advanced levels of proficiency showed faster processing of translated idioms, compared to control phrases. Importantly, the data from Chapter 6 on Swedish idioms included only those items where no unequivocal recognition point existed until the end of the idiom, minimising any opportunity for participants to make active predictions about upcoming material. The specific word combinations used throughout these studies cannot have been seen before in English (or at least are very unlikely to have been seen), so a lexical argument, whereby some form of stored template exists to aid idiom recognition, is untenable. Equally, an argument where an underlying superlemma specifies the lexical items that make up an idiom is problematic, unless we consider this to be language non-specific. Two main possibilities, discussed throughout the empirical chapters, therefore explain the results.

The first possibility is that since idioms are accessed via their individual component words, language non-selective lexical access and automatic translation processes are responsible for the effects. Many models of idiom processing posit the obligatory activation of component words (Sprenger et al., 2006; Smolka et al., 2007), and as shown by the results of several studies by Thierry and colleagues for Chinese-English bilinguals (Thierry & Wu, 2007; Wu, Cristino, Leek & Thierry, 2013; Wu & Thierry, 2010), encountering English words triggers aspects of the phonology and orthography of the translation equivalent Chinese character. Formal mediation therefore means that reading L2 words activates L1 translation equivalent words, which could trigger a known lexical combination in the L1, leading to overall facilitation. Certainly this may explain the differential results seen in Study 2, where Chinese-English bilinguals showed fast processing of form (Experiment 2) but no associated advantage for the figurative meaning of translated idioms (Experiment 3). On the other hand, it is assumed that activation of a known L1 lexical combination should also automatically trigger consideration of the underlying figurative meaning, especially since the context supported this, so it is difficult to explain why this should not have happened.

Two other objections suggest that the process is not necessarily solely lexical/translation based. Firstly, the fragility of idioms in the studies discussed above (Holsinger, 2013; Holsinger & Kaiser, 2010; McGlone et al., 1994) suggests that even minor changes are enough to negate some aspects of the formulaic advantage. In fact, McGlone et al. (1994) showed that although idiom variants showed no processing advantage in terms of how quickly they were read, participants had no difficulty at all in correctly interpreting the figurative meaning. By analogy, modifying the formal properties by translating the idioms should impair processing of form but not negate the figurative meaning, which is the opposite to what was observed.

Secondly, comparable studies (e.g. Wolter & Yamashita, 2014) looking at translated forms of collocations have found no evidence of facilitation. Logically, if a ballistic activation/automatic translation mechanism was at work, any well-known, formulaic combinations should be activated in the same way, but this did not seem to be true. However, it should be noted that the proficiency level of bilingual participants does seem to be important here. Wolter and Yamashita (2014) found no facilitation for translated collocations, either for moderate or advanced proficiency participants. Ueno (2009), however, did show facilitation for translated collocations amongst the highest proficiency participants in her study, although her results were in general quite inconsistent. Similarly, the Chinese participants in Study 2 (who were in general at an intermediate-high level of proficiency) showed no evidence that figurative meanings were understood for translated idioms, even though they seemed to recognise the form, while the Swedish participants in Study 3 (who had very advanced proficiency) showed lexical facilitation and had no difficulty understanding any of the figurative phrases (L1 translation, congruent or L2 only). Hence looking across the studies, the higher proficiency of the Swedish participants seemed to demonstrate more complete priming (activation of L2-L1 translation equivalents *and* activation of underlying concepts via L2 forms). Given the variability in results in the literature, however, automatic activation of L1 equivalents may tell only some of the story as to why the translated idioms showed such consistent facilitation for form.

The second possibility is that the conceptual underpinning of idioms is responsible for the cross-language effects. Two variations of this are possible: either a view based on the Revised Hierarchical Model (Kroll & Stewart, 1994) whereby shared concepts can be accessed via either L1 or L2 forms, or a lemma-mediation model (Jiang, 2000), whereby learning words in L2 involves the acquisition of new forms but L1 conceptual information specific to the word is copied over and in many cases 'fossilised'. In either model, this means that L1 or L2 forms should trigger the same underlying conceptual associations; idioms are therefore triggered because these concepts contribute to the overall figurative meaning (although it could be argued that this should only be the case for decomposable idioms). In effect, underlying conceptual associations are driving lexical activation in both languages, regardless of the specific form in which the idioms are presented.

Research into the types of semantic/associative priming that can be generated by L2 forms suggests that this may explain why results are seen for idioms but not other formulaic combinations such as collocations. Cross-language semantic priming has been very inconsistently reported in terms of what can be observed via masked/unmasked priming tasks, and the effects of direction (L1 to L2, L2 to L1 or both) and language dominance (Basnight-Brown & Altarriba, 2007; Chen & Ng, 1989; Singh, 2014; Wang, 2007; Williams, 1994). In particular, Williams and Cheung (2011; also de Groot & Nas, 1991) suggested that core semantic elements show cross-language priming, whereas more associative, schematic relationships are generally specific to either the L1 or L2, since they must be acquired through experience with a language specific word form combination. This is consistent with a more general view that the way in which concepts are mapped to word forms will not be equivalent in all languages (Bialystok, 2007; Pavlenko, 2009; Tokowicz & Degani, 2013), since L2 words tend to be associated with a more narrow range of semantic mappings because of the restricted contexts in which they are encountered (Finkbeiner & Nicol, 2003). Brysbaert and Duyck (2010) suggest that it is therefore important to make a distinction between language-dependent and language independent semantics. It seems like a plausible suggestion then that idioms will show cross language effects precisely because they tap into shared concepts, whereas other formulaic types, such as collocations, do not, since associations are more schematic and experiential, rather than representing core underlying conceptual links.

For idioms, it seems logical that the associations between words are more the result of conceptual mappings than simply having encountered the specific lexical combinations previously. Such a conclusion is based on the evidence discussed throughout this thesis: idioms are not particularly frequent (but are perhaps unusually salient), and do represent more than the ‘sum of their parts’, in that correct interpretation requires more than the straightforward combination of individual words, meaning that they exist as rich, complex semantic units, or at least as unitary entries in a conceptual sense. Hence if the concepts underpinning lexical items are shared, encountering the form of an idiom (even in the L2) should trigger the same underlying semantic/conceptual information. Again, this was not seen in Study 2 (Experiment 3), but was the case for the higher proficiency Swedish-English bilinguals in Study 3. Thus it appears that if participants are at a high enough level of proficiency, encountering L2 words can trigger L1 equivalent forms, leading to activation of a known lexical sequence, and underlying concepts, leading to conceptual mediation of a known idiom. Referring back to the modified dual route model presented in the discussion of Study 1, this means that non-native participants were able to access idioms directly, either via a translation mechanism or via direct conceptual mapping. In Study 3, Swedish-English bilinguals appear to access both routes at the same time: the results show that there was lexical facilitation leading to faster recognition, and conceptual mediation leading to unproblematic understanding of the phrase as a whole. In contrast, lower proficiency participants show evidence only of lexical activation, as in Study 2, where the final words were facilitated (Experiment 2) but the figurative uses of phrases were harder for participants to understand than the literal uses (Experiment 3), or no effect at all, in the case of translations of non-idiomatic collocations (Wolter & Yamashita, 2014).

The results of all four studies, and the theoretical implications discussed so far, address a number of important issues in the literature. I deal with these in turn in the following sections.

Formulaic processing in the L2

It has been suggested that L2 speakers have a more compositional approach to language processing than native speakers, since the literal meanings of individual words are more salient than idiomatic or figurative interpretations (Cieślicka, 2006, 2012; Cieślicka et al., 2014; Cieślicka & Heredia, 2011; Siyanova-Chanturia, Conklin & Schmitt, 2011). Importantly, the consistent priming effects found for translated idioms in Studies 1 to 3 argue against this. In all cases there was a clear processing advantage for 'correct' L1 forms, reflected in both response times (Study 1) and early reading behaviour (Studies 2 and 3). In contrast, even the highest proficiency non-natives (the Swedish-English bilinguals in Study 3) showed limited evidence that L2 only idioms were processed quickly. This supports the idea, raised in the discussion section of Chapter 4, that the difference is not in the approach but rather in the available resources, whereby non-native speakers simply do not have the lexical/associative links available to trigger fast processing of formulaic units. When such links exist, as in the case of combinations that are formulaic in the L1, formulaic processing is possible and appears to occur by default, at least in terms of the processing advantage for expected lexical forms. Previous studies (e.g. Jiang &

Nekrasova, 2007; Isobe, 2011; Underwood, Schmitt & Galpin, 2004) also suggest that for advanced non-native speakers, processing of L2 material does become more formulaic as proficiency increases, presumably reflecting the strengthening of links between items as a result of increased exposure. The results for the Swedish native speakers in Study 3 would support this, with some indication that L2 only idioms were processed quickly, at least in terms of figurative meaning activation and for the most well-known items. This suggests that as these participants were of a very high level of proficiency, they had begun to develop formulaic representations for at least some idioms in their L2.

This distinction between an apparent advantage for translated items and a limited advantage for L2 only idioms relates back to one of the claims of needs only analysis (Wray, 2002), that in the L1 links between component words need only be retained, while in L2 these must be built up from scratch. It is important to reiterate that such a process is presumably only the case for certain semantically anomalous items, such as compounds or idioms, so we must be careful how far we extend this to L1 acquisition in general, but it seems reasonable to suggest that this might at least partly explain the pattern of results seen for translated idioms and L2 idioms seen throughout Studies 1 to 3. A broader question relates to why it is that idioms should be acquired 'holistically' in the L1 in the first place. It is one thing to suggest that a limited set of idioms (e.g. *by and large*) or opaque compounds (e.g. *teddy bear*) might be initially acquired as unanalysed wholes during the process of L1 development (i.e. children hear and repeat idiomatic or non-compositional sequences without really understanding or analysing them at first). Whether or not these are then broken down further at a later point, the initial 'chunk' has been initially registered as a whole unit.

It seems unlikely, however, that this could be said for all idioms, especially given the relative lack of frequency that they show in natural language (i.e. many idioms might not be encountered until later in life, when a greater level of cognitive and linguistic awareness might mean that analysis is more likely). An alternative view may be that their non-compositional nature gives them an added degree of salience: the first encounter with a non-compositional phrase (and the consequent confusion or lack of understanding) carries a degree of novelty that renders it more salient and memorable than it would otherwise be, leading to better retention of the whole phrase. If we adopt this approach, however, then there is no reason that the same should not be true for non-native speakers: encountering a phrase like *kick the bucket* for the first time should be highly disruptive, even if the individual words are already known. Such disruption should in turn increase the degree of salience of the phrase, leading to better retention and formulaic processing for any subsequent encounters. It is beyond the scope of the present discussion to explore this, admittedly speculative, account of L1/L2 idiom acquisition in more detail, but the importance of the results of Studies 1 to 3 are clear: formulaic processing in the L2 is not fundamentally different than for native speakers, and any word combinations that are ‘known’ as formulaic sequences in L1 or L2 show similarly privileged lexical processing, regardless of the language of presentation. In general, this argues strongly against a view of idioms as ‘whole units’ in a lexical sense, and perhaps supports some of the hybrid models that advocate spreading activation from component words as the main driver of the idiom superiority effect (e.g. Sprenger et al., 2006).

L1 influence and the role of congruency

Whilst idioms have shown clear cross-language priming effects (e.g. in Studies 1 to 3), the results for other types of formulaic unit in previous studies have been inconsistent, with some finding facilitation for congruent collocations (e.g. Wolter & Gyllstad, 2011, 2013; Yamashita & Jiang, 2010) and others showing either no effect for translated collocations (e.g. Wolter & Yamashita, 2014) or variable results (Ueno, 2009). It is important to note that all of the studies showing cross-language facilitation of congruent collocations agree that L2 proficiency/exposure is a key driver of processing, hence both L1 knowledge and L2 specific experience contribute to the processing advantage. It is possible that when facilitation has been observed, the learners in these studies had simply come across the word combinations in English often enough to begin to form dedicated L2 representations, especially since the items used were fairly common collocations (e.g. *final year*, *good news*). However this cannot have been the case for the items in Wolter and Yamashita (2014) or Ueno (2009), since they were translations of L1 only items. If we compare the null result for translations of L1 only collocations in Wolter and Yamashita (2014) with the results of Studies 1 to 3, it may be that idioms represent more cohesive units than non-idiomatic collocations, and are more amenable to transfer even when no congruent L1 form has been seen before. Further, it is interesting to note that the findings of this thesis run contrary to the 'classic' findings by Kellerman (1977, 1983, 1986), where learners were found to be inherently unwilling to transfer idiomatic meanings from L1. Importantly, subsequent research, much of which was discussed in Chapter 6 (e.g. Charteris-Black, 2002; Irujo, 1986; Odlin, 1991) has called this claim into question. In particular, studies making use of comprehension based tasks seem to show that any available information is employed that can be used to aid processing, much like the approach that has been proposed for constraint based models for native speaker idiom processing (Libben & Titone, 2008; Smolka, Rabanus & Rösler, 2007; Titone & Connine, 1999). It therefore seems logical to suggest that all L1 idioms are potentially ‘congruent’, in the sense that all speakers are prepared to consider that L1 idioms may be transferable to the L2 in some cases. If the same word combination is encountered in a non-native-language, this (unconscious) assumption of possible equivalence is validated, and existing connections between words are used to facilitate processing.

It is also highly likely that task demands contributed to the variation in results. Amongst the studies that do find an effect of congruence, two used a phrase level acceptability task (Wolter & Gyllstad, 2013; Yamashita & Jiang, 2010), which encourages the kind of metalinguistic consideration that is not generally a part of ‘normal’ language processing, and one used a primed lexical decision task (Wolter & Gyllstad, 2011), akin to that used in Study 1. In comparison, Wolter and Yamashita (2014) used a double lexical decision task, where both words appeared on screen at the same time, one above the other. Each study provides a justification for the methodology chosen that is appropriate for the particular investigation, but such variation may well explain the different results. Again, the value of eye-tracking as an investigative technique is highlighted, since this allowed me to observe multiple aspects of form and meaning processing in the same task, and allowed me to present items in as natural a context as possible.

Comparing idioms to other formulaic types

An explanation worth exploring is whether we should concede some degree of ‘special’ formulaic status to idioms, most likely because they have such clear conceptual underpinnings. Alternatively, it may be that formulaicity exists at multiple levels of representation at the same time. This idea of multiple representation fits well with the model of the Heteromorphic Distributed Lexicon outlined in Wray (2002), whereby a range of formulaic word strings, components and morphemes are all stored to a greater or lesser degree, depending on the exact nature of the unit in question. It is also reflected in experimental evidence from e.g. Tremblay and Baayen (2010), who suggested that their ERP and behavioural results showed that lexical bundles were stored and accessed as a series of incremental parts *and* as whole units (i.e. they showed independent frequency effects for two, three and four word combinations).

On the suggestion that idioms are a special case, they seem to show effects that other types of formulaic sequence do not. As well as the translation effects seen throughout Studies 1 to 3, Experiments 5 and 6 (Chapter 7) suggested that idioms show a different kind of advantage from other formulaic types, both in their canonical forms and when the component words are split apart. This is likely due to idioms being formulaic on multiple levels, being conventionalised lexical sequences, having fixed structures *and* having conceptually whole entries. This means that even when the expected lexical frame is compromised, as in in Experiment 6, some degree of facilitation is still observed since an underlying concept exists to unify the component words.

In comparison, frequency defined collocations may be formulaic only at a lexical level. In support of such a contention, Biber (2009) suggested that collocation strength, as measured by traditional metrics like mutual information, was not a particularly good measure of formulaicity, but was instead simply a way of registering co-occurrences between lexical items. Since lexical co-occurrence is presumed to be an experiential and language-specific property, in line with the suggestion of Williams and Cheung (2011), compromising the canonical form of non-idiomatic collocations (either by breaking apart the components or by translating them) therefore removes the formulaicity. Binomials seem to fall somewhere in the middle: a conventionalised lexical template based on previous experience is available, just like collocations, but this is more often than not underpinned by a strong and central semantic relationship. This strong semantic relation underpins the facilitation seen for binomial pairs in non-formulaic configurations in Experiment 6, since the analysis showed that this rather than measures such as predictability drove the priming effect. It is also important to note that this facilitation of the component words occurred regardless of priming direction, which is an interesting finding, and different to idioms. This suggests that the effects in Experiment 6 for idioms, where priming only occurred in the forward configuration, and binomials, where priming occurred in both directions dependent on the strength of the semantic relationship, may be due to different underlying mechanisms.

Again, positing that formulaicity exists at multiple levels would explain these effects. Idioms represent formulaic configurations at a lexical level, meaning that in their citation form they are highly predictable (Libben & Titone, 2008; Tabossi et al., 2005); at a structural level, which may also be considered as the location of an abstract ‘superlemma’ specifying the relationships between components and any allowable transformations (Konopka & Bock, 2009; Sprenger et al., 2006); and at a conceptual level, which is assumed to be language non-specific. In contrast, binomials and collocations show lexical formulaicity (as seen in Experiment 5), because they represent frequently encountered forms. They may also have some level of structural formulaicity, especially in the case of binomials, where the default or at least conventional ordering of components is specified. However, since both types are compositional, whereby the overall semantic meaning of the whole is simply the combination of the parts rather than denoting a separate concept, no conceptual level of formulaicity exists (as seen in Figure 7.1), other than in the sense that individual concepts are linked via standard semantic and associative connections.

For idioms, this multi-level distribution means that they are robust and highly flexible, at least in terms of the ways in which the component words can be maintained and recombined. Interestingly, this runs counter to the limited amount of other studies that have shown no processing advantage for idioms in non-canonical configurations (e.g. McGlone et al., 1994). One explanation may again be the nature of the task used. McGlone et al. (1994) used a self-paced line-by-line reading task, hence their reading times were based on fairly broad measure (the whole phrase and the surrounding material). Experiment 6, on the other hand, was able to specifically consider the effect of seeing a formulaic ‘partner’ on the reading time for a single word later in the sentence, and showed a significant degree of lexical facilitation. Such a discrepancy in how the data were measured may well explain the difference in conclusions between these two studies.

Encountering the first word of an idiom therefore generates multiple expectancies: a lexical expectancy of a possible continuation, especially if the preceding context makes this likely; a structural expectation based on an underlying formulaic frame, specified at the superlemma level (any known structure of which the word is a part); and a conceptual expectation based on all possible meanings of the word, i.e. encountering *spill* generates consideration of all possible meanings, including the idiomatic "reveal a secret". Cacciari (2014) suggests that such a view is unrealistic, given the number of idioms that start with the same initial words (e.g. she states that the Dictionary of American Idioms lists 132 idioms that start with the verb *take*). However, similar processes occur in the resolution of ambiguous/polysemous words (Swinney, 1979; Lupker, 2009) whereby multiple candidates are activated from the earliest stages of processing, so it is not unreasonable to extend this to larger units, especially considering the additional contribution of frequency and context in this process. Certainly the results in Studies 3 and 4 support this, since idioms with no recognition point prior to offset and no preceding contextual bias showed a clear and consistent advantage for native speakers. Consideration of possible idiomatic continuations based on the first word therefore seems to be a plausible explanation. Importantly, in constructing the stimulus items in all studies in this thesis I tried to avoid generic and 'light' verbs like *make* or *do*, and the vast majority of idioms in all studies contained lexical verbs that form part of only a limited set of formulaic items. For idioms containing 'light' verbs, it may well be that Cacciari's (2014) objection is correct, but these may actually represent a relatively small and atypical set of items.[26]

Related to the differences in how formulaic types are represented at multiple levels, it is worth revisiting the findings in Experiment 6, where separation of formulaic components showed facilitation for idioms but not other units (notwithstanding the semantic effects seen for binomials). An area to explore here is why adjective-noun

[26] Wittenberg and Piñango (2011) provide results showing that light verb constructions actually incur a processing cost compared to non-light counterparts. They suggest that this is evidence for underspecification, i.e. that light verbs form a particular subset of the lexical entry for any given verb. They conclude that light verb interpretation is built in real time and is dependent on many factors, including context, hence it is not the same as the 'retrieval' of other idioms making use of non-light verbs. For this reason, the objection raised by Cacciari might not apply to the majority of idioms that do not include light or otherwise delexicalised verbs.

collocations such as *abject poverty* (which in theory should be easily separable) showed no such facilitation, and even showed some evidence of disruption when separated. One helpful avenue to explain this might be the idea of concgrams, introduced by Cheng and colleagues (Cheng, Greaves & Warren, 2006; Cheng, Greaves, Sinclair & Warren, 2009). A concgram is defined as a co-occurring set of words that is not necessarily fixed in terms of direction or constituency, i.e. a word pair AB could occur in the reversed form (BA) or in non-contiguous contexts (A…B) and still be considered to be a formulaic unit. The example given in Cheng et al. (2009, based on a discussion in Sinclair, 2007) is *hard work*, which collocates in the sense that the words co-occur very often but not always in this canonical sequence (i.e. *hard work*, *work hard* and *work really hard* would all be variations of the same concgram, although it should be noted that the different variants may not all denote exactly the same sense, i.e. *hard work* and *work hard* are related but don't necessarily mean the same thing: one can *work hard* at something without it being *hard work*, or something can be *hard work* without someone *working hard* at it). Analysis of this collocation as only instances of *hard work* would therefore be deficient in a number of ways, since it would discount viable variants and also restrict the parameters of the semantic domain.

The importance of this to Study 4 is clear: if we consider and define formulaic sequences as only those that are reliably fixed in a non-variable frame, we are removing a lot of the natural flexibility and creativity that surrounds the use of multi-word sequences. In the case of the collocations in Experiment 6, certainly some can be used in separated contexts (in particular the adjective-noun combinations, such as *abject poverty*, which could be realised in forms such as *the poverty was abject* or *abject levels of poverty*). Other items, in particular the noun-noun combinations (*post office*, *light bulb*) as well as some of the adjective-noun items (*ancient history*) are harder to transform without changing the meaning. Since collocations here were defined primarily by frequency of the canonical forms, the flexibility or otherwise of each item was not included as a selection criteria. It seems likely that if we were to subdivide items into fixed collocations (where either reversing the form or inserting words effectively changes the meaning) and concgrams (that allow for AB, BA, A….B and B….A configurations) then a different set of results might emerge, although it would be important to take into account the relative frequencies of different realisations for any given item. The same might also be true of idioms, with those that can reasonably be passivised showing a different pattern to those which never occur in a passive form, or those that often undergo insertion of additional adjectives showing a greater degree of delayed priming than those that are predominantly used in a single fixed form. It may also be that compositional binomials, which are inherently amenable to reversal, showed bidirectional priming effects in Experiment 6 for precisely this reason. Certainly this would be an avenue to consider further in future studies as a way to explore the complex interplay of syntactic flexibility and semantic integrity in formulaic sequences.

Again, it seems likely that multiple levels of formulaicity may be at play. Previously encountered sequences of words show clear lexical effects in their canonical frames (true for the idioms, binomials and collocations in Experiment 5). The structural level of formulaicity may manifest in how it restricts what transformations are possible, meaning that binomials, with their default ordering of components, and collocations, which in many cases are not separable without compromising the meaning, are more restricted in how they can be deployed than idioms. Once again, the fact that idioms are also formulaic at an underlying conceptual level means that they can be accessed via the deepest level of formulaicity, so even in non-standard forms they remain accessible in a way that non-idiomatic formulaic types may not be.

An as yet undiscussed area of relevance to the results presented throughout this thesis is the contribution of the right hemisphere of the brain, which is often implicated in formulaic and figurative language processing. A long-standing finding in impairment studies is that left hemisphere damage leads to impaired single word and syntactic processing with relatively spared processing of idioms and figurative language, while right hemisphere damage tends to lead to relatively unimpaired combinatorial language abilities (syntax, single word processing) and instead difficulty with appropriate use of formulaic expressions (Van Lancker & Kempler, 1987; Van Lancker Sidtis, 2004, 2012a). One important element of this may be that while the left hemisphere activates highly salient, core meanings of single words, the right hemisphere is thought to activate much coarser, more distant semantic relationships (Faust & Mashal, 2007; Jung-Beeman, 2005). A distributed view of how idioms are represented (Cacciari & Tabossi, 1988; Holsinger, 2013) characterises idioms as closely interconnected sets of single words, rather than strictly speaking 'whole units' at a lexical level. In this regard, it may be the case that preserved processing of idioms and other formulaic expressions in right hemisphere damaged patients reflects a relatively unimpaired ability to make such diverse and tangential connections, while the more 'central' connections between words that form a part of core linguistic processing in the left hemisphere are impaired. Certainly consideration of the pragmatic importance of idioms (and other formulaic units, such as speech formulae and other social/interpersonal routines) would further implicate right hemisphere involvement in activation and processing.

One problem with using the fine/coarse coding view to explain the current findings and others in the idiom literature is that right hemisphere associations are thought to be activated more slowly and maintained over a longer period of time (Faust & Mashal, 2007), so the speed advantage seen so consistently in the idiom literature would be difficult to explain. Still, converging evidence from impairment studies and investigations such as those presented here can offer a valuable new direction in our study of formulaic language and the complex ways in which it is realised at multiple levels in the mental lexicon. This offers a number of ways to extend studies of formulaic language, in particular Study 4. For example, if idioms, binomials and collocations show fundamentally different properties, we might expect left and right hemisphere damaged patients to show differential patterns of performance, reflecting the different levels at which formulaicity is realised in the brain.

8.2 Limitations and future directions

It is important to recognise the limitations of research endeavors, including in the studies presented here. As with any consideration of a broad topic, there are areas of both the formulaic language and bilingualism literature that have received limited attention. I have tried to focus on the issues in both areas that are the most central to my research. However it is important to note that the findings of the empirical chapters should be situated in a much wider context.

Just as the literature on the topic is quite extensive, formulaic language itself covers a broad range of phenomena, and this thesis has concentrated only on a limited class of multi-word units (idioms, then binomials and collocations). Importantly, the distinctions between different types of formulaic language are not always clear cut. In particular, the definitions for some of the classifications in Experiments 5 and 6 (Chapter 7) may leave room for disagreement. For example, some of the collocations might better be considered as spaced compounds (e.g. *post office*), which may display particular qualities of their own (e.g. Cutter et al., 2014). As with any study of formulaic language, formulaic types need to be clearly demarcated for the purposes of comparison because, as is apparent throughout the literature (e.g. Titone et al., 2015; Van Lancker Sidtis, 2012a; Wray, 2002), formulaicity exists on multiple continua at once. Thus, it is important to consider the many and varied characteristics in the design of any large scale study. By doing this, findings as they relate to lexically fixed, highly predictable word combinations can be integrated into wider models of formulaic language.

Further, multiple semantic properties have been shown in the idiom literature to make an important contribution to processing. It is possible that they make a much more important contribution than I have accounted for in this thesis. For example, several researchers make clear distinctions between idiom types in terms of decomposability (Cacciari & Glucksberg, 1991; Gibbs & Nayak, 1989; Gibbs, Nayak & Cutting, 1989), with clear predictions about what implications this has for processing. Other studies (e.g. Libben & Titone, 2008) distinguish multiple factors such as global decomposability, individual noun and verb decomposability, transparency and meaningfulness, suggesting that many potential semantic factors could come into play. With this in mind, it could be that the fairly simple measure of compositionality used in my studies is not sensitive enough to account for the full range of semantic properties that can influence idiom processing. However, a consistent effect was that compositionality, as it was assessed here, was relatively unimportant in terms of lexical recognition (i.e. the initial facilitation seen for an expected word over an unexpected word), and only in later processes such as integration of overall meaning into context did this variable become significant. Broadly, this agrees with the prior literature, where semantic factors are only important when a task explicitly requires a semantic judgement (e.g. Libben & Titone, 2008). Future studies might usefully consider compositionality in a more detailed way (with the caveat that the more detailed the consideration of such factors, the more complex the experimental design becomes), but my results here suggest that its contribution to the underlying processes of idiom recognition and activation of known forms is limited. Other factors such as transparency and the semantic contributions of individual words might reveal more detailed patterns, but it is beyond the scope of this thesis to explore these in any more detail here.

On the bilingualism side, formulaic language represents an underexplored aspect of the bilingual processing literature (notwithstanding the interest in applied linguistic research and language pedagogy). An important conclusion with regards to bilinguals is that the current findings clearly implicate both L2 experience and underlying L1 knowledge in how idioms are processed. To my knowledge, translating formulaic language as a way to investigate bilingual processing has only been done in the few studies discussed in Chapters 4, 5 and 6 (e.g. Ueno, 2009; Wolter & Yamashita, 2014), and more often studies have focused on the influence of first language patterns, in particular as this relates to how congruent items are processed in the L2 (e.g. Titone et al., 2015; Wolter & Gyllstad, 2011, 2013; Yamashita & Jiang, 2010). Using incongruent translations provides a fruitful way to build on such findings, as it unequivocally shows that L1 knowledge is a vital aspect of how words are processed in the L2. However, given the lack of comparable studies, one of the challenges when approaching this topic from the perspective taken in this thesis has been that little theoretical guidance is available, other than what can be inferred from the bilingual single word processing literature and the literature on monolingual formulaic language. This is apparent in a number of areas, not least in the use of eye-tracking to investigate multi-word units, as discussed in Chapter 3. A major contribution of this thesis is therefore to propose new directions for research into formulaic language and particularly its status in the bilingual lexicon, both in terms of a methodology for data analysis and in terms of a model for how it is represented across languages.

Other limitations relate more specifically to the methods used and the nature of the investigations. While the results show a clear pattern for visual presentation of highly familiar idioms, they can say little about how this might manifest in other modalities (auditory presentation or production), where different cues (acoustic features, prosodic contours) might contribute to recognition and processing. However, the idiom literature in general has considered processing in a range of contexts, and often findings are congruent across methodologies, so there seems good reason to assume that the effects seen here would hold in other presentation modes and experimental tasks. Similarly, as summarised by Wray (2009), task demands in experimental situations may tap into only a small part of the true processing that underpins natural language. Hence formulaicity as it is used to support fluency in real-world online communication may play only a limited role in experimentally elicited processing, and any results should therefore be considered accordingly. Finding ways of overcoming all of these limitations to build a more integrated and complete model of the formulaic lexicon is a fruitful avenue for future exploration. In this regard eye-tracking was a highly appropriate choice of methodology since it represents one of the most natural ways to collect linguistic data. That is, it requires participants to read sentences in a relatively natural fashion as opposed to requiring something like a button press in response to a meta-linguistic judgement. It also allows for multi-word units to be presented as continuous units as they would be encountered under 'normal' real-world circumstances, rather than word by word, which would be required in an ERP study.[27]

8.3 Final conclusions

This thesis has presented evidence against the fixed and recurrent nature of formulaic language being the sole driver of its fast processing. In the case of idioms, frequency cannot be the main reason that they are recognised and understood quickly. Study 4 exemplifies this: idioms had the lowest mean frequency of all formulaic units, yet showed the largest formulaic advantage relative to control phrases. Equally, Studies 1 to 3 show that even in translated forms, which are necessarily new and unfamiliar on a formal level, idioms show a clear and consistent processing advantage. They are recognised more quickly than comparable control sequences, and in the case of higher proficiency Swedish-English bilinguals, they are understood effortlessly in sentence contexts. All of this suggests that there is something underlying the associations of

[27] In ERP studies involving visual language processing it is generally the case the material is presented one word at a time in a process called rapid serial visual presentation (RSVP). This is necessary to isolate the electrical signal that is a response to each word from the signals relating to things like eye-movements, which would contaminate the signal if material was presented in a 'normal' sentence context. Whilst ERP studies are therefore very useful, they necessarily split formulaic sequences into component words, which might be considered a major limitation.

words that is not language specific and which is not simply learned through multiple repeated encounters, and that idioms represent a deeper, more robust connection between component words than certain other types of word combination.

Returning to how formulaic language is best accommodated within general theories of language organisation, a usage-based account provides the most flexible way to accommodate the data. Results from the first part of Study 4 show unequivocally that in their citation forms, idioms, binomials and collocations, as they are defined here, show quantitatively faster processing than equally plausible control phrases, even when nothing in the preceding sentence context exists to bias a formulaic continuation. There is clearly some way in which these particular lexical combinations are stored in memory, with the fact that they are fundamentally ‘known’ from previous encounters being their primary defining feature. However, Pinker and Ullman (2002, p.462) make the following assertion, with which it is difficult to disagree: “Nothing in linguistics prevents theories from appealing to richer conceptions of memory than simple rote storage. Neither does neural network modelling prohibit structured or abstract representations, combinatorial operations, and subsystems for different kinds of computation.” Idioms seems to exemplify this perfectly: they are examples of rote storage in some cases, where an arbitrary sequence of words is memorised and possibly never considered in an analytical way, but they also exhibit rich and pervasive semantic, structural and pragmatic properties that contribute to their cohesion and robust representation in the monolingual and bilingual lexicon.

References

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

Allen, D. & Conklin, K. (2013). Cross-linguistic similarity and task demands in Japanese-English bilingual processing. *PLoS ONE*, 8(8), 1-14.

Allen, D., Conklin, K. & Van Heuven, W. (2015). Making Sense of the Sense model: Translation priming with Japanese-English bilinguals. *The Mental Lexicon*, 10(1), 32-52.

Altarriba, J., Kroll, J., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory and Cognition*, 24(4), 477-492.

Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, 16, 285-311.

Arcara, G., Lacaita, G., Mattaloni, E., Passarini, L., Mondini, S., Beninca, P., & Semenza, C. (2012). Is "hit and run" a single word? The processing of irreversible binomials in neglect dyslexia. *Frontiers in Psychology*, 3(11), 1-11.

Arnon, I., & Cohen-Priva, U. (2013). More than Words: The Effect of Multi-word Frequency and Constituency on Phonetic Duration. *Language and Speech*, 56(3), 349-371.

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

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology*, 58A, 1065-1086.

Ashby, M. (2006). Prosody and idioms in English. *Journal of Pragmatics*, 38(10), 1580-1597.

Baayen, R. H. (2009). languageR: Data sets and functions with "Analysing linguistic data: A practical introduction to statistics". R Package version 0.955.

Baldwin, T., & Nam Kim, S. (2010). Multiword Expressions. In N. Indurkhya & F. Damerau (Eds.), *Handbook of Natural Language Processing* (2nd Edition), pp. 267-292. Boca Raton/London: CRC Press.

Bannard, C., & Lieven, E. (2009). Repetition and Reuse in Child Language Learning. In R. Corrigan, E. Moravcsik, H. Ouali & K. Wheatley (Eds.), *Formulaic Language: Volume II: Acquisition, Loss, Psychological reality, Functional Explanations*, pp. 297-321. Amsterdam: John Benjamins.

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

Barfield, A. & Gyllstad, H. (2009) (Eds*). Researching collocations in another language – Multiple interpretations*. Basingstoke and New York: Palgrave Macmillan.

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

Basnight-Brown, D., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & Cognition*, 35(5), 953-965.

Bates, D. M., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R., Singmann, H., & Dai, B. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-7.

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

Bialystok, E. (2007). Cognitive Effects of Bilingualism: How Linguistic Experience Leads to Cognitive Change. *International Journal of Bilingual Education and Bilingualism*, 10(3), 210-223.

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

Blais, M.-J., & Gonnerman, L. M. (2013). Explicit and implicit semantic processing of verb–particle constructions by French–English bilinguals. *Bilingualism: Language and Cognition*, 16(4), 829-846.

Bley-Vroman, R. (2002). Frequency in production, comprehension, and acquisition. *Studies in Second Language Acquisition*, 24, 209–213.

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

Bod, L. (2001). *Sentence memory: Storage vs. computation of frequent sentences*. Paper presented at the CUNY 2001.

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

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

Bonk, W., & Healy, A. (2005). *The company words keep: Priming effects without semantic or associative links through collocation*. Paper presented at the 46th Annual Meeting of the Psychonomics Society, Toronto.

Brenders, P., van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology*, 109(4), 383-396.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition*, 13(03), 359-371.

Bulut, T. & Çelik-Yazici, I. (2004). Idioms processing in L2: through rose-colored glasses. *The Reading Matrix*, 4, 105-116.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261-290.

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

Bybee, J. (2008). Usage-based Grammar and Second Language Acquisition. In P. Robinson, & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, pp. 216–236. New York: Routledge.

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 degrees of constituency: The reduction of don't in English. *Linguistics*, 37, 575-596.

Cacciari, C. (2014). Processing multiword idiomatic strings: Many words in one? *The Mental Lexicon*, 9(2), 267-293.

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

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, 27(6), 668-683.

Caillies, S., & Butcher, K. (2007). Processing of Idiomatic Expressions: Evidence for a New Hybrid View. *Metaphor and Symbol*, 22, 79–108.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, 56(1), 103-128.

Cappelle, B., Shtyrov, Y., & Pulvermuller, F. (2010). Heating up or cooling up the brain? MEG evidence that phrasal verbs are lexical units. *Brain and Language*, 115(3), 189-201.

Carrol, G., & Conklin, K. (2014a). Eye-tracking multi-word units: some methodological questions. *Journal of Eye Movement Research*, 7(5), 5, 1-11.

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

Carrol, G., & Conklin, K. (2015). Cross language priming extends to formulaic units: Evidence from eye-tracking suggests that this idea "has legs". *Bilingualism: Language and Cognition*, published online by Cambridge University Press, April 20th 2015. DOI: 10.1017/S1366728915000103.

Carrol, G. & Conklin, K. (under review). No word is an island: Exploring the mechanisms of formulaic language processing.

Carrol, G., Conklin, K., & Gyllstad, H. (under review). Found in translation: The influence of L1 on the processing of idioms in L2.

Carroll, P., & Slowiaczek, M. (1986). Constraints on semantic priming in reading: A fixation time analysis. *Memory and Cognition*, 14(6), 509-522.

Chafe, W. (1968). Idiomaticity as an Anomaly in the Chomskyan Paradigm. *Foundations of Language*, 4(2), 109-127.

Charteris-Black, J. (2002). Second Language Figurative Proficiency: A Comparative Study of Malay and English. *Applied Linguistics*, 23, 104–133.

Chen, H., & Ng, M. (1989). Semantic facilitation and translation priming effects in Chinese–English bilinguals. *Memory and Cognition*, 17, 454–462.

Chen, X., Tang, Y., Xiao, Z., Wang, J., Zhang, M., Zhang, Z., & Zhang, T. (2010). N400 elicited by incongruent ending words of Chinese idioms in healthy adults. *Chinese Medical Journal*, 123(6), 686-689.

Cheng, W., Greaves, C., & Warren, M. (2006). From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics*, 11(4), 411-433.

Cheng, W., Greaves, C., Sinclair, J. M., & Warren, M. (2009). Uncovering the Extent of the Phraseological Tendency: Towards a Systematic Analysis of Concgrams. *Applied Linguistics*, 30(2), 236-252.

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

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

Chomsky, N. (1995). Language and Nature. *Mind*, 104, 1-61.

Chung, K., Code, C., & Ball, M. (2004). Lexical and non-lexical speech automatisms in aphasic Cantonese speakers. *Journal of Multilingual Communication Disorders*, 2, 32–42.

Cieślicka, A. B. & Heredia, R. R. (2011). Hemispheric asymmetries in processing L1 and L2 idioms: Effects of salience and context. *Brain and Language*, 116, 136-150.

Cieślicka, A. B. (2002). Comprehension and Interpretation of Proverbs in L2. *Studia Anglica Posnaniensia: An International Review of English Studies*, 37, 173-200.

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

Cieślicka, A. B. (2013). Do nonnative language speakers chew the fat and spill the beans with different brain hemispheres? Investigating idiom decomposability with the divided visual field paradigm. *Journal of Psycholinguistic Research*, 42, 475–503.

Ciéslicka, A. B., & Heredia, R. R. (2011). Hemispheric asymmetries in processing L1 and L2 idioms: Effects of salience and context. *Brain and Language*, 116, 136-150.

Ciéslicka, A. B., Heredia, R. R. & Olivares, M. (2014). It's All in the Eyes: How Language Dominance, Salience, and Context Affect Eye Movements During Idiomatic Language Processing. In M. Pawlak & L. Aronin (Eds.), *Essential Topics in Applied Linguistics and Multilingualism*, pp. 21-41. Switzerland: Springer.

Clark, H. H. (1973). The language as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335–359.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye Movements in Reading Words and Sentence. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye Movements: A Window on Mind and Brain*, pp. 341-372. Oxford: Elsevier.

Code, C. (1994). Speech automatisation production in aphasia. *Journal of Neurolinguistics*, 8, 135–148.

Code, C. (2005). First in, last out? The evolution of aphasic lexical speech automatisms to agrammatism and the evolution of human communication. *Interaction Studies*, 6, 311-334.

Colombo, L. (1993). The comprehension of ambiguous idioms in context. In C. Cacciari & P. Tabossi (eds.), *Idioms: Processing, structure, and interpretation*, pp. 163–200. Hillsdale, NJ: Erlbaum.

Colombo, L. (1998). Role of Context in the Comprehension of Ambiguous Italian Idioms. In D. G. Hillert (Ed.), *Sentence Processing: A Cross-Linguistic Perspective*, Volume 31, pp.405-425. New York: Academic Press.

Columbus, G. (2010). Processing MWUs: Are MWU Subtypes Psycholinguistically Real? In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication*, pp. 194-210. London: Continuum.

Conklin, K., & Schmitt, N. (2008). Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics*, 29, 72-89.

Conklin, K., & Mauner, G. (2005). *Investigating Bilingual Lexical Access: Processing French-English Homographs in Sentential Contexts*. Paper presented at the The 4th International Symposium on Bilingualism, Arizona State University.

Conklin, K., & Schmitt, N. (2012). The Processing of Formulaic Language. *Annual Review of Applied Linguistics*, 32, 45-61.

Costa, A., Santesteban, M., & Cano, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, 94(1), 94-103.

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory and Cognition*, 30(6), 958-968.

Cronk, B., & Schweigert, W. (1992). The comprehension of idioms: The effects of familiarity, literalness, and usage. *Applied Psycholinguistics*, 13, 131-146.

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2014). Preview benefit in English spaced compounds. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 40, 1778–1786.

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, 57–71.

Dagut, M. & Laufer, B. (1985). Avoidance of Phrasal Verbs - A Case for Contrastive Analysis. *Studies in Second Language Acquisition*, 7, 73-80.

Dahl, R. (1953) The Great Automatic Grammatizor. In R. Dahl, *Someone Like You*. US: Knopf.

Dahlmann, I. (2009). *Towards a multi-word unit inventory of spoken discourse*. Unpublished PhD thesis, University of Nottingham.

Davies, M. (2004). *BYU-BNC*. (Based on the British National Corpus from Oxford University Press) [Online]. Available at: http://corpus.byu.edu/bnc/.

De Cat, C., Klepousniotou, E., & Baayen, R. H. (2015). Representational deficit or processing effect? An electrophysiological study of noun-noun compound processing by very advanced L2 speakers of English. *Frontiers in Psychology*, 6, 77, 1-17.

De Groot, A. & Nas, G. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, 30, 90–123.

De Groot, A., Delmaar, P., & Lupker, S. (2000). The processing of interlingual homographs in a bilingual and multilingual task: support for non-selective access to bilingual memory. *Quarterly Journal of Experimental Psychology*, 53, 397-428.

De Rosario-Martinez, H. (2013). Phia: Post-hoc interaction analysis. R package version 0.1-5.

Dijkstra, T. (2007). The multilingual lexicon. In M.G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics*, pp. 251-265. Oxford: Oxford University Press.

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

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*, 42, 445-464.

Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, 1, 51-66.

Dimitropoulou, M., Dunabeitia, J. A., & Carreiras, M. (2011). Masked translation priming effects with low proficient bilinguals. *Memory and Cognition*, 39(2), 260-275.

Duffy, S., Morris, R., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429-446.

Duñabeitia, J., Perea, M., & Carreriras, M. (2007). The role of the frequency of constituents in compound words: Evidence from Basque and Spanish. *Psychonomic Bulletin & Review*, 14(6), 1171-1176.

Duñabeitia, J., Perea, M., & Carreiras, M. (2010). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology*, 57, 98-107.

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

Durrant, P. (2008). *High Frequency Collocations and Second Language Learning*. Ph.D. dissertation, University of Nottingham.

Durrant, P., & Doherty, A. (2012). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, 6(2), 125-155.

Ehrlich, S., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20, 641-655.

Ellis, N. C. (1994). Vocabulary acquisition: The implicit in and outs of explicit cognitive mediation. In N. C. Ellis (Ed.), *Implicit and Explicit Learning of Language*, pp. 211-282. London: Academic Press.

Ellis, N. C. (2008). Phraseology: The periphery and the heart of language. In F. Meunier & S. Granger (eds.), *Phraseology in language learning and teaching*, pp. 1-13. Amsterdam: John Benjamins.

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

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

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade gener-ation 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.

Estes, Z., & Jones, L. L. (2009). Integrative priming occurs rapidly and uncontrollably during lexical processing. *Journal of Experimental Psychology: General*, 138(1), 112-130.

Faust, M., & Mashal, N. (2007). The role of the right cerebral hemisphere in processing novel metaphoric expressions taken from poetry: a divided visual field study. *Neuropsychologia*, 45(4), 860-870.

Federmeier, K. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44, 491-505.

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.

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, 24, 369-383.

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

Forster, K., & Jiang, N. (2001). The nature of the bilingual lexicon: Experiments with the masked priming paradigm. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 72-83). Malden, MA: Blackwell.

Foster, P. (2001). Rules and routines: A consideration of the role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*, pp. 75-95. Harlow: Longman.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science*, 6, 78–84.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31(5), 862-877.

Garrod, S., O'Brien, E., Morris, R., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 250-257.

Gibbs, J. R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory and Cognition*, 8, 149–156.

Gibbs, R. W. (1991). Semantic Analyzability in children's understanding of idioms. *Journal of Speech and Hearing Research*, 34, 613–620.

Gibbs, J. R. W. (2006). Metaphor Interpretation as Embodied Simulation. *Mind and Language*, 21(3), 434-458.

Gibbs, J. R. W., Bogdanovich, J., Sykes, J., & Barr, D. (1997). Metaphor in Idiom Comprehension. *Journal of Memory and Language*, 37, 141-154.

Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to Kick the Bucket and Not Decompose: Analyzability and Idiom Processing. *Journal of Memory and Language*, 28, 576–593.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, 8, 183–206.

Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics*, 31, 919-929.

Giora, R. (2002). Literal vs. figurative language: Different or equal? *Journal of Pragmatics*, 34, 487-506.

Giora, R. (2003). *On Our Mind: Salience, context and figurative language*. Oxford: Oxford University Press.

Glucksberg, S. (1993). Idiom meaning and allusional content. In C. Cacciari and P. Tabossi (eds.), *Idioms: Processing, Structure and Interpretation*, pp. 3-26. New Jersey: Lawrence Erlbaum.

Google Translate. Available from: http://translate.google.co.uk/. Retrieved 17/10/2012 (Study 1)

Google Translate. Available from: http://translate.google.co.uk/. Retrieved 10/06/2013 (Study 2)

Grant, L., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics*, 25, 38–61.

Grant, L., & Nation, P. (2006). How many idioms are there in English? *International Journal of Applied Linguistics*, 151, 1–14.

Gyllstad, H., & Wolter, B. (in press). Collocational processing in the light of a phraseological continuum model: Does semantic transparency matter? *Language Learning*.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439-483.

Hamblin, J., & Gibbs, J. R. W. (1999). Why You Can't Kick the Bucket as You Slowly Die: Verbs in Idiom Comprehension. *Journal of Psycholinguistic Research*, 28(1), 25–39.

Hargevik, S. & Ljung, M. (1989). *Say when och 1.284 andra engelska idiom* [Say when and 1,284 other English idioms]. Solna: Esselte studium.

Hillert, D. G. & Swinney, D. (2001). The processing of fixed expressions during sentence comprehension In A. Cienki, B. Luka & M. Smith (Eds.), *Conceptual Structure, Discourse and Language*, pp. 107-122. Stanford, CA: CSLI Publications.

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Holsinger, E. & Kaiser, E. (2010). *Effects of Context on Processing (non)-Compositional Expressions*. 84th Annual Meeting of the Linguistics Society of America Baltimore, MD.

Holsinger, E., & Kaiser, E. (2013). Processing (Non)Compositional Expressions: Mistakes and Recovery. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 39(3), 866-878.

Holsinger, E. (2013). Representing Idioms: Syntactic and Contextual Effects on Idiom Processing. *Language and Speech*, 56, 373-394.

Hopper, P. (1987). Emergent Grammar. *Berkeley Linguistics Society*, 13, 139-157.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: does cross-language activation survive a change of script? *Cognition*, 106(1), 501-511.

Huang, M. (2009). Solving the riddle of metaphor: A salience-based model for metaphorical interpretation in a discourse context. . In V. Evans & S. Pourcel (Eds.), *New directions in cognitive linguistics*, pp. 107-126. Amsterdam: John Benjamins.

Hübinette, L. & Odenstedt, B. (1988). *Ord och inga visor – 2000 svenska idiom i engelsk översättning* [Straight talking – 2000 Swedish idioms with English translations]. Lund: Studentlitteratur

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

Inhoff, A. (1984). Two Stages of Word Processing during Eye Fixations in the Reading of Prose. *Journal of Verbal Learning and Verbal Behavior*, 23, 612-624.

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

Irujo, S. (1986). A piece of cake: learning and teaching idioms. *ELT Journal*, 40, 236–242.

Irujo, S. (1993). Steering clear: Avoidance in the production of idioms. *IRAL*, 31, 205–219.

Isobe, Y. (2011). Representation and Processing of Formulaic Sequences in L2 Mental Lexicon: How Do Japanese EFL Learners Process Multi-word Expressions? *JACET Kansai Journal*, 13, 38–49.

Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E. van der Linden, A. Schenk & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives*, pp. 133-166. Hillsdale, NJ: Erlbaum.

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

Jarema, G. (2006). Compound representation and processing: a cross-language perspective. In G. Libben & G. Jarema (Eds.), *The Representation and Processing of Compound Words*, pp. 45-70. Oxford: Oxford University Press.

Jiang, N. (2000). Lexical Representation and Development in a Second Language. *Applied Linguistics*, 21(1), 47-77.

Jiang, N., & Forster, K. I. (2001). Cross-Language Priming Asymmetries in Lexical Decision and Episodic Recognition. *Journal of Memory and Language*, 44(1), 32-51.

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

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2013). *Meaning overrides frequency in idiomatic and compositional multiword chunks*. Paper presented at the 35th Annual Conference of the Cognitive Science Society, Austin, Texas.

Jones, L. L., & Estes, Z. (2012). Lexical priming: Associative, semantic, and thematic influences on word recognition. In S. Adelman (Ed.), *Word Recognition* (Vol. 2), pp. 44-72. Hove: Psychology Press.

Juhasz, B. J. (2007). The influence of semantic transparency on eye movements during English compound word recognition. In R. Van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain*, pp. 373-389. Amsterdam: Elsevier.

Juhasz, B. J., Pollatsek, A., Hyönä, J., Drieghe, D., & Rayner, K. (2009). Parafoveal processing within and between words. *The Quarterly Journal of Experimental Psychology*, 62, 1356 –1376.

Juhasz, B. J., Starr, M., Inhoff, A., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*, 94, 223-244.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Science*, 9(11), 512-518.

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

Kecskés, I. (2000). A cognitive-pragmatic approach to situation-bound utterances. *Journal of Pragmatics*, 32, 605-625.

Kellerman, E. (1977). Towards a characterization of the strategy of transfer in second language learning. *Interlanguage Studies Bulletin*, 2, 58-145.

Kellerman, E. (1983). Now you see it, now you don't. In S. Gass, & L. Selinker (Eds.), *Language transfer in language learning*, pp. 112–134. Rowley, MA: Newbury House.

Kellerman, E. (1986). An eye for an eye: crosslinguistic constraints on the development of the L2 lexicon. In E. Kellerman & M. Sharwood Smith (Eds.), *Crosslinguistic Influence in Second Language Acquisition*, pp. 35-48. Oxford: Pergamon Press.

Kellerman, E. (1987). *Aspects of Transferability in Second Language Acquisition*. PhD dissertation, University of Nijmegen.

Kiss, G., Armstrong, C., Milroy, R., & Piper, J. (1973*). An associative thesaurus of English and its computer analysis*. Edinburgh: Edinburgh University Press.

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

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

Kroll, J., & de Groot, A. (1997). Lexical and conceptual memory in the bilingual: mapping form to meaning in two languages. In A. de Groot & J. Kroll (Eds.), *Tutorials in bilingualism: psycholinguistic perspectives*, pp. 169-199. Mahwah, NJ: Lawrence Erlbaum.

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.

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

Kroll, J. F., van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Biling (Camb Engl)*, 13(3), 373-381.

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

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, 23(7-8), 1089-1132.

Kutas, M., & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, 4, 463-470.

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

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2014). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-11.

Laufer, B. (2000). Avoidance of Idioms in Second Language: The Effect of L1-L2 Degree of Similarity. *Studia Linguistica*, 54, 186-196.

Laufer, B. & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language Learning*, 54, 3, 399-436.

Levelt, W. (1989). *Speaking: From Intention to Articulation*. Massachusetts: MIT Press.

Libben, G. (1998). Semantic Transparency in the Processing of Compounds: Consequences for Representation, Processing, and Impairment. *Brain and Language*, 61, 30-44.

Libben, M., & Titone, D. (2008). The multidetermined nature of idiom processing. *Memory and Cognition*, 36(6), 1103-1121.

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

Lin, M., & Leonard, S. (2012). *Dictionary of 1000 Chinese Idioms* (Revised Edition). New York: Hippocrene Books.

Lin, P. (2010). The phonology of formulaic sequences: A review. In D. Wood (Ed.), *Perspectives on formulaic language in acquisition and communication*, pp. 174-193. London: Continuum.

Lin, P. (2012). Sound Evidence: The Missing Piece of the Jigsaw in Formulaic Language Research. *Applied Linguistics*, 33(3), 342-347.

Lin, P., & Adolphs, S. (2009). Sound Evidence: Phraseological Units in Spoken Corpora. In A. Barfield & H. Gyllstad (Eds.), *Researching Collocations in Another Language*, pp. 34-48. New York: Palgrave Macmillan.

Liontas, J. (2001). That’s all Greek to me! The comprehension and interpretation of modern Greek phrasal idioms. *The Reading Matrix*, 1, 1-32.

Liu, Y., Li, P., Shu, H., Zhang, Q., & Chen, L. (2010). Structure and meaning in Chinese: An ERP study of idioms. *Journal of Neurolinguistics*, 23, 615–630.

Lohmann, A. (2012). A processing view on order in reversible and irreversible binomials. *Vienna English Working Papers*, 21(1), 25-50.

Lupker, S. (2007). Representation and processing of lexically ambiguous words. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics*, pp. 159-174. Oxford: Oxford University Press.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.

MacGregor, L. J., & Shtyrov, Y. (2013). Multiple routes for compound word processing in the brain: evidence from EEG. *Brain and Language*, 126(2), 217-229.

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

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

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

McGlone, M., Glucksberg, S., & Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Processes*, 17, 167–190.

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

Molinaro, N., Canal, P., Vespignani, F., Pesciarelli, F., & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? A reading time and ERP study of collocational complex prepositions. *Language and Cognitive Processes*, 28(6), 762-788.

Mollin, S. (2012). Revisiting binomial order in English: ordering constraints and reversibility. *English Language and Linguistics*, 16(01), 81-103.

Mondini, S., Sarema, G., Lusatia, C., Borane, C. & Semenza, C. (2002). Why is ‘Red Cross’ different from ‘Yellow Cross’? A neuropsychological study on noun-adjective agreement within Italian compounds. *Brain and Language*, 81, 621-634.

Moon, J., & Jiang, N. A. N. (2012). Non-selective lexical access in different-script bilinguals. *Bilingualism: Language and Cognition*, 15(01), 173-180.

Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.

Mueller, R. A. G., & Gibbs, R. W. (1987). Processing idioms with multiple meanings. *Journal of Psycholinguistic Research*, 16(1), 63-81.

Murray, W., & Rowan, M. (1998). Early, mandatory pragmatic processing. *Journal of Psycholinguistic Research*, 27, 1-22.

Myers, J., Cook, A., Kambe, G., Mason, R., & O'Brien, E. (2000). Semantic and episodic effects on bridging inferences. *Discourse Processes*, 29, 179-200.

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31, 9–13.

O'Brien, E., Shank, D., Myers, J., & Rayner, K. (1988). Elaborative inferences during reading: do they occur on-line? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 410-420.

Odlin, T. (1991). Irish English idioms and language transfer. *English World-Wide*, 12, 175-193.

Omazic, M. (2008). Processing of idioms and idiom modifications. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*, pp. 67–79. Amsterdam: John Benjamins Publishing Company.

On-line Chinese Tools Chinese/English Dictionary. Available from: http://www.mandarintools.com/worddict.html. Retrieved 17/10/2012 (Study 1)

On-line Chinese Tools Chinese/English Dictionary. Available from: http://www.mandarintools.com/worddict.html. Retrieved 10/06/2013 (Study 2)

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

Pace-Sigge, M. (2013). The concept of lexical priming in the context of language use. *ICAME Journal*, 37, 149-174.

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.

Paulmann, S., Chareeb-Ali, Z., & Felser, C. (2015). Neurophysiological markers of phrasal verb processing: Evidence from L1 and L2 speakers. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing*, pp. 245-267. Cambridge: Cambridge University Press.

Pavlenko, A. (2009). Conceptual representations in the bilingual lexicon and second language vocabulary learning. In A. Pavlenko (Ed.), *The Bilingual Mental Lexicon: Interdisciplinary Approaches*, pp. 125-160. Tonawanda, NY: Multilingual Matters.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communicatio*n, pp. 191-225. London: Longman.

Pesciarelli, F., Gamberoni, T., Ferlazzo, F., Lo Russo, L., Pedrazzi, F., Melati, E., & Cacciari, C. (2014). Is the comprehension of idiomatic sentences indeed impaired in paranoid Schizophrenia? A window into semantic processing deficits. *Frontiers in Human Neuroscience*, 8(799), 1-14.

Peterson, R., Burgess, C., Dell, G., & Eberhard, K. (2001). Dissociation Between Syntactic and Semantic Processing During Idiom Comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27(5), 1223-1237.

Phillips, N. A., Segalowitz, N., O'Brien, I., & Yamasaki, N. (2004). Semantic priming in a first and second language: evidence from reaction time variability and event-related brain potentials. *Journal of Neurolinguistics*, 17, 237-262.

Pickering, M. J., Frisson, S., McElree, B. & Traxler, M. (2004). Eye movements and semantic composition. In M. Carreriras & C. Clifton (Eds.), *On-line Study of Sentence*

Comprehension: Eyetracking, ERPs and Beyond, pp. 33-50. New York: Psychology Press.

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

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

Pollatsek, A., & Hyona, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes,* 20, 261-290.

Pollatsek, A., Hyona, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 820-833.

R Core Team. (2009). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

R Core Team (2014). R: A language and environment for statistical computing (version 3.1.2). R Foundation for Statistical Computing, Vienna, Austria.

Radach, R., Inhoff, A. W., Glover, L., & Vorstius, C. (2013). Contextual constraint and n + 2 preview effects in reading. *The Quarterly Journal of Experimental Psychology*, 66(3), 619-633.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology*, 55A, 1339–1362.

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

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

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(6), 1290-1301.

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

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

Reichle, E., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioural and Brain Sciences*, 26, 445-526.

Rhodes, S. M., & Donaldson, D. I. (2008). Association and not semantic relationships elicit the N400 effect: electrophysiological evidence from an explicit language comprehension task. *Psychophysiology*, 45(1), 50-59.

Roehm, D., Bornkessel-Schlewiesky, I., Roesler, F., & Schlewiesky, 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.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, 25(5), 762-776.

Rosselli, M., Ardila, A., Jurado, M. B., & Salvatierra, J. L. (2012). Cognate facilitation effect in balanced and non-balanced Spanish-English bilinguals using the Boston Naming Test. *International Journal of Bilingualism*, 18(6), 649-662.

Scarborough, D., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, 3(1), 1-17.

Schmitt, N. (2005). Formulaic language: Fixed and varied. *Estudios de Lingüística Inglesa Aplicada*, 6, 13–39.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An Introduction. In N. Schmitt (Ed.), *Formulaic Sequences*, pp. 1-22. Amsterdam: John Benjamins.

Schweigert, W. (1986). The comprehension of familiar and less familiar idioms. *Journal of Psycholinguistic Research*, 15, 33-45.

Schweigert, W. (1991). The Muddy Waters of Idiom Comprehension. *Journal of Psycholinguistic Research*, 20, 305–314.

Schweigert, W. & Moates, D. (1988). Familiar Idiom Comprehension. *Journal of Psycholinguistic Research*, 17, 281-296.

Sereno, S., O'Donnell, P., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate bias effect. *Journal of Experimental Psychology: Human Perception and Performance*, 32(2), 335-350.

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

Simon, H., Zhang, W., Zang, W., & Peng, R. (1989). STM capacity for Chinese words and idioms with visual and auditory presentations. In H. Simon (Ed.), *Models of Thought* (Vol. 2), pp. 68-75. New Haven and London: Yale University Press.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. M. (2007). *Collocation reviewed*. Unpublished manuscript, Tuscan Word Centre, Italy.

Singh, L. (2014). One world, two languages: cross-language semantic priming in bilingual toddlers. *Child Development*, 85(2), 755-766.

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

Siyanova-Chanturia, A. (2013). Eye-tracking and ERPs in multi-word expression research: A state-of-the-art review of the method and findings. *The Mental Lexicon*, 8(2), 245-268.

Siyanova-Chanturia, A., & Martinez, R. (2014). The Idiom Principle Revisited. *Applied Linguistics*. Published online January 26th, 2014.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2), 251–272.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. (2011). Seeing a phrase 'time and again' matters: The role of phrasal Frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 37(3), 776-784.

Smolka, E., Rabanus, S., & Rösler, F. (2007). Processing Verbs in German Idioms: Evidence Against the Configuration Hypothesis. *Metaphor and Symbol*, 22, 213–231.

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

Spivey, M., & Marian, V. (1999). Cross talk between native and second languages: partial activation of an irrelevant lexicon. *Psychological Science*, 10, 281-284.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In M. G. Gaskell (Ed.*), The Oxford Handbook of Psycholinguistics*, pp. 327–342. Oxford: Oxford University Press.

Sturt, P. (2003). The time course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542-562.

Swinney, D. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-659.

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

Tabossi, P., Arduino, L., & Fanari, R. (2011). Descriptive norms for 245 Italian idiomatic expressions. *Behavioural Research Methods*, 43(1), 110-123.

Tabossi, P., Fanari, R., & Wolf, K. (2005). Spoken idiom recognition: meaning retrieval and word expectancy. *Journal of Psycholinguistic Research*, 34(5), 465-495.

Tabossi, P., Fanari, R., & Wolf, K. (2008). Processing idiomatic expressions: effects of semantic compositionality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 313-327.

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

The Phrase Finder website. Available at: www.phrases.org.uk. Retrieved 16/22/2015.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign language comprehension. *Proceedings of the National Academy of Science*, USA, 104, 12530–12535.

Titone, D., & Connine, C. (1994). Comprehension of Idiomatic Expressions: Effects of Predictability and Literality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20(5), 1126-1138.

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

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting Bilingual and Monolingual Idiom Processing. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing*, pp. 171–207. Cambridge: Cambridge University Press.

Tokowicz, N., & Degani, T. (2013). Lexical Concepts. In P. Robinson (Ed.), *The Routledege Encyclopedia of Second Language Acquisition*, pp. 386-388. Abingdon: Routledge.

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

Traxler, M., Foss, D., Seely, R., Kaup, B., & Morris, R. (2000). Priming in Sentence Processing: Intralexical Spreading Activation, Schemas, and Situation Models. *Journal of Psycholinguistic Research*, 29(6), 581-595.

Tremblay, A., & Baayen, H. (2010). Holistic processing of regular four-word sequences: A behavioural and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication*, pp. 151–173. London: Continuum.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks. *Language Learning*, 61, 569–613.

Ueno, T. (2009). *An investigation of the relationship between the development of bilingual semantic organisation and interactive connectivity across languages*. Unpublished PhD Thesis, Trinity College, Dublin.

Ullman, M. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, 30(1), 37-69.

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

Van Berkum, J., Hagoort, P., & Brown, C. M. (1999). Semantic Integration in Sentences and Discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657-671.

Van Heuven, W., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39, 458-483.

Van Lancker Sidtis, D. (2004). When novel sentences spoken or heard for the first time in the history of the universe are not enough: toward a dual-process model of language. *International Journal of Language and Communication Disorders*, 39(1), 1-44.

Van Lancker Sidtis, D. (2012a). Formulaic language and language disorders. *Annual Review of Applied Linguistics*, 32, 62–80.

Van Lancker Sidtis, D. (2012b). Two-track mind: Formulaic and novel language support a dual-process model. In M. Faust (Ed.), *The Handbook of the Neuropsychology of Language*, pp. 342–367. Chichester: Wiley-Blackwell.

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 Sidtis, D., & Postman, W. A. (2006). Formulaic expressions in spontaneous speech of left- and right-hemisphere-damaged subjects. *Aphasiology*, 20(5), 411-426.

Van Lancker, D., Canter, G. J., & Terbeek, D. (1981). Disambiguation of ditropic sentences: Acoustic and phonetic cues. *Journal of Speech and Hearing Research*, 24, 330–335.

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 subjects. *Brain and Language*, 91, 47-48.

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

Vietri, S. (2014). *Idiomatic Constructions in Italian: A Lexicon-Grammar approach.* Amsterdam: John Benjamins.

Vocab Profile English – Compleat Lexical Tutor. Available at: http://www.lextutor.ca/vp/eng/. Retrieved 13/06/2013.

Wang, X. (2007). Does Semantic Activation Spread Across Languages? An Experimental Study with Chinese–English Bilinguals. *Coyote Papers*, 15, 90–99.

Warren, H. (1994). *Oxford Learner's Dictionary of English Idioms*. Oxford: Oxford University Press.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50, 1-25.

Whaley, C. (1978). Word-non-word classification time. *Journal of Verbal Learning and Verbal Behavior*, 17, 143-154.

Williams, J. (1994). The relationship between word meanings in the first and second language: Evidence for a common, but restricted, semantic code. *European Journal of Cognitive Psychology*, 6, 195-220.

Williams, J., & Cheung, A. (2011). Using priming to explore early word learning. In P. Trofimovich & K. McDonough (eds.), *Applying priming methods to L2 learning, teaching and research*, pp. 73-104. Amsterdam: John Benjamin's.

Wittenberg, E., & Piñango, M. M. (2011). Processing light verb constructions. *The Mental Lexicon*, 6(3), 393-413.

Wolter, B., & Gyllstad, H. (2011). Collocational Links in the L2 Mental Lexicon and the Influence of L1 Intralexical Knowledge. *Applied Linguistics*, 32, 430–449.

Wolter, B., & Gyllstad, H. (2013). Frequency of Input and L2 Collocational Processing: A Comparison of Congruent and Incongruent Collocations. *Studies in Second Language Acquisition*, 35, 451–482.

Wolter, B., & Yamashita, J. (2014). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*. Published online by Cambridge University Press, May 15, 2014. DOI: 10.1017/S0142716414000113.

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

Wray, A. (2004). 'Here's one I prepared earlier': Formulaic language learning on television. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use*, pp. 249-268. Amsterdam: John Benjamins.

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

Wray, A. (2009). Formulaic Sequences and Language Disorder. In M. Ball, M. Perkins, N. Müller & S. Howard (Eds.), *The Handbook of Clinical Linguistics*, pp. 184-197. Oxford: Blackwell.

Wray, A. (2011). Formulaic Language as a Barrier to Effective Communication with People with Alzheimer's Disease. *The Canadian Modern Language Review*, 67(4), 429-458.

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

Wray, A. (2014). Why Are We So Sure We Know What a Word Is? In J. Taylor (Ed.), *The Oxford Handbook of the Word*. Published online March 2014. DOI: 10.1093/oxfordhb/9780199641604.013.032

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

Wu, Y. J., & Thierry, G. (2010). Chinese–English bilinguals reading English hear Chinese. *The Journal of Neuroscience*, 30, 7646–7651.

Wu, Y. J., Cristino, F., Leek, C., & Thierry, G. (2013). Non-selective lexical access in bilinguals is spontaneous and independent of input monitoring: evidence from eye tracking. *Cognition*, 129, 418–25.

Wulff, S. (2008). *Rethinking idiomaticity - a Usage-based Approach*. London/New York: Continuum.

Wulff, S. (2013). Words and idioms. In G. Trousdale & T. Hoffmann (Eds.), *The Oxford Handbook of Construction Grammar*, pp. 274-289. Oxford: Oxford University Press.

Yamashita, J. & Jiang, N. (2010). L1 Influence on the Acquisition of L2 Collocations: Japanese ESL Users and EFL Learners Acquiring English Collocations. *TESOL Quarterly*, 44, 647-668.

Yeganehjoo, M. & Thai, Y. N. (2012). Lexical Access in Production of Idioms by Proficient L2 Learners. *The Southeast Asian Journal of English Language Studies*, 18, 87-104.

Zhang, H., Yang, Y., Gu, J., & Ji, F. (2013). ERP correlates of compositionality in Chinese idiom comprehension. *Journal of Neurolinguistics*, 26, 89–112.

Zhang, T., Van Heuven, W. J., & Conklin, K. (2011). Fast automatic translation and morphological decomposition in Chinese–English bilinguals. *Psychological Science*, 22, 1237–42.

Zhang, X. (2013). The I Don't Know Option in the Vocabulary Size Test. *TESOL Quarterly*, 47, 790–811.

Zhao, X., Li, P., Liu, Y., Fang, X., & Shu, H. (2011). *Cross-Language Priming in Chinese-English Bilinguals with Different Second Language Proficiency Levels*. Paper presented at the 33rd Annual Conference of the Cognitive Science Society, Austin, Texas.

Zhou, S., Zhou, W., & Chen, X. (2004). Spatiotemporal Analysis of ERP During Chinese Idiom Comprehension. *Brain Topography*, 17, 27–37.

Appendices

Appendix 1a. Vocabulary items used in the modified Vocabulary Size Test (Nation & Beglar, 2007).

Level	*Words used*
1	Poor, shoe
2	Patience, circle
3	Dinosaur, dash
4	Quiz, vocabulary
5	Compost, peel
6	Threshold, cavalier
7	Bristle, gimmick
8	Eclipse, authentic
9	Whim, octopus
10	Upbeat, crowbar

N.B. The same items were used in studies 1, 2 and 3. *Level* relates to the BNC word lists, so level 1 = the 1000 most frequent words in English, level 2 = the next 1000 most frequent words, etc.

Appendix 1b. Stimulus items used in Study 1 (Chapter 4).

English phrase *Idiom (control)*	*Chinese phrase* *Chengyu*	 *Idiom (control)*	 *Meaning (approximate)*
Stab in the back (head)	半信半疑	Half believe and half doubt (judge)	Not quite convinced
Armed to the teeth (bones)	不三不四	Neither three nor four (five)	Dubious or shady
Wet behind the (bins)	酒肉朋友	Wine and meat friends (smells)	Fair weather friends
An ace up your (jeans)	鷄毛蒜皮	Chicken feathers and garlic skins (plant)	Trivial things
Have egg on your (suit)	打草驚 蛇	Beat the grass and scare the snake (beast)	Act rashly and give yourself away
On the edge of your seat (plate)	爱财如命	Love money like life (gold)	Be very greedy
Pass the buck (beef)	三長兩 短	Three long and two short (drawn)	Unforeseen disasters or problems
Get blood from a stone (coal)	張牙舞爪	Bare fangs and show claws (anger)	Get ready to fight
Spill the beans (chips)	白雲蒼狗	White clouds change into grey dogs (boys)	Life changes in strange ways
Call the shots (staff)	畫蛇添足	Draw a snake and add feet (hair)	Ruin something by over-meddling
Jump the gun (bed)	馬不停蹄	A horse doesn't stop its hooves (sprint)	Continuous, non-stop
Have a head for heights (faults)	單槍匹馬	One gun and a horse (knife)	All by yourself
A chip off the old block (fence)	不學無術	Have neither learning nor skill (brain)	Be very ignorant
Be strapped for cash (coins)	不共戴天	Not share the same sky (tin)	Have deep hatred for someone
Barking up the wrong tree (wall)	異口同聲	Different mouths but one sound (note)	Everyone saying the same thing
Draw a blank (queen)	接二連三	Connect two and three (eight)	One after another
The coast is clear (warm)	七手八脚	Seven hands and eight feet (dogs)	Too many people trying to do something
Raise the roof (rope)	不辨菽麥	Can't tell beans from wheat (crumbs)	Ignorant
Rock the boat (desk)	不知好歹	Don't know good from bad (odd)	Don’t know what’s good for you
Bare your soul (core)	抱薪救火	Bring sticks to put out a fire (dish)	Make a situation worse

Appendix 2a. Stimulus items used in Study 2 (Chapter 5).

Experiment 2

English idioms	*Controls*	*Chinese idioms*	*Controls*
A chip off the old block	wall	Half believe and half doubt	judge
A pain in the neck	back	Neither three nor four	five
A piece of cake	pie	Draw a snake and add feet	hair
Armed to the teeth	bones	Beat the grass to scare the snake	beast
At the end of the day	hour	Move the tiger away from the mountain	jungle
Bare your soul	core	Chicken feathers and garlic skins	plants
Barking up the wrong tree	bush	One gun and a horse	knife
Be on cloud nine	eight	Can't tell beans from wheat	corn
Below the belt	knee	Doesn't know good from bad	odd
Egg on your face	suit	Eyes bright like torches	mirrors
Get blood from a stone	rock	Three long and two short	small
Get cold feet	hands	Add oil and vinegar	coconut
Left in the dark	cold	Different mouths but one sound	voice
Not my cup of tea	milk	Bare fangs and show claws	anger
On the edge of your seat	chair	Have neither learning nor skill	brains
On the other hand	side	Cover your ears to steal a bell	bike
Raise the roof	wall	Without shirt or shoes	socks
Rock the boat	ship	Love money like life	gold
Spill the beans	chips	Call a deer a horse	sheep
Stab in the back	neck	Wine and meat friends	people
Strapped for cash	coins	Kill the chicken to scare the monkey	donkey
The coast is clear	safe	Chase the wind and grasp at shadows	sunlight
Throw in the towel	bucket	Won't share the same sky	cup
Turn back the clock	years	White clouds change into grey dogs	cats
Under the weather	illness	Keep your mouth shut like a bottle	packet
Wet behind the ears	eyes	Seven mouths and eight tongues	voices

Experiment 3

English idioms	*Chinese idioms*
On the edge of your seat	Three long and two short
Rock the boat	Add oil and vinegar
A chip off the old block	Call a deer a horse
Below the belt	Bare fangs and show claws
Stab in the back	Have neither learning nor skill
Spill the beans	Cover your ears to steal a bell
At the end of the day	Without shirt or shoes
Turn back the clock	Love money like life
A piece of cake	Different mouths but one sound
Under the weather	Wine and meat friends
Flog a dead horse	Half believe and half doubt
Egg on your face	Neither three nor four
Left in the dark	Draw a snake and add feet
Not my cup of tea	Beat the grass to scare the snake
A pain in the neck	Lure the tiger away from the mountain
Throw in the towel	Chicken feathers and garlic skins
Get cold feet	One gun and a horse
The coast is clear	Can't tell beans from wheat
Raise the roof	Doesn't know good from bad
On the other hand	Eyes bright like torches

N.B. Additional idioms used in Experiment 2/3 not defined in Appendix 1b:

Move the tiger away from the mountain = draw out an enemy; *Eyes bright like torches* = focused, far sighted; *Add oil and vinegar* = embellish a story; *Cover your ears to steal a bell* = fool yourself; *Without shirt or shoes* = very scruffy; *Call a deer a horse* = deliberately misrepresent; *Kill the chicken to scare the monkey* = put on a show of strength; *Chase the wind and grasp at shadows* = make groundless accusations; *Keep your mouth shut like a bottle* = not say a word.

Appendix 2b. Mixed effects model outputs referenced in Chapter 5.

In all models significance values are estimated by the R package lmerTest (version 2.0-11; Kuznetsova, Brockhoff & Christensen, 2014): *** $p < .001$, ** $p < .01$, * $p < .05$, $^{+}$ $p < .10$. For likelihood of skipping a logistic linear mixed effects model was used and for fixation count a generalised linear model with poisson regression was used. For all duration measures linear mixed effects models were used and only non-skipped items were included; values were log-transformed in all cases.

Experiment 2

Table 1. Omnibus linear mixed effects models for word level data, corrected to assume a lower bound of 100ms for all skipped items.

	First fixation duration			First pass reading time			Total reading time		
Fixed effects:	β	SE	t	β	SE	t	β	SE	t
Intercept	5.53	0.03	160.76***	5.68	0.13	42.87***	5.68	0.11	50.38***
Group: English	-0.27	0.05	-5.81***	-0.37	0.06	-6.43***	-0.47	0.08	-5.98***
Language: English	0.01	0.04	0.12	-0.02	0.04	-0.46	-0.04	0.06	-0.61
Type: Control	0.10	0.04	2.84**	0.12	0.04	2.70**	0.15	0.06	2.64**
Group*Language	-0.22	0.05	-4.42***	-0.21	0.05	-3.90***	-0.33	0.07	-5.03***
Group*Type	-0.09	0.05	-1.77	-0.11	0.06	-1.80	-0.14	0.08	-1.87
Language*Type	-0.14	0.05	-2.82**	-0.11	0.06	-1.88	-0.10	0.07	-1.53
Group*Language*Type	0.28	0.07	4.12***	0.25	0.08	3.22**	0.37	0.09	4.00***
Control predictors:									
Word length (letters)	n/a	n/a	n/a	0.04	0.01	3.44**	0.06	0.07	3.12**
LogFrequency	n/a	n/a	n/a	-0.02	0.01	-2.43*	n/a	n/a	n/a

Table 2. Chinese native speakers, all items, word level data (skipped items removed from durational analysis).

	Likelihood of skipping			First fixation duration			First pass reading time			Total reading time			Total fixation count		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	z
Intercept	3.47	1.51	2.30*	5.55	0.04	135.22***	5.78	0.16	36.74***	5.73	0.15	38.29***	0.54	0.08	6.53***
Language: English	0.19	0.87	0.21	0.02	0.04	0.50	0.00	0.04	0.05	-0.02	0.07	-0.37	-0.07	0.08	-0.84
Type: Control	-4.10	0.97	-2.08*	0.08	0.04	2.32*	0.09	0.05	2.03*	0.12	0.05	2.19*	0.08	0.07	1.16
Language*Type	2.11	1.59	1.33	-0.13	0.05	-2.65**	-0.09	0.06	-1.66	-0.09	0.07	-1.39	-0.08	0.10	-0.76
Control predictors:															
Word length (letters)	-1.69	0.37	-4.61***	n/a	n/a	n/a	0.05	0.01	3.20**	0.05	0.02	2.18*	n/a	n/a	n/a
LogFrequency	n/a	n/a	n/a	n/a	n/a	n/a	-0.03	0.01	-3.01**	n/a	n/a	n/a	n/a	n/a	n/a
Random effects	Variance			Variance			Variance			Variance			Variance		
Subject	0.669			0.020			0.035			0.069			0.012		
Item	0.000			0.000			0.000			0.021			0.066		
Subject | Language	0.485			0.006			0.002			0.002			0.000		
Subject | Type	3.489			0.001			0.010			0.014			0.000		
Residual	n/a			0.126			0.160			0.235			n/a		

Table 3. English native speakers, all items, word level data (skipped items removed from durational analysis).

	Likelihood of skipping			First fixation duration			First pass reading time			Total reading time			Total fixation count		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	z
Intercept	-0.65	0.69	-0.94	5.41	0.12	44.03***	5.40	0.03	183.68***	5.57	0.04	144.18***	0.29	0.06	5.00***
Language: English	1.63	0.36	4.53***	-0.08	0.03	-2.30*	-0.14	0.04	-3.37***	-0.27	0.05	-5.23***	-0.56	0.09	-6.10***
Type: Control	0.27	0.40	0.68	0.03	0.03	0.99	0.03	0.04	0.76	0.02	0.05	0.45	-0.02	0.08	-0.30
Language*Type	-2.22	0.49	-4.50***	-0.01	0.04	-0.36	-0.02	0.05	-0.46	0.10	0.06	1.69	0.44	0.12	3.66***
Control predictors:															
Word length (letters)	-0.43	0.13	-3.31***	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
LogFrequency	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Random effects	Variance			Variance			Variance			Variance			Variance		
Subject	0.011			0.002			0.004			0.007			0.004		
Item	0.144			0.002			0.004			0.007			0.000		
Subject | Language	0.327			0.001			0.002			0.003			0.003		
Subject | Type	0.333			0.004			0.004			0.015			0.006		
Residual	n/a			0.087			0.120			0.192			n/a		

Table 4. Chinese native speakers, word and phrase level data, all items, interaction with familiarity.

	Likelihood of skipping			First fixation duration (word)			First pass reading time (word)			Total reading time (word)			Total reading time (phrase)			Regression duration (word)		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	3.16	2.13	1.48	5.54	0.10	53.34***	5.78	0.18	31.27***	5.91	0.16	37.36***	7.02	0.21	32.95***	0.68	0.22	28.71***
Language: English	-0.63	1.63	-0.39	0.02	0.11	0.22	0.07	0.12	0.59	0.13	0.16	0.79	-0.28	0.14	-1.94*	0.03	0.23	0.14
Type: Control	-0.02	0.00	-0.00	-0.11	0.14	-0.81	-0.09	0.16	-0.59	0.05	0.20	0.23	-0.06	0.15	-0.42	0.07	0.27	0.26
Familiarity	-0.01	0.23	-0.24	0.00	0.02	0.11	0.01	0.02	0.41	0.02	0.02	0.75	-0.01	0.02	-0.55	-0.00	0.03	-0.02
Language*Type	0.01	0.00	0.00	0.01	0.15	0.09	0.05	0.17	0.32	-0.06	0.21	-0.29	0.08	0.16	0.51	-0.11	0.29	-0.39
Language*Fam	0.12	0.26	0.45	0.00	0.02	0.01	-0.02	0.02	-0.73	-0.04	0.03	-1.65	-0.01	0.02	-0.51	-0.03	0.04	-0.89
Type*Fam	0.15	0.01	0.00	0.03	0.02	1.43	0.03	0.02	1.20	0.01	0.03	0.38	0.01	0.02	0.62	0.01	0.04	0.15
Language*Type*Fam	-0.15	0.01	-0.00	-0.02	0.03	-0.64	-0.02	0.03	-0.60	0.00	0.04	0.05	-0.01	0.03	-0.48	0.02	0.05	0.37
Control predictors																		
Word length (letters)	-1.48	0.37	-3.98***	n/a	n/a	n/a	0.04	0.02	2.88**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
LogFrequency	n/a	n/a	n/a	n/a	n/a	n/a	-0.04	0.01	-3.24**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Idiom length (words)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.09	0.03	3.08**	n/a	n/a	n/a
Random effects	Variance			Variance			Variance			Variance			Variance			Variance		
Subject	0.221			0.020			0.020			0.070			0.091			0.129		
Item	0.000			0.001			0.070			0.022			0.039			0.045		
Subject | Language	0.124			0.005			0.003			0.003			0.008			0.033		
Subject | Type	7.859			0.001			0.014			0.014			0.003			0.013		
Residual	n/a			0.125			0.235			0.235			0.139			0.444		

Table 5. Chinese native speakers, word and phrase level data, all items, interaction with compositionality (English ratings).

	Likelihood of skipping			First fixation duration (word)			First pass reading time (word)			Total reading time (word)			Total reading time (phrase)			Regression duration (word)		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	2.78	1.92	1.44	5.62	0.08	74.53***	5.82	0.17	34.01***	6.07	0.14	43.71***	6.95	0.22	31.24***	6.40	0.19	33.76***
Language: English	1.24	2.63	0.47	0.04	0.15	0.30	-0.13	0.17	-0.81	-0.03	0.27	-0.10	-0.03	0.29	-0.10	0.08	0.37	0.22
Type: Control	-0.77	4.75	-0.16	0.08	0.10	0.80	0.09	0.11	0.78	0.10	0.14	0.72	0.08	0.10	0.82	0.08	0.18	0.45
Compositionality	0.18	0.32	0.58	-0.02	0.02	-0.99	-0.01	0.02	-0.71	-0.02	0.03	-0.53	0.00	0.03	0.03	-0.03	0.04	-0.75
Language*Type	-5.05	6.54	-0.77	-0.15	0.21	-0.73	0.10	0.23	0.41	0.21	0.28	0.75	0.36	0.22	1.64	0.38	0.39	1.01
Language*Comp	-0.27	0.59	-0.46	-0.00	0.04	-0.12	0.03	0.04	0.86	-0.01	0.07	-0.14	-0.06	0.07	-0.94	-0.04	0.09	-0.40
Type*Comp	-1.08	1.58	-0.69	0.00	0.02	0.07	0.00	0.03	0.06	0.01	0.03	0.18	-0.02	0.03	-0.60	0.01	0.05	0.16
Language*Type*Comp	1.94	1.90	1.02	0.01	0.05	0.10	-0.05	0.06	-0.82	-0.07	0.07	-1.10	-0.09	0.05	-1.66	-0.11	0.09	-1.21
Control predictors																		
Word length (letters)	-1.69	0.37	-4.58***	n/a	n/a	n/a	0.05	0.02	3.17**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
LogFrequency	n/a	n/a	n/a	n/a	n/a	n/a	-0.03	0.01	-2.88**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Idiom length (words)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.09	0.03	3.11**	n/a	n/a	n/a
Random effects	Variance			Variance			Variance			Variance			Variance			Variance		
Subject	0.669			0.021			0.035			0.069			0.087			0.128		
Item	0.000			0.000			0.000			0.024			0.038			0.044		
Subject | Language	0.468			0.001			0.002			0.003			0.002			0.035		
Subject | Type	3.776			0.000			0.010			0.013			0.007			0.014		
Residual	n/a			0.127			0.160			0.235			0.139			0.443		

Table 6. Chinese native speakers, word and phrase level data, all items, interaction with compositionality (Chinese ratings).

	Likelihood of skipping			First fixation duration (word)			First pass reading time (word)			Total reading time (word)			Total reading time (phrase)			Regression duration (word)		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	4.02	3.20	1.25	5.55	0.15	38.21***	5.64	0.25	22.75***	5.83	0.27	21.02***	7.22	0.29	24.83***	6.43	0.37	17.47***
Type: Control	-4.81	7.32	-0.66	-0.05	0.19	-0.27	0.07	0.23	0.29	0.46	0.27	1.70	0.14	0.19	0.76	0.24	0.38	0.62
Compositionality	-0.30	0.43	-0.70	0.00	0.02	0.04	0.01	0.03	0.49	0.03	0.05	0.67	-0.04	0.04	-1.02	-0.03	0.06	-0.44
Type*Comp	0.02	0.88	0.02	0.02	0.03	0.72	0.01	0.04	0.13	-0.06	0.05	-1.28	-0.02	0.03	-0.61	-0.02	0.07	-0.34
Control predictors																		
Word length (letters)	-1.83	0.73	-2.51	n/a	n/a	n/a	0.05	0.02	2.73**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
LogFrequency	n/a	n/a	n/a	n/a	n/a	n/a	-0.03	0.02	-1.93*	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Idiom length (words)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.09	0.03	2.60*	n/a	n/a	n/a
Random effects	Variance			Variance			Variance			Variance			Variance			Variance		
Subject	5.914			0.021			0.039			0.090			0.104			0.153		
Item	0.000			0.000			0.000			0.026			0.026			0.042		
Subject | Type	56.76			0.000			0.018			0.045			0.011			0.061		
Residual	n/a			0.121			0.167			0.225			0.113			0.446		

Table 7. Chinese native speakers, word and phrase level data, all items, interaction with plausibility.

	Likelihood of skipping			First fixation duration (word)			First pass reading time (word)			Total reading time (word)			Total reading time (phrase)			Regression duration (word)		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	2.90	2.05	1.42	5.69	0.09	64.48***	5.89	0.19	31.75***	6.12	0.15	41.21***	7.11	0.22	32.16***	6.29	0.20	30.68***
Language: English	1.18	2.66	0.44	-0.05	0.19	-0.25	0.25	0.21	1.17	-0.26	0.31	-0.83	-0.36	0.25	-1.45	-0.40	0.43	-0.92
Type: Control	-15.8	26.8	-0.59	-0.21	0.14	-1.49	-0.18	0.16	-1.18	-0.07	0.21	-0.35	0.13	0.16	0.79	0.21	0.29	0.73
Plausibility	0.11	0.29	0.37	-0.04	0.02	-1.75	-0.01	0.02	-0.48	-0.03	0.04	-0.84	-0.03	0.03	-0.99	-0.00	0.05	-0.06
Language*Type	11.9	26.9	0.44	0.12	0.24	0.52	-0.01	0.26	-0.02	0.29	0.35	0.83	-0.08	0.26	-0.31	0.23	0.48	0.47
Language*Plaus	-0.22	0.46	-0.47	0.03	0.04	0.80	-0.03	0.04	-0.88	0.05	0.06	0.76	0.02	0.04	0.49	0.05	0.08	0.66
Type*Plaus	2.88	6.21	0.46	0.08	0.04	2.15*	0.08	0.04	1.86	0.06	0.06	0.94	-0.03	0.05	-0.69	-0.03	0.08	-0.37
Language*Type*Plaus	-2.51	6.23	-0.40	-0.08	0.05	-1.61	-0.06	0.06	-1.09	-0.10	0.08	-1.26	0.02	0.06	0.36	-0.04	0.11	-0.39
Control predictors																		
Word length (letters)	-1.62	0.38	-4.22***	n/a	n/a	n/a	0.04	0.02	2.77**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
LogFrequency	n/a	n/a	n/a	n/a	n/a	n/a	-0.04	0.01	-3.33***	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Idiom length (words)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.08	0.03	2.69**	n/a	n/a	n/a
Random effects				Variance			Variance			Variance			Variance			Variance		
Subject	0.360			0.020			0.035			0.069			0.088			0.128		
Item	0.000			0.001			0.000			0.024			0.040			0.048		
Subject | Language	0.117			0.005			0.002			0.002			0.007			0.034		
Subject | Type	4.022			0.001			0.010			0.014			0.002			0.015		
Residual	n/a			0.126			0.159			0.235			0.140			0.443		

Table 8. English native speakers, word and phrase level data, all items, interaction with familiarity.

	Likelihood of skipping			First fixation duration (word)			First pass reading time (word)			Total reading time (word)			Total reading time (phrase)			Regression duration (word)		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-1.61	0.89	-1.82	5.37	0.07	75.47***	5.41	0.05	119.70***	5.60	0.06	97.16***	6.57	0.17	38.30***	6.06	0.11	55.81***
Language: English	0.64	1.24	0.52	-0.08	0.11	-0.67	-0.14	0.13	-1.08	-0.36	0.17	-2.14*	-0.62	0.17	-3.77***	-0.40	0.24	-1.62
Type: Control	1.13	0.69	1.65	0.03	0.05	0.51	-0.01	0.06	-0.13	0.01	0.08	0.17	-0.03	0.07	-0.50	-0.09	0.10	-0.86
Familiarity	0.29	0.16	1.87^{+}	-0.00	0.01	-0.34	-0.00	0.02	-0.19	-0.01	0.02	-0.68	-0.03	0.02	-1.33	-0.03	0.03	-0.90
Language*Type	-0.39	1.61	-0.24	0.10	0.15	0.68	0.10	0.17	0.59	-0.04	0.22	-0.17	0.08	0.19	0.42	-0.18	0.31	-0.58
Language*Fam	-0.02	0.23	-0.07	0.00	0.02	0.13	0.00	0.03	0.14	0.02	0.03	0.74	0.00	0.03	0.06	-0.00	0.05	-0.03
Type*Fam	-0.32	0.20	-1.58	0.00	0.02	0.17	0.02	0.02	0.76	0.01	0.03	0.19	0.02	0.02	0.96	0.03	0.04	0.85
Language*Type*Fam	-0.10	0.30	-0.35	-0.02	0.03	-0.72	-0.03	0.03	-0.92	0.02	0.04	0.45	0.01	0.04	0.20	0.04	0.06	0.67
Control predictors																		
Word length (letters)	-0.40	0.13	-3.07**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
LogFrequency	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Idiom length (words)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.09	0.03	3.26**	n/a	n/a	n/a
Random effects	Variance			Variance			Variance			Variance			Variance			Variance		
Subject	0.145			0.002			0.004			0.007			0.088			0.105		
Item	0.131			0.002			0.004			0.007			0.029			0.018		
Subject | Language	0.472			0.001			0.002			0.003			0.006			0.031		
Subject | Type	0.438			0.004			0.005			0.016			0.013			0.014		
Residual	n/a			0.087			0.121			0.192			0.156			0.392		

Table 9. English native speakers, word and phrase level data, all items, interaction with compositionality (English ratings).

	Likelihood of skipping			First fixation duration (word)			First pass reading time (word)			Total reading time (word)			Total reading time (phrase)			Regression duration (word)		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-1.02	1.00	-1.02	5.34	0.06	89.41***	5.39	0.07	72.98***	5.43	0.09	57.62***	6.47	0.21	31.10***	5.88	0.16	37.86***
Language: English	1.81	1.19	1.52	-0.11	0.14	-0.78	-0.19	0.17	-1.12	-0.31	0.22	-1.42	-0.76	0.27	-2.84**	-0.68	0.32	-2.13*
Type: Control	0.29	1.02	0.28	0.01	0.08	0.18	0.03	0.09	0.37	0.19	0.12	1.58	-0.01	0.10	-0.07	0.26	0.17	1.55
Compositionality	0.11	0.19	0.57	-0.02	0.01	-0.74	0.00	0.02	0.20	0.04	0.02	1.59	0.01	0.03	0.18	0.03	0.03	0.96
Language*Type	-0.78	1.77	-0.44	-0.13	0.18	-0.77	-0.04	0.21	-0.17	0.35	0.26	1.33	0.16	0.21	2.13	0.27	0.37	0.72
Language*Comp	-0.05	0.28	-0.19	0.01	0.03	0.31	0.01	0.04	0.32	0.01	0.05	0.12	0.01	0.06	0.16	0.04	0.08	0.48
Type*Comp	-0.00	0.24	-0.02	0.00	0.02	0.23	-0.00	0.02	-0.08	-0.04	0.03	-1.52	0.01	0.02	0.23	-0.07	0.04	-1.81
Language*Type*Comp	-0.35	0.42	-0.83	0.03	0.04	0.68	0.00	0.05	0.07	-0.06	0.06	-0.89	-0.06	0.05	-1.13	-0.01	0.09	-0.08
Control predictors																		
Word length (letters)	-0.44	0.13	-3.34***	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
LogFrequency	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Idiom length (words)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.09	0.03	3.33**	n/a	n/a	n/a
Random effects	Variance			Variance			Variance			Variance			Variance			Variance		
Subject	0.010			0.002			0.004			0.006			0.088			0.103		
Item	0.132			0.002			0.004			0.007			0.031			0.019		
Subject | Language	0.323			0.000			0.002			0.003			0.006			0.031		
Subject | Type	0.334			0.004			0.004			0.016			0.014			0.013		
Residual	n/a			0.087			0.121			0.191			0.156			0.390		

Table 10. English native speakers, word and phrase level data, all items, interaction with plausibility.

	Likelihood of skipping			First fixation duration (word)			First pass reading time (word)			Total reading time (word)			Total reading time (phrase)			Regression duration (word)		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-2.04	1.15	-1.77	5.52	0.09	67.77***	5.53	0.08	66.07***	5.70	0.11	52.71***	6.67	0.18	36.66***	6.15	0.17	35.24***
Language: English	0.75	1.55	0.48	-0.32	0.16	-1.96*	-0.45	0.20	-2.27*	-0.52	0.25	-2.06*	-0.95	0.14	-6.99***	-0.66	0.38	-1.76
Type: Control	0.69	1.44	0.48	0.00	0.11	0.02	-0.01	0.14	-0.11	-0.19	0.18	-1.08	-0.24	0.11	-2.11*	0.13	0.26	0.50
Plausibility	0.36	0.22	1.63	-0.04	0.02	-2.22*	-0.04	0.02	-1.67	-0.04	0.03	-1.30	-0.04	0.02	-2.11*	-0.04	0.04	-1.00
Language*Type	0.30	2.03	0.15	0.22	0.19	1.12	0.30	0.23	1.28	0.64	0.30	2.15*	0.48	0.16	2.93**	0.16	0.43	0.36
Language*Plaus	-0.01	0.29	-0.03	0.06	0.03	1.89*	0.07	0.04	1.85	0.06	0.05	1.24	0.05	0.03	2.04*	0.04	0.07	0.61
Type*Plaus	-0.08	0.38	-0.22	0.01	0.03	0.21	0.01	0.04	0.28	0.07	0.05	1.25	0.06	0.03	2.40*	-0.05	0.07	-0.63
Language*Type*Plaus	-0.44	0.46	-0.96	-0.05	0.04	-1.16	-0.06	0.05	-1.25	-0.13	0.06	-2.03*	-0.06	0.04	-1.71	0.02	0.09	0.24
Control predictors																		
Word length (letters)	-0.42	0.13	-3.26**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
LogFrequency	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Idiom length (words)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.09	0.03	3.29**	n/a	n/a	n/a
Random effects				Variance			Variance			Variance			Variance			Variance		
Subject	0.011			0.002			0.003			0.007			0.089			0.105		
Item	0.115			0.001			0.004			0.007			0.029			0.021		
Subject | Language	0.341			0.000			0.001			0.016			0.005			0.014		
Subject | Type	0.327			0.005			0.005			0.003			0.014			0.032		
Residual	n/a			0.087			0.120			0.191			0.155			0.390		

Experiment 3

Table 11. Chinese native speakers, all items.

	First pass reading time			Total reading time			Total fixation count			Regression path duration		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	t
Intercept	6.58	0.19	34.60***	6.92	0.20	33.93***	1.46	0.21	7.03***	5.98	0.21	28.22***
Language: English	-0.35	0.09	-4.00***	-0.40	0.09	-4.62***	-0.37	0.09	-4.08***	-0.47	0.09	-5.17***
Type: Literal	-0.14	0.06	-2.30*	-0.12	0.04	-2.72**	-0.10	0.04	-2.52*	-0.13	0.04	2.91**
Language*Type	0.00	0.08	0.05	-0.02	0.06	-0.27	-0.01	0.06	-0.21	0.02	0.06	0.29
Control predictors:												
Idiom length (words)	0.09	0.03	2.62**	0.11	0.04	3.06**	0.11	0.04	2.92**	0.19	0.04	4.93***
Random effects	Variance			Variance			Variance			Variance		
Subject	0.034			0.076			0.058			0.052		
Item	0.032			0.046			0.049			0.052		
Subject | Language	0.002			0.007			0.007			0.003		
Subject | Type	0.001			0.004			0.000			0.004		
Residual	0.319			0.146			n/a			0.155		

Table 12. English native speakers, all items.

	First pass reading time			Total reading time			Total fixation count			Regression path duration		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	t
Intercept	6.10	0.15	41.61***	6.43	0.18	36.03***	1.16	0.17	7.01***	5.39	0.18	29.24***
Language: English	-0.51	0.07	-7.40***	-0.76	0.08	-9.89***	-0.64	0.08	-8.20***	-0.52	0.08	-6.22***
Type: Literal	-0.11	0.06	-1.94*	-0.16	0.04	-3.66***	-0.14	0.05	-3.04**	-0.08	0.05	-1.73
Language*Type	0.12	0.08	1.59	0.18	0.06	3.07**	0.16	0.08	2.09*	0.09	0.07	1.34
Control predictors:												
Idiom length (words)	0.06	0.03	2.42*	0.10	0.03	3.09**	0.10	0.03	3.33***	0.18	0.03	5.52***
Random effects	Variance			Variance			Variance			Variance		
Subject	0.026			0.064			0.041			0.095		
Item	0.012			0.031			0.022			0.030		
Subject | Language	0.000			0.003			0.003			0.022		
Subject | Type	0.005			0.001			0.000			0.000		
Residual	0.306			0.172			n/a			0.207		

Table 13. Chinese native speakers, all items, interaction with familiarity.

	First pass reading time			Total reading time			Total fixation count			Regression path duration		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	t
Intercept	6.35	0.27	23.88***	6.90	0.24	28.82***	1.50	0.24	6.37***	6.06	0.25	23.78***
Language: English	0.04	0.23	0.17	-0.20	0.18	-1.12	-0.26	0.18	-1.43	-0.51	0.19	-2.73**
Type: Literal	0.16	0.28	0.57	-0.13	0.19	-0.67	-0.16	0.19	-0.84	-0.21	0.20	-1.07
Familiarity	0.03	0.03	0.98	-0.00	0.02	-0.14	-0.01	0.02	-0.65	-0.01	0.02	-0.59
Language*Type	-0.39	0.30	-1.29	-0.0	0.21	-0.97	-0.05	0.21	-0.22	0.02	0.22	0.10
Language*Fam	-0.07	0.04	-1.94*	-0.04	0.03	-1.62	-0.03	0.03	-1.13	0.00	0.03	0.13
Type*Fam	-0.05	0.04	-1.09	0.00	0.03	0.05	0.01	0.03	0.32	0.01	0.03	0.42
Language*Fam*Type	0.07	0.05	1.40	0.04	0.03	1.24	0.01	0.03	0.38	0.00	0.03	0.11
Control predictors:												
Idiom length (words)	0.10	0.03	2.87**	-0.20	0.03	3.43***	0.12	0.04	3.37**	0.19	0.04	5.00***
Random effects	Variance			Variance			Variance			Variance		
Subject	0.035			0.076			0.058			0.053		
Item	0.028			0.041			0.042			0.051		
Subject | Language	0.002			0.008			0.009			0.005		
Subject | Type	0.001			0.004			0.000			0.004		
Residual	0.319			0.145			n/a			0.156		

Table 14. Chinese native speakers, all items, interaction with compositionality (English ratings).

	First pass reading time			Total reading time			Total fixation count			Regression path duration		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	t
Intercept	6.49	0.25	25.97***	6.95	0.26	27.19***	1.50	0.25	5.92***	5.83	0.26	22.04***
Language: English	-0.33	0.10	-0.83	-0.62	0.38	-1.62	-0.40	0.39	-1.04	-0.23	0.40	-0.59
Type: Literal	-0.03	0.16	-0.17	-0.48	0.11	-4.43***	-0.43	0.11	-4.04***	-0.27	0.11	-2.39*
Compositionality	0.03	0.04	0.70	-0.03	0.03	-0.76	-0.02	0.03	-0.58	0.03	0.04	0.81
Language*Type	-0.39	0.42	-0.91	0.63	0.29	2.14*	0.48	0.34	1.43	0.26	0.30	0.84
Language*Comp	-0.01	0.09	-0.06	0.06	0.09	0.65	0.02	0.09	0.17	-0.06	0.09	-0.60
Type*Comp	-0.03	0.04	-0.77	0.10	0.03	3.63***	0.09	0.03	3.34***	0.04	0.03	1.34
Language*Comp*Type	0.10	0.10	0.95	-0.17	0.07	-2.40*	-0.13	0.08	-1.61	-0.06	0.07	-0.86
Control predictors:												
Idiom length (words)	0.09	0.04	2.50*	0.11	0.04	3.06**	0.11	0.04	3.05**	0.20	0.04	5.07***
Random effects	Variance			Variance			Variance			Variance		
Subject	0.034			0.075			0.056			0.052		
Item	0.034			0.048			0.047			0.052		
Subject | Language	0.002			0.008			0.008			0.003		
Subject | Type	0.001			0.004			0.000			0.004		
Residual	0.319			0.144			n/a			0.155		

Table 15. Chinese native speakers, Chinese items only, interaction with compositionality (Chinese ratings).

	First pass reading time			Total reading time			Total fixation count			Regression path duration		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	t
Intercept	6.70	0.45	14.86***	7.45	0.45	16.65***	2.08	0.43	4.75***	5.99	0.50	12.05***
Type: Literal	-0.07	0.49	-0.14	-0.76	0.29	-2.63**	-0.84	0.30	-2.77**	0.07	0.29	0.26
Compositionality	-0.06	0.07	-0.73	-0.12	0.07	-1.66	-0.12	0.07	-1.74	-0.01	0.08	-0.07
Type*Comp	-0.01	0.08	-0.15	0.11	0.05	2.24*	0.13	0.05	2.47*	-0.04	0.05	-0.73
Control predictors:												
Idiom length (words)	0.13	0.05	2.82*	0.14	0.05	2.78**	0.13	0.05	2.56*	0.20	0.06	3.47***
Random effects	Variance			Variance			Variance			Variance		
Subject	0.024			0.078			0.055			0.054		
Item	0.021			0.046			0.044			0.061		
Subject | Type	0.004			0.001			0.000			0.000		
Residual	0.397			0.127			n/a			0.125		

Table 16. Chinese native speakers, all items, interaction with literal plausibility.

	First pass reading time			Total reading time			Total fixation count			Regression path duration		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	t
Intercept	6.77	0.39	17.35***	7.53	0.37	20.23***	2.08	0.36	5.73***	6.77	0.37	18.31***
Language: English	-0.35	0.49	-0.72	-0.70	0.46	-1.51	-0.47	0.46	-1.03	-0.85	0.46	-1.85
Type: Literal	-0.40	0.33	-1.19	-0.38	0.23	-1.66	-0.31	0.23	-1.37	-0.42	0.23	-1.80
Literal plausibility	-0.04	0.06	-0.64	-0.12	0.06	-2.08*	-0.13	0.06	-2.17*	-0.16	0.06	-2.72**
Language*Type	0.50	0.48	1.03	0.52	0.33	1.57	0.36	0.35	1.02	0.81	0.34	2.40*
Language*LitPlaus	0.01	0.09	0.06	0.07	0.08	0.78	0.03	0.08	0.39	0.08	0.08	1.01
Type*LitPlaus	0.05	0.07	0.79	0.05	0.04	1.16	0.04	0.5	0.94	0.06	0.05	1.27
Language*LitPlaus*Type	-0.09	0.09	-1.04	-0.10	0.06	-1.64	-0.07	0.07	-1.08	-0.15	0.06	-2.35*
Control predictors:												
Idiom length (words)	0.10	0.04	2.71**	0.12	0.04	3.29***	0.12	0.03	3.33***	0.20	0.04	5.61***
Random effects	Variance			Variance			Variance			Variance		
Subject	0.033			0.076			0.057			0.052		
Item	0.032			0.041			0.039			0.040		
Subject | Language	0.002			0.007			0.007			0.003		
Subject | Type	0.001			0.004			0.000			0.004		
Residual	0.320			0.146			n/a			0.155		

Table 17. English native speakers, all items, interaction with familiarity.

	First pass reading time			Total reading time			Total fixation count			Regression path duration		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	t
Intercept	6.14	0.16	37.61***	6.52	0.19	34.86***	1.23	0.18	6.94***	5.40	0.20	27.49***
Language: English	-0.63	0.24	-2.67**	-0.60	0.20	-3.03**	-0.53	0.26	-2.09*	-0.52	0.21	-2.45*
Type: Literal	-0.16	0.10	-1.61	-0.19	0.07	-2.62**	-0.20	0.8	-2.60**	-0.10	0.08	-1.22
Familiarity	-0.01	0.02	-0.61	-0.03	0.02	-1.52	-0.02	0.02	-1.16	-0.00	0.02	-0.11
Language*Type	0.55	0.36	1.55	0.08	0.27	0.30	0.17	0.37	0.45	0.04	0.29	0.12
Language*Fam	0.03	0.04	0.65	-0.01	0.03	-0.23	-0.00	0.04	-0.02	0.00	0.04	0.03
Type*Fam	0.02	0.03	0.54	0.01	0.03	0.57	0.03	0.03	0.98	0.01	0.03	0.28
Language*Fam*Type	-0.08	0.06	-1.26	0.01	0.05	0.14	-0.02	0.06	-0.31	0.00	0.05	0.06
Control predictors:												
Idiom length (words)	0.06	0.03	2.37*	0.09	0.03	2.98**	0.10	0.03	3.24**	0.18	0.03	5.48***
Random effects	Variance			Variance			Variance			Variance		
Subject	0.026			0.063			0.041			0.095		
Item	0.012			0.031			0.021			0.031		
Subject | Language	0.000			0.003			0.003			0.023		
Subject | Type	0.005			0.001			0.000			0.000		
Residual	0.307			0.172			n/a			0.208		

Table 18. English native speakers, all items, interaction with compositionality (English ratings).

	First pass reading time			Total reading time			Total fixation count			Regression path duration		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	t
Intercept	6.27	0.19	33.78***	6.44	0.23	28.31***	1.16	0.20	5.68***	5.34	0.23	23.14***
Language: English	-1.03	0.32	-3.26**	-0.81	0.36	-2.24*	-0.78	0.38	-2.04*	-0.44	0.36	-1.20
Type: Literal	-0.38	0.15	-2.56*	-0.35	0.11	-3.11**	-0.30	0.12	-2.50*	-0.33	0.12	-2.68**
Compositionality	-0.03	0.03	-1.22	-0.01	0.03	-0.17	-0.00	0.03	-0.12	0.01	0.03	0.20
Language*Type	0.31	0.39	0.80	0.20	0.29	0.67	0.34	0.42	0.81	0.18	0.32	0.55
Language*Comp	0.13	0.07	1.70	0.01	0.08	0.14	0.03	0.09	0.38	-0.02	0.08	-0.23
Type*Comp	0.07	0.04	1.97[+]	0.05	0.03	1.85	0.04	0.03	1.43	0.07	0.03	2.18*
Language*Comp*Type	-0.05	0.09	-0.56	-0.01	0.07	-0.13	-0.05	0.10	-0.48	-0.03	0.08	-0.36
Control predictors:												
Idiom length (words)	0.06	0.03	2.24*	0.10	0.03	3.01**	0.10	0.03	3.35***	0.18	0.03	5.57***
Random effects	Variance			Variance			Variance			Variance		
Subject	0.026			0.065			0.034			0.096		
Item	0.009			0.033			0.021			0.030		
Subject | Language	0.000			0.003			n/a[++]			0.023		
Subject | Type	0.005			0.001			n/a[++]			0.000		
Residual	0.306			0.171			n/a			0.206		

[++]Model including random slopes for language and type failed to converge, therefore random intercepts only were included.

Table 19. English native speakers, all items, interaction with literal plausibility.

	First pass reading time			Total reading time			Total fixation count			Regression path duration		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	t
Intercept	6.80	0.30	22.55***	7.09	0.30	23.76***	1.67	0.27	6.05***	6.04	0.33	18.12***
Language: English	-1.41	0.39	-3.58***	-1.32	0.38	-3.52***	-1.03	0.40	-2.57**	-0.78	0.42	-1.85
Type: Literal	-0.50	0.32	-1.52	0.09	0.24	0.37	0.09	0.26	0.33	-0.54	0.26	-2.06*
Literal plausibility	-0.14	0.05	-2.64**	-0.13	0.05	-2.71**	-0.10	0.05	-2.31*	-0.14	0.05	-2.50*
Language*Type	0.98	0.47	2.07*	0.69	0.35	1.95*	0.72	0.46	1.56	0.89	0.39	2.29*
Language*LitPlaus	0.17	0.07	2.40*	0.11	0.07	1.66	0.08	0.07	1.13	0.06	0.08	0.78
Type*LitPlaus	0.08	0.06	1.20	-0.05	0.05	-1.03	-0.05	0.05	-0.87	0.09	0.05	1.78
Language*LitPlaus*Type	-0.16	0.09	-1.83	-0.09	0.07	-1.32	-0.10	0.08	-1.12	-0.15	0.07	-2.12*
Control predictors:												
Idiom length (words)	0.06	0.03	2.37*	0.10	0.03	3.74	0.10	0.03	4.09***	0.19	0.03	6.16***
Random effects	Variance			Variance			Variance			Variance		
Subject	0.026			0.064			0.040			0.095		
Item	0.009			0.019			0.011			0.024		
Subject | Language	0.000			0.003			0.003			0.023		
Subject | Type	0.005			0.001			0.000			0.000		
Residual	0.306			0.170			n/a			0.206		

Appendix 3a. Stimulus item used in Study 3 (Chapter 6).

English Idiom (control)	*Congruent Idiom (control)*	*Swedish Idiom (control)*	*Meaning (approximate)*
Bite (grab) the bullet	Bear (grow) fruit	Born (left) in the hall	Not experienced
Blow (need) a fuse	Bite (burn) your tongue	Chew (use) foam	Be furious/foam at the mouth
Break (hurt) the bank	Bend (read) the rules	Come (focus) on shame	Come to nothing
Chew (use) the fat	Bide (use) your time	Confess (change) colour	Show your hand
Clear (wash) the decks	Break (crack) the ice	A (game) cow on the ice	A problem
Cook (check) the books	Break (end) the silence	Crawl (pray) to the cross	Eat humble pie
Cross (mind) your fingers	Burn (lose) your boats	Cream (sauce) on the mash	The cherry on the cake
Cut (count) your losses	Bury (find) the hatchet	Cry (use) rivers	Cry buckets
Drop (miss) the ball	Call (match) your bluff	Get (miss) the kick	Be fired
Face (play) the music	Clear (check) the air	Give (sell) him the basket	Give someone the elbow
Find (hurt) your feet	Draw (leave) a blank	Give (sell) the iron	Get a move on
Foot (read) the bill	Drown (express) our sorrows	Hang (give) lip	Be downhearted
Hit (fix) the roof	Eat (know) your words	Hard (new) bandages	A tough struggle
Hold (take) the fort	Fall (slip) from grace	Harvest (collect) victims	Claim victims
Hold (lead) your horses	Gain (clear) ground	Hold (never) box	Keep on talking
Jump (take) the gun	Gather (produce) dust	Hot (keen) on the porridge	Be over-eager
Keep (mind) your head	Have (deserve) a point	Lose (apply) the suction	Lose heart
Kick (drop) the bucket	Keep (like) the peace	Make (buy) a painting	Make a mistake
Know (bring) the ropes	Learn (finish) your lesson	Neck (back) over head	Head over heels
Lose (count) your marbles	Lick (dress) your wounds	Play (taste) monkey	Mess about
Make (paint) a scene	Lose (pull) the thread	Pull (cut) logs	Snore loudly

Mark (hear) your words	Lose (hurt) your head	Shoulder (carry) his coat	Step into someone’s shoes
Pick (have) a fight	Meet (win) your match	Similar (tasty) as berries	Identical/very similar
Pick (use) your brains	Meet (call) your maker	Sit (stay) inside	Do time in prison
Pop (shout) the question	Miss (pass) the point	Smell (hear) cat	Be suspicious
Pull (grab) your leg	Pass (use) the time	Stand (focus) on the nose	Come a cropper
Pull (control) your weight	Play (cook) with fire	Step (load) in the piano	Commit a faux pas
Push (make) your luck	Show (paint) your face	Suck (grow) on the frames	Have very little money
Risk (hurt) your neck	Steal (like) the show	Take (risk) battle	Pick a fight
Rock (crash) the boat	Stretch (move) your legs	Take (be) it piano	Take it easy
Save (ruin) the day	Swallow (regain) your pride	Take (need) screw	Do the trick/succeed
Smell (hear) a rat	Sweeten (swallow) the pill	The (main) red thread	The principle argument
Spill (drop) the beans	Take (tell) a joke	The (new) whole ballet	The whole lot
Stand (keep) your ground	Take (lose) shape	Throw (find) water	Urinate
Take (make) your pick	Tighten (change) your belt	Toil (eat) dog	Work very hard
Toe (mark) the line	Tread (lose) water	Turn (cook) the steak	See things the other way round
Turn (move) the tables	Try (fix) your luck	Under (into) the ice	Gone to the dogs/run down
Waste (lose) your breath	Turn (find) the screw	Understand (hear) the gallop	Get the idea/understand
Watch (clean) your step	Wait (miss) your turn	Walk (move) away	Pass away/die
Weather (monitor) the storm	Watch (mend) the clock	Weak (small) comfort	Cold comfort/little comfort

Appendix 3b. Mixed effects model outputs referenced in Chapter 6.

Table 1: Separate mixed effects model analysis for Swedish speakers (top) and English native speakers (bottom), phrase level measures. For condition, native language is taken to be the baseline.

Swedish NS	First pass reading time			Total reading time			Fixation count		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z
Intercept	6.43	0.16	40.01	7.67	0.19	39.58	1.61	0.08	20.97
Type	-0.04	0.05	-0.84	-0.10	0.04	-2.65**	-0.08	0.04	-1.98*
Condition: English	0.00	0.06	0.05	-0.19	0.06	-2.96**	-0.20	0.06	-3.25**
Condition: Congruent	-0.07	0.06	-1.09	-0.15	0.06	-2.43*	-0.14	0.06	-2.28*
Type*Condition: English	-0.04	0.07	-0.57	0.05	0.05	0.97	0.04	0.06	0.60
Type*Condition: Congruent	0.07	0.07	0.99	0.00	0.05	0.03	-0.01	0.06	-0.10
Control predictors:									
Word 1 Length	n/a	n/a	n/a	-0.02	0.01	-1.88	n/a	n/a	n/a
Word 1 Frequency (log)	-0.04	0.01	-3.20***	-0.02	0.01	-2.50*	n/a	n/a	n/a
Word 2 Length	0.04	0.02	2.66**	n/a	n/a	n/a	n/a	n/a	n/a
Word 2 Frequency (log)	n/a	n/a	n/a	-0.04	0.02	-2.46*	n/a	n/a	n/a
Random effects	Variance			Variance			Variance		
Item	0.025			0.044			0.036		
Subject	0.045			0.108			0.100		
Subject | Type	0.001			0.002			0.000		
Subject | Condition: English	0.000			0.006			0.006		
Subject | Condition: Swedish	0.000			0.001			0.003		
Residual	0.327			0.160			n/a		

English NS	First pass reading time			Total reading time			Fixation count		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z
Intercept	6.20	0.12	50.89	6.22	0.07	89.60	0.99	0.06	15.86
Type	-0.16	0.04	-3.85***	-0.20	0.04	-4.43***	-0.17	0.06	-2.99**
Condition: Swedish	0.01	0.06	-0.12	0.11	0.06	1.78	0.09	0.06	1.46
Condition: Congruent	-0.05	0.05	-0.94	-0.04	0.06	-0.70	-0.02	0.07	-0.35
Type*Condition: Swedish	0.15	0.05	2.71**	0.41	0.06	7.31***	0.42	0.08	5.58***
Type*Condition: Congruent	0.04	0.05	0.77	0.07	0.06	1.21	0.06	0.08	0.74
Control predictors:									
Word 1 Length	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Word 1 Frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Word 2 Length	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Word 2 Frequency (log)	-0.03	0.01	-2.15*	n/a	n/a	n/a	n/a	n/a	n/a
Random effects	Variance			Variance			Variance		
Item	0.021			0.036			0.021		
Subject	0.038			0.075			0.043		
Subject | Type	0.004			0.009			0.002		
Subject | Condition: English	0.022			0.006			0.004		
Subject | Condition: Swedish	0.010			0.003			0.003		
Residual	0.180			0.188			n/a		

Table 2: Separate mixed effects model analysis for Swedish native speakers, word level measures. For condition, native language is taken to be the baseline.

	Likelihood of skipping			First fixation duration			First pass reading time			Total reading time			Regression path duration		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-2.91	0.89	-3.29	5.49	0.04	148.85	5.60	0.13	41.87	6.02	0.19	31.91	6.56	0.23	28.51
Type: Idiom	1.91	0.65	2.96**	-0.01	0.03	-0.32	-0.01	0.04	-0.32	-0.11	0.05	-2.27*	-0.08	0.06	-1.31
Condition: English	2.43	0.70	3.50***	-0.07	0.04	-1.86	-0.04	0.05	-0.82	-0.23	0.07	-3.26**	-0.32	0.08	-4.05***
Condition: Congruent	1.15	0.74	1.54	-0.08	0.0	-2.29*	-0.09	0.05	-1.93*	-0.24	0.07	-3.58***	-0.27	0.08	-2.69**
Type*Condition: English	-1.84	0.68	-2.71**	0.02	0.05	0.45	-0.04	0.05	-0.74	0.05	0.07	0.73	0.03	0.09	0.36
Type*Condition: Congruent	-0.37	0.73	-0.50	0.03	0.04	0.51	0.03	0.05	0.65	0.06	0.07	0.91	-0.04	0.08	-0.52
Control predictors:															
Word 2 Length	-0.43	0.13	-3.32***	n/a	n/a	n/a	0.05	0.01	3.64***	0.07	0.02	3.77***	0.04	0.02	1.94*
Word 2 Frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	-0.03	0.01	-2.52*	-0.03	0.02	-2.16*	-0.04	0.02	-2.00*
Random effects	Variance			Variance			Variance			Variance			Variance		
Item	0.899			0.005			0.015			0.032			0.041		
Subject	0.399			0.018			0.029			0.059			0.154		
Subject | Type	0.079			0.000			0.007			0.001			0.007		
Subject | Condition: English	0.122			0.002			0.006			0.018			0.014		
Subject | Condition: Swedish	0.129			0.002			0.006			0.009			0.012		
Residual	n/a			0.113			0.138			0.251			0.396		

Table 3: Separate mixed effects model analysis for English native speakers, word level measures. For condition, native language is taken to be the baseline.

	Likelihood of skipping			First fixation duration			First pass reading time			Total reading time			Regression path duration		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-0.52	0.39	-1.35	5.41	0.07	74.60	5.47	0.08	71.28	5.46	0.06	90.33	5.65	0.08	69.00
Type: Idiom	0.60	0.22	2.69**	-0.06	0.03	-1.75	-0.06	0.03	-1.73	-0.11	0.04	-2.49*	-0.18	0.06	-3.04**
Condition: Swedish	-0.76	0.30	-2.58**	0.04	0.04	1.15	0.05	0.04	1.11	0.07	0.05	1.34	0.14	0.07	2.13*
Condition: Congruent	0.20	0.25	0.80	0.01	0.04	0.33	0.01	0.04	0.23	0.03	0.06	0.37	0.01	0.07	0.12
Type*Condition: Swedish	-0.72	0.37	-1.94*	0.06	0.04	1.59	0.11	0.04	2.37*	0.34	0.06	5.85***	0.52	0.08	6.74***
Type*Condition: Congruent	-0.40	0.31	-1.31	0.01	0.04	0.20	-0.01	0.05	-0.18	0.00	0.06	0.00	-0.01	0.08	-0.07
Control predictors:															
Word 2 Length	-0.18	0.07	-2.71**	n/a	n/a	n/a	0.05	0.01	3.64***	0.07	0.02	3.77***	0.04	0.02	1.94*
Word 2 Frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	-0.03	0.01	-2.52*	-0.03	0.02	-2.16*	-0.04	0.02	-2.00*
Random effects	Variance			Variance			Variance			Variance			Variance		
Item	0.203			0.003			0.003			0.013			0.023		
Subject	0.289			0.020			0.021			0.058			0.110		
Subject | Type	0.006			0.001			0.001			0.001			0.006		
Subject | Condition: English	0.010			0.008			0.012			0.007			0.013		
Subject | Condition: Swedish	0.007			0.012			0.010			0.018			0.016		
Residual	n/a			0.078			0.094			0.161			0.275		

Table 4: Omnibus mixed effects model estimates for all eye-tracking measures, component parts (word level measures for the second word). For condition, Congruent is taken to be the baseline.

	Likelihood of skipping			First fixation duration			First pass reading time			Total reading time			Fixation count			Regression path duration		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	z	β	SE	t
Intercept	0.44	0.42	1.05	5.39	0.05	104.17	5.36	0.08	68.39	2.51	0.14	38.09	-0.16	0.19	-0.85	5.60	0.17	32.75
Group: Swedish	-1.59	0.38	-4.23***	0.21	0.04	5.16***	0.26	0.04	6.03***	0.53	0.07	7.51***	0.61	0.10	5.87***	0.67	0.09	7.50***
Type: Idiom	-0.35	0.24	-1.43	0.04	0.03	1.45	0.02	0.04	0.68	0.04	0.05	0.86	0.05	0.09	0.53	0.03	0.06	0.47
Condition: English	-0.41	0.29	-1.41	0.04	0.03	1.36	0.04	0.04	1.11	0.11	0.06	2.01*	0.14	0.10	1.46	0.11	0.07	1.44
Condition: Swedish	-0.25	0.30	-0.83	0.02	0.03	0.63	-0.00	0.04	-0.01	0.04	0.06	0.66	0.09	0.10	0.98	0.02	0.07	0.11
Group*Type	0.53	0.41	1.31	-0.05	0.04	-1.20	-0.05	0.05	-0.95	-0.13	0.06	-2.00*	-0.18	0.12	-1.49	-0.13	0.08	-1.78
Group*Condition: English	0.13	0.44	0.29	-0.07	0.04	-1.62	-0.06	0.05	-1.26	-0.16	0.06	-2.53*	-0.16	0.12	-1.36	-0.20	0.07	-2.76**
Group*Condition: Swedish	0.03	0.47	0.07	-0.09	0.04	-2.13*	-0.05	0.05	-1.08	-0.14	0.06	-2.14*	-0.12	0.12	-1.06	-0.15	0.08	-2.01*
Type*Condition: English	0.69	0.33	2.08*	-0.04	0.04	-0.95	-0.01	0.05	-0.26	-0.08	0.06	-1.25	-0.16	0.13	-1.24	-0.04	0.08	-0.51
Type*Condition: Swedish	0.41	0.34	1.22	-0.05	0.04	-1.25	-0.01	0.05	-0.27	-0.05	0.06	-0.71	-0.06	0.13	-0.43	0.01	0.07	0.10
Group*Type*Condition: English	-1.96	0.67	-2.90**	0.04	0.06	0.77	0.02	0.06	0.33	0.15	0.09	1.70	0.30	0.17	1.79	0.14	0.10	1.36
Group*Type*Condition: Swedish	-0.26	0.60	-0.43	0.08	0.06	1.42	0.06	0.06	0.91	0.16	0.09	1.80	0.15	0.16	0.90	0.15	0.10	1.47
Control predictors:																		
Word 2 Length	-0.35	0.07	-2.11***	n/a	n/a	n/a	0.02	0.01	3.36***	0.04	0.01	2.96**	0.08	0.02	4.44***	0.03	0.01	1.98*
Word 2 Frequency (log)	n/a	n/a	n/a	-0.02	0.01	-3.79***	-0.02	0.01	-3.90***	-0.04	0.01	-3.32**	-0.04	0.01	-2.75**	-0.03	0.01	-2.41*
Random effects																		
Item	0.404			0.002			0.003			0.020			0.015			0.027		
Subject	0.648			0.009			0.010			0.035			0.047			0.065		
Subject | Type	0.054			0.001			0.003			0.002			0.000			0.006		
Subject | Condition: English	0.098			0.001			0.001			n/a^{+}			0.001			0.004		
Subject | Condition: Swedish	0.328			0.003			0.003			n/a^{+}			0.001			0.008		
Residual	n/a			0.079			0.104			0.190			n/a			0.255		

$^{+}$Model including random slopes for condition failed to converge, therefore only random slopes for type were included

Appendix 4a: Stimulus items used in Study 4 (Chapter 7).

Experiment 5

Idiom	*Control 1*	*Control 2*	*Binomial*	*Control 1*	*Control 2*
behind the scenes	between	bushes	aches and pains	spasms	spasms
below the belt	about	line	arms and legs	hands	feet
bite the bullet	load	packet	art and design	music	music
break the bank	hurt	wall	black and white	green	green
break the ice	crack	lock	boys and girl	men	men
bury the hatchet	find	cable	bread and butter	cheese	meat
caught the sun	seen	flu	brother and sister	cousin	cousin
changed your tune	learned	tyre	deaf and dumb	blind	blind
chewing the fat	using	rind	doctors and nurses	surgeons	surgeons
dropped the ball	stopped	plate	fish and chips	beans	rice
eat your words	know	beans	food and drink	cups	plates
fit the bill	see	role	gold and silver	diamond	diamond
found his feet	hurt	ring	goods and services	items	items
hang his head	mind	shirt	horse and rider	pony	pony
hold the fort	take	door	husband and wife	mothers	sons
hold your horses	lead	drinks	iron and steel	gold	gold
jump the gun	take	wall	king and queen	prince	prince
jump the queue	join	fence	knife and fork	spoon	spoon
look the part	get	best	ladies and gentlemen	children	children
lose his marbles	count	memories	law and order	rules	rules
make your mark	show	sign	left and right	back	back
mark his words	hear	work	live and learn	think	think
missed the boat	cracked	train	live and work	move	write
pass the time	use	house	male and female	mixed	mixed
pick a fight	have	shirt	mum and dad	son	son
pick your brains	use	gift	name and address	number	number
playing with fire	cooking	dolls	nice and easy	slow	slow
popped the question	shouted	balloon	north and south	east	east
pull my leg	grab	arm	nuts and bolts	screws	screws
push his luck	make	body	oil and gas	coal	coal
rock the boat	crash	table	out and about	here	busy
runs the show	saw	shop	peace and quiet	calm	calm
saved the day	ruined	cash	pick and choose	select	select
seen the light	found	film	plain and simple	easy	easy
set the scene	paint	clock	read and write	spell	spell
smell a rat	hear	fire	rich and poor	sick	noble
spill the beans	drop	chips	salt and pepper	spices	spices
stole the show	liked	phone	sick and tired	bored	bored
stood his ground	kept	child	soap and water	towels	towels
stretch my legs	rest	back	son and daughter	friend	friend

tighten your belt	changed	hands	tea and coffee	juice	juice
turn the tables	move	wheels	time and money	people	people
twist his arm	hold	leg	trial and error	bias	appeal
wasting your breath	losing	lives	warm and dry	safe	safe
watch your step	clean	child	wind and rain	snow	snow

Associated collocations	*Control 1*	*Control 2*	*Unassociated collocations*	*Control 1*	*Control 2*
ancient history	distant	stories	abject poverty	total	agony
angry mob	large	gang	academy award	additional	prize
apartment building	exhibition	structure	ancestral home	traditional	house
card game	chess	show	anecdotal evidence	additional	account
classic example	decent	version	animal rights	people's	homes
clean clothes	fresh	things	annual report	regular	message
clear sky	pretty	sea	approval ratings	support	scores
cruise ship	small	vessel	back burner	rear	cooker
crystal ball	silver	vase	baking dish	cooking	bowl
current affairs	modern	actions	ballot box	voting	tin
cutting edge	nasty	side	colour scheme	paint	choice
daily paper	regular	update	complex series	diverse	string
estate agent	housing	keeper	cosmic rays	stellar	dust
express train	fastest	coach	cruel joke	nasty	trick
feather dusters	yellow	pillows	direct result	straight	change
feature film	recent	movie	finance bill	monetary	law
final exam	last	task	foreign debt	overseas	plan
football match	evening	final	former student	previous	neighbour
housing estate	forest	records	full text	new	book
inner self	ideal	dreams	great concern	large	worry
killer whale	large	shark	heavy rain	steady	snow
kitchen sink	upstairs	cloth	human health	animal	growth
light bulb	plant	meters	likely effects	normal	results
lunch box	snack	tin	low risk	small	chance
luxury items	special	things	menial task	boring	role
market research	extra	surveys	mental picture	abstract	portrait
married couple	lovely	person	music hall	dancing	place
modern art	recent	stuff	narrow range	better	piece
nuclear reactor	modern	station	price index	cost	guide
parallel lines	equal	strips	private homes	modern	grounds
parish church	modern	records	public opinion	general	thought
parking meter	payment	machine	quick break	small	rest
post office	new	centre	real impact	huge	result
pretty girl	elegant	view	rescue mission	safety	attempt
research student	language	concept	rough surface	poor	coating
roast beef	nice	goose	separate occasions	earlier	attempts
science fiction	comic	books	serious injury	nasty	outcome
sentence structure	general	patterns	short stay	brief	tour

shallow water	normal	ground	slow motion	fast	moving
shopping list	holiday	guide	special unit	specific	team
storm cloud	smoke	alert	spirit world	ghost	realm
table tennis	live	games	stone floor	new	surface
trade union	local	people	warm welcome	good	greeting
tragic death	awful	finish	wild horses	crazy	ponies
trusted friend	caring	ally	winning streak	victory	cycle

Experiment 6

Semantic pair	*Forward*			*Backward*		
	Semantic	*Control*	*Target*	*Semantic*	*Control*	*Target*
apple-fruit	fruit	bread	apple	apple	toffee	fruit
boat-dock	dock	shop	boat	boat	bike	dock
bread-baker	baker	grocer	bread	bread	milk	baker
bullet-pistol	pistol	weapon	bullet	bullet	arrow	pistol
cake-icing	icing	juice	cake	cake	fruit	icing
chair-seat	seat	path	chair	chair	path	seat
church-vicar	vicar	writer	church	church	school	vicar
circle-square	square	frame	circle	circle	frame	square
coal-fuel	fuel	warmth	coal	coal	bread	fuel
country-land	land	space	country	country	village	land
cow-horse	horse	house	cow	cow	wolf	horse
dagger-sword	sword	club	dagger	dagger	rifle	sword
dirt-dust	dust	side	dirt	dirt	side	dust
dress-skirt	skirt	scarf	dress	dress	scarf	skirt
fence-wall	wall	door	fence	fence	door	wall
field-farm	farm	shop	field	field	stream	farm
floor-carpet	carpet	basket	floor	floor	edge	carpet
fly-spider	spider	beetle	fly	fly	bee	spider
foot-shoe	shoe	coat	foot	foot	back	shoe
hand-glove	glove	coat	hand	hand	head	glove
head-hat	hat	shoe	head	head	hand	hat
kettle-steam	steam	smoke	kettle	kettle	bottle	steam
leg-knee	knee	side	leg	leg	arm	knee
letter-envelope	envelope	container	letter	letter	parcel	envelope
lion-jungle	jungle	river	lion	lion	goat	jungle
milk-cream	cream	juice	milk	milk	juice	cream
money-bank	bank	house	money	money	letter	bank
mouth-face	face	head	mouth	mouth	back	face
music-tune	tune	line	music	music	piece	tune
numbers-maths	maths	words	numbers	numbers	letters	maths
paint-draw	draw	watch	paint	paint	watch	draw
plane-pilot	pilot	owner	plane	plane	coach	pilot
plate-dish	dish	towel	plate	plate	knife	dish

police-detective	detective	secretary	police	police	people	detective
rabbit-hare	hare	mouse	rabbit	rabbit	horse	hare
sheep-lamb	lamb	deer	sheep	sheep	deer	lamb
shelf-cupboard	cupboard	dresser	shelf	shelf	hook	cupboard
sleep-dream	dream	idea	sleep	sleep	rest	dream
sport-game	game	type	sport	sport	type	game
steak-meat	meat	fruit	steak	steak	fruit	meat
sweets-sugar	sugar	coffee	sweets	sweets	fruit	sugar
tool-hammer	hammer	helmet	tool	tool	item	hammer
water-bridge	bridge	street	water	water	field	bridge
wheels-bike	bike	stool	wheels	wheels	frame	bike

Idioms	*Forward*			*Backward*		
	Idiom	*Control*	*Target*	*Idiom*	*Control*	*Target*
behind the scenes	behind	under	scenes	scenes	parts	behind
below the belt	below	above	belt	belt	shirt	below
bite the bullet	bite	grab	bullet	bullet	arrow	bite
break the bank	break	work	bank	bank	wall	break
break the ice	break	crack	ice	ice	lock	break
bury the hatchet	bury	mend	hatchet	hatchet	tools	bury
caught the sun	catch	seen	sun	sun	time	catch
changed your tune	changed	updated	tune	tune	song	changed
chewing the fat	chew	use	fat	fat	rind	chew
dropped the ball	drop	missed	ball	ball	plate	drop
eat your words	eat	rest	words	words	phrase	eat
fit the bill	fit	fix	bill	bill	role	fit
found his feet	find	reach	feet	feet	shoes	find
hold the fort	hold	keep	fort	fort	castle	hold
hold your horses	hold	lead	horses	horses	ponies	hold
jump the gun	jumped	take	gun	gun	knife	jumped
jump the queue	jump	join	queue	queue	line	jump
look the part	look	check	part	part	best	look
lose his marbles	lost	fail	marbles	marbles	memories	lost
make your mark	made	grow	mark	mark	sign	made
mark my words	mark	spotted	words	words	piece	mark
missed the boat	missed	skip	boat	boat	train	missed
pass the time	pass	fail	time	time	ideas	pass
pick a fight	pick	have	fight	fight	quarrel	pick
pick your brains	pick	use	brains	brains	skill	pick
playing with fire	play	cooking	fire	fire	doll	play
popped the question	popped	burst	question	question	thought	popped
pull my leg	pull	grab	leg	leg	feet	pull
push his luck	pushing	make	luck	luck	body	pushing
rock the boat	rock	crash	boat	boat	table	rock
runs the show	runs	saw	show	show	play	runs

save the day	saved	collect	day	day	week	saved
seen the light	seen	found	light	light	lamp	seen
set the scene	set	made	scene	scene	part	set
smell a rat	smell	hear	rat	rat	fire	smell
spill the beans	spilled	drop	beans	beans	chips	spilled
stole the show	steal	take	show	show	race	steal
stood his ground	standing	waiting	ground	ground	floor	standing
stretch my legs	stretch	reach	legs	legs	feet	stretch
tighten your belt	tighten	change	belt	belt	skirt	tighten
turn the tables	turn	move	tables	tables	wheels	turn
twist his arm	twist	pull	arm	arm	leg	twist
wasting your breath	waste	miss	breath	breath	lives	waste
watch your step	watch	clean	step	step	pace	watch

Binomials	*Forward*			*Backward*		
	Binomial	*Control*	*Target*	*Binomial*	*Control*	*Target*
aches and pains	aches	spasms	pains	pains	spasms	aches
arms and legs	arms	hands	legs	legs	feet	arms
art and design	art	music	design	design	music	art
black and white	black	green	white	white	green	black
boys and girls	boys	lads	girls	girls	men	boys
bread and butter	bread	cheese	butter	butter	cheese	bread
brother and sister	brother	cousin	sister	sister	cousin	brother
deaf and dumb	deaf	blind	dumb	dumb	blind	deaf
doctors and nurses	doctors	surgeons	nurses	nurses	surgeons	doctors
fish and chips	fish	beans	chips	chips	rice	fish
food and drink	food	cups	drink	drink	cups	food
goods and services	goods	items	service	service	items	goods
horse and rider	horse	pony	rider	rider	pony	horse
husband and wife	husband	mother	wife	wife	son	husband
iron and steel	iron	gold	steel	steel	gold	iron
king and queen	king	prince	queen	queen	prince	king
knife and fork	knife	spoon	fork	fork	spoon	knife
ladies and gentlemen	ladies	children	gentlemen	gentlemen	children	ladies
law and order	law	rules	order	order	rules	law
left and right	left	back	right	right	back	left
live and learn	live	work	learn	learn	think	live
live and work	live	move	work	work	write	live
male and female	male	mixed	female	female	mixed	male
mum and dad	mum	son	dad	dad	son	mum
name and address	name	number	address	address	number	name
nice and easy	nice	slow	easy	easy	slow	nice
north and south	north	east	south	south	east	north
nuts and bolts	nuts	screws	bolts	bolts	screws	nuts
oil and gas	oil	coal	gas	gas	coal	oil

out and about	out	here	about	about	busy	out
peace and quiet	peace	calm	quiet	quiet	calm	peace
pick and choose	pick	select	choose	choose	select	pick
plain and simple	plain	easy	simple	simple	easy	plain
read and write	read	spell	write	write	spell	read
rich and poor	rich	sick	poor	poor	noble	rich
salt and pepper	salt	spices	pepper	pepper	spices	salt
sick and tired	sick	bored	tired	tired	bored	sick
soap and water	soap	towels	water	water	towels	soap
son and daughter	son	friend	daughter	daughter	friend	son
tea and coffee	tea	juice	coffee	coffee	juice	tea
time and money	time	people	money	money	people	time
trial and error	trial	game	error	error	appeal	trial
warm and dry	warm	safe	dry	dry	safe	warm
wind and rain	wind	snow	rain	rain	snow	wind

Associated	*Forward*			*Backward*		
collocations	*Coll*	*Control*	*Target*	*Coll*	*Control*	*Target*
ancient history	ancient	distant	history	history	stories	ancient
angry mob	angry	large	mob	mob	gang	angry
apartment building	apartment	exhibition	building	building	structure	apartment
card game	card	chess	game	game	show	card
classic example	classic	decent	example	example	version	classic
clean clothes	clean	fresh	clothes	clothes	things	clean
clear sky	clear	pretty	sky	sky	sea	clear
cruise ship	cruise	small	ship	ship	vessel	cruise
crystal ball	crystal	silver	ball	ball	vase	crystal
current affairs	current	modern	affairs	affairs	actions	current
cutting edge	cutting	nasty	edge	edge	side	cutting
daily paper	daily	regular	paper	paper	update	daily
estate agent	estate	housing	agent	agent	keeper	estate
express train	express	fastest	train	train	coach	express
feather dusters	feathers	yellow	duster	duster	pillow	feathers
feature film	feature	recent	film	film	movie	feature
final exam	final	last	exam	exam	task	final
football match	football	evening	match	match	final	football
housing estate	housing	forest	estate	estate	records	housing
inner self	inner	inner	self	self	dreams	inner
killer whale	killer	large	whale	whale	shark	killer
kitchen sink	kitchen	upstairs	sink	sink	cloth	kitchen
light bulb	light	plant	bulb	bulb	meter	light
lunch box	lunch	snack	box	box	tin	lunch
luxury items	luxury	special	items	items	things	luxury
market research	market	extra	research	research	surveys	market
married couple	married	lovely	couple	couple	person	married

modern art	modern	recent	art	art	stuff	modern
nuclear reactor	nuclear	modern	reactor	reactor	station	nuclear
parallel lines	parallel	equal	lines	lines	strips	parallel
parish church	parish	modern	church	church	records	parish
parking meter	parking	payment	meter	meter	machine	parking
pretty girl	pretty	elegant	girl	girl	view	pretty
research student	research	language	student	student	concept	research
roast beef	roast	nice	beef	beef	goose	roast
science fiction	science	comic	fiction	fiction	books	science
sentence structure	sentences	general	structure	structure	patterns	sentences
shallow water	shallow	normal	water	water	ground	shallow
shopping list	shopping	holiday	list	list	guide	shopping
storm cloud	storm	smoke	cloud	cloud	alert	storm
table tennis	table	live	tennis	tennis	games	table
trade union	trade	local	union	union	people	trade
tragic death	tragic	awful	death	death	finish	tragic
trusted friend	trusted	caring	friend	friend	ally	trusted

Unassociated	*Forward*			*Backward*		
collocations	*Coll*	*Control*	*Target*	*Coll*	*Control*	*Target*
abject poverty	abject	total	poverty	poverty	agony	abject
academy award	academy	additional	award	award	prize	academy
ancestral home	ancestral	traditional	home	home	house	ancestral
anecdotal evidence	anecdotal	additional	evidence	evidence	account	anecdotal
animal rights	animal	people	rights	rights	homes	animal
annual report	annual	regular	report	report	message	annual
approval ratings	approval	support	ratings	ratings	scores	approval
back burner	back	rear	burner	burner	cooker	back
baking dish	baking	cooking	dish	dish	bowl	baking
ballot box	ballot	voting	box	box	tin	ballot
colour scheme	colour	paint	scheme	scheme	choice	colour
complex series	complex	diverse	series	series	string	complex
cosmic rays	cosmic	stellar	rays	rays	dust	cosmic
cruel joke	cruel	nasty	joke	joke	trick	cruel
direct result	direct	straight	result	result	change	direct
finance bill	finance	monetary	bill	bill	law	finance
former student	former	previous	student	student	neighbour	former
foreign debt	foreign	overseas	debt	debt	plan	foreign
full text	full	new	text	text	book	full
great concern	great	large	concern	concern	worry	great
heavy rain	heavy	steady	rain	rain	snow	heavy
human health	human	animal	health	health	growth	human
likely effects	likely	normal	effects	effects	results	likely
low risk	low	small	risk	risk	chance	low
menial task	menial	boring	task	task	role	menial

mental picture	mental	abstract	picture	picture	portrait	mental
music hall	music	dancing	hall	hall	place	music
narrow range	narrow	better	range	range	piece	narrow
price index	price	cost	index	index	guide	price
private homes	private	modern	homes	homes	grounds	private
quick break	quick	small	break	break	rest	quick
real impact	real	huge	impact	impact	result	real
rescue mission	rescue	safety	mission	mission	attempt	rescue
rough surface	rough	poor	surface	surface	coating	rough
separate occasions	separate	earlier	occasions	occasions	attempts	separate
serious injury	serious	nasty	injury	injury	outcome	serious
short stay	short	brief	stay	stay	tour	short
slow motion	slow	fast	motion	motion	moving	slow
special unit	special	specific	unit	unit	team	special
spirit world	spirit	ghost	world	world	realm	spirit
stone floor	stone	new	floor	floor	surface	stone
warm welcome	warm	good	welcome	welcome	greeting	warm
wild horses	wild	crazy	horses	horses	ponies	wild
winning streak	winning	victory	streak	streak	cycle	winning

Appendix 4b. Mixed effects model outputs referenced in Chapter 7.

Table 1: Mixed effects model estimates for idioms only (Experiment 5).

Idioms	Phrase									Word														
	First Run			Dwell			Fix count			Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	t	B	SE	t	β	SE	z	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	5.92	0.06	94.4	6.25	0.08	82.68	0.92	0.05	19.7	-2.51	0.70	-3.61	5.33	0.04	121.36	5.36	0.05	117.99	5.33	0.03	165.20	5.78	0.09	65.44
Condition: C1	0.02	0.04	0.49	0.06	0.05	1.27	0.14	0.05	3.06 **	-0.64	0.30	-2.12 *	-0.02	0.03	-0.68	-0.03	0.04	-0.92	0.06	0.03	1.70	0.04	0.06	0.59
Condition: C2	-0.02	0.04	-0.50	0.01	0.05	0.14	0.10	0.05	2.10 *	-0.99	0.33	-2.98 **	-0.04	0.03	-1.33	-0.05	0.04	-1.45	0.05	0.03	1.51	-0.01	0.06	-0.11
Control predictors:																								
W1 Length	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
W1 Freq (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
W2 Length	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
W2 Freq (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.14	0.07	2.08 *	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Phrase Freq (log)	-0.06	0.01	-5.11 ***	-0.05	0.01	-3.92 ***	n/a	n/a	n/a	n/a	n/a	n/a	-0.02	0.01	-2.19*	-0.02	0.01	-3.02 **	n/a	n/a	n/a	-0.04	0.02	-1.93*
Phrase %	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
W2 Cloze	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.01	0.00	1.81^{+}	-0.01	0.00	-1.91*	-0.01	0.00	-1.88^{+}	n/a	n/a	n/a	n/a	n/a	n/a
Fam (centred)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	-0.03	0.01	-1.76^{+}	-0.03	0.02	-2.12*	n/a	n/a	n/a	n/a	n/a	n/a
Comp (centred)	0.02	0.01	1.79^{+}	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	-0.03	0.01	-2.74**	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance			Variance			Variance			Variance		
Subject	0.041			0.040			0.013			0.533			0.010			0.010			0.009			0.017		
Item	0.003			0.034			0.022			0.149			0.001			0.001			0.003			0.055		
Subject | Control 1	0.001			0.008			0.002			0.546			0.000			0.000			0.001			0.003		
Subject | Control 2	0.000			0.006			0.000			0.616			0.001			0.001			0.003			0.002		
Residual	0.161			0.219			n/a			n/a			0.069			0.075			0.138			0.305		

Table 2: Mixed effects model estimates for binomials only (Experiment 5).

Binomials	Phrase									Word														
	First Run			Dwell			Fix count			Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	5.77	0.06	90.09	6.23	0.12	53.43	0.87	0.05	17.54	-1.69	0.22	-7.81	5.09	0.10	51.96	5.08	0.11	48.40	5.15	0.13	39.03	5.60	0.18	30.72
Condition: C1	-0.00	0.05	-0.08	-0.05	0.09	-0.54	0.11	0.05	2.20*	-0.38	0.22	-1.74^{+}	0.01	0.03	0.29	0.01	0.04	0.38	0.10	0.04	2.41*	-0.01	0.11	-0.11
Condition: C2	-0.00	0.05	-0.02	-0.01	0.08	-0.07	0.13	0.05	2.71**	-0.69	0.23	-3.01**	-0.01	0.04	-0.16	0.01	0.04	0.29	0.12	0.04	3.32**	0.03	0.10	0.30
Control predictors:																								
W1 Length	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
W1 Freq (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
W2 Length	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.05	0.03	2.01
W2 Freq (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.02	0.01	2.25*	0.03	0.01	2.36*	0.02	0.01	1.69^{+}	n/a	n/a	n/a
Phrase Freq (log)	n/a	n/a	n/a	-0.04	0.02	-2.10*	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	-0.05	0.02	-2.00*
Phrase %	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Cloze	-0.15	0.06	-2.32*	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	-0.11	0.05	-2.30*	-0.12	0.05	-2.40*	n/a	n/a	n/a	n/a	n/a	n/a
Ratio	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
AssForward	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
AssBackward	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance[a]			Variance			Variance			Variance			Variance		
Subject	0.039			0.044			0.016			0.407			0.011			0.013			0.017			0.025		
Item	0.006			0.041			0.025			0.291			0.006			0.007			0.007			0.061		
Subject | Control 1	0.009			0.009			0.001			n/a			0.001			0.001			0.010			0.019		
Subject | Control 2	0.008			0.010			0.001			n/a			0.004			0.005			0.003			0.006		
Residual	0.184			0.224			n/a			n/a			0.077			0.090			0.184			0.280		

a: model with random slopes by condition failed to converge, therefore random intercepts only model was fitted.

Table 3: Mixed effects model estimates for collocations only (Experiment 5).

Collocations	Phrase									Word														
	First Run			Dwell			Fix count			Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	t	β	SE	t	β	SE	z	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	5.43	0.11	49.03	5.69	0.14	41.25	0.51	0.12	4.31	-0.49	0.39	-1.25	5.28	0.04	129.37	5.35	0.08	66.65	5.51	0.03	80.21	5.83	0.09	66.25
Type (Ass:Non)	0.08	0.04	2.04*	0.22	0.08	2.86**	0.14	0.07	2.05*	-0.21	0.26	-0.84	-0.01	0.03	-0.45	0.01	0.03	0.20	0.10	0.06	1.55	0.08	0.09	0.95
Condition: C1	0.08	0.03	2.40*	0.11	0.06	1.71+	0.08	0.05	1.58	-0.48	0.24	-2.06*	-0.03	0.03	-0.92	-0.01	0.04	-0.19	-0.00	0.05	-0.08	-0.04	0.06	-0.66
Condition: C2	0.09	0.03	2.50*	0.16	0.07	2.30*	0.12	0.05	2.52*	-0.53	0.25	-2.15*	-0.05	0.03	-1.40	-0.04	0.04	-1.06	0.03	0.05	0.63	0.02	0.06	0.36
Type*Control 1	-0.08	0.05	-1.62	-0.09	0.06	-1.50	-0.01	0.07	-0.13	0.25	0.31	0.80	0.01	0.03	0.17	-0.02	0.04	-0.62	-0.04	0.05	-0.77	0.05	0.06	0.82
Type*Control 2	-0.01	0.05	-0.24	-0.07	0.06	-1.06	0.01	0.07	0.23	-0.10	0.32	-0.31	0.05	0.03	1.56	0.05	0.04	1.28	0.04	0.05	0.80	0.08	0.06	1.26
Control predictors:																								
W1 Length	0.04	0.01	4.26***	0.05	0.02	3.20**	0.05	0.02	3.15**	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
W1 Freq (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
W2 Length	0.02	0.01	2.08*	n/a	n/a	n/a	n/a	n/a	n/a	-0.21	0.07	-3.28**	n/a	n/a	n/a	0.02	0.01	2.42*	n/a	n/a	n/a	n/a	n/a	n/a
W2 Freq (log)	-0.02	0.01	-2.48*	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	-0.01	0.01	-2.04*	n/a	n/a	n/a	n/a	n/a	n/a
Phrase Freq (log)	n/a	n/a	n/a	-0.03	0.01	-1.99*	n/a	n/a	n/a	n/a	n/a	n/a	-0.02	0.01	-2.53*	-0.02	0.01	-2.24*	n/a	n/a	n/a	n/a	n/a	n/a
Phrase %	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Cloze	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
MI	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	-0.01	0.01	-2.14*	-0.02	0.01	-2.51*
AssForward	n/a	n/a	n/a	0.19	0.22	2.24*	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
AssBackward	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance			Variance			Variance			Variance		
Subject	0.031			0.028			0.012			0.310			0.007			0.009			0.014			0.014		
Item	0.006			0.081			0.053			0.343			0.003			0.004			0.051			0.127		
Subject | Type	0.001			0.000			0.000			0.052			0.001			0.001			0.001			0.000		
Subject | Control 1	0.000			0.003			0.002			0.133			0.002			0.003			0.000			0.006		
Subject | Control 2	0.001			0.005			0.000			0.184			0.003			0.002			0.006			0.006		
Residual	0.195			0.236			n/a			n/a			0.073			0.099			0.206			0.284		

Table 4: Mixed effects model estimates for semantic pairs only (Experiment 6). Tables show all items with fixed effects of type and direction (top), all items with fixed effects of type only (middle), and all items with no categorical fixed effects (bottom).

Semantic pairs	Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-2.46	0.55	-4.44	5.32	0.4	128.50	5.48	0.10	57.47	5.39	0.03	155.61	5.69	0.13	43.61
Type: Control	-0.17	0.18	-0.94	0.02	0.02	1.15	0.02	0.02	1.08	0.05	0.03	1.56	0.04	0.04	0.84
Direction: Backward	0.07	0.18	0.38	0.03	0.02	1.26	0.01	0.02	0.56	-0.01	0.04	-0.20	0.04	0.04	0.80
Type*Direction	-0.05	0.25	-0.19	-0.02	0.03	-0.73	-0.01	0.03	-0.18	-0.00	0.05	-0.10	0.03	0.06	0.49
Control predictors:															
Length	n/a	n/a	n/a	-0.02	0.01	-2.76**	-0.02	0.01	-2.45	n/a	n/a	n/a	-0.07	0.02	-3.77***
Frequency (log)	0.12	0.06	2.06*	n/a	n/a	n/a	-0.02	0.01	-1.73^{+}	n/a	n/a	n/a	n/a	n/a	n/a
InterWords	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.11	0.03	3.29**
Association	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.133			0.009			0.003			0.017			0.038		
Item	0.054			0.001			0.009			0.009			0.051		
Subject | Type	0.003			0.000			0.000			0.002			0.001		
Subject | Direction	0.068			0.000			0.000			0.004			0.003		
Residual	n/a			0.068			0.082			0.185			0.303		

Semantic pairs (no direction)	Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-2.28	0.51	-4.48	5.33	0.04	132.29	5.50	0.09	62.03	5.39	0.03	177.36	6.02	0.21	28.24
Type: Control	-0.19	0.13	-1.47	0.01	0.01	0.92	0.02	0.02	1.35	0.05	0.02	2.16*	0.05	0.03	1.73^{+}
Control predictors:															
Length	n/a	n/a	n/a	-0.02	0.01	-2.78**	-0.02	0.01	-2.49*	n/a	n/a	n/a	-0.08	0.02	-4.10***
Frequency (log)	0.11	0.06	1.87^{+}	n/a	n/a	n/a	-0.02	0.01	-2.04*	n/a	n/a	n/a	-0.03	0.02	-1.88^{+}
InterWords	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.12	0.03	3.40***
Association	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.232			0.010			0.011			0.017			0.040		
Item	0.053			0.001			0.003			0.009			0.052		
Subject | Type	0.005			0.000			0.000			0.000			0.000		
Residual	n/a			0.068			0.082			0.186			0.303		

Semantic pairs (no fixed effect)	Skip			First Fix			First Run			Dwell			Reg		
	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-1.49	0.12	-12.74	5.25	0.02	282.42	5.37	0.05	115.50	5.43	0.03	187.54	5.76	0.13	44.48
AssFor	0.60	0.56	1.08	-0.10	0.07	-1.47	-0.13	0.08	-1.75^{+}	-0.19	0.12	-1.63	-0.29	0.15	-1.91^{+}
Length	n/a	n/a	n/a	n/a	n/a	n/a	-0.02	0.01	-2.26*	n/a	n/a	n/a	-0.07	0.02	-3.92***
Frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
InterWords	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.12	0.03	3.37***
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.263			0.010			0.011			0.017			0.037		
Item	0.060			0.001			0.003			0.010			0.051		
Residual	n/a			0.069			0.082			0.187			0.304		

Table 5: Mixed effects model estimates for idioms only (Experiment 6).

Idioms	Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-0.03	0.33	-0.10	5.32	0.04	142.32	5.28	0.02	232.34	5.37	0.03	171.74	5.54	0.05	119.52
Type: Control	-0.30	0.21	-1.39	-0.02	0.02	-0.92	-0.01	0.02	-0.24	0.02	0.03	0.55	0.03	0.04	0.65
Direction: Backward	-0.69	0.18	-3.85***	0.01	0.02	0.28	0.00	0.02	0.04	0.05	0.03	1.71^{+}	0.07	0.04	1.62
Type*Direction	0.57	0.25	2.26*	0.01	0.03	0.28	-0.00	0.03	-0.09	-0.01	0.04	-0.34	-0.00	0.06	-0.04
Control predictors:															
Length	-0.28	0.06	-4.60***	-0.01	0.01	-1.96*	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Phrase frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Cloze	0.01	0.00	2.15*	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Phrase %	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Comp (centred)	n/a	n/a	n/a	0.01	0.01	1.95^{+}	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.325			0.007			0.007			0.011			0.024		
Item	0.013			0.001			0.003			0.008			0.028		
Subject | Type	0.040			0.000			0.001			0.000			0.009		
Subject | Direction	0.001			0.000			0.001			0.000			0.002		
Residual	n/a			0.069			0.081			0.155			0.271		

Table 6: Mixed effects model estimates for idioms in forward configurations only (Experiment 6).

Idioms (forward only)	Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-0.18	0.40	-0.45	2.29	0.03	200.19	5.28	0.02	255.47	5.44	0.04	133.95	5.67	0.06	89.34
Type: Control	-0.07	0.23	-0.32	-0.05	0.03	-1.80+	-0.01	0.02	-0.24	-0.04	0.04	-1.05	-0.08	0.06	-1.34
Control predictors:															
Length	-0.29	-0.07	-3.80***	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Cloze	1.23	0.41	2.98**	-0.10	0.06	-1.76+	n/a	n/a	n/a	-0.20	0.09	-2.15*	-0.35	0.13	-2.73**
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.332			0.005			0.005			0.003			0.027		
Item	0.000			0.000			0.002			0.009			0.026		
Subject | Type	0.036			0.001			0.002			0.003			0.008		
Residual	n/a			0.063			0.077			0.136			0.255		

Table 7: Mixed effects model estimates for binomials only (Experiment 6).

Binomials	Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-1.44	0.19	-7.59***	5.25	0.02	223.24	5.27	0.03	205.23	5.45	0.06	91.58	5.17	0.10	53.54
Type: Control	-0.21	0.21	-0.99	-0.01	0.02	-0.41	-0.02	0.03	-0.61	-0.04	0.05	-0.87	0.03	0.04	0.78
Direction: Backward	0.02	0.17	0.13	0.03	0.02	0.13	-0.00	0.02	-0.19	0.02	0.03	0.49	0.03	0.04	0.69
Type*Direction	0.27	0.24	1.09	0.04	0.03	1.35	0.05	0.03	1.59	-0.03	0.04	-0.06	-0.00	0.06	-0.04
Control predictors:															
Length	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.03	0.02	1.91^{+}
Frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Phrase frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	-0.02	0.01	-2.14*	n/a	n/a	n/a
Cloze	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Inter words	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Association	0.67	0.36	1.89^{+}	-0.09	0.04	-2.10*	-0.11	0.05	-2.34*	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.228			0.007			0.007			0.009			0.031		
Item	0.131			0.001			0.001			0.001			0.025		
Subject | Type	0.121			0.003			0.002			0.002			0.005		
Subject | Direction	0.019			0.001			0.001			0.000			0.008		
Residual	n/a			0.071			0.085			0.145			0.340		

Table 8: Mixed effects model estimates for all collocations only (Experiment 6).

Collocations (all)	Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-1.22	0.52	-2.34	5.31	0.04	119.97	5.48	0.06	91.64	5.77	0.09	66.01	5.71	0.15	39.05
Subset: Unassociated	-0.21	0.20	-1.07	0.08	0.03	3.14**	0.10	0.03	3.16**	0.21	0.05	4.56***	0.26	0.07	3.64***
Type: Control	-0.09	0.18	-0.49	0.07	0.03	2.86**	0.07	0.03	2.45*	0.15	0.04	3.47***	0.18	0.06	3.21**
Direction: Backward	-0.11	0.21	-0.56	0.03	0.02	1.47	0.04	0.03	1.47	0.10	0.04	2.69**	0.08	0.05	1.63
Subset* Type	-0.06	0.26	-0.24	-0.09	0.03	-2.84**	-0.10	0.04	-2.71**	-0.19	0.05	-3.59***	-0.18	0.07	-2.72**
Subset*Direction	-0.08	0.27	-0.29	-0.03	0.03	-1.12	-0.03	0.03	-0.97	-0.08	0.05	-1.58	-0.06	0.06	-1.03
Type*Direction	0.05	0.26	0.18	-0.03	0.03	-1.14	-0.03	0.03	-0.79	-0.05	0.05	-1.09	-0.01	0.06	-0.18
Subset*Type*Direction	-0.20	0.39	-0.52	0.07	0.04	1.79^{+}	0.07	0.05	1.62	0.14	0.07	2.13*	0.10	0.08	1.25
Control predictors:															
Length	-0.19	0.04	-4.38***	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.03	0.01	2.22*
Frequency (log)	0.08	0.04	1.89^{+}	-0.01	0.00	-3.52*	-0.03	0.00	-5.35***	-0.04	0.01	-5.54***	-0.04	0.01	-4.03***
Phrase frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Inter words	n/a	n/a	n/a	n/a	n/a	n/a	-0.01	0.01	-1.72^{+}	-0.05	0.01	-4.32***	n/a	n/a	n/a
Cloze	n/a	n/a	n/a	0.01	0.00	2.68**	0.09	0.04	2.50*	0.22	0.05	4.03***	0.25	0.08	3.21**
Association	n/a	n/a	n/a	0.24	0.09	2.79**	0.27	0.10	2.73**	0.30	0.14	2.08*	0.43	0.19	2.22*
Mutual information	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.401			0.008			0.009			0.013			0.027		
Item	0.123			0.000			0.002			0.006			0.050		
Subject | Type	0.018			0.000			0.001			0.005			0.001		
Subject | Direction	0.162			0.000			0.000			0.001			0.001		
Residual	n/a			0.070			0.090			0.177			0.286		

Table 9: Mixed effects model estimates for associated collocations only (Experiment 6).

Collocations (Associated)	Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-0.08	0.35	-0.23	5.31	0.06	84.36	5.36	0.07	71.82	5.35	0.12	43.21	5.65	0.20	28.43
Type: Control	-0.09	0.18	-0.50	0.08	0.03	2.89**	0.08	0.03	2.40*	0.14	0.05	3.14**	0.17	0.06	2.95**
Direction: Backward	-0.12	0.22	-0.52	0.03	0.02	1.37	0.04	0.03	1.45	0.08	0.04	2.22*	0.06	0.05	1.31
Type*Direction	0.00	0.27	0.01	-0.03	0.03	-1.08	-0.02	0.04	-0.73	-0.05	0.05	-1.04	-0.01	0.06	-0.21
Control predictors:															
Length	-0.26	0.06	-4.18	n/a	n/a	n/a	n/a	n/a	n/a	0.02	0.01	1.83^{+}	0.04	0.02	2.12*
Frequency (log)	n/a	n/a	n/a	-0.01	0.01	-2.21*	-0.02	0.01	-2.29*	-0.02	0.01	-2.00*	-0.04	0.02	-2.54*
Phrase frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Inter words	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Association	n/a	n/a	n/a	0.24	0.09	2.71**	0.26	0.10	2.50*	0.28	0.14	1.93^{+}	0.44	0.19	2.28*
Cloze	n/a	n/a	n/a	0.10	0.04	2.64**	0.11	0.04	2.42*	0.20	0.06	3.22**	0.22	0.09	2.50*
Mutual information	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.246			0.008			0.009			0.009			0.015		
Item	0.157			0.001			0.002			0.005			0.049		
Subject | Type	0.025			0.000			0.002			0.006			0.007		
Subject | Direction	0.299			0.000			0.003			0.001			0.003		
Residual	n/a			0.069			0.089			0.175			0.273		

Table 10: Mixed effects model estimates for unassociated collocations only (Experiment 6).

Collocations (Unassociated)	Skip			First Fix			First Run			Dwell			Reg		
Fixed effects:	β	SE	z	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	-2.08	0.75	-2.78	5.41	0.05	104.72	5.50	0.01	69.07	5.93	0.14	42.63	6.47	0.17	38.08
Type: Control	-0.05	0.21	-0.24	-0.02	0.02	-1.10	-0.03	0.02	-1.36	-0.03	0.03	-0.99	0.00	0.04	0.01
Direction: Backward	-0.09	0.21	-0.41	-0.00	0.02	-0.11	0.00	0.02	0.12	0.00	0.03	0.01	-0.02	0.04	-0.47
Type*Direction	-0.19	0.28	-0.66	0.04	0.03	1.46	0.05	0.03	1.64	0.09	0.04	1.99*	0.09	0.06	1.61
Control predictors:															
Length	-0.12	0.06	-1.97*	n/a	n/a	n/a	0.01	0.01	1.92^{+}	0.02	0.01	2.20*	n/a	n/a	n/a
Frequency (log)	0.10	0.06	1.71^{+}	-0.02	0.01	-3.22	-0.03	0.01	-4.85***	-0.04	0.01	-4.47***	-0.06	0.01	-4.00***
Phrase frequency (log)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Inter words	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	-0.07	0.02	-4.39***	-0.07	0.02	-2.84**
Cloze	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.26	0.11	2.41*	0.32	0.16	1.98*
Mutual information	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Random effects:	Variance			Variance			Variance			Variance			Variance		
Subject	0.782			0.008			0.009			0.015			0.038		
Item	0.074			0.000			0.001			0.004			0.052		
Subject | Type	0.022			0.002			0.003			0.003			0.002		
Subject | Direction	0.072			0.000			0.000			0.002			0.000		
Residual	n/a			0.070			0.089			0.177			0.291		