Native and Non-native Idiom Processing: Same Difference

Dissertation
Zur Erlangung des akademischen Grades
Doktor der Philosophie
In der Philosophischen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von

Sara Donnell Beck (geb. Wallsworth), M.A.

Aus Milwaukee, WI, USA

2020

Gedruckt mit Genehmigung der Philosophischen Fakultät
der Eberhard Karls Universität Tübingen

Dekan: Prof. Dr. Jürgen Leonhardt

Hauptberichterstatterin: Prof. Dr. Andrea Weber
Mitberichterstatter: Prof. Dr. Claudia Friedrich, Prof. Dr. Britta Stolterfoht, Dr. Simone Sprenger

Tag der mündlichen Prüfung: 17. Juli 2020

Universitätsbibliothek Tübingen: TOBIAS-lib

Native and Non-native Idiom Processing: Same Difference

Sara Donnell Beck

ACKNOWLEDGMENTS

My grandmother always told me: “Dessert first!” This being a lesson I have taken to heart, I would like to begin this dissertation with the sweetest part: my acknowledgments. Getting to this point was neither a short nor a straight-forward path. When I left the USA nearly 10 years ago for a gap year between my B.A. and whatever was to come next, I never imagined staying to complete my graduate studies and building a life here. Tübingen has become my academic and personal home city in the 9 years to follow, first as I completed my M.A. in English Linguistics (applied) looking to teach academic English, then even more so during a transition from lecturer to PhD student. Though, I must admit that even during my M.A., I did not picture myself doing experimental research, and I didn’t yet know the value of what a PhD journey could bring. Many of the decisions that brought me here were, very much against my nature, made quickly and spontaneously, and I am filled with gratitude for the people who have helped guide me to take these opportunities as they came along and to make them my own. I will surely leave out many people who have helped me over the years, so for those of you reading this who have supported me in any way, large or small: thank you so very much!

First and foremost, thank you to my advisor Andrea Weber. Thank you for not letting me pass up this chance, which I most naively almost turned down. While many young researchers plan and strive for such opportunities, I seemingly stumbled into an ideal situation in which I transformed from a student linguist and teacher into a confident psycholinguist and researcher. I learned valuable tools from you about how to design and run experiments, how to give and take critical and constructive feedback, how to organize and run a lab, how to conduct myself at a professional conference, and what it means to mentor young researchers. On a personal and professional level, it has been a joy to work with you. Thank you.

I am also indebted to my other colleagues in our SFB Project: Claudia Friedrich, my second advisor, and Ruth Keßler, my fellow PhD student and collaborator. While our departments and locations differ, I always feel welcome in your lab. Ruth, thank you for making research collaborations with you so easy and enjoyable. It has been a pleasure sharing coffee over research and data with you.

I am also grateful to the rest of the SFB 833. Instead of having to go it alone, I was surrounded during this process by other like-minded students and researchers who were always willing to help. I appreciate the financial means provided by the SFB to travel abroad to the USA to collect data on several occasions as well as to travel around the world to present my research at major conferences. This travel allowed me to receive critical feedback that moved my research forward. I am also grateful for the countless workshops (programming, statistics in R, among others) and DoKo organized by other PhD students. These workshops and meetings were both helpful on both a practical and emotional

level; there is indeed strength in numbers. I also acknowledge that without Beate Starke and her wonderful administrative support organizing workshops (among many other things) would have been a much more difficult task. Also, to Sigrid Beck, thank you for giving me a start as a student assistant during my M.A. in Tübingen, and thank you for your continued support within and outside of the SFB into the present, even when my research has led me in other directions.

Thank you to my officemates and colleagues (present and past), Thanh Lan Truong, Ann-Kathrin Grohe, and Maria Lazareva. Thanh Lan began as an outstanding student assistant and grew into a trusted and respected colleague. Most of all, thank you for the breaks and the laughs. Maria, thank you especially for your help organizing and managing workshops (and the Russian chocolate, of course). I also want to thank my other colleagues in the department who have offered anything from lunch or coffee dates, organizational help, to advice within the department and beyond over the years. Among those: Simón Ruiz, Karin Klett, Vera Hohaus, Katharina Luther, and Shawn Raisig.

Thank you to all the student assistants who have helped in so many ways throughout this process. Your hours spent gathering and testing participants, translating words and phrases, going through countless PDF and Excel files, fixing my basic programming mistakes, scanning texts and documents, and providing valuable experience in teamwork have not gone unnoticed. Thank you to Johanna Geyer, Marc Stoyanopoulos, Eliane Leberer, Ida Sanders, Brock Schardin, Stacie Boswell, Sören Koch, Sabrina Bucher, Achim Skuta, and any others who have come and gone.

Finally, I would like to thank my families, which I am lucky to have in abundance around the world. First, my parents, whose financial and emotional support during my undergrad (and beyond) allowed me to travel abroad and sustained me as I made a new home in Germany. Without you, I could not possibly be where I am today. To the strong women role models in my life, in particular my mother Sheila and my grandmother Marylin, who modeled what strong, opinionated women look like. Thank you to my brother, who was there to sympathize with my laments about statistics and R, and my sister, who has always supplied much needed humor. Thank you to the Beck family in Germany who have willingly added and accepted an American into their extremely Swabian folds. And last, and anything but least, to my husband Florian: thank you so very much. Thank you for your unwavering support and partnership during this very long process. Thank you for the small and large acts of kindness that have sustained me throughout this very long journey.

From the bottom of my heart: **thank you all.**

TABLE OF CONTENTS

We have the ability to speak in riddles. These riddles are neither constructed nor interpreted in the normal way. Yet we use them so readily that we are usually unaware of their special character—unless we have the misfortune not to be a native speaker. We call these special riddles idioms.

Johnson-Laird (1993)

GENERAL INTRODUCTION

1 GENERAL INTRODUCTION

Understanding what is meant in language comprehension may be easier *done* than *said.* Typically, word-meanings are combined to create a corresponding phrasal meaning based on the meaning of the individual words—or at least that's what it seems *at first glance*. However, in everyday communication, it is often the case that words and phrases are used to express meaning beyond both the individual word meanings and their compositional meaning. For example, if two students are discussing the difficulty of a history quiz, and one comments that she thought it was *a piece of cake*, one can infer that she found the quiz to be very easy. In this case, the meaning of the word "cake" does not directly contribute to meaning of the phrase in a way typical of language composition. Rather, the occurrence of the entire phrase triggers the meaning of this idiom, namely, something that is easy. This seemingly atypical way of constructing meaning is needed for idioms. However, despite the special status of this idiomatic expression, language users will easily and readily be able to interpret the meaning of the phrase quickly and possibly even without noticing that something is different.

But how atypical is figurative language, really? In short: not at all. Figurative language describes language which expresses a meaning that differs from the compositional, literal meaning (e.g., Gibbs & Colston, 2012). Its occurrence is neither special, nor often particularly creative. In addition to idioms, other types of figurative language include irony (describing a very hard quiz as *no problem*, possibly accompanied by an auditory cue such as laughter, in order to express the opposite meaning), metaphor (*the quiz was a beast*), simile (*the quiz was like a beast*), metonymy (*the quiz ate me alive*), proverbs (*smooth seas never made a skilled sailor*), and puns (*learning history is old news, anyway*). Moreover, much of this figurative language is also formulaic, in that it occurs in a specific form. For idioms, this fixedness is one of its defining features (e.g., *a piece of cake* is easy while *a piece of pie* is usually just a sweet dessert), as idioms must be both formulaic and figurative. While some have tried to quantify the use of such expressions (e.g., Glucksberg, 1989; Pollio, Barlow, Fine, & Pollio, 1977) it's clear that they are pervasive in language and therefore serve an important communicative function.

In answering the question "Why don't people say what they mean?", Colston (2015) suggests that figurative language must express some additional meaning beyond the associated figurative meaning. In the example above with the student, by expressing the idea with an idiom (the quiz was *a piece of cake*) rather than simply saying that the quiz was easy, she may wish to express a sentiment beyond the simplicity of the quiz and add humor or light-heartedness to what can be a stressful situation. In other words, idioms may simply provide speakers with a concise way of expressing a large amount of information. This idea is not novel; Sinclair (1991) suggests that language, and

communication as a whole, is built using the 'Idiom Principle'. According to this idea, communication takes place by combining primarily preconstructed choices, or phrases, that constitute a single choice, even though phrases may appear to be analyzable into their individual words. Wray (2002) echoes this idea considering multi-word units or chunks an economic way to communicate and even suggests that these chunks may be retrieved as a single unit from the mental lexicon. Thus, the economy of using idioms may be both in production as well as comprehension, and listeners and speakers can benefit from their usage. In any case, idioms express complex meanings in a precise manner, and listeners do not seem to struggle with these complex meanings.

However, though seemingly inconspicuous for a native (L1) adult listener, the processes taking place during idiom comprehension are decidedly complex. Understanding the figurative meaning of idioms involves linguistic and cultural experience, which not all speakers have. Yet, learning and using idioms and other formulaic phrases are a critical part of linguistic competence (e.g., Tomasello, 2003). If idiomatic meanings are represented or even retrieved from the lexicon in a manner different from literal language composition (i.e., as a single unit), then meaning must be learned by contact with the phrase, and it cannot be constructed on its own. Learners with less language experience, for example non-native (L2) speakers, or even child (L1) speakers, may struggle to comprehend idioms for this reason. Additionally, if typical language processing involves meaning composition on the basis of individual words, listeners must forgo these processes and make online decisions during listening in order for these phrases to indeed be an efficient manner of communication. This struggle between the figurative meaning of the whole idiom and the literal meaning of the individual constituent words opens fundamental questions about the way idiom processing occurs and offers additional challenges to L2 listeners. Such questions about idiom processing apply both to native speakers and non-native speakers, who may have altogether different processing strategies based on their experience with languages in general as well as their experience with the L2 language.

The current dissertation will address issues of idiom processing primarily in adult native and non-native speakers. Specifically, it attempts to answer the overarching question:

How do L1 and L2 idiom processing compare?

After introducing the general background on the issues that may contribute to the processes under investigation, the primary goals of the thesis will be outlined followed by the research methods used. Finally, the individual chapters will be briefly discussed.

1.1 L1 AND L2 PROCESSING

This thesis will look closely at idiom processing in native (L1) speakers as well as non-native (L2) speakers. In order to clearly discuss these speaker groups, some terms as used in this work must first

be defined. While these two speaker groups may be distinguished in different ways depending on the particular research questions at hand, the two terms as used here will be defined as follows: native speakers (L1 speakers) are those speakers of a language who grew up speaking the target language in a monolingual setting; non-native speakers are those speakers of a language who did not grow up speaking the target language in their household, but learned the language outside of this setting later on (e.g., in a classroom). L2 will also be used interchangeably with bilingual speakers and L2 learners to refer to the same group. Additionally, while some studies in this work will refer to L2 learning and/or acquisition, the primary focus and intent is to look closely at L2 processing. When discussing L2 learning or acquisition in regard to idioms, what is meant is not the process of attaining the meaning at hand, which will not be in focus, but rather how L2 users (i.e. learners) are able to process and comprehend idioms in their current stage of proficiency.

Language processing is carried out in an automatic, subconscious manner. For native speakers, this process requires little to no conscious effort when communication occurs in a normal manner. Critical in this process is how words are stored and retrieved in the mental lexicon. Many linguists assume that the mental lexicon consists of individual entries of words which are stored and linked for easy access and retrieval during processing. This idea is straight-forward when considering simple objects and individual words such as CAT and BAG. During language processing the meanings of such words are quickly retrieved and used to build phrasal- and sentence-meaning efficiently, even where language is novel and creative. But when meanings are based on fixed and figurative phrases (e.g., *let the cat out of the bag*, meaning to reveal a secret), more questions than answers about this organization arise. Considering idioms in the mental lexicon: are idioms stored as individual units (i.e., as large words), or are they built from the meanings and/or associations with the individual words? This question harkens to fundamental issues in language processing and the way meaning is built in a general manner. In particular, is meaning retrieval based on experience and usage (e.g., Tomasello, 2003) or is it a strictly generative words-and-rules approach (e.g., Chomsky, 1965; Pinker, 1991)? The way idiom meanings are stored and retrieved directly impacts the way meaning is built and how much flexibility there is in this process, and the consequence of a possible departure of figurative meaning in idioms from the way literal language is processed could mean fundamental differences in the timing and access to literal and figurative meaning. This issue will be addressed in more detail in the following sections.

In addition to the issue of storage and retrieval, compositional processes are also a crucial component when considering idiom processing. In addition to the retrieval of individual word-meanings, a sentence is understood during listening and reading based on the parsing of the other syntactic and semantic components of the phrase, and possibly even the larger context. In this process, native speakers may build up meaning immediately in an incremental manner (e.g., Bott & Sternefeld,

2017) or they may delay and even reconsider composition on the basis of larger units (e.g., Sigrid Beck & Tiemann, 2016). The former suggests a strict adherence to the syntactic and semantic structures being presented as they are unfolding, and any diversions from 'normal' sentential structures and meanings (i.e., garden-path sentences, figurative meaning) may cause even native speakers processing delays. The latter, however, allows for more flexibility in processing, and there is some evidence that native speakers may be able to avoid these delays in some cases (e.g., Carrol & Conklin, 2014; Tabossi, Fanari, & Wolf, 2009). Whether this meaning construction occurs in a top-down or bottom-up fashion, however, does not diminish the fact that native speakers build even complex sentence meanings quickly and with little effort. Idioms even sometimes disobey syntactic rules without being ungrammatical (i.e., *to trip the light fantastic*, meaning to ballroom dance). The lack of difficulty in comprehending familiar idioms in addition to the sheer volume of their usage in everyday language (e.g., Pollio et al., 1977) suggests that idioms may be neither special nor problematic in their processing, at least for native speakers. For non-native speakers, research is only beginning to outline whether this is also the case. The details of research on idiom processing in native and non-native speakers is discussed in Section 1.2.

1.1.1 L2-Specific language processing challenges

L2 learners face additional challenges during language processing that do not apply to first language users. While these are both widely researched and varied, some of them particularly relevant to idiom processing are: the interference of the L1 in L2 processing and how these two languages interact in general; challenges associated with integrating contextual meaning in online meaning construction; reduced inhibition abilities of unnecessary information; and individual differences such as proficiency. The first issue at hand is complex, as it also involves the way meaning is retrieved in a second language. While the organization of the L1 mental lexicon is still under discussion, the L2 lexicon is decidedly more difficult to describe, as it involves mapping L1 *and* L2 words and concepts. One view that is still widely accepted in the field as that the L2 lexicon is entirely separate from the L1 lexicon, and concepts are shared on a conceptual level, a view proposed by Kroll and Stewart (1994) in the Revised Hierarchical Model. According to this model, only proficient users of an L2 are able to successfully deactivate the L1 lexicon, and less-proficient speakers may suffer from more L1 interference. In the case of idioms, as well as other vocabulary that may not be as well-engrained in the L2 lexicon where L1 and L2 concepts do not align, this may prove to be a particular challenge in processing. Additionally, the model makes the assumption that associations first begin between translations from the L2 to the L1 and can only later become strong enough that they map directly between the L2 lexicon and the conceptual level. While the model has since developed further and makes fewer claims about changes following proficiency (see Kroll, van Hell, Tokowicz, & Green, 2010) interference and directional challenges remain an issue for unskilled L2 learners dealing with less frequent vocabulary

items. A detailed discussion of the effect of the L1 on L2 idiom processing are discussed in Section 1.2.3.

Another challenge L2 speakers face is the possibility of a more limited capacity to undertake complex processing procedures during online comprehension. This includes the integration of contextual information as well as the inhibition of unnecessary information. In the case of the former, there is a growing body of research that suggests that L2 speakers are less able to make predictions about upcoming language than native speakers (e.g., Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013; Grüter, Lew-Williams, & Fernald, 2012; Kaan, 2014). This may be due to a deficit in processing resources in the L2 as lexical access and contextual integration involve more effort in an L2 (e.g., Foucart, Martin, Moreno, & Costa, 2014). However, in spite of possible shortcomings, it may be the case that L2 speakers rely heavily on context as a general strategy during processing to make up for a lack of L2 skill and knowledge (e.g., Cieślicka & Heredia, 2011; Gernsbacher, Varner, & Faust, 1990), and this interplay between resources and comprehension strategies may serve as a contention point. These integrative issues are examined in terms of idiom processing in **Chapters 4** and **5**. In the case of inhibitory processes, a similar issue may be to blame. Arriving at an idiom's figurative interpretation involves inhibiting the literal meaning of an idiom and its constituent words to at least a small extent. Less skilled comprehenders, such as L2 learners, are decidedly less able to make use of this skill (e.g., Gernsbacher et al., 1990; Gernsbacher & Faust, 1991). While the current work does not look at inhibition directly, L2 speakers' access to literal compared to figurative meaning runs throughout the body of this dissertation, and **Chapters 5B** and **5C** shed light on this particular issue when context deems specific interpretations inappropriate.

Finally, one critical issue for L2 learners and their ability to comprehend and process the L2 is their proficiency. While the current work focuses primarily on advanced L2 speakers for pragmatic reasons (i.e., they are familiar with the idioms used in our studies), the issue of proficiency will be an accompanying theme throughout this text. Each of the issues already discussed here has the potential to be influenced by proficiency (e.g., Elston-Güttler, Paulmann, & Kotz, 2005), and individual differences in proficiency may even contribute to inconsistent or varied results. Critically, even when looking at highly proficient non-native speakers, some challenges based on lack of L2 knowledge may be present, and this topic is addressed more directly in terms of cognitive effort and brain activation during idiom comprehension in **Chapter 7**.

1.2 IDIOMATIC PROCESSING

1.2.1 Idioms

Before addressing idiom processing in native and non-native speakers, idioms must first be more precisely defined and discussed as a category of linguistic expressions. While the classic definition of idioms defines them simply as expressions with a figurative meaning that cannot be inferred based on the individual constituents (e.g., Chomsky, 1965), researchers have continually defined and redefined what idioms are and are not (see e.g. Nunberg, Sag, & Wasow, 1994). This process has been necessary, as the classic definition has excluded a number of phrases considered by many to be idioms, and following definitions have expanded on this idea in order to be more inclusive and/or more precise in their target phrases. Fernando (1996) calls idioms "conventionalized multiword expressions, often but not always non-literal" (p. 1), thereby including both phrases without clear figuration as well as idioms with a more classically compositional meaning. Moon (1998) even more vaguely suggests that idioms may be a broad term for "many kinds of multi-word item[s]" (p.4) and uses idiom as a blanket term rather than being overly specific. Wray (2000) takes a more psycholinguistic approach to the definition, in which idioms are multi-word sequences which seem to be "prefabricated: that is, stored and retrieved whole from memory at the time of use" (p. 465). The reason definitions vary so widely may be due to the extreme diversity between these phrases. Nunberg et al. (1994) get to the heart of the issue in suggesting that idiomaticity involves a number of, but often not all, properties which include conventionality, inflexibility, figuration, proverbiality, informality, and affect. Rather than seeing idioms as a homogenous group, idioms can be defined in broad terms with properties that often, though need not, apply to all idioms. Though, a phrase which has many of these properties may be easily classified as an idiom.

In the current dissertation, the idioms under investigation can be broadly defined as multi-word sequences which are generally formulaic in nature and have figurative meanings. This definition still leaves a wide range of diversity, and **Chapter 2** will specifically address this diversity in establishing a database of idioms with a number of norms for controlling this diversity in the experiments presented throughout this work. However, at the center of the current investigations will be the figurative and formulaic nature of idioms. These properties are of interest because the figurative meaning of idioms, unlike (non-dead) metaphors, is fixed in nature. And, also generally fixed, are the constituent parts that activate this meaning. The figurative nature of idioms means that it belongs to the category of figurative expressions along with metaphor, irony, metonymy, etc., but the formulaic nature also puts idioms in the category of multi-word expressions (also called formulaic language, multi-word units, etc.) along with binomials and clichés. For this reason, the possible dichotomy between figurative and literal language is of particular interest for idioms. Whether idiomatic processing more uniquely resembles that of figurative language or other formulaic sequences may give insight into whether figurativeness

itself is a marked feature in language processing and whether figurative and literal language processing are indeed different in nature. Considering L2 speakers, the same questions apply, but as L2 learners face their own challenges during processing, whether or not their processing of idioms resembles L1 idiom processing can more broadly give more insight into how idioms can be handled during learning. As specific issues surrounding L1 and L2 idiom processing will also be discussed in depth in each individual chapter, a short literature review on L1 and then L2 idiom processing follows.

1.2.2 L1 idiom processing

Idiom processing (in focus here: comprehension) has been most widely researched in terms of adult native speakers of a language. Pragmatically, this group of speakers is familiar with and uses a large number of idioms on a daily basis, whereas non-native speakers may have considerably less knowledge of idioms. For L1 speakers, two main lines of investigation have broadly led research in the field: 1) how access to literal and figurative meaning occur (i.e. timing) and 2) how idiom processing compares to the processing of similar novel phrases. While most of the research discussed in this dissertation concerns the first question, the second is also critical in understanding what factors may impact idiom processing.

Early models of idiom processing assumed three general possibilities concerning the timing of access to figurative compared to literal meaning during online processing: literal-first, figurative-first, or simultaneous processing. These step-wise models assume that figurative and literal meaning are processed independently and are therefore somehow different in nature. For an idiom such as *to spill the beans*, meaning figuratively to reveal a secret, such models assume that access to the figurative meaning and the literal compositional meaning, in which the legumes were part of an accident, are both accessed separately. Literal-first models date back to the Standard Pragmatic Model (e.g., Searle, 1979) in which non-literal language occurs indirectly because of the involvement of complex implicatures. First, literal language is automatically and obligatorily computed, and only after it is deemed inappropriate is a figurative interpretation attempted. Following this model, figurative language, including idioms, belongs to a class of exceptional and inherently difficult expressions. This view was directly endorsed in the Idiom List Hypothesis (Bobrow & Bell, 1973) in which idioms are stored in a separate domain in the mental lexicon that must be activated when an idiomatic mode of processing is necessary. The consequence of this theory is that literal interpretations should be available faster than figurative ones, a result unsubstantiated by studies that followed (e.g., Cacciari & Tabossi, 1988; Gibbs, 1980; Swinney, 1979). On the other hand, if figurative meaning is available faster than literal ones, it's possible that figurative meaning is accessed first, as laid forward in the Direct Access Hypothesis (Gibbs, 1980). Under this hypothesis, literal computation and meaning activation need not occur after the figurative meaning has been successfully activated, though an abundance of counter-evidence showing literal constituent activation suggests otherwise (e.g., Cacciari & Tabossi, 1988;

Libben & Titone, 2008; Rommers, Dijkstra, & Bastiaansen, 2013; Sprenger, Levelt, & Kempen, 2006; Titone & Connine, 1994b). Finally, following the Lexical Representation Hypothesis, literal and figurative meanings may be accessed simultaneously (Swinney, 1979). Following this dual access, figurative meaning is available first, as it involves direct retrieval of a word-like meaning, whereas literal computation involves structural and lexical components taking more time. Here, literal computation is also deemed unnecessary when the figurative meaning is accessed quickly and successfully integrated. Problematic for all of these hypotheses is the fact that both types of meanings are consistently available in L1 online processing, and this access may be more flexible than these models allow. Additionally, all of these models assume that idioms' meanings do not interact with the semantics of their constituent words.

Hybrid-accounts for processing have replaced these step-wise models due to their unsubstantiated assumptions about meaning access and inflexibility. Cacciari and Tabossi (1988) put forward the Configuration Hypothesis, which suggests that processing begins in a literal manner until a sufficient portion of the idiom has been encountered that deems it recognizable. This recognition point is known as the idiomatic key. As the idiomatic key may occur in different places in each idiom, literal meaning construction may also follow slightly different patterns for different idioms. However, once the idiomatic key has been reached, the literal meaning is no longer necessary. For an idiom like *to let the cat out of the bag*, the word OUT serves as the idiomatic key, and literal processing need not continue (see **Chapter 2**). This contrasts with an idiom such as *to lose one's cool*, meaning figuratively to lose control of one's emotions, in which the final word COOL is the idiomatic key. In this case, literal processing proceeds necessarily until the entire idiom has been encountered. Tabossi and Zardon (1993) later add to this concept to show that in the presence of biasing context, recognition and abandonment of literal meaning may begin even earlier. However, later studies suggest that this model only holds when updated to allow for some form of continued literal processing (e.g., Cacciari, 2014; Cutting & Bock, 1997; Libben & Titone, 2008; Titone & Connine, 1994b).

Sprenger et al. (2006) proposed another hybrid account, The Hybrid Model of Idiom Representation or the Superlemma Hypothesis, which was originally used to describe production. In this hypothesis, idioms have representations at the lexical-syntactic level called *superlemmas* in addition to the individual lemmas associated with the constituent parts of the idiom. Activation of the idiom begins immediately and competes with activation of the individual parts. Like the Configuration Hypothesis, activation of the idiomatic meaning may increase until it becomes stronger than the literal activation, and in this manner, a flexible activation of both figurative and literal constituent meaning can be considered during processing. In the way that meaning builds over time, it is similar to the Configuration Hypothesis, however, they critically differ as the Superlemma Hypothesis allows immediate figurative activation and continued activation of literal constituent words.

In addition to these hybrid accounts, there is a variety of research that suggests that idiom processing may differ based on the individual linguistic properties of an idiom. In the Idiom Decomposition Hypothesis, Gibbs and Nayak (1989) suggested that idioms are subject to compositional processes, and this differs based on the decomposable nature of an idiom, or the extent to which individual constituent meanings contribute to the figurative meaning of an idiom. However, composition is always attempted to a certain extent. The expectation following this account is that non-decomposable idioms should be more difficult to process because of failed compositional attempts. Online studies, though, have failed to confirm this result (e.g., Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Libben & Titone, 2008). Additional research, however, has shown that some properties do directly affect the speed of access of either literal, figurative, or both interpretations, though they may not be fully responsible for differential processing. In particular, research has considered properties such as familiarity (e.g., Cronk, Lima, & Schweigert, 1993; Titone & Connine, 1994a; Titone & Libben, 2014), word-frequency (e.g., Cronk et al., 1993; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Titone & Libben, 2014), literality (e.g., Holsinger & Kaiser, 2013; Titone & Connine, 1994a), and predictability (e.g., Cacciari & Tabossi, 1988; Titone & Connine, 1994b), among others, many of these which will be discussed in detail in the coming chapters. By looking at the combined effects of familiarity with an idiom, word-frequency, and decomposability in idiom processing, Libben and Titone (2008) devised the Constraint-Based Model of Idiom Processing. Following such a model, familiar idioms with semi-lexicalized meanings can be retrieved as units from memory early during comprehension, and these meanings continually interact with literal compositional analysis of the phrase. Following this model, individual properties can affect the speed of access or availability of different meanings during online processing. Thus, literal and figurative meaning interact here, too, in an ongoing processing during comprehension.

While the favored hybrid models of idiom processing account for the often fast access to figurative meaning compared to literal meaning, a comparison of speed of processing of idioms and novel phrases shows a similar effect suggesting that fast processing may not be a uniquely figurative property of idioms, but rather a formulaic property. Studies consistently show that idioms are read faster and with fewer fixations than similar novel phrases (e.g., *at the end of the day* vs. *at the end of the war*; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Tabossi et al., 2009; Underwood, Schmitt, & Galpin, 2004). Furthermore, these results often hold for both figurative and literal readings of idioms (e.g., Siyanova-Chanturia, Conklin, & Schmitt, 2011). Such findings lend further support for models involving cumulative recognition effects that may even be frequency-based such as the Configuration Hypothesis and even the Superlemma Hypothesis. However, teasing apart advantages associated with figurativeness from fixedness is neither necessary nor the focus of the current body of work, as the idioms dealt with here have both properties to a certain extent, and hybrid models of representation

explain advantages well in both cases. What is clear, however, is that idioms are processed quickly by native speakers, and figurative meaning in addition to literal constituent meaning are both available quickly in online processing.

1.2.3 L2 Idiom processing

Considering non-native idiom processing, the challenges for comprehending idioms are well-documented (e.g., Abel, 2003a; Charteris-Black, 2002; Cieślicka, 2006; Kecskés, 2000; Matlock & Heredia, 2002). For a non-native speaker, an idiom must first be learned and recognized in order for it to be understood. When asked if an L2 speaker wants to join in to *chew the fat*, meaning figuratively to talk or gossip about the affairs of others, this request would lead to possible confusion for a speaker who has never before heard the phrase, which may occur with little clarifying context. Critically, however, once an idiom has been learned, it seems that online processing difficulties do not entirely disappear. For instance, the advantage of fast processing of idioms and quick access to figurative meaning is not a given for L2 speakers, and figurative meaning appears to be slower for L2 speakers than L1 speakers (e.g., Cieślicka, 2006; Siyanova-Chanturia, Conklin, & Schmitt, 2011). Models of L2 idiom processing are varied, as they attempt to explain the challenges facing L2 speakers during the comprehension of known idioms based on fundamental differences in processing. Since the current dissertation aims to look at whether L1 and L2 processing is indeed fundamentally different, a look at where they may differ will be the focus of this section.

One major line of investigation in L2 idiom processing is whether processing may differ based on a reliance on first language knowledge. A lack of knowledge, for instance, may cause learners to use their L1 conceptual knowledge to fill in gaps in L2 knowledge (e.g., Abel, 2003a; Kecskés, 2000). Abel (2003a) proposed in her Model of Dual L2 Idiom Representation that this lack of experience causes learners to attempt to decompose idioms, even when they are not decomposable and to rely on conceptual representations applied to metaphors in their first language (e.g., Lakoff & Johnson, 1980). Such a comprehension strategy leads to both errors in understanding as well as a slower process. Another way L2 learners may use their first language in L2 idiom processing is via cross-language facilitation and translation effects. Both studies by Irujo (1986) and Liontas (2002) investigated comprehension and production of idioms with differing levels of translation. Using offline methods, both authors found that idioms with direct translations were easiest to comprehend and produce, while those with partial and mismatching translations and meanings were difficult and often avoided altogether by L2 speakers. Since then, online studies have similarly investigated this effect in collocations (e.g., Wolter & Gyllstad, 2011) and idioms (e.g., Carrol & Conklin, 2014; Carrol, Conklin, & Gyllstad, 2016) and facilitation of L1 idioms has often extended to L2 idioms. Such results question whether and how idiomatic processing may be connected and/or modulated through the L1 lexicon,

specifically whether idioms have separate or shared L1 and L2 representations. While conclusions remain unclear at this point, this topic is further explored in **Chapter 3**.

Another critical question at hand for L2 speakers is whether they rely more heavily on literal meaning even during figurative language comprehension. Following suit with Giora's (1997) Graded Salience Hypothesis, Cieślicka (2006) suggests in her Literal Saliency Model that literal meaning has an increased salience in L2 processing in general. Because of this saliency, literal meanings are processed obligatorily first, resulting in a literal-first by default processing strategy for L2 speakers. In contrast, L1 speakers, having often encountered these figurative expressions, rely on the figurative meaning of an idiom as the most salient meaning. Thus, while this hypothesis suggests that processing is fundamentally different for these two speaker groups, the mechanism remains the same. Evidence for this model comes from a cross-modal priming study in which L2 listeners heard idioms (e.g., *kick the bucket*) and responded to figuratively (DIE) and literally related (PAIL) targets, and participants showed faster responses to literally related targets. Notably, two issues with the study leave room for further investigation: first, the semantic relationship between the literally related target and the constituent word from the idiom is a direct semantic associate (e.g., Nelson, McEvoy, & Schreiber, 1998) and a larger priming effect can be expected compared to an abstract concept like figurative meaning; second, following the first point, without a native-speaker group for comparison, differences in priming cannot be clearly attributed to L2 processing differences alone. However, though not explicitly mentioned in the original study, this model leaves open the possibility that saliency may change with language proficiency, and so, too, may a literal processing priority. This theory is one of the most widely discussed possibilities for differences in L1 and L2 idiom processing, and more recently Arnon and Christiansen (2017) have suggested that differences in L1 and L2 processing may be due to a difference in reliance on the individual constituent words in the L2 compared to larger chunks in the L1. This possible difference in literal priority in L2 processing will be addressed in several studies in the current work (see **Chapters 3**-**6**).

Finally, it is also possible that differences are generally experience-based, and this may be evident in research comparing idioms and novel phrases in L2 learners, which typically associate fast idiom processing with experience and familiarity with the idioms. Unlike the clear evidence for an idiom-superiority effect (i.e., faster idiom processing compared to novel phrases), evidence for non-native speakers is mixed. In reading studies involving both L1 and L2 speaker groups, some authors have found evidence that both groups display faster reading times of idioms and fewer fixations in their terminal constituent words (e.g., Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Underwood et al., 2004). However, other studies have provided conflicting evidence (e.g., Siyanova-Chanturia, Conklin, & Schmitt, 2011). For example, Siyanova-Chanturia, Conklin, and Schmitt (2011) did not find that idioms were read faster than novel phrases nor did eye-movements

show fewer fixations in L2 readers. Additionally, when comparing figurative and literal interpretations of the idioms, figurative readings showed processing costs in L2 readers where no differences were found in L1 readers. A later study by Siyanova-Chanturia, Conklin, and van Heuven (2011), however, suggest that these differences may indeed be due to experience with the language, as is the case with frequency effects in binomial phrases. Testing proficiency effects in idiom-processing, however, is a difficult task as exposure to idioms and their figurative meaning are both an important criteria for testing the L2 participants. In the current study, however, the idea of exposure in learning as well as proficiency are addressed in **Chapters 5** and **6**.

Overall, based on the current body of research in L2 idiom processing, it is still unclear whether differences in the timing of access to figurative and literal meaning as well as a possible lack of idiom advantage effects provide enough evidence to conclude that L1 and L2 speakers process idioms in a different manner. By investigating instances where these speaker groups may show differences in controlled online studies, a more rounded approach to answering this question can be achieved.

1.3 THE CURRENT RESEARCH

In a series of studies, the current dissertation aims to answer the main question:

How do L1 and L2 processing compare?

Answering this question may have important implications for the way idioms are used and understood in non-native interactions as well as how they may be taught to non-native speakers more effectively. In order to answer this larger question, the most defining aspects of idioms will be taken into consideration: their figurativeness, and the resulting dichotomy of figurative and literal meaning, and their fixed, formulaic nature. In particular, the following issues will be addressed:

1. Access to literal and figurative meaning in controlled environments:
    a) Do both native and non-native speakers have access to the figurative meaning of idioms during online processing?
    b) Do both native and non-native speakers have access to the literal constituent meaning of idioms during online processing?
    c) Do both native- and non-native speakers compute the literal phrasal meaning of idioms?
    d) Does the timeline of meaning access differ between native and non-native speakers?
2. Flexibility in meaning access and factors that may influence access to both types of meaning:
    a) Does global (experimental or situational) context impact access to meaning?
    b) Does linguistic context impact access to meaning?
    c) Do idiom-specific properties influence access to meaning?

   d) Do L2-specific challenges (e.g., cross-linguistic influence) impact access to meaning?
3. Processes associated with the formulaic nature of idioms:
   a) Does the predictability of idioms impact access to figurative meaning?
   b) Do listeners predict words within idioms? And if so, how does that impact literal word-activation?
4. Learning and memory for idioms:
   a) Are idioms learned and recalled differently from literal phrases?
   b) Are there L2 specific challenges that may interfere with idiom-learning (i.e., competing unfamiliar cues)?
5. Brain activation during idiom comprehension:
   a) Is idiom and literal language comprehension different on a neurophysiological level ?
   b) Do L1 and L2 speakers differ on a physiological level during idiom comprehension?
   c) Does proficiency impact L2 idiom comprehension?

By addressing these issues individually, a larger picture of idiom processing in native and non-native speakers can be viewed more clearly, and a comparison of L1 and L2 idiom processing can be established.

1.3.1 Methods

The research reported in this dissertation adopts a multidisciplinary approach and makes use of several different methods. The variety of methods used helps address the multifaceted questions associated with idiom processing, as each method offers different types of information. The methods used are described very briefly below.

1.3.1.1 Offline methods

For the purpose of the database in **Chapter 2**, offline data collection measures including ratings and cloze-probability testing were used. As the goal of the database is to estimate knowledge of idioms and their properties, ratings have been the standard measure in the field (e.g., Citron et al., 2016; Titone & Connine, 1994a). Following Preston and Colman (2000), and keeping the scale as simple and comparable to previous studies as possible (e.g., Titone & Connine, 1994a), a 7-point Likert scale was used to measure the subjective responses of participants on linguistic aspects of idioms. Averages from Likert scales can be used to give a general quantitative assessment of how idioms compare to one another across variables, and the values can be applied to analyses in individual studies for further investigation. Additionally, cloze probability tests were used to identify the predictability of idioms. This type of test has been used for decades (e.g., Taylor, 1953) to measure the predictability of words within given contexts. Offline measures such as these are necessary to identify the best items to be used in online studies. However, one issue with these methods is participant reliability, which can vary

widely in these tasks (e.g., Hubers, Cucchiarini, Strik, & Dijkstra, 2019; Tabossi, Arduino, & Fanari, 2011). Therefore, these measures must be combined with sensible selection of variables by experimenters.

1.3.1.2 Cross-modal lexical priming

Cross-modal lexical priming is used to examine lexical meaning activation (literal and figurative) during idiom comprehension in listening in studies in **Chapters 3**, **4**, and **5**. Following this method, participants listen to an auditory stimulus (prime) and respond to a lexical decision (whether or not the presented string is a word) on a visual target that may be presented at different times during or following listening. In this dissertation, primes were idioms in sentence contexts (e.g., John *kicked the bucket*.), and targets were words related to either the figurative meaning of the idiom (e.g., DIE), the literal meaning of a constituent word (e.g., PAIL), or an unrelated word (e.g., MILK). By comparing reaction times between related and unrelated targets, facilitatory priming effects can be determined. In addition to the established effects this method has consistently produced in the field of language processing (e.g., Swinney, 1979; Tabossi, 1988; Tabossi, Colombo, & Job, 1987), it has also shown robust effects in idiom processing literature (Cacciari & Tabossi, 1988; Cieślicka, 2006; Tabossi & Zardon, 1993; van Ginkel & Dijkstra, 2019). One possible issue with this method in idiom processing research is that semantic priming effects are often shown between comparable types of word pairs. In idiomatic studies, this priming can be examined in literal constituent word priming (e.g., BUCKET primes PAIL), while figurative priming (e.g., *kick the bucket* primes DIE) is a more abstract relationship, which may have slightly different connotations and associations between listeners. For that reason, it may be expected that literal effect sizes show larger priming effects than figurative ones, and a direct comparison of these effect sizes should not solely be used to draw conclusions about differences in figurative and literal processes.

1.3.1.3 Eye-tracking visual world paradigm

Eye-tracking, in particular the visual world paradigm, is used in **Chapter 4** to examine the time course of literal constituent prediction and activation. In this method, participants are presented with a visual display containing visual images or words while listening to an auditory stimulus. Typically, one object or word on the display is the target object from the auditory stimulus, and the others are related or unrelated distractors. By recording participants' eye-movements, the visual attention given to objects on the visual display reflect visual attention during listening (see e.g., Huettig, Rommers, & Meyer, 2011 for an overview). The study used in this dissertation applies the same method from Keßler, Weber, and Friedrich (submitted) in which participants listen to an incomplete idiom (e.g., *let the cat out of the*…) and the visual display contains the correct completion (e.g., BAG), a word literally related to the correct completion (e.g., BASKET), and two distractor words (e.g., ARM, STOMACH). By recording the fixations of participants on each of the four display words, analyzed by looks to the

region containing each word, the time course of prediction or integration of the final constituent word can be determined (see e.g., Pickering & Gambi, 2018). These looks can also be compared with looks to the literally related and unrelated targets to consider literal activation of the final constituent word. Unlike cross-modal priming, eye-tracking can provide evidence of lexical activation during listening throughout the entire recording time-course. This may provide more evidence for the processing effects at different times in the course of listening, compared to individual time-points in cross-modal priming. However, one drawback of this method as applied in the current work is the presence of the visual target during listening. As activation may also occur due to the visual presence of a target word, this method is best used in combination with data using other methods such as cross-modal priming or EEG.

1.3.1.4 Self-paced reading

Self-paced reading is used in **Chapter 5** to determine the impact of context on literal and figurative readings of idioms. This method measures reading times of participants by presenting experimental items (sentences) in a self-directed word-by-word or phrase-by-phrase manner. Participants actively click a keyboard key or button to advance to the next part of the sentence at a natural pace. The moving window version of this method, as used in the current dissertation, reveals individual sentence segments (phrases) one at a time, and masks the previous and future segments with dashes in a linear fashion (see e.g., Jegerski, 2014). By varying sentences minimally between conditions, the response times in areas of interest can be compared. Longer reading times are interpreted to indicate increased processing effort, and this method has been used to look at a number of ambiguities in sentence processing for both native (e.g., MacDonald, 1994; Trueswell & Kim, 1998) and non-native speakers (e.g. Dussias & Cramer Scaltz, 2008; Pan & Felser, 2011). In idiom studies, it has also addressed the possible ambiguity of figurative and literal readings of idioms (e.g., Holsinger & Kaiser, 2013). The method, however, has some drawbacks, including the fact that it does not completely represent normal reading, which would allow for regressions and non-linear segments during reading. A more natural phrase-by-phrase approach, compared to word-by-word, allows for slightly more flexibility in reading. However, fewer data points are available for analysis when using phrases compared to words. Additionally, reading time effects can spill over to later regions making analysis and interpretation of results sometimes difficult. Nevertheless, as this method can easily be applied to both native and non-native readers, it provides a sensible approach when looking for possible challenges beyond the phrasal level in idiom processing.

1.3.1.5 Training and recognition paradigm

Chapter 6 adapts a training and recognition paradigm to test memory processes in idiom learning. Following this method, an experiment consists of two individual sessions. In the first session, participants complete a training or learning task, and the second session consists of testing the learning

via recognition tasks. While studies using such paradigms may vary widely in training and testing tasks, depending on the subject under investigation, the testing phase is typically conducted at least 24 hours later in order to allow for consolidation processes to occur (e.g., Takashima & Bakker, 2017). By comparing the correct recall of different conditions based on what was presented in training, this method allows for comparisons either between training types and their effects on items or, as in the current work, between item-types following training. The current dissertation asked participants to learn the meanings of new phrases which were presented for a limited time on a computer screen and tested whether learning was different between item-types (figurative and literal phrases). Testing consisted of recognition of the phrases (YES/NO via button-press), recognition of the phrasal meaning (multiple choice task), and, for L2 learners, recall of translations for individual words (open recall). While there is a plethora of research on idiom learning in classroom and naturalist settings for L1 learners (e.g., Reuterskiöld & Van Lancker Sidtis, 2012) and L2 learners (e.g., Boers, 2001; Steinel, Hulstijn, & Steinel, 2007), there is a lack of research on idiom learning in an experimental setting, and this method to fills such a gap. One drawback of the method as applied here, however, is that participants often perform well on recognition tasks (e.g., Tiv, 2016), and using different testing tasks may indeed produce different results (e.g., Horowitz & Manelis, 1973).

1.3.1.6 Functional magnetic resonance imaging (fMRI)

A neurolinguistic approach is taken in **Chapter 7** by using functional magnetic resonance imaging to compare literal and figurative language processing as well as L1 and L2 idiom processing. Using this method, participants are presented with visual stimuli (sentences) and asked to read sentences as well as complete an unrelated task. During reading, changes in blood flow in the brain can be detected; increases in blood flow are associated with increased activation and cognitive effort. This paradigm has been used in linguistic studies to determine areas of the brain associated with language processing, and in idiom processing, to determine whether idioms, as well as other types of figurative language, are processed differently than literal language (e.g., Mashal, Faust, Hendler, & Jung-Beeman, 2008; Mashal, Vishne, & Laor, 2014; A. M. Rapp, 2018; Winner & Gardner, 1977). Blood flow in the brain is measured using a magnetic field, and images are normalized and averaged across participants to determine areas activated during comprehension. Findings in the field of idiom processing have shown a wide range of differences in activation patterns (see e.g., A. M. Rapp, 2018), but there is to-date no other research comparing L1 and L2 activation during idiom processing using this technique. Unlike the other methods used in this dissertation, fMRI has the limitation that it does not provide a timeline for activation during processing; rather, activation for a large self-selected time window during comprehension must be averaged for examination. Thus, this method can only point out overall differences in brain activation without associating this activation with a time during comprehension.

1.3.1.7 Statistical analyses

All statistical analyses conducted by this author alone used linear mixed effects regression models to analyze the data in each experiment[1]. For this purpose, the program R (R Core Team, 2013) was used with the help of the lme4 package (Bates, Maechler, Bolker, & Walker S., 2015) and RePsychLing (Baayen, Bates, Kliegl, & Vasishth, 2015). Models include both fixed and random effects with random slopes and intercepts (e.g., Baayen, 2008). This type of model accounts for a large amount of variation in the data (i.e., subject- and item-based) simultaneously and has the ability to incorporate many fixed and random effects at once. Additionally, as only one analysis must be conducted, as opposed to two in ANOVAS, it reduces ambiguities in interpretations where separate analyses show conflicting results (e.g., Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Although this method has widely become the standard method of analysis in psycholinguistics, the field is developing rapidly, and there is still little uniformity of process. The current dissertation applies linear mixed effects regression in each chapter based on the latest information available at the time. For example, while the earliest papers included applied a maximal random effects structure as recommended by Barr, Levy, Scheepers, and Tily (2013), later papers reduced the random effects structure by taking measures to ensure that such models were not overparameterized as suggested by Bates, Kliegl, Vasishth, and Baayen (2015) and Matuschek et al. (2017). For this reason, the processes are described in detail for each individual experiment and do include slight variations in their procedures.

1.3.2 Outline

In addressing the main question of how L1 and L2 idiom processing compare, the current dissertation set out to achieve the following goals in order to answer the research question of how L1 and L2 idiom processing compare:

I. Establish a large working database of English idioms containing L1 and L2 (L1 German) norming data to draw from.
II. Assess where critical differences between L1 and L2 idiom processing may lie and either collect experimental data from both groups where there is a paucity of data or collect experimental data from L2 speakers for comparison with existing L1 data.
III. Determine whether L1 and L2 speakers use different processing strategies and/or mechanisms during idiom comprehension.

[1] Analyses done in partnership with other authors used methods deemed appropriate by the additional authors in discussion with this author for the experimental methods involved. These include ANOVAs, running T-tests, and growth-curve analyses.

To this end, the current dissertation includes a series of individual studies addressing each of the smaller research questions outlined at the beginning of this chapter. The organization of these studies by chapter is presented here:

Chapter 2 outlines the German-English Database of Idiom Norms (DIN). Directly addressing the first goal of this thesis, this database was created by selecting 300 (American) English idioms and providing descriptive norms based on a series of online studies. The database includes ratings from both native speakers of English as well as non-native (German L1) speakers of English as these two groups also serve as the participants in studies in the chapters to follow. The idioms also function as materials for the other studies in this dissertation, and the norms are used in both experimental designs as well as in the analyses of results.

Chapter 3 consists of two cross-modal priming studies on native and non-native speakers respectively. These studies aim to compare L1 and L2 meaning access during online processing by asking whether both groups show online access to figurative (idiomatic) and literal constituent meaning. Additionally, by varying the translatability of the idioms from English into German, it tests whether translation facilitates L2 idiom processing. Based on evidence of both figurative and literal facilitatory priming, this chapter serves as a baseline both in method and result for later variations in the dissertation.

Chapter 4 investigates properties typically associated with the formulaic nature of idioms in language learners rather than focusing on their figurativeness. Whereas existing L1 research shows that native speakers access idiomatic meaning early in predictable idioms (e.g., Tabossi & Zardon, 1993) and are able to predict idiom constituents during listening (e.g., Keßler et al., submitted), this chapter address a lack of research in learners on the same topics. In the first series of cross-modal priming studies, L2 learners' access to figurative meaning based on an idiom's predictability is tested. The same method and primarily the same materials from **Chapter 3** are adapted to ask whether L2 learners have access to the figurative meaning before the entire idiom is presented. In the second series of studies, we test whether learners (both L1 and L2) have the ability to predict words within idioms, and how this affects access to literal constituent meanings. As both L1 and L2 learners were tested, the target language was German for practical reasons. Although the idioms were not taken from the database in **Chapter 2**, the same series of norming studies were run on the German idioms for comparability. This study employed eye-tracking using the visual world paradigm to address these questions.

Chapter 5 addresses the impact of two different kinds of contexts on access to meaning in idiom processing: experimental and linguistic context. The first study repeats the cross-modal priming paradigm already used in the dissertation to again fill in a gap in L2 idiom processing research. It asks whether learners' idiom processing is also flexible to global linguistic context. In particular: Are L2 listeners affected by the number of idioms present in the experimental context? The second study is a

series of experiments (L1 and L2) addressing both linguistic context as well as the literal interpretability (literality) of an idiom. In this set of experiments, self-paced reading is used to identify whether context can facilitate both figurative and literal readings of idioms, and how an idiom's literality interacts with this process. Notably, this study addresses the literal meaning of idioms on a phrasal level rather than a constituent level.

Chapter 6 aims to look more closely at idiom learning by examining memory for idioms after a controlled learning environment. Using a learning and testing paradigm, memory for equivalent figurative and literal phrases are compared in L1 and L2 learners using recognition tests in order to determine whether figurativeness provides a memory advantage already early in learning. Additionally, competition between unfamiliar words and the learning of these phrases is tested in L2 learners. By addressing possible differences and challenges in learning idioms compared to literal phrases, the results shed light on possible differences in processing and storage of idiomatic and literal phrases.

Chapter 7 investigates the neurophysiological differences in processing idioms and literal phrases in an fMRI study. This study uses a direct comparison of L1 and L2 brain activation during listening to address whether different physiological processes occur when L1 and L2 listeners comprehend idioms as well as whether differences within these groups occur based on the figurativeness of the phrase. The study also addresses the issue of language proficiency during idiom processing.

While each chapter answers small questions with insights into possible differences in L1 and L2 idiom processing, the results are synthesized and discussed in a broader sense in the conclusion.

2 SHEDDING LIGHT ON IDIOMATIC DIVERSITY: ENGLISH-GERMAN DATABASE OF IDIOM NORMS (DIN)[2]

Abstract

The DIN database is a useful tool for American English idiom research using L1 (English) and L2 (German L1) speakers as participants. The database is publicly available in the form of an Excel file and is attached as an external supplement. The database contains 300 English idioms and descriptive norms for English and German. Each idiom contains descriptive data such as a paraphrase of the meaning, syntactic structure, length (in words and letters), a German equivalent idiom, if available, a literal English translation of the matching German idiom, and a categorization of the level of translatability. Additionally, this database includes ratings on other idiomatic norms from both L1 (American English) and L2 (German L1) users such as familiarity, meaningfulness, and literality. Additional L1 data include decomposability ratings and predictability. The norming data was collected in several studies using OnExp software. The database also includes comparable data for idioms from the study conducted by Titone and Connine (1994). The database serves both as one of the only databases considering non-native speaker judgements[3] as well as the basis for most studies included in the current work.

[2] The data from this chapter are found in a digital attachment as well as available publicly at https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/englisches-seminar/sections/english-linguistics/lehrstuhl-prof-dr-andrea-weber/staff/sara-d-beck/din/ as well as via the CLARIN-D repository and are cited in upcoming texts as the database of project B9 (Beck &Weber, 2016) of the SFB 833 in Tübingen.

[3] Since this study was conducted and placed online, one additional study has considered and analyzed native and non-native speaker judgements on (Dutch) idioms. L1 data is available in Hubers et al. (2019), and L2 data has been submitted for publication.

2.1 IDIOMATIC PROPERTIES

Idioms can be categorized as belonging to both the class of language that is formulaic (e.g., multi-word units, formulaic language, fixed expressions, etc.) as well as the class of language known as non-literal or figurative (e.g., metaphor, irony, etc.). While early research treated idioms as a homogenous group of non-literal expressions and assumed that idiomatic processing occurred in a single manner differing from literal language (e.g., Bobrow & Bell, 1973; Swinney & Cutler, 1979), research in recent decades has taken the heterogenous nature of idiomatic expressions into account, and found that idiomatic processing is also directly affected by the differences between idioms (e.g., Tabossi & Zardon, 1993; Titone & Connine, 1994b; Titone & Libben, 2014). Theories of idiom processing have evolved from step-wise models (e.g., Bobrow & Bell, 1973; Swinney & Cutler, 1979) into more complex hybrid models that account for an idiom's predictability (Cacciari & Tabossi, 1988), decomposability (Gibbs, Nayak, & Cutting, 1989) literality, and familiarity (Titone & Libben, 2014) among other factors. Titone and Libben (2014) even show that these factors interact and affect online processing at different times during listening.

In spite of these individual idiomatic differences, there is still little publicly available data on such idiomatic norms, even in a well-researched language such as English. Furthermore, the databases available (e.g., Libben & Titone, 2008; Titone & Connine, 1994a) typically include only native speaker norms, though research on such idioms is often undertaken on non-native speakers (e.g., Cieślicka, 2006; Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011). While there has been some recent research considering non-native speaker judgements on idioms (e.g., Abel, 2003b; Carrol, Littlemore, & Gillon Dowens, 2017), a comprehensive database remains to be seen. This chapter describes the database consisting of 300 American English idioms and general information on their syntactic structure, length (word and letters), and average word frequency as well as ratings on a variety of properties concerning native speakers of American English (familiarity, meaningfulness, literality, decomposability, predictability) as well as German non-native speakers of English (translatability, familiarity, meaningfulness, literality).

The purpose of this database is not to investigate how each property affects idiom processing individually, but rather to collect information on the most prominent factors known to affect idiom processing on a large scale for further research on idiom processing in American English considering both native and non-native (German L1) speakers of English. The importance of each property in terms of idiom processing is first discussed, and then the series of studies undertaken for the collection of this data are described including brief discussions describing the results. The database is attached as a separate datafile.

2.1.1 Familiarity with the form (familiarity) and meaning (meaningfulness)

Familiarity of an idiom can be described as how well known an idiom is to a language user. However, this familiarity can be described in several ways. Here, the overall familiarity with the form of an idiom itself, also known as the subjective frequency of the idiom, will be called familiarity, while a familiarity with the meaning of the idiom will be called meaningfulness. These two properties have been defined in a number of ways, with familiarity often equated with a combination of these two or even as familiarity with the meaning (e.g., Abel, 2003a). In the current study, we equate familiarity with a subjective or experience-based account of an idiom's frequency (see e.g., Carrol et al., 2017; Libben & Titone, 2008). Although frequency of occurrence in the language is typically one of the greatest predictors for speed of processing in other types of formulaic language (see e.g., Siyanova-Chanturia & Pellicer-Sanchez, 2018), familiarity has been shown to be such a predictor for idioms (e.g., Titone & Libben, 2014). In fact, idioms tend to have very low frequencies in the language while still being very well known to language users. For example, the well-used example *kick the bucket* occurs only 60 times using this word order in the COCA corpus compared to the similar novel expression *kick the ball* at 151 occurrences (Davies, 2008-), and yet the idiom is one of the most well-used and known examples when discussing English idioms. For adult L1 speakers, knowing such prominent idioms seems inevitable and processing them takes little effort. However, for language learners, only exposure facilitates familiarity, and the same knowledge of L1 idioms cannot be assumed in language learners (neither L1 nor L2).

In native speakers, familiarity has been the subject of research in both novel metaphors (e.g., Blasko & Connine, 1993) as well as idioms (e.g., Schweigert, 1986; Schweigert & Moates, 1988), and results from these studies show evidence of improved comprehension and reading times in familiar idioms compared with unfamiliar idioms. Importantly, Schweigert (1986) found that familiar and "less familiar" idioms are both recognized as idioms while "unfamiliar" idioms are not—a finding that suggests that familiarity is a necessary condition for idiomatic processing and interpretation as well as any advantages such processing may provide (e.g., Tabossi et al., 2009). Titone and Connine (1994a) created a more precise measure of what is meant when dealing with familiarity by separating familiarity with meaningfulness, as one may know that an idiom exists without being familiar with the meaning of an idiom. More specifically, familiarity should represent how often an idiom has been encountered (i.e., subjective frequency), and meaningfulness whether or not the person is familiar with the idiomatic meaning of the phrase. Though these two measures highly correlate (.904, $p < .001$), the distinction may still prove useful in identifying how well particular idioms are known, and both this distinction and the definitions used in the previous study (Titone & Connine, 1994a) will be used in the current study.

The impact of idiom familiarity on processing has been less researched in terms of language learners. Levorato and Cacciari (1992) found that familiarity with an idiom increased the chance that children abandoned a literal interpretation, but did not necessarily increase the likelihood of a correct idiomatic interpretation. Thus, for L1 language learners, familiarity helps idiom recognition but not necessarily comprehension. In L2 research, as figurative competence is generally assumed, studies have typically adopted highly familiar idioms (according to L1 ratings) in the assumption that this should improve chances of recognition and comprehension (e.g., S. D. Beck & Weber, 2016a; Cieślicka, 2006) or confirmed familiarity via pre-study testing of the idioms used. Studies in chapters from this work concerning L2 idiom processing (see **Chapters 3**, **4**, and **6**) adopt L2 rather than L1 familiarity and meaningfulness to identify appropriately familiar idioms.

2.1.2 Literality

The potential of an idiom to be interpreted literally is known as literality. While idioms are typically known to mean something other than the compositional meaning of the literal constituents (i.e., they have a figurative interpretation), some idioms have a greater capacity to be interpreted literally than others. For example, *kick the bucket* has the literal interpretation of "to strike a pail with your foot" (Titone & Connine, 1994a) and therefore has a high degree of literality. On the other hand, it is improbable that the phrase *to lose one's cool* (figuratively to lose control of one's emotions) would occur in a literal sense, as it is difficult to determine what the phrase might mean without the help of a specific context. Notably, like decomposability, literality should also be considered in degrees. While the above examples may denote the extremes of high- and low-literality respectively, many more fall somewhere in between these extremes. Not only can grammaticality play a role here, but also the frequency of an idiom's occurrence in both the literal and figurative sense may contribute to their literality (e.g., Milburn & Warren, 2019).

An idiom's literality has been shown to affect L1 speed of processing. For instance, Mueller and Gibbs (1987) found that idioms with a high literality were classified as idioms faster than those with only a figurative meaning. This finding was later corroborated by a reading study by Cronk and Schweigert (1992) in which high literality idioms were read faster than those with a low literality. Both studies attributed this increase in speed to multiple entries in the lexicon (i.e., one literal and one figurative entry). Additionally, low-literality idioms may be less subject to literal constituent activation in processing (e.g., Titone & Connine, 1994b), a finding which suggests that literality may be one driver of idiom processing strategies in native speakers. L2 research considering the literality of an idiom is scarce, and a cross-modal priming study using literality by Cieślicka (2006) did not find any differences in literal or figurative activation based on literality. However, more research is needed in order to determine the limits of literality and its effect on activation in both L1 and L2 speakers. In the

current study, this property is rated in both L1 and L2 speakers and will serve as a theoretical basis for the studies in **Chapter 4**.

2.1.3 Predictability

The predictability of an idiom is the likelihood of correctly completing an incomplete idiomatic phrase (e.g., Titone & Connine, 1994a) and it is often tied to the formulaic nature of idiomatic expressions (e.g., Carrol & Conklin, 2019; Tabossi et al., 2009). An idiom with high predictability is one that is recognized or completed correctly based on the beginning of the phrase before the final word, while low predictable idioms are not recognizable before the final word. It has been shown that predictability is sometimes correlated with familiarity (Titone & Connine, 1994a), but they remain separately important properties as they are not mutually exclusive. Of particular note, studies focusing on predictability within idioms have given way to the concept of the idiomatic key. The idiomatic key is the point in the idiom at which it is recognizable to listeners or readers as an idiomatic expression (Cacciari & Tabossi, 1988). Based on this idea, predictability can also be framed in terms of the placement of the idiomatic key: high predictable idioms have keys early in the phrase while low predictable idioms' keys are found at or near the end (see e.g., **Chapter 3**).

Predictability is also a key measure for idiomatic properties as it may interact with other properties during processing such as literal activation of idiom constituent or phrasal meanings. Titone and Connine (1994b) found, for instance, that highly predictable idioms (e.g., *bury the hatchet*) showed more priming for figuratively related targets (FORGIVE) than low predictable idioms (e.g., *hit the sack* and SLEEP) when target words were presented before the final constituent word, in line with the idea of early and late idiomatic keys. However, they also found that literal priming of constituent words in idioms with a low potential for literal interpretation (low literality) disappears (e.g., *burn the midnight oil* and FUEL). This implies that the predictability of an idiom may also drive processing strategies in native speakers. Additionally, beyond studies examining only L1 adults, of interest is whether or not L2 language learners are able to predict to the same extent as L1 adults and even learners. While Keßler et al. (submitted) find that L1 adults are able to predict and activate the meanings of literal constituents during idiom processing as shown in an eye-tracking and EEG study, additional challenges in L2 learners' abilities to predict during language processing (e.g., Kaan, 2014) leaves open questions about whether L2 learners can predict in the same manner. Predictability in this paper will be measured based on L1 cloze-probability testing and, later in the current work, both predictability and specifically the role of the idiomatic key (see **Chapter 3**) will be examined more in depth.

2.1.4 Decomposability

Idiom decomposability refers to the relation of the individual lexical components to an idiom's figurative meaning. It is precisely this characteristic of idioms that seems to divide idioms into two

(e.g., Nunberg et al., 1994) or more (Gibbs & Nayak, 1989) classes. Non-decomposable idioms are a class of idioms whose individual components don't overtly play a role in the idiomatic meaning of the phrase. *Chew the fat* (figuratively, to casually talk about current affairs or others) and *kick the bucket* (figuratively, to die) are two such idioms. Conversely, decomposable idioms are those whose lexical parts do contribute directly to the figurative meaning. For example, Gibbs et al. (1989) argues that *pop the question* is a decomposable idiom because, when taken apart, it appears that the words contribute to the meaning. The verb *pop* refers to "the act of uttering [the question]" while the question is interpreted as a "marriage proposal" (p. 577). The authors argue further, though, that decomposability, like many other characteristics of idioms, can be measured in degrees; not all decomposable idioms are equally decomposable. Thus, another division common for the decomposability of idioms is into normally and abnormally decomposable idioms. The previous example, *pop the question*, is a normally decomposable idiom since there is a clear and obvious contribution of the individual parts to the literal meaning, but *carry the torch* (figuratively, to be in love with someone) is abnormally decomposable as the parts contribute to the meaning via the conventional metaphor of torches as warm feelings (Gibbs et al., 1989, pp. 577-578).

Studies on the decomposability of idioms have played an important role in idiom processing research in both the L1 and the L2 (e.g., Abel, 2003a, 2003b; Gibbs et al., 1989). Decompositional models of idiom comprehension assume that idioms are semantically decomposed during processing, and therefore decomposable idioms should have a processing advantage over non-decomposable idioms (e.g., Gibbs, 1980; Gibbs & Nayak, 1989). While these claims have not been confirmed in later studies (e.g., Titone & Connine, 1999), decomposability may play a role in L1 tasks involving competing meanings (e.g., Libben & Titone, 2008). L2 research on the matter is even more complex, as Abel (2003a) suggests that L1 and L2 speakers differ in their judgements and even ability to decompose idioms. L2 speakers may rely on decomposition as a comprehension strategy and therefore rate even non-decomposable idioms as decomposable. For this reason, only L1 judgements on decomposability are collected in the current study. Furthermore, decomposability does not drive the theoretical questions in the current work, but it serves as a comparable measure to previous databases and remains available as a possible mode of investigation where unexpected differences between L1 and L2 speakers occur.

2.2 DATA COLLECTION

300 American English idioms were collected and used in the current study. For the database, all quantitative data was collected using ratings or other simple tasks in a series of online experiments using free software from OnExp (Onea, 2011). Descriptive data for each idiom was either summarized from the objective properties of the idioms (e.g., syntactic structure, length, etc.) or gathered from

textual or informant sources (e.g., constituent word frequency, translatability). The study was divided into three individual tasks with separately collected data as follows:

1. Familiarity, meaningfulness, literality (L1 and L2)
2. Decomposability (L1)
3. Predictability (L1)

These variables were collected separately for important practical reasons that may have influenced the data. The first three variables collected in Experiment 1 involved the idiomatic phrase without definitions. The task was designed so that L1 (American English) and L2 (German L1) participants would use the same experiment with no differences. Experiment 2 was designed only for native speakers, as already mentioned above, as non-native speakers have proven unreliable in their ratings on decomposability (e.g., Abel, 2003a, 2003b). Additionally, this Experiment was not combined with Experiment 1 for concerns over experiment length as this task involved more concentration and consideration per item. Finally, Experiment 3 was designed for native speakers as a separate task in order to present the task as one not explicitly about idioms and excluded non-native speakers to avoid responses based on the variability of the L2 experience. Unlike Experiments 1 and 2, Experiment 3 also included phrases not meant to be idioms, and the instructions do not reference idioms in order to achieve a more natural measure of the phrase predictability than may be given if participants know that they should complete the idiom. The collection of data for each individual study is described briefly below.

2.2.1 Materials

Of the 300 American English idioms, 171 were taken from the database of Titone and Connine (1994a). The rest were selected based on their use in other studies (e.g., Libben & Titone, 2008) or taken directly from idiom dictionaries aimed at native speakers (e.g., Ayto, 2010) and learners of English (e.g., Francis, 2006). American English idioms were the focus of the database as this group of speakers serves as the main source of L1 speakers for the studies in the rest of this work.

2.2.1.1 Variables determined by the experimenter

Descriptive information for each idiom was collected in order to provide as much information as possible for future studies. These descriptors involved general information such as meaning (definitions), basic syntactic structure, length (number of letters and words), constituent word frequency, and L2-relvant information such as translatability (description below). Where possible, the information was collected from source texts (i.e., meanings taken from earlier studies or dictionaries). Simple syntactic structures were summarized with focus on the head phrase-type so that studies wishing to adopt similarly structured idioms can use these markers. Information needing additional resources and/or informants is described below.

2.2.1.2 Constituent frequency

Frequency for constituent words was collected from the SUBTLEX corpus of American English (Brysbaert & New, 2009) which considers subtitles in American English and includes 51 million words. The value for each word was collected from the total count in the database transformed into occurrences per million words, and then transformed into the log value for easy comparison between items. The individual word values for all constituent words were then averaged across the number of words in the idiom, and this average log-transformed value is displayed in the database.

2.2.1.3 Translatability

The translatability between German and English is recorded and measured using a number of indicators. First, idioms were translated into their German counterparts with the help of idiom dictionaries (e.g., Schemann & Knight, 2011), online resources (e.g., leo.org, linguee.com) as well as highly proficient L2 English (German L1) student assistant informants. Where an equivalent idiom was determined, this is listed as the "Matching German Idiom" in the database. For classification purposes, a literal translation of the German equivalent idiom (word-for-word) is also given in the database. Based on the equivalent idiom and its literal translation, the translatability, as categorized by Liontas (2002) and Irujo (1993), was determined according to how closely the idioms match in their word-for-word translations. Where translation of content words from German to English produced the same idiom with an equivalent meaning (e.g., English: *the tip of the iceberg*, German: *die Spitze des Eisbergs*), idioms were classified as Lexical Level (LL) idioms. Semi-Lexical (SL) idioms match in meaning and have partial matches in content word translations (e.g., English: to get the ball rolling, German: *den Stein ins Rollen bringen/to get the stone rolling*). Idioms that do not overlap in meaning or form when translating were classified as Post-Lexical (PL) idioms.

As there was not a database of equivalent idioms from English to German that covered all of the idioms in the current database, some of the idioms produced either multiple idiomatic matches or unclear results. For example, the idiom *to rise to the bait* was translated into German as *anbeißen* (to bite). The German translation, though metaphorical in nature, is less clearly idiomatic than the original English idiom, and this is noted in the column titled "Translatability Notes". Additionally, any cases where an idiom overlaps in German and English in form but not meaning that was noted during translation is recoded under "False Friends/Other Notes". While the translations and notes encompass the knowledge accrued by this author, they in no way encompass all of the possible translations, and are meant to consider the most known equivalents.

2.2.2 Experiment 1: Familiarity, meaningfulness, literality

Familiarity, meaningfulness, and literality were measured by collecting ratings from native and non-native (German L1) speakers of English.

2.2.2.1 Participants

136 native speakers of American English (average age=33 *SD*= 10.398, 77 male) and 120 non-native speakers (L1 German, average age = 24, *SD*= 4.836, 21 male) participated in the study. The recruiting of native speakers was done via the University of Maryland's Language Science Center in the USA, and the recruiting of non-native speakers was done online via email and Facebook from the University of Tübingen in Germany. L1 participants were paid online via Amazon's Mechanical Turk (MTurk) and L2 participants were compensated for their time via participation in a gift card lottery.

2.2.2.2 Materials

The 300 idioms were divided into 6 lists (50 idioms per list), and a minimum of 20 L1 and L2 participants saw each list.

2.2.2.3 Procedure

Participants first consented to participation in the study and then saw explanations including examples for each measure individually. Participants were then asked to rate the idioms individually on their *familiarity*, *meaningfulness*, and *literality* on a scale from 1-7 (1 being not at all, and 7 being extremely familiar, meaningful, or literal). The study took about 10 minutes to complete. Screen shots from the study can be found in Appendix A.

Instructions were adapted from previous databases (e.g., Titone & Connine, 1994a). *Familiarity* was described as how often a person has come across the idiom in any discourse situation (written and spoken). A rating of 1 indicated that the participant had never before encountered the idiom, and 7 indicated that the participant encountered the idiom very frequently. *Meaningfulness* was described as how meaningful the idiom is to the participant in the sense of how well understood the figurative meaning of the idiom is. The ends of the scale indicated 1 that the idiom was not at all meaningful and 7 that the idiom was very meaningful. *Literality* was defined as the extent to which the idiom can be interpreted literally (i.e., whether a literal interpretation of the phrase is well-formed and probable). A rating of 1 indicated that there is no possible literal interpretation for the idiom, and 7 indicated that there is a very clear, well-formed literal interpretation.

2.2.2.4 Analysis and results

The responses for each idiom were compiled and their mean ratings from all participants as well as the standard error of the mean are recorded in the database. Idioms covered a wide range of values for each measure, and L1 and L2 intuitions differed in familiarity and meaningfulness. The average means for each descriptor and the standard deviation is displayed in Table 2-1. L1 familiarity ratings ranged from 2.8 (*swing the lead*) to 6.83 (*speak your mind*), with an overall average of 5.74 compared to a wider range for L2 ratings between 1.60 (*gild the lily*) to 6.95 (*play with fire*) and a lower average of 4.09. Paired t-tests confirmed that these averages significantly differed from one another ($t = 17.39$, $p < .001$).

This result was similar for meaningfulness with L1 ratings ranging from 2.08 (*carry the can*) to 6.87 (*out of the blue*) compared to the L2 range from 1.15 (*gild the lily*) to 6.95 (*read between the lines*). The overall difference for meaningfulness was also significant according to a t-test ($t = 12.38$, $p < .001$). Ratings for literality were generally similar overall between the two groups. L1 literality ratings ranged from 1.52 (*eat one's words*) to 6.41 (*play the game*) and L2 ratings ranged from 1.40 (*steal someone's thunder*) to 6.35 (*play with fire*). A t-test confirmed that L1 and L2 ratings do not significantly differ from one another ($t = 0.56$, $p = .579$).

Table 2-1. Average ratings for all 300 idioms by language group

	Familiarity	**Meaningfulness**	**Literality**
L1	5.74 (0.94)	5.89 (0.82)	3.97 (1.05)
L2	4.09 (1.34)	4.85 (1.21)	3.92 (1.03)

Pearson correlations between each norming descriptor were calculated across all ratings for the idioms per speaker group first individually. Overall, all calculated correlations, displayed in Table 2-2, were highly significant in both L1 and L2 participants. For both groups, the correlation between familiarity and meaningfulness was the strongest (Pearson's $r > .75$). Literality, however differed between speaker groups. For L1 speakers, literality had a poor correlation with both familiarity and meaningfulness (Pearson's $r \leq 18$), whereas it correlated more strongly for L2 speakers (Pearson's $r \geq .30$). These results suggest that L2 speakers are increasingly more likely to determine that an idiom has a literal meaning if it is more familiar in both frequency and meaning.

Table 2-2. L1 and L2 correlations between ratings

	L1		**L2**	
	Familiarity	**Meaningfulness**	**Familiarity**	**Meaningfulness**
Meaningfulness	0.76***		0.78***	
Literality	0.16***	0.18***	0.30***	0.38***

Note: ***p<.001

Additionally, further Pearson correlations considering the average ratings for each idiom per norming descriptor between L1 and L2 participants were conducted. All correlations between speaker groups were also highly significant ($p < .001$), with literality showing the strongest correlation (Pearson's $r = .70$) compared to familiarity (Pearson's $r > .43$) and meaningfulness (Pearson's $r > .42$). These correlations between groups are critical in that they further confirm the results of the t-tests between L1 and L2 ratings suggesting that L1 and L2 speakers have similar intuitions about an idiom's literality. Thus, when considering this norm for studies concerning both groups, L1 ratings may indeed

be a good measure if idioms are highly familiar to L2 speakers. Additionally, although the correlations for familiarity and meaningfulness are relatively strong, there appear to be critical differences in idiom knowledge and experience, as also demonstrated by the significant differences in familiarity and meaningfulness ratings. These differences suggest that L2 indicators for these norms (either familiarity or meaningfulness) are crucial in studies using L2 speaker groups.

2.2.3 Experiment 2: Decomposability

Idiom decomposability is measured in this rating study by native speakers of English.

2.2.3.1 Participants

136 native speakers of American English (average age= 33, SD= 11.67, male= 77) participated in the study. Participants were recruited online via Amazon's MTurk website and were paid to take part in the study. L2 ratings on decomposability were not collected as previous research has shown that L2 participants use decomposition as a method for comprehension, even when idioms are not decomposable, and therefore do not provide accurate ratings for this measure (e.g., Abel, 2003b).

2.2.3.2 Materials

The 300 idioms were divided into 6 lists (50 idioms per list), and a minimum of 20 participants saw each list.

2.2.3.3 Procedure

After providing consent, participants were given detailed instructions and examples on how to rate decomposability. While definitions on this variable vary between authors (e.g., Gibbs & Nayak, 1989; Titone & Connine, 1994a), we adapted the definition from Titone and Connine (1994a) for easy comparison. *Decomposability* was defined as the extent to which the individual words contribute to the overall figurative meaning of the idiom. Participants were asked to decide whether the idiom is decomposable by selecting YES or NO, and then to indicate how difficult they found the decision on a scale from 1-7 (1 indicating very easy and 7 indicating very difficult). The indication of difficulty for the task was intended as a measure of certainty. The entire task took about 20 minutes. Screen shots from the study can be found in Appendix A.

2.2.3.4 Analysis and results

Participants' YES/NO responses were compiled into the percentage of responses indicating that the idiom is decomposable. Based on the classifications from Abel (2003b), idioms with 85% of responses indicating either YES or NO were classified as decomposable (D) or non-decomposable (ND), respectively. If idioms did not reach this threshold, they were classified as neither (NA). Both the percentage of YES responses and the classification are listed in the database. Additionally, the ratings on the difficulty of the decision were averaged for each idiom, and this mean as well as the standard

error of the mean is provided in the database. While the difficulty of the decision cannot be directly compared to the earlier databases as it was not measured, it serves as an indication of certainty in the case that ratings differ either from earlier studies or experimenter expectations.

Table 2-3. Decomposability classifications and average responses per category

Decomposability	**n**	**YES Responses**	**Difficulty (*sd*)**
decomposable	67	85%	2.79 (0.51)
non-decomposable	51	15%	3.04 (0.50)
neither	182	52%	3.40 (0.05)

Overall, 67 idioms were classified as decomposable, 51 as non-decomposable, and 182 did not meet the threshold for either. This distribution is displayed in Table 2-3 including also the average percentage of decomposable classifications (YES responses) and difficulty rating (from 1-5) for each category. As a whole, the idioms rated and classified as decomposable were easiest for participants, while the unclassified idioms were the most difficult. The distribution of mean difficulty ratings by the percentage of YES responses is displayed in Figure 2-1. While one may use the percentage of responses as scalar indication of how decomposable the idiom is, the variability of these responses may be an argument against such use. Additionally, as a majority of the idioms could be placed neither as clearly decomposable or non-decomposable, using only decomposability as a norm for studies of idiom processing may be difficult based on these great disagreements in classification.

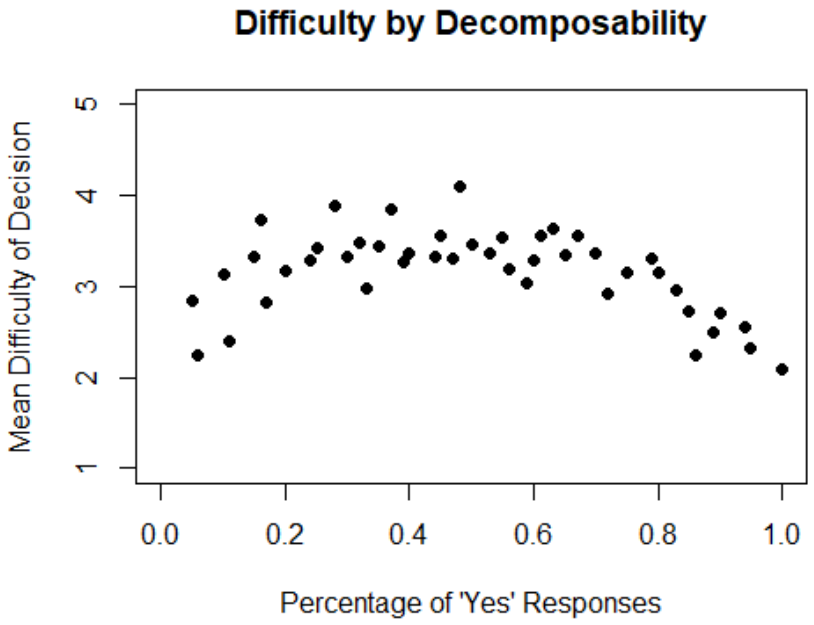


Figure 2-1. Average difficulty of decision ratings by percent of YES decomposable responses

2.2.4 Experiment 3: Predictability

Idiom predictability is measured in this study using cloze-probability responses in order to identify the placement of the idiomatic key of the idiom (Cacciari & Tabossi, 1988).

2.2.4.1 Participants

183 native speakers of American English (average age= 33, SD= 9.92, male= 80) participated in the online study. Participants were recruited online via Amazon's MTurk website and were paid to take part in the study. Native speakers were used for this measure so as to ensure that predictability was more dependent on the predictive properties of the idiomatic phrase rather than differences in exposure via L2 learning of the idiom.

2.2.4.2 Materials

For this study, idioms were first divided into 4 lists, and final constituent words were removed from the idioms (e.g., *throw money out the …[window]*) as in Titone and Connine (1994a). Each list consisted of 160 items including fillers, all of which were incomplete phrases. A total of 100 filler items were developed for this purpose with differing lengths (e.g., *do the*…, *ladies and*…, and *have a problem with*...). If idioms were correctly identified in one of the four lists by 70% of participants, they were compiled and included in a new list in an additional round of testing with one additional word removed (e.g., *throw money out …[the window]*). Unlike in the previous study, which looked at idioms with only the final word removed, this process was repeated several times to allow for a more precise pinpointing of the idiomatic key. This process was repeated until idioms were no longer correctly identified. In total, 8 lists were presented to at least 20 participants each: 4 lists in Round 1, 3 lists in Round 2, and 1 list in Round 3.

2.2.4.3 Procedure

After giving consent, participants were informed that they would be doing a "complete the phrase" task. Participants were asked to complete the presented phrase naturally with the first word or words that come to mind which also complete the phrase in a meaningful manner. A screen shot with the examples and instructions can be found in Appendix A. The task took about 20 minutes to complete.

2.2.4.4 Analysis and results

Correct responses to idioms were analyzed after each individual round as correct or incorrect. In the case of items that were nearly correct, two native speaker judges agreed on a correct or incorrect judgement. A threshold of 70% of correct completions was set in order to determine the placement of the idiomatic key (Cacciari & Tabossi, 1988; Titone & Connine, 1994b). If the idiom was not completed correctly in the first round with a single constituent removed, the percentage of correct completions is listed in "Idiomatic Key, Predictability Percentage" in the database, and the idiomatic key is identified as the final constituent word. If the idiom was correctly identified, the idiomatic key

is marked with the symbol | following the key word under "Idiomatic Key" and the percentage of correct completions with the presentation of this word is listed in "Idiomatic Key, Predictability Percentage" in the database.

Table 2-4. Summary of total idioms and norming values from L1 and L2 participants

		L1 Ratings				L2 Ratings			
Predictability	***n***	**Familiarity**		**Meaningfulness**		**Familiarity**		**Meaningfulness**	
final word	233	5.80	(0.88)	5.62	(1.01)	4.81	(1.23)	4.09	(1.37)
penultimate word	39	6.25	(0.37)	6.20	(0.41)	4.83	(1.02)	3.96	(1.11)
earlier	28	6.21	(0.41)	6.10	(0.51)	5.18	(1.27)	4.27	(1.42)

Note: *n* stands for the total number and the value in parenthesis is the standard deviation.

Of the 300 idioms, 233 were unpredictable before the final word, 39 were predictable at the penultimate word, and 28 were predictable at least 2 words before the end. See Table 2-4 for a summary of the results and average ratings for each of these groups. Comparisons between groups were not undertaken, as the group of unpredictable idioms is much greater than the other two groups. However, it appears that the most predictable idioms are also very familiar and meaningful. While these only represent L1 intuitions on predictability, this study set a high threshold of 70% predictability compared to 50% in earlier studies (e.g., Titone & Connine, 1994b). This higher threshold of 70% may better represents the intuitions of more language users, and hopefully include less-experienced language users, such as L2 or L1 learners.

2.3 CONCLUSION

The current study provides an important addition to the available descriptive norms on American English idioms. In addition to updating existing databases with more current ratings (e.g., Titone & Connine, 1994a), it adds the novelty of L2 ratings for appropriate norms. Critically, some but not all of the measures correlated between L1 and L2 participants. Research that considers more than one speaker group must therefore also consider norms from the participant groups being tested rather than relying on L1 norms alone. Including L2 ratings allows researchers to more carefully select appropriate idioms for studies (e.g., high-familiarity for testing idiomatic processing or low-familiarity for testing acquisition and learning) and it opens up more possibilities for data analysis that may account for variability (i.e., including such norms in LMER models).

Finally, the online availability of the database as well as the raw data serves as an open source for future L1 and L2 (German L1) research in idiomatic processing and may also serve as an example to researchers looking to expand on L2 information about such idioms in other languages. While there are a number of studies dealing with L2 learner groups (e.g., Cieślicka, 2006; Siyanova-Chanturia,

Conklin, & Schmitt, 2011), many studies continue to utilize their own individual ratings in pre-studies because of a lack of availability of norming data on idioms for L2 learners or a lack of idioms appropriate for specific tasks with norms, even in a language as studied as English. The current data are also stored in the publicly available and searchable Tübingen CLARIN-D Repository (https://talar.sfb833.uni-tuebingen.de/about/), which specializes in web-based services for the Humanities and Social Sciences (see https://www.clarin-d.net/en/). By archiving the data here, the data will be available in a sustainable, long-term non-proprietary format, and can be easily found and accessed by future researchers studying idiomatic processing and beyond.

3 BILINGUAL AND MONOLINGUAL IDIOM PROCESSING IS CUT FROM THE SAME CLOTH: THE ROLE OF THE L1 IN LITERAL AND FIGURATIVE MEANING ACTIVATION[4]

Sara D. Beck & Andrea Weber

Abstract

The present study examines non-native (L2) and native (L1) listeners' access to figurative idiomatic meaning and literal constituent meaning in two cross-modal priming experiments. Proficient German learners of English (L2) and native speakers of American English (L1) responded to English target words preceded by English idioms embedded in non-biasing prime sentences in a lexical decision task. English idioms differed in levels of translatability: *Lexical level idioms* had word-for-word translation equivalents in German, while *post-lexical level idioms* had matching idiomatic concepts in German but could not be translated word for word. Target words either related to the figurative meaning of the idiom or related to the literal meaning of the final constituent word of the idiom (e.g., *to pull someone's leg*, literal target: *walk*, figurative target: *joke*). Both L1 and L2 listeners showed facilitatory priming for literally and figuratively related target words compared to unrelated control target words, with only marginal differences between the listener groups. No effect of translatability was found; that is, the existence of word-for-word translation equivalents in German neither facilitated nor hindered meaning activation for German L2 listeners. The results are interpreted in the context of L1 and L2 models of idiom processing as well as further relevant translation studies.

[4] This chapter is adapted from the published version of the following article: Beck, Sara D.; Weber, Andrea (2016): Bilingual and Monolingual Idiom Processing Is Cut from the Same Cloth: The Role of the L1 in Literal and Figurative Meaning Activation. In *Frontiers in Psychology* 7, p. 1350. DOI: 10.3389/fpsyg.2016.01350.

3.1 INTRODUCTION

While understanding idioms is usually easy for native (L1) listeners of a language, non-native (L2) listeners often find recognizing and understanding them to be a stumbling block. Like other facets of figurative language, idioms are both complex and pervasive in language; Pollio et al. (1977) as cited in (Cooper, 1999) estimated an average of 7,000 idioms a week for L1 speakers based on the occurrence of figurative language analyses of political debates, psychology texts, novels, and psychotherapy sessions. Though idioms are not a homogenous group, researchers can generally agree that idioms are multi-word expressions with limited variation in syntactic structure. Additionally, the meaning of an idiom typically differs from the literal meanings of the individual constituent words (see e.g., Liu, 2008). The fixed nature of idioms might suggest that non-native speakers can learn and use idioms quickly, but the conventional misalignment between figurative and literal meaning poses a particular challenge. For example, when a speaker expresses that she is *in hot water*, one can interpret from the figurative meaning that she is in trouble rather than assuming that she is literally submersed in heated water, as in a hot bath or hot springs. What might be obvious to a native listener might not be to a non-native listener, and the expression could confuse rather than inform an L2 listener about the situation. L2 proficiency would, however, benefit significantly from mastery of L2 idiomatic expressions: not only would it make L2 speakers sound more native-like (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006) it would also free up processing capacities since fixed multi-word expressions are known to be easier to process than novel phrases (Conklin & Schmitt, 2008; Pawley & Syder, 1983; Siyanova-Chanturia, 2013). In fact, an ability to produce and comprehend idiomatic expressions is one of the measurements used to determine the level of English of non-native users (see e.g., "Common European Framework of Reference for Languages," 2001). Compared to the broad research on L1 idiom processing, a number of questions remain for the L2. In the studies presented here, we addressed two major questions: How does L2 access to figurative meaning of an idiom and literal meaning of constituent words compare to L1 access? And, what is the role of the L1 in this process? More specifically, can L1 similarities (i.e., translatability) ease L2 processing? To address these questions, we tested online processing of figurative and literal meaning in idioms in L1 and L2 listeners. Additionally, we explored the impact of the L1 on the L2 by addressing the translatability of the idioms from the L2 into the L1 and tested idioms that were either directly translatable or not translatable.

L1 idiom processing research, though diverse in focus, has put weight on the comparisons between processing literal and figurative meaning as well as processing idioms and novel phrases. The fact that idioms often have both a literal and a figurative interpretation has provided for a rich field of research (Cacciari & Corradini, 2015; Cacciari & Tabossi, 1988; Libben & Titone, 2008; Tabossi, Fanari, & Wolf, 2008). While the figurative meaning of an idiom is conventionalized and known to native speakers, its literal interpretation can be either logical, nonsensical, or somewhere in between.

Though it is possible that someone is bathing in the example of being *in hot water* (with an idiomatic or figurative interpretation denoting "in trouble"), in the idiom *to be on cloud 9* (with a figurative interpretation of "being very happy") there is not a likely, logical interpretation in the real world in which a person can be found on a cloud called "9." Furthermore, when considering the literal interpretation of an idiom, research can remain on the phrasal level or can consider access to the literal meaning of the constituent parts. When again considering the idiom *in hot water*, we can focus on access to the figurative interpretation, "in trouble," access to the whole interpretation of the literal phrase, "to be in heated water such as a bath or hot springs," or we can consider access to the meanings of the individual constituent words such as "hot" or "water."

In the presence of such diverse processing possibilities, various theoretical approaches to the processing of idioms concerning access to literal constituent meaning and figurative meaning have been put forward in the last decades. Early approaches treated idioms homogenously and suggested two individual modes of processing for literal and figurative meaning. These approaches suggested that processing occurred in stages which had, generally, three possibilities: literal meaning first (e.g., Bobrow & Bell, 1973), figurative meaning first (e.g., Gibbs, 1980), and simultaneous processing (e.g., Swinney & Cutler, 1979). These stage approaches to idiomatic processing often assumed figurative meaning as part of retrieval, while literal processing involves composition. Evidence from psycholinguistic studies quickly showed that, in some cases, figurative meaning is available more quickly than literal meaning (Cacciari & Tabossi, 1988; Ortony, Schallert, Reynolds, & Antos, 1978; Swinney & Cutler, 1979), discounting purely literal-first processing approaches and pushing for simultaneous processing models or more complex models. Furthermore, such models of processing have been criticized as idioms have been more systematically categorized and labeled based on their diverse properties (Gibbs, Nayak, & Cutting, 1989; Nunberg, Sag, & Wasow, 1994; Titone & Connine, 1994b) such as literality (Titone & Connine, 1994b), predictability (Cacciari & Tabossi, 1988), decomposability (Gibbs et al., 1989), familiarity (Tabossi et al., 2009), saliency (Giora, 1997), and more recently emotional valence and arousal (Citron et al., 2016). Theoretical approaches and experimental methods have adapted to incorporate psycholinguistic findings on such idiomatic properties. For instance, both familiarity (see e.g., Libben & Titone, 2008 for an overview) and a strongly biasing figurative context (Cacciari & Corradini, 2015; Cacciari, Padovani, & Corradini, 2007; Colombo, 1993) have been shown to affect access to figurative meaning. In fact, Giora (1997, 2002) suggests that saliency—conventionality, frequency, familiarity, contextually supported—rather than an individual factor is most relevant in idiomatic processing. Instead of differentiating between literal and non-literal processing, she separates salient processing from non-salient processing: Saliency facilitates processing as salient meanings are retrieved immediately and directly, while non-salient meanings are retrieved through default language processing and integration processes. Thus,

access to figurative meaning is dependent on saliency rather than individual idiomatic properties. In contrast, Gibbs and Nayak (1989) adapted a theory of comprehension via meaning mappings after finding that decomposability, or the extent to which the meaning of individual word components in idioms contributes to the overall figurative meaning, also affects this access. In a timed response task in which participants decided whether a phrase was meaningful or not, participants were faster when responding to decomposable idioms than non-decomposable idioms, suggesting that composition via meaning mappings helps comprehension in decomposable idioms, while the missing possibility to do so in non-decomposable idioms slows down comprehension.

Cacciari and Tabossi (1988) tested the effect of the presence of an idiomatic key—an aspect of predictability (see Titone & Connine, 1994b)—a point at which the configuration of words is recognized as an idiom and found that this aspect of predictability also impacts the availability of figurative meaning (e.g., in the idiom *give the cold shoulder*, the word cold is the idiomatic key as listeners can recognize the configuration as an idiom at this point in the idiom). Presented with target words related to the idiomatic meaning and the literal meaning of an idiomatic phrase, participants' reaction times for figurative targets elicited faster responses only in highly predictable idioms while literal targets elicited faster responses only in not highly predictable idioms. Thus, the speed of access to figurative meaning can differ based on the placement of the key within the word—the sooner it occurs, the faster idiomatic meaning is available—after which point the configuration is recognized as an idiom and meaning is accessed through retrieval rather than composition. In a more recent study, Titone and Libben (2014) took several of these idiomatic differences into consideration in a study examining the retrieval of idioms and their figurative meaning. In their cross-modal priming study, participants made a lexical decision on target words related to the figurative meaning of the idiom in one of four positions (in two experiments). Idioms differed in familiarity, decomposability, literality, and final word predictability. They found that differing idiomatic properties modulate meaning over time in idiomatic processing, namely that high literality can hinder idiomatic processing before phrase offset; familiarity can facilitate processing at the phrase offset, and high decomposability can hinder processing 1000ms after phrase offset. Titone and Libben interpret these results in support of a hybrid model of idiom processing in which all available information is used to facilitate processing; the result is both direct retrieval of figurative meaning and composition.

While psycholinguistic evidence generally supports processing models that take into account the heterogeneity of idioms and contexts, there are still open questions concerning their relative importance concerning access to figurative meaning in comparison to literal constituent meaning. However, research has consistently shown that not only is access to figurative meaning available online, but in some circumstances, it can have an advantage over literal constituent meaning. Additionally, idiomatic processing has been consistently found to be faster than processing of

comparative novel phrases (Tabossi et al., 2009). Though it's unclear whether idiom processing is essentially different from literal language processing, native speakers seem to have *mastered the art* of idiom processing.

Though there is considerably less headway in L2 idiom processing, like L1 research, L2 research addresses both the comparison of access to figurative and literal meaning and the comparison of idiomatic processing to novel language have been in focus. However, there is more variation in the results exploring overall access to figurative meaning. Using eye-tracking methods, researchers such as Conklin and Schmitt (2008) and Underwood, Schmitt, and Galpin (2004) and Siyanova-Chanturia, Conklin, and Schmitt (2011) investigated access to figurative meaning in comparison to novel phrases. Conklin and Schmitt (2008) found that L1 and advanced L2 English users read idioms, belonging to formulaic language or multi-word expressions, more quickly than comparable novel phrases whether used figuratively or literally. Underwood et al. (2004) only partially confirmed this advantage. In a reading task in which idioms were embedded in extended contexts, both L1 and L2 English users fixated less on terminal words in idiomatic phrases than in comparable novel phrases; however, they did not find the same results for total fixation length as L1 English users did not need to look as long at final words in idiomatic phrases compared to novel phrases and L2 English users showed no significant difference in the length of the gaze. These results suggest a more complex picture of the processing of idioms and other formulaic language that might be heavily influenced by disadvantages particular to L2 language use. Neither of these studies, though, are able to draw conclusions concerning access to literal in comparison to figurative meaning. In contrast, in a study run by Siyanova-Chanturia et al. (2011), L1 and L2 English users read idioms in a biasing story context used figuratively (*at the end of the day*—eventually) literally (*at the end of the day*—in the evening) in comparison to a matched novel phrase (at the end of the war) and found that L1 users read idiomatic phrases, both literal and figurative uses, more quickly than novel phrases while proficient L2 users did not show differences. They did find, however, that the figurative meanings of idioms required more time to retrieve than the literal interpretation in L2 users only.

Further evidence for the priority of literal meaning in L2 idiom processing is presented by Cieślicka (2006). Following ideas about L1 language processing presented by Giora (1997), Cieślicka proposed that literal meanings are always most salient for L2 users, regardless of their frequency and familiarity, as literal words and phrases are more likely to be used and encountered by language learners. Additionally, Cieślicka examined whether the saliency of literal meaning differs with varying degrees of literality, the degree to which the idiom can be interpreted literally. In a cross-modal priming study comparing reaction times to visual targets related to literal constituents of idioms (*bury the hatchet*—AXE) and targets related to the figurative interpretation of the idiom (*bury the hatchet*—FORGIVE) in advanced Polish learners of English, Cieślicka found more facilitatory priming for

literally related targets than for figuratively related targets in both idioms with a highly literal interpretation and those without a very literal interpretation. This is used as support for the *Literal Salience Model,* which suggests a priority for literal meaning based solely on L2 users' superior experience with literal meanings—saliency—compared to L2 users' limited experience with figurative meaning in their L2. Thus, unlike previous findings in L1 research, literal meaning seems to have a processing priority over figurative meaning, even in known idioms.

Among consistent research indicating that L2 language users face a variety of linguistic challenges when encountering figurative language, research in L2 idiom processing has also begun to question the effects the L1 can have on L2 figurative processing, particularly access to meaning via translation and cross-language facilitation. Research on cross-language facilitation for single word translation equivalents in bilinguals has consistently shown priming effects (Carrol & Conklin, 2014; Chen & Ng, 1989; de Groot & Nas, 1991; Duyck & Brysbaert, 2004; Sunderman & Kroll, 2006). Chen and Ng (1989), for example, found semantic facilitation using translation equivalent words as well as pictures in Chinese-English bilinguals. In line with these results, Kroll, van Hell, Tokowicz, and Green (2010) point out that translation is conceptually mediated, and advanced L2 learners need not translate words, but rather access concepts immediately; thus, any facilitation might be available via a conceptual language non-selective level. Although, Brysbaert and Duyck (2010) warn that this mediation may be more difficult in the case of abstract words (see also de Groot, 1992), a potential problem for figurative language. Multi-word units, for example collocations, have also shown similar effects to single words. Wolter and Gyllstad (2011) found facilitation in English word pairs forming word-for-word translations of word pairs in Swedish compared to English-only word pairs in bilinguals of Swedish and English. Thus, it seems that a conceptual level might also mediate some multi-word units. However, there is considerably less psycholinguistic research for cross-linguistic effects in idioms. In a lexical decision task run by Carrol and Conklin (2014), highly proficient Chinese speakers of English responded faster to targets completing the final word of transliterated (word-for-word translations) Chinese idioms (draw a snake and add—FEET) than to matched controls (put it in your—DISH) and English idioms (on the edge of your—SEAT) just as L1 English users responded quickly to targets completing English idioms compared to matched controls and the Chinese idioms. Since the transliterated Chinese idioms do not have English equivalents, Carrol and Conklin suggest that access to the Chinese idiom can occur, as in their study, via a lexical route, though, their proposed dual-route also allows for access via a conceptual level of idiomatic meaning when idioms are equivalent in both languages. Their conclusion is also supported by models of late bilingual comprehension suggesting language non-selective conceptions that connect to both the L1 and L2 (see e.g., Kroll & Stewart, 1994; Kroll et al., 2010). However, as Carrol and Conklin did not find effects for L2 idioms, it remains to be

seen as to whether familiar L2 idioms might also show facilitation effects comparable to those found for single words and collocations.

Research focusing on access to meaning in idioms with variable translatability from the L1 to the L2 also suggests facilitation in comprehension, production, and processing. Irujo (1986) examined the production and comprehension of English idioms of differing levels of translatability from Spanish to English in advanced Venezuelan learners of English in an offline study including a written task with multiple choice questions for recognition, an open-ended definition-writing task for comprehension, a discourse-completion task for recall, and a translation task for production. Identical idioms (one-to-one, word-for-word translations) were the easiest to comprehend and produce while different idioms (equivalent concepts not available via word-for-word translation) were the most difficult; negative interference in the form of transfer occurred in the production of partially-matching idioms (equivalent concepts and partially-matching translation). Irujo concluded that both production and comprehension can be aided in an L2 by using L1 knowledge. In a timed production task from Liontas (2002) in third-year learners of Spanish, French, and German, translatability was found to be a predicting factor for speed and accuracy of production with and without context. Additionally, Liontas also found that translation is one of the most common strategies used by L2 users in comprehending idioms based on learners' written reflections. Liontas (2002; 2015) uses his findings to propose a two-stage comprehension model: prediction, eased by idioms which are the same in a learner's L1 and L2, followed by confirmation or replacement and/or reconstruction. Thus, the figurative meaning of matching idioms should be easier than non-matching or partially-matching idioms, and possible even allow for faster availability of figurative meaning. In a study comparing bilingual idiom processing, Titone, Columbus, Whitford, Mercier, and Libben (2015) looked at translatability (in this study called cross-language overlap) in French-English bilinguals—some with English L1 and some with French L1—by asking participants to decide whether sentences containing idioms were meaningful or not in a word-by-word reading task. The idioms also had differing levels of translatability, and, in some conditions, the final word was presented in French rather than English. An increase in translatability facilitated a decrease in response times for French final word idioms, but not for English final word idioms. Additionally, accuracy increased overall as translatability increased. Their analyses, including also interactions for other idiomatic properties such as familiarity, predictability, and decomposability, Titone et al. (2015) took these results as support for a hybrid model of processing, like monolingual idiom processing, in which listeners use all available information to facilitate processing.

The present cross-modal priming studies aim to look more closely at access to figurative meaning in comparison to literal meaning as well as the influence of the L1 on L2 idiom comprehension. It does seem that L2 users, like L1 users, have access to figurative meaning, though the constraints seem to be more particular than in the L1, and the priority of literal language over

figurative language in the L2 is well-established. Translatability also appears to make a difference in comprehension and production of L2 idioms, and cross-linguistic effects on translations of idioms have been found, at least on a lexical level. While both of these aspects have been examined separately, they have not yet been considered together in an online study.

This cross-modal priming study includes two experiments designed to test the online availability of literal and figurative meaning in L1 and L2 listeners. In Experiment 1, native German listeners who were highly proficient in English were presented with English idioms as auditory primes. Listeners heard idioms placed at the end of short, non-biasing sentences before they had to decide whether a visually presented target word was an existing word of English or not. Target words were either related to the literal meaning of a constituent word of the idiom or to the figurative meaning, and both were compared to unrelated control targets. Half of the idioms had matching translations from English to German (called lexical level idioms) and half had non-matching translations (called post-lexical level idioms). Based on the L2 research presented above, we expected literally related targets to have faster RTs than their unrelated controls. If L2 listeners also have online access to figurative meaning, then we should find facilitatory priming, that is, faster reaction times to figuratively related targets compared to matched controls. However, based on Cieślicka's (2006) findings, it is also conceivable for the priming effect to be greater for literal targets than figurative ones. Additionally, if the offline findings of Liontas (2002) in addition to the online findings of Titone et al. (2015) hold true for this type of task, we expected the priming effect to be greater in lexical level idioms than in post-lexical level idioms. In Experiment 2, monolingual native listeners of American English were presented with the same English stimuli as in Experiment 1. Based on consistent evidence that L1 listeners have online access to figurative meaning in idioms, and evidence that literal constituents also undergo processing, we expected to find facilitatory priming for both literally and figuratively related targets. Additionally, we predicted that translatability should not make a difference for our monolingual English participants, since they did not know any German.

3.2 EXPERIMENT 1

3.2.1 Method

3.2.1.1 Participants

Sixty-five native speakers of German (48 female and 17 male; average age 24.5, range from 18 to 42) were paid a small fee to participate in the experiment. Participants learned English later in life in instructional settings and most were students of English at the University of Tübingen at the time of testing. All participants identified themselves as skilled speakers of English. Participants reported at least five years of English instruction in school and or university and averaged 4.9 on a 7-point scale

(1 corresponds to very poor, 7 to native-like) in their self-proficiency ratings. Seven of the participants were left-handed. None reported any hearing or visual impairments.

3.2.1.2 Materials

Sixty-four English idioms were selected from the English-German Database of Idiom Norms (DIN). The DIN database includes 300 English idioms and their English L1 and German L2 ratings and encoding on a variety of features. Selected idioms shared a VP syntactic structure, always beginning with a verb and ending with a noun (such as in *kick the bucket*) and were chosen based on ratings for additional attributes shown to affect processing (see e.g., Titone & Connine, 1994a; Titone & Connine, 1994b; Titone & Libben, 2014) such as familiarity—both frequency of encounter and familiarity with the meaning—decomposability, literality—which can be defined as an idiom's potential for a literal interpretation—and word-component frequency—based on averages of individual words taken from SUBTLEXus (Brysbaert & New, 2009). Target idioms had high ratings of familiarity for German L2 speakers (on a scale of 1-7, an average of 5.2 for familiarity of 5.7 for meaningfulness) in order to assure that German L2 participants in the present study would be familiar with the idioms presented. Additionally, idioms were selected with a medium literality rating (an average of 4.3) so as not to bias the experiment with highly literal idioms nor with idioms without possible literal interpretations. See Table 3-1 for a summary of descriptive idiom norms.

Table 3-1. Idiom norming distribution

	LL				**PL**			
	min	**max**	**mean**	**(sd)**	**min**	**max**	**mean**	**(sd)**
L2 Meaningfulness	5	6.9	6	(0.56)	4.25	6.3	5.47	(0.50)
L2 Familiarity	4.05	6.95	5.31	(0.80)	4.2	6.05	5	(0.52)
L2 Literality	2.6	5.48	4.25	(0.90)	2.6	5.55	4.33	(0.77)
Decomposable (% yes)	0.15	0.9	0.54	(0.22)	0.06	0.95	0.53	(0.25)
Constituent Frequency	1.44	3.84	3.42	(0.44)	3.04	3.92	3.54	(0.22)
Idiom Length (words)	2	7	3.78	(1.24)	3	6	3.16	(0.57)

Note: LL =lexical-level idioms, PL=post-lexical level idioms

Idioms varied systematically in translatability based on two of the levels of translatability laid out by Irujo (1986), Liontas (2002), and Titone et al. (2015). Irujo and Liontas used a three-level scale, and Titone et al. rightfully added to the scalarity of translatability using five levels, also including a non-language overlap condition in which there is neither an overlap in word-for-word translation nor a corresponding idiom with the same meaning in addition to adding multiple levels with partial overlap. The 32 English idioms in this study were directly translatable in that matching idioms and figurative meaning results from a word-for-word translation of the idiom from English to German (lexical level

idioms) and 32 idioms were not directly translatable in that a word-for-word translation from English to German does not produce a matching idiom with equivalent meaning (post-lexical level). The post-lexical level idioms selected do, however, have idioms with equivalent meanings in German, and all idioms should be familiar based on the ratings discussed above (see Table 3-2 for examples). Both levels of translatability selected could be matched closely on the measures listed above, although slight differences were present in the lists. Between lists, idioms did not significantly differ in frequency of encounter, literality, decomposability, or word-component frequency, but there was a small but significant difference in familiarity with the meaning (means 6.0, 5.47; student's *t*-test, t=3.991 df =62, p<.001) and a significant difference in idiom word length (means 3.781, 3.156; Welch's *t*-test, df =43.756, p<.05), with translatable idioms being more familiar and longer (see Table 3-1). Idioms which only occur in the L2 and not in the L1, were not included in spite of providing a starker contrast in translatability as they could not be well-matched based on the selection criterion laid out above. However, Irujo and Liontas' results still predict that these levels will show differences in processing. All idioms were placed at the end of sentences with non-biasing contexts with as little additional information as possible, as in the sentence *John <u>let the cat out of the bag</u>*. Each sentence with an idiom prime was followed by one of four target words distributed across four experimental lists.

Table 3-2. Levels of translatability

English Idiom	**Type**	**Equivalent German Idiom**	**Figurative meaning**
to lend (someone) an ear	LL	*Jemandem sein Ohr leihen* someone his ear lend	to listen or pay attention to someone
to kick the bucket	PL	*den Löffel abgeben* the spoon give away	to die

The four target words for each idiom were literally or figuratively related words and their respective unrelated control words. Target words semantically related to the literal meaning were based on the last content word of the idiom and chosen from the Nelson, McEvoy, and Schreiber (1998) association norms database. For example, for the idiom *to pull my leg (*primed in the sentence *John likes to <u>pull my leg</u>.),* the target word *WALK* was chosen. The unrelated control word was matched for orthographic complexity and length (unrelated *MILK* compared to literally related *WALK*). Targets related to the figurative meaning of the idiom were chosen based on relation to the overall meaning of the idiom. For the same idiom, the target *JOKE* was chosen, as *to pull my leg* has the meaning "to make fun of someone in a friendly way." Similarly, figurative control targets were also controlled for orthographic complexity and length (unrelated SHIP compared to figuratively related JOKE). Target words were also controlled for average lexical frequency. As there are no existing databases for association of figurative meaning, targets were selected by these authors (see Appendix B for a list of idioms and targets used).

In addition, an online ratings study using OnExp (Onea, 2011) was conducted in which 26 native speakers of American English rated the level to which the chosen figurative target reflects the meaning of the idiom in a fifteen-minute survey. The sixty-four target idioms were paired with their figuratively related targets from Experiment 1, and twenty additional filler idioms were added with targets as to elicit the full range of possible responses from very related, somewhat related, to very unrelated. Participants rated the relatedness of a target to the figurative meaning of an idiom using a 5-point scale ranging from *not at all related* (1) to *highly related* (5). Participants were not given the meaning of the idiom and were therefore also provided with an option to select *I don't know the meaning of this phrase* rather than giving a rating. Overall, figuratively related targets received high association ratings: post-lexical idioms received an average rating of 3.87 and lexical level idioms 4.01. In a two-sampled t-test, there was no significant difference in association ratings ($t(62) = .73$, $p > .46$).

In addition to the 64 sentences with target idioms, an additional set of 104 filler sentences were selected. Twenty filler trials contained additional idioms, and 84 trials consisted of non-figurative sentences, meaning one-half of the trials in the experiment contained idioms. The twenty sentences with filler idioms were followed by non-words in order to ensure that a sentence containing an idiom was not always followed by a word. The remaining 84 filler trials consisted of varying sentence structures so as to provide a variety of sentence types. Filler sentences were followed visually by word or non-word targets; half of the word-targets were related, and half were unrelated. In total, word and non-word targets were equally present in the trials, and all English non-words were also non-words in German. Each of the total of 168 trials (64 experimental trials and 104 filler trials) consisted of an auditory sentence prime followed by a visual target. Thirty-two of the experimental trials contained a post-lexical level idiom and 32 contained a lexical level English idiom. The experiment was presented to participants in one of four randomized lists. Filler items remained the same while lists were counterbalanced for experimental target type.

The experiment was performed using Presentation®, software (Version 17.2, www.neurobs.com). Experimental sentences were recorded by a female speaker of American English (first author) in an experimental lab setting. All auditory sentences were the same for each participant, however, four counterbalanced lists were randomly allocated to participants so that conditions were evenly distributed across lists. The four lists allowed targets to be present once for each condition. Each list began with four practice trials followed in a randomized order by 64 experimental and 104 filler trials, and participants had the option of a short self-directed concentration break after trial 84. Each list was presented to an equal number of participants.

3.2.1.3 Procedure

Participants were tested individually in a quiet room. First, participants were given instructions in English on the lexical decision using Google Forms. Participants saw written instructions and had to answer a short series of questions to ensure uniform instructions and understanding. This was followed by a short verbal confirmation of the instructions. Participants were instructed that they would hear sentences directly followed by the appearance of an English word or non-word on the computer screen in front of them. They were instructed to listen to sentences and then decide whether the string of letters on the screen was a word of English or not. Participants were asked to make their decisions by pressing a green button with their dominant hand for 'YES' and a red button with the non-dominant hand for 'NO' as quickly and accurately as possible. They were also told that it was important both to listen and to respond to the visual targets, as they would be asked about what they heard after the lexical decision task. This instruction was included to ensure their continued attention throughout the experiment.

Once participants understood the instructions and answered the confirmation questions correctly, they participated in the priming study. The participants listened over closed headphones and the visual targets were presented on a laptop screen as capital white letters on a black background in size 20 font. The targets appeared on the screen right after the offset of the auditorily presented prime sentences. Reaction times were measured from the onset of the presentation of the visual target words. Participants had 2500 ms to respond, and 1000 ms after a key response, the following trial began.

The experiment concluded with a short comprehension test and a language background questionnaire. As participants were told that they would be asked about the auditory sentences following the experimental trials, participants took a short yes/no comprehension test following the experiment. This was included to ensure that participants listened to the sentences (see Cieślicka, 2006). The test consisted of a list of 60 sentences, of which 30 were heard in the experiment. Participants had to check 'YES' or 'NO' to the question of whether they heard the sentences in the experiments (participants averaged 73% correct). The entire experiment lasted about 25 minutes.

3.2.1.4 Results

Only trials with correct responses to target words were analyzed. L2 participants answered an average of 93% of targets correctly. Such a high level of performance attests the high level of L2 proficiency of our participants. Thirty-six responses with RTs longer than 1800 ms or shorter than 250 ms were considered outliers and were removed from these analyses (0.9% of the total data). Additionally, seven targets (CELERY, CHIME, EEL, NITPICK, SEAM, SHRUB, TALKATIVE) were answered correctly in 50% or less of the total trials and were excluded from the results (44 responses or 1.1% of the total data).

We used R (R Core Team, 2015) and lme4 (Bates, Maechler, Bolker, & Walker S., 2015) to perform a linear mixed effects analysis of the relationships between *figurativeness*, *relatedness*, and *lexicality* on reaction times. Table 3-3 reports the mean RTs and inverse RTs (see Baayen, 2008) measured from target onset for each condition. The mean RTs across *figurativeness* and *relatedness* are shown in Figure 3-1 with error bars representing the standard error. After excluding residual data points in the inverse RTs, a total of 3.3% of the data were not included in the final analysis.

Table 3-3. L2 Mean reaction times (in ms)

	Lexical				**Post-Lexical**			
	Figurative		**Literal**		**Figurative**		**Literal**	
	RT	**Inverse RT**	**RT**	**Inverse RT**	**RT**	**Inverse RT**	**RT**	**Inverse RT**
Related	686	-1.567	644	-1.669	667	-1.614	628	-1.700
Unrelated	682	-1.586	666	-1.620	695	-1.556	661	-1.633

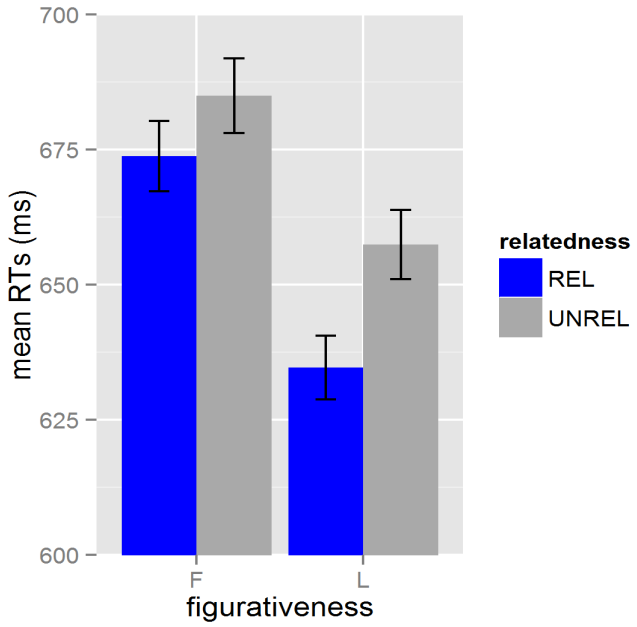


Figure 3-1. Mean RTs (in ms) for German L2 listeners in Experiment 1

Note: *Figurativeness* is represented by F (figurative) and L (literal) and *relatedness* is represented by REL (related) and UNREL (unrelated).

A set of LMER models were built with inverse RTs as the dependent measure and fixed factors, centered around 0, were coded and included as follows: *figurativeness* (literal: -0.5 or figurative: 0.5), *relatedness* (related: 0.5 or unrelated: -0.5), and *lexicality* (lexical level: -0.5 or post-lexical level: 0.5).

Items and participants were included as random factors, and the maximal structure suggested by Barr, Levy, Scheepers, & Tily, 2013, was not included in the final model as model comparisons of individual slope adjustments—due to convergence issues with a fully maximal structure—suggested that this approach is unwarranted for this data set (see e.g., Bates, Kliegl et al., 2015). *P*-values were calculated using likelihood ratio tests of the full model with the effect in question against a model excluding the effect in question in order to achieve the most parsimonious model using a backward, stepwise selection procedure. The values of the first full model can be found in Table 3-4.

After stepwise selection, the final model included *figurativeness* and *relatedness* as fixed factors (*lexicality* and all interactions were excluded because they were not significant: all *t*-values < 1.5, *p*-values > .10) and items and participants as random effects. The effect of *relatedness* (b_{coded}= 0.044, p < .05) shows that related targets were significantly faster than unrelated targets ($\chi^2(1) = 4.02$, $p < 0.05$). The effect of *figurativeness* (b_{coded}= 0.0717, p < .001) demonstrated that figurative targets were significantly slower than literal targets ($\chi^2(1) = 11.03$, $p < 0.001$).

Table 3-4. L2 Full model output

Fixed Effects & Controls	**Effect Size**	**SE**	**t-Value**	**Pr(>|t|)**
Intercept	-1.605	0.031	-51.03	<2e-16***
Figurativeness (coded)	0.072	0.022	-3.428	<0.001***
Relatedness (coded)	-0.044	0.022	2.013	0.045*
Lexicality (coded)	-0.015	0.022	-2.280	0.500
Figurativeness*Relatedness (coded)	0.048	0.042	1.139	0.256
Figurativeness*Lexicality (coded)	0.019	0.042	0.457	0.648
Relatedness*Lexicality (coded)	-0.048	0.043	-1.106	0.270
Figurativeness*Relatedness*Lexicality (coded)	-0.06983	0.08494	-0.822	0.411824
Random Effects	**Variance**	**SD**		
Target	0.02401	0.1549		
Subject	0.05667	0.238		
Residual	0.07528	0.2744		

Note: *.p<.10 *p<.05 **p<.01 ***p<.001*

3.2.2 Discussion

This analysis shows that facilitatory priming was observed for both figuratively and literally related targets for L2 listeners. Even in the absence of biasing context, listeners responded more quickly to related than unrelated targets. These results are consistent with those found by Cieślicka (2006). However, the effect of *figurativeness* applies across figuratively related and -unrelated targets,

suggesting that the figurative targets were more difficult overall for L2 listeners, and a comparison of solely the figuratively and literally related targets should take this difference into consideration. These results suggest that non-native listeners have online access to the figurative meaning of familiar L2 idioms in addition to access to the meaning of individual component words.

We found, however, no effects or interactions involving *lexicalilty*, suggesting that the translatability of the idioms from the L1 to the L2 did not have a direct impact on processing as the model and research presented by Liontas (2002, 2015) would have predicted. Thus, while translatability may well affect comprehension strategies, there is no evidence from this analysis that suggests an extension into online processing. However, looking descriptively at the priming effects, one can see that priming effects for LL idioms are quite small in comparison to the PL idioms. The observed differences in the idiom norms do not account for this perceived difference. The difference in familiarity should predict an ease in processing for lexical-level idioms, and therefore shorter reaction times (see e.g., Titone, Libben 2014), the opposite result from our observation. The difference in idiom length might predict more processing effort for post-lexical level idioms, supporting the results seen descriptively. However, upon closer inspection of the data, post-lexical level idioms include only six idioms that have key words occurring before the final word in the idiom, while lexical-level idioms include ten (see DIN Database). Again, this difference in predictability, if significant for processing, should give a processing advantage to the post-lexical level idioms rather than the lexical-level idioms. Thus, though insignificant, this perceived difference in priming cannot be explained by the typical idiomatic factors associated with priming differences.

As these results are consistent with the results of Cieślicka (2006) and inconsistent with predictions from Irujo (1986), Liontas (2002), and Titone et al. (2015), it is important to make comparisons with L1 listeners. Cieślicka's study used similar results to support evidence that L2 listeners rely more heavily on literal meaning than L1 listeners; however, her study lacked an L1 comparison. In order to make conclusions about whether these results apply only to L2 listeners, we conducted a second experiment using the same materials and methods on L1 listeners.

3.3 EXPERIMENT 2

3.3.1 Method

The method was the same as in Experiment 1.

3.3.1.1 Participants

Forty native speakers of American English were paid a small fee to participate in the experiment. All participants grew up in English-speaking households and did not speak or know any German. Participants were students at the University of Maryland in College Park, MD, USA. L1 participants

ranged from 19 to 31 years old with an average of 22.5 years old. Five of these participants were left-handed. There were 30 female and 10 male L1 participants. Participants reported no hearing or visual impairments.

3.3.1.2 Materials

The materials were the same as in Experiment 1.

3.3.1.3 Procedure

The procedure was the same as in Experiment 1, except that the experiment was conducted in the laboratory at the University of Maryland Language Science Center.

3.3.1.4 Results

The analyses include again only correct responses to targets. L1 participants correctly answered an average of 97% of targets. Eleven responses with RTs longer than 1800 ms or shorter than 250 ms were considered outliers and were removed from these analyses (0.004% of the total data). The data from one participant was also excluded because she did not follow the instructions. The comparable averages for correct responses in Experiments 1 (93%) and 2 (97%) further confirm the high level of proficiency of our L2 participants in Experiment 1.

As in Experiment 1, we used R (R Core Team, 2015) and lme4 (Bates et al., 2015) to perform a linear mixed effects analysis of the relationships between *figurativeness*, *relatedness*, and *lexicality* on reaction times. Table 3-5 reports the mean RTs and inverse RTs measured from target onset for each condition. Figure 3-2 shows mean RTs for *figurativeness* and *relatedness*, with error bars representing standard error. A total of 3.7% of the data were not included in the final analysis after exclusion of any remaining residuals.

Table 3-5. L1 Mean reaction times (in ms)

	Lexical				**Post-Lexical**			
	Figurative		**Literal**		**Figurative**		**Literal**	
	RT	**inverse RT**	**RT**	**inverse RT**	**RT**	**inverse RT**	**RT**	**inverse RT**
Related	601	-1.791	587	-1.831	595	-1.830	584	-1.856
Unrelated	603	-1.788	631	-1.744	622	-1.732	606	-1.791

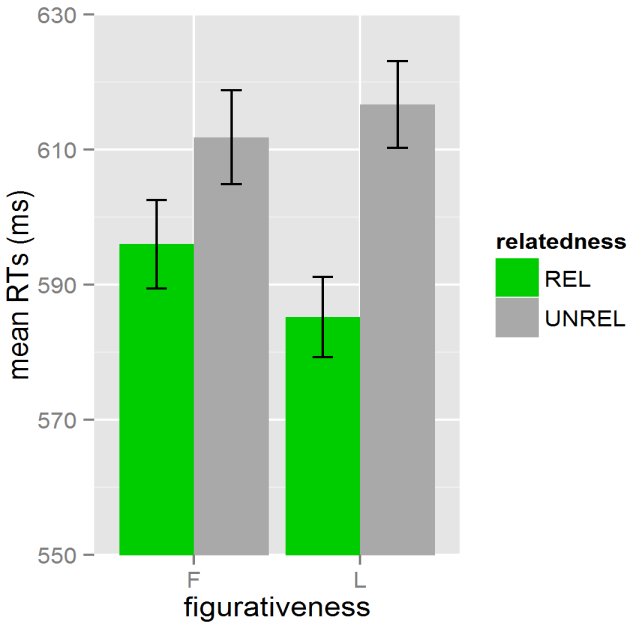


Figure 3-2. Mean RTs (in ms) for English L1 listeners in Experiment 2

Note: *Figurativeness* is represented by F (figurative) and L (literal) and *relatedness* is represented by REL (related) and UNREL (unrelated).

Using inverse RTs as the dependent measure and the fixed factors, coded and centered around 0, of *figurativeness* (literal: -0.5 or figurative: 0.5), *relatedness* (related: 0.5 or unrelated: -0.5), and *lexicality* (lexical level: -0.5 or post-lexical level: 0.5) as well as items and participants as random factors (see e.g., Barr et al., 2013), a set of LMER models were constructed. *P*-values were calculated using likelihood ratio tests of the full model with the effect in question against a model excluding the effect in question in order to achieve the most parsimonious model using a backward, stepwise selection procedure. Table 3-6 displays the full output from our original model.

Only *relatedness* was included in the final model as a fixed factor (all interactions, *figurativeness*, and *lexicality* were excluded because they were not significant: all *t*-values < 2, *p*-values > .05) in addition to items and participants as random factors. The three-way interaction approaches, but does not reach significance. The effect of *relatedness* (b_{coded}= -0.0607, p < .01) shows that responses to related targets were significantly faster than responses to unrelated targets ($\chi^2(1)$= 9.1494, p= 0.002488).

Table 3-6. L1 Full model output

Fixed Effects & Controls	Effect Size	SE	t-Value	Pr(>|t|)
Intercept	-1.793	0.052	-34.361	<2e-16***
Figurativeness (coded)	0.023	0.019	1.183	0.238
Relatedness (coded)	-0.060	0.020	-3.09	0.002**
Lexicality (coded)	-0.014	0.020	-0.721	0.472
Figurativeness*Relatedness (coded)	0.034	0.039	0.887	0.376
Figurativeness*Lexicality (coded)	0.046	0.039	1.189	0.236
Relatedness*Lexicality (coded)	-0.040	0.039	-1.013	0.312
Figurativeness*Relatedness*Lexicality (coded)	-0.130	0.078	-1.669	0.096
Random Effects	**Variance**	**SD**		
Target	0.014	0.119		
Subject	0.102	0.32		
Residual	0.096	0.3103		

Note: *.p<.10 *p<.05 **p<.01 ***p<.001*

3.3.2 Discussion

As in Experiment 1, facilitatory priming was found for both figuratively and literally related targets in L1 listeners. Even in the absence of biasing context, listeners responded more quickly to related than unrelated targets, as expected based on previous findings (Cacciari & Tabossi, 1988; Ortony et al., 1978; Swinney & Cutler, 1979). While the priming effect, however, is numerically larger for literally related targets than for figuratively related targets, this difference is not significant. There are a multitude of explanations for this potential difference in priming effects. Cacciari and Tabossi's (1988) study suggest that figurative priming only increases to the level of literal priming for idioms with high levels of predictability. Our study did not account for predictability, and instead relied on controlling factors such as familiarity and syntactic structure, as familiarity and predictability correlate (Titone & Connine, 1994a) and syntactic structure can also be an indication of predictability (i.e., idioms beginning with common verbs such as "take the…" and "get the…" will be similarly unpredictable before the final word). Colombo (1993) had similar findings in a cross-modal priming study and suggested that the difference in nature of the semantic relationship of figuratively and literally related targets to the idioms might explain the differences. While our literal targets are single word associations, our figurative targets are semantic relations of the idiomatic phrase to a related word.

Colombo and Williams (1990) suggest that priming effects are less stable in semantic relations in comparison to associative relations (as cited in Colombo, 1993).

The results differed from those in Experiment 1 as there was no effect for *figurativeness* in the L1. Though the visual data in Figure 3-2 displays figurative targets as generally slower than literal targets also for L1 participants, this effect did not reach significance in Experiment 2. This suggests that the effect found in L2 listeners is likely a consequence of the difficulty of the figurative words. While the target lists were controlled for frequency based on the COBUILD frequencies per million and syllable structure provided in WebCelex (2001), this process cannot exclusively control for L2 difficulty. Furthermore, we did not place restrictions on word class or concreteness of the term, as it was most important that our subjects identify the figuratively related target with the meaning of the idiom, and idiomatic meaning is often expressed most closely with abstract words rather than the more concrete words from the literal word associations (e.g., BEHAVE vs. FOOTBALL). It is possible that the differing word classes or levels of concreteness of the word might also have impacted difficulty; while L1 listeners were less challenged by this variation, L2 listeners appear to have reacted differently. As Brysbaert and Duyck (2010) pointed out, L2 translation studies consistently find that translation of abstract words differ from concrete words (see also Peterson, Burgess, Dell, & Eberhard, 2001). Thus, the slower responses to figuratively related targets compared to literally related targets might be an expression of an effect of concreteness.

Another difference that can be observed between the experiments is that the L1 listeners in Experiment 2 had overall faster reaction times than the L2 listeners in Experiment 1. This reflects common findings that suggest that L2 comprehension is more difficult and typically slower (e.g., see review in Cutler, 2012) and does not necessarily reflect a low level of proficiency of the L2 participants. Rather, the similarity in L1 and L2 behavior displayed in the priming effects for relatedness confirms a high level of proficiency of the L2 participants. Both L1 and L2 listeners displayed access to figurative meaning in addition to literal meaning, though priming effects are somewhat smaller for L2 participants.

Although the lack of significance in the effect of *lexicality* implies that L1 listeners were also not impacted by the translatability of the idioms from English to German, an observation of the data from both groups still suggests that the priming effect is stronger for post-lexical idioms than lexical level idioms. As our descriptive data of idiom norms did not provide an explanation for this observation an additional data analysis is included. For this reason, in addition to the observation that our three-way interaction in the L1 model approaches significance, we expanded our models in another analysis of the data to include the available idiomatic norms as fixed effects. In addition to the fixed factors, coded and centered around 0, of *figurativeness* (literal: -0.5 or figurative: 0.5), *relatedness* (related: 0.5 or unrelated: -0.5), and *lexicality* (lexical level: -0.5 or post-lexical level: 0.5) on inverse reaction times,

we added individual effects for *L2 meaningfulness*, *L2 familiarity*, *decomposability* (based on percent positively rated decomposable), *L2 literality*, *idiom constituent frequency* and *idiom length*. In addition to the three-way interactions between *figurativeness*, *relatedness*, and *lexicality* and the individual two-way interactions, interactions between lexicality and each of the idiomatic norm measures were also included in L1 and L2 models. Items and participants as random factors were also included. Our results remained stable, and we found no significant interactions with *lexicality* and the individual idiomatic norm measures in neither the L1 nor the L2 models. However, in the L2 models, decomposability (b= 0.081, p = .090) approaches significance. Additionally, L2 meaningfulness as a fixed effect approaches significance (b= 0.064, p = .079) in the L1 model. This suggests that for L2 listeners, decomposability might have an effect on processing, and for L1, meaningfulness could play a role. However, their lack of significance and the stability of our initial results further supports our controls across these measures. Interestingly, based on the differences in averages between the two groups (significant differences in familiarity and word-length), we would have predicted a further boost for processing in lexical-level idioms (see e.g., Titone & Libben 2014). However, as all other effects remain consistent across both experiments and listener groups, we can conclude that there is not a fundamental difference connected with translatability that we can ascertain between the two sets of idioms that overtly affected processing.

In a final additional post-hoc analysis, we combined both data sets into one analysis to confirm whether our findings remain consistent, and to confirm any differences between the groups. Our full model again used inverse RTs as the dependent measure and the fixed factors, coded and centered around 0, of *figurativeness* (literal: -0.5 or figurative: 0.5), *relatedness* (related: 0.5 or unrelated: -0.5), *lexicality* (lexical level: -0.5 or post-lexical level: 0.5), and *language* (L1: 0.5 or L2: -0.5) as well as items and participants, which included random slopes (including language across subjects), as random factors (see e.g., Barr et al., 2013). *Relatedness* remains significant in the full analysis (b= -.056, p < .01), confirming our initial results. An additional effect for *language* was present (b= -.197, p < .01), validating our observation that L2 listeners performed significantly slower than L1 listeners. Although *figurativeness* also appears as a significant factor in this full analysis (b= .046, p < .05), a highly significant interaction between language and figurativeness (b= -.050, p < .001) motivates a split of the data naturally presented in this paper in each of the two experiments. Thus, our final analysis can also confirm an overall difference in performance between our two listener groups in the experiments individually presented.

3.4 GENERAL DISCUSSION

The primary purpose of this study was to extend the body of evidence concerning L2 idiom processing and access to figurative meaning in comparison to literal constituent meaning and to test whether this

access is impacted by the L1 via translatability of idioms. While the results of Experiment 2 can be used as a control for comparison to the L2 results, the findings in Experiment 1 are also a reflection of current findings in psycholinguistic literature concerning L1 idiom processing. The availability of both literal constituent meaning in addition to the figurative meaning confirms findings by Cacciari and Tabossi (1988), Gibbs et al. (1989), Sprenger, Levelt, and Kempen (2006), and Titone and Connine (1994b), for example. When examining these results against the performance of our L2 participants in Experiment 1, there does not seem to be evidence for an L2 mode of idiom processing that significantly differs from the way idioms are processed in an L1.

In addressing access to figurative meaning in comparison with literal meaning in particular, the results of these experiments can be interpreted in light of current models of L1 and L2 idiom processing. The findings of this study generally do not support early stage models for L1 idiom processing, as they present idiom processing as semantically empty (see e.g., Cacciari, 2014). The speed and availability of the literal constituent meaning in addition to activation of the figurative meaning suggests that neither was literal composition aborted, as suggested by the early models presented by Swinney and Cutler (1979) and Gibbs (1980), nor is it likely that this meaning only became available after obligatory composition and rejection of the literal meaning, as proposed by the standard pragmatic view supported by Bobrow and Bell (1973).

Taking the effects of individual idiom properties into consideration, our data are largely compatible with several L1 and L2 processing theories, while casting doubt on others. Although theories on decomposability are somewhat compatible, the major theories considering this property still present problems. The Idiom Decomposition Hypothesis proposed by Gibbs and Nayak (1989) suggests that literal composition aids idiomatic processing; thus, decomposable idioms are comprehended more quickly than non-decomposable idioms. While we cannot make any claims about advantages of one or the other (and this tenant of the hypothesis has been criticized by conflicting evidence—see e.g. Cutting & Bock, 1997; Libben & Titone, 2008) since the idioms in the present study were of an medium level of decomposability, it is clear that some level of composition is taking place. Thus, while this theory fits our data more accurately than theories that assume an either-or approach to composition and retrieval in the processing of idioms, we cannot present evidence for all aspects of this approach. Additionally, as Cieślicka (2006) points out, the role the literal meaning of the individual words play in constructing figurative meaning is somewhat vague. A similar L2 model proposed by Abel (2003) in her Model of Dual Idiom Representation is generally compatible with our findings; though, this model differs as it explains representation rather than processing. This model assumes that both non-decomposable and frequently encountered idioms are represented in the mental lexicon by idiom entries on the conceptual level, as seen in the Idiom List Hypothesis, while decomposable idioms are represented via lexical entries of the individual constituent words. Abel argues that encountering

an idiom often enough allows storage to occur, though, decomposability will also facilitate a faster retrieval if no entry has yet been created. Our study included idioms that were highly familiar, not only to L1 speakers, but also to L2 speakers in the target language group, suggesting that many of the idioms should be available merely via direct retrieval—a notion already deemed problematically vague as the availability of literal meaning is also consistent even in these highly familiar idioms. What's not clear is whether literal composition continues in the case of retrieval, in which case, we would have expected to find strong figurative activation to occur in opposition to strong literal activation, a notion our data do not represent. However, if retrieval accompanies further composition, our data can generally support this model of representation.

The Configuration Hypothesis (Cacciari & Tabossi, 1988), suggesting that literal word processing occurs until an idiomatic key is reached, is somewhat compatible with our findings. The assumption that both literal constituent meaning as well as figurative meaning can be activated is confirmed by our findings. Cacciari's (2014) new look at this hypothesis also addressed the problem of literal constituent availability in retrieval-based models of idiom processing. Cacciari points to evidence presented by Peterson et al. (2001) that suggest that individual word meanings, though activated, might only be done so for purposes of syntactic control and not integrated into the meaning of the sentence. This view is also supported by the ERP study presented by Rommers, Dijkstra, and Bastiaansen (2013) in which predictable words in idioms were replaced by semantically similar words, and while this produced N400 effects for predictable literal phrases, the effect was not present for idioms. Thus, assuming activation of literal meaning is possible in spite of retrieval in predictable idioms, the process posed by the Configuration Hypothesis is supported by our data. However, the difference confirmed in predictability (translatable idioms had more predictable idioms than non-translatable idioms) could not be confirmed by the data, though, the advantage for predictable idioms in comparison to non-predictable idioms was not addressed in this study.

Two alternative propositions for this simultaneous activation, also supported by our data, are presented by hybrid models such as the Superlemma Hypothesis (Sprenger et al., 2006) and the Constraint-Based Model of Idiom Processing (Libben & Titone, 2008; Titone & Connine, 1999; Titone & Libben, 2015). In Sprenger et al.'s production-based hypothesis, continued processing of constituent parts is not discounted after an idiom is identified. Based on psycholinguistic evidence from production tasks that constituent and phrasal processing occur simultaneously, this hypothesis suggests that activation occurs for individual word lemmas and spreads to super lemmas at the lexical-syntactic level representing the meaning of the idiom. Thus, this hybrid model sees processing as a web of meaning taking place on several levels, allowing literal constituent meaning to spread and reach figurative phrasal meaning once enough information is present, very much like the Configuration Hypothesis. However, as Cacciari (2014) points out, this hypothesis assumes that idiomatic meaning is built up

beginning with the first word of an idiom rather than retrieved upon cue. As participants reacted at the offset of the idiom, our data does not support one view more than the other. The Constraint-Based Model answers some of these uncertainties in its account that meaning is built over time using all available information; thus, while some meaning may be available upon encountering the first word, it is likely that full activation will only occur in the presence of more available information (e.g., context). Rather, comprehension will use all available information, as it is made available, causing different information to be available at different times. Our data supports this view as literal constituent meaning seems to be available faster than figurative phrasal meaning; however, both are available online. Additionally, this model is also supported by our analyses including idiomatic norms. Though none of the idiomatic norms reached significance as fixed effects, we conclude that this is due to the similarity of our idiomatic items and predict that these norms might reach significance, as trends were already present, in the presence of more idiomatic diversity. Thus, an account of processing that also takes these norms into consideration is highly compatible with our data.

While our data replicates the findings of Cieślicka's (2006) study, we do not propose that our findings entirely support the Literal Salience Model. Like Giora's (1997) Graded Salience Hypothesis, this model suggests that saliency always has processing priority, regardless of the figurativeness or literality of a given phrase. In this case, L2 listeners should always respond faster to literal meaning, as it is generally more salient for L2 listeners. While our L2 results also produced this general effect, it is repeated in our L1 findings—something we would not expect if we assume that L1 listeners should often find the figurative meaning to be more salient (Giora, 1997; 2002). While the availability of meaning does confirm that literal constituent meaning plays a role in processing idioms, we are not prepared to assume a necessary priority. Though the eye-tracking evidence from Siyanova-Chanturia et al. (2011) also supports this notion generally, it is rather frequency—an important aspect of saliency—that is important, an idea that does not preclude literal priority, particularly for advances L2 users. It should also be noted that Cieślicka's (2006) study did not include an L1 group for comparison, and thus makes the assumption that L1 processing is fundamentally different than L2 research based on different findings in studies conducted on solely L1 groups (Cacciari & Tabossi, 1988, for example). We would argue that the similarity of behaviors in Experiments 1 and 2 suggests that L2 processing generally mirrors L1 processing.

The model addressing translatability, specifically, Liontas' (2002; 2015) Idiom Diffusion Model of Second Languages is also not supported by our data. We expected to find a significant difference in the priming effect of post-lexical level idioms compared to lexical level idioms in Experiment 1 and not in Experiment 2. We found no evidence that processing was affected by translatability. Additionally, the similarity of L1 and L2 behavior further poses problems for this theory. One possible reason our data might not reflect this model is that our listeners were highly

proficient, and Liontas (2002) focused on third year learners. Just as Abel (2003) suggests that L2 idiom entries occur on a conceptual level after L2 users encounter them over time, Duyck and Brysbaert (2004) suggest that proficient L2 users directly map L2 words to conceptual meaning when there is a direct overlapping of L1 and L2 words. Therefore, the available evidence that single words are translated as we hear them (see e.g., Blumenfeld & Marian, 2007) may not affect the activation of figurative meaning in the L1 if the L2 listeners directly access meaning. We might then still expect to find an effect of translation in less-proficient L2 listeners as they may not have direct mapping from the L2 to the conceptual figurative meaning. However, it is difficult to test this as an examination of idiomatic processing requires that listeners are familiar with the figurative meaning in order to test its availability—something we would not expect of less-proficient listeners. Additionally, we did not test the full scale of translatability presented by Liontas or Titone et al. (2015), as we omitted a partial overlap present in both studies and all of our idioms had matching counterparts in both languages, unlike in the study from Titone et al. While a control group with no overlap might have increased the distance in translatability, these idioms tended to have a low L2 familiarity, an important selection criterion for our experimental idioms. We omitted intermediate levels, however, to ensure that enough distance in translatability was present to still impact processing. Therefore, while offline evidence presented by Liontas as well as Irujo (1986) makes the case for facilitation in comprehension and production and Titone et al. (2015) find an online effect in code-switched idioms, we were not able to replicate these results in online processing for proficient L2 listeners. Experimental materials and task-type may have played a role in these findings (see e.g., Libben & Titone, 2008).

3.5 CONCLUSION

In summary, the results of the present study show that both L1 and L2 listeners show access to figurative meaning as well as literal constituent meaning in the absence of a clearly biasing context. Additionally, we did not find that the translatability of idioms from listeners' L1 to their L2 had a measurable impact on processing. We take these results as evidence that highly proficient L2 listeners process figurative meaning in a way that is not entirely unique from L1 listeners—namely, it is impacted by the same factors that L1 idiom processing is. While our intent was not to generalize about models of processing or representation, we can interpret our results as supporting models that maintain direct mapping of L2 words to a conceptual language non-selective level (e.g., Abel, 2003; Carrol & Conklin, 2014; Brysbaert & Duyck, 2010; Kroll et al., 2010) and include the possibility for both retrieval and composition (e.g., Libben & Titone, 2008; Titone & Connine, 1999). Thus, for highly proficient L2 users, as the results of Titone et al. (2015) suggest, idiom processing follows the same routes as L1 idiom processing. This conclusion is also in line with Matlock and Heredia (2002), who suggest that beginning learners first access the literal meaning and translate it into their L1, then try to get the literal meaning before they can access the figurative language. Over time, learners are able to

bypass the first two steps. L2 listeners may be slower or have more difficulty processing some figurative language, however, as they become more proficient, we conclude that any differences in processing are likely due to general L1 and L2 differences rather than a distinct manner of processing.

4 THE FORMULAIC NATURE OF IDIOMS: PREDICTABILITY AND PREDICTION IN IDIOMATIC PROCESSING

This chapter includes two separate studies stemming from the formulaic nature of idioms: first, the predictability of an idiom and second, the ability of language learners to predict constituent words during idiom processing. Both aspects stem from the fixed nature of idioms, but they focus on different aspects of meaning activation. The first study addresses the idiomatic property predictability. Described in detail in **Chapter 2**, predictability describes whether an idiom can be successfully completed before the final constituent word is presented. In this study, how this affects L2 access to figurative meaning is investigated, and this is done so in two experiments. In the second study, language learners' abilities to successfully complete an idiom early during listening is investigated along with the timeline for literal constituent activation. Here, both L1 and L2 learners are investigated in the German language. While these two studies arguably look at different processes during comprehension, both are the consequence of the fixed nature of idiomatic phrases (see e.g., Wray, 2002) and consider the interplay between the whole of the idiomatic chunk and how this interacts with idiomatic meaning as well as the meaning of the parts.

4A UNLOCKING THE KEY TO L2 IDIOMATIC PROCESSING: NON-NATIVE LISTENERS' IDIOMATIC PROCESSING IS NOT IMMEDIATELY AFFECTED BY THE IDIOMATIC KEY

Abstract

This study aimed to investigate the role of predictability in idiomatic processing. Predictability of idioms has widely been associated with the idiomatic key or idiomatic recognition point of an idiom (e.g., Cacciari & Tabossi, 1988). While previous research has shown that L1 listeners have access to the figurative meaning of an idiom once the idiomatic key has been encountered (e.g., Tabossi & Zardon, 1993), this study addresses a research gap in this area for L2 listeners. In two cross-modal priming experiments, L2 listeners were auditorily presented idioms in neutral sentences and responded to a lexical decision task on words related to the figurative meaning of the idioms or to a literal constituent of the idiom immediately after the offset of the idiomatic key. In the first experiment, no priming was found for either the literal or figurative condition. In the second experiment, the task was simplified in order to rule out difficulties caused by task demands, and participants saw targets 400ms after the offset of the key, and the sentence stopped playing after target presentation. Again, no significant priming could be established for either condition. The results of this experiment, compared with earlier findings showing priming at sentence offset (Beck & Weber, 2016a, 2016b) suggest that L2 listeners access to figurative meaning is not immediate and more precise experimental methods and measures may be needed to identify the presence of predictability effects.

4.1 INTRODUCTION

One of the defining factors of idioms, setting them apart from other types of figurative language, is their formulaic nature (see e.g., Wray, 2002). While idioms do allow for a certain amount of flexibility in syntactic structure (e.g., Kyriacou, Conklin, & Thompson, 2019; Nunberg, Sag, & Wasow, 1994), the constituents are generally fixed, and there is a preferred canonical form (e.g., McGlone, Glucksberg, & Cacciari, 1994). The consequence of this fixed nature is that idioms which have been frequently encountered can also be highly predictable, reducing processing costs. Research done on native (L1) speakers has confirmed this advantage by comparing idioms to similar novel language and finding evidence of eased processing in a number of ways including, but not limited to, faster reaction times in semantic judgement tasks (Tabossi, Fanari, & Wolf, 2009), faster reading times (Conklin & Schmitt, 2008), and fewer and shorter fixations on idioms during reading (e.g., Underwood, Schmitt, & Galpin, 2004). In addition to these faster recognition processes, native speakers are able to access the idiomatic meaning even before being exposed to the final constituent of the idiom in cases where an idiom has been shown to be highly predictable (Cacciari & Tabossi, 1988; Rommers, Dijkstra, & Bastiaansen, 2013; Tabossi & Zardon, 1993). What is unclear, however, is the extent that such advantages also apply to highly proficient non-native speakers. Though it has been established that non-native speakers, like native speakers, have online access to both figurative and literal-constituent meaning in idioms at idiom offset (e.g., Beck & Weber, 2016a), research has yet to clearly establish the boundaries of this access to meaning and whether some of the same processing advantages apply. In particular, there is some evidence that idiomatic processing as a whole is slower for L2 speakers compared to L1 speakers (e.g., S. D. Beck & Weber, 2016a; Siyanova-Chanturia, Conklin, & Schmitt, 2011). If this is the case, idiom predictability may not play a role in L2 idiom processing. In the studies presented here, we look at the aspect of predictability, and whether or not non-native speakers are able to access the idiomatic meaning of highly-predictable idioms before the offset of the final idiomatic constituent.

Prior to the study of predictability in idioms, idiomatic processing was assumed to be a different and separate process from literal language processing. Following this assumption, processing idiomatic phrases was generally assumed to be a step-wise procedure. In the traditions of standard pragmatic views of language processing (e.g., Searle, 1979), a literal-first approach requires that literal meaning is processed first, and only when this meaning is rejected or in the presence of exceptional circumstances (i.e., biasing context) is figurative language directly retrieved (e.g., Bobrow & Bell, 1973). Conversely, a figurative-first approach suggests that the figurative meanings of idioms must be obligatorily retrieved immediately before literal meanings are activated, again based on a meaning and integration mismatch (e.g., Gibbs, 1980; Schweigert & Moates, 1988). A third possibility following the two-step approach is that figurative and literal processing occur simultaneously. Because figurative

retrieval is faster than literal computation, this hypothesis predicts faster access to figurative meanings (e.g., Swinney & Cutler, 1979). However, it is not the case that any of these step-wise procedures explains the variable timing of access to figurative meaning all the time, so no single hypothesis following such approaches can account for the variation found in the timing of access to meaning between idioms (e.g., Gibbs, Nayak, & Cutting, 1989; Libben & Titone, 2008; Titone & Connine, 1994a).

One source of such variation between idioms is predictability. The seminal study on predictability of idioms was undertaken by Cacciari and Tabossi (1988) in which the notion of the idiomatic key was derived. Based on a series of cross-modal priming studies, the authors found that an idiom's figurative and literal constituent meanings were available at different times depending on the point at which an idiom became recognizable as such (e.g., *to be in seventh heaven* is recognized after "seventh", and *go to the devil* after "devil"). In a lexical decision task, the priming of words related to the figurative meaning of an idiom and a literal constituent were compared with unrelated control words (e.g., figurative: HAPPY, literal: SAINT, control: UMBRELLA). If an idiom was recognizable by most native speakers before the final constituent (based on cloze-probability testing for idiomatic completions), only the figurative meaning and not the literal meaning was available at the offset of the idiom. Conversely, if the idiom was not predictably completed as an idiom before the final constituent, then the literal constituent meaning was activated while figuratively related words were not. An additional study confirmed that both meanings were available in the latter case when tested 300ms after the offset of the idiom. These results formed the basis for the authors' Configuration Hypothesis. Under this theory of processing, idioms are made up of configurations of words, and processing is necessarily literal until the occurrence of the recognition point within the idiom is reached, known as the idiomatic key. Once the idiomatic key has been encountered, or enough information has been gathered to recognize the string as idiomatic, literal processing ceases, and access to figurative meaning occurs. The concept of the idiomatic key crucially implies that processing of idioms occurs in a probabilistic, or predictability-based manner. Though, since meaning activation was always tested following the idiomatic key, lexical models of processing could not be entirely excluded based on their findings.

Tabossi and Zardon (1993) followed up on this hypothesis by testing whether the idiomatic meaning was available prior to the idiomatic key or not. In two cross-modal priming studies, the authors first tested predictable idioms with two content words following the verb. The target words were presented following the verb, the first content word, and the second content word. As predicted by the Configuration Hypothesis, figuratively related words showed priming effects following the first and second content word, with no differences between these two placements. In a second study, in order to negate the possibility of a lexical view of processing, unpredictable idioms were chosen and tested

following the same experimental design. For these idioms, the meaning was only available after the second content words, confirming the importance of the idiomatic key in processing.

The findings of Cacciari and Tabossi (1988) and Tabossi and Zardon (1993) suggest that step-wise models of idiomatic processing are insufficient in describing L1 idiom processing. Rather, hybrid models are necessary that better describe the flexibility in meaning access that native speakers display. The Configuration Hypothesis was later extended by Titone and Connine (1994a) to suggest that literal meaning need not be terminated upon activation. Rather, the idiom representation is linked with the composite literal meaning, therefore still allowing for activation of the literal constituents and even phrases as well as competition in meanings under certain conditions. This adjustment is in tune with more recent research on idioms as well as other types of formulaic language in which access to the whole does not preclude meaning access to the literal constituents' meaning (e.g., Beck & Weber, 2016a) and structure (e.g., Peterson, Burgess, Dell, & Eberhard, 2001).

The flexibility assumed of native speaker idiomatic processing, however, is not necessarily presumed to mirror non-native idiomatic processing. Instead, a fundamental difference in L1 and L2 processing is often assumed, namely, that non-native speakers will always first process literal meaning by default. Cieślicka (2006) conducted a cross-modal priming study on non-native listeners in which figuratively or a literally related targets were presented either at the offset of the phrase or at the offset of the penultimate word in the phrase. Overall, the results indicated more priming for literally related targets and only marginal results for figurative targets. There were also no significant differences in figurative priming between the two positions. These results were interpreted to suggest that literal processing indeed has priority over figurative processing and suggested that the salience of literal language (see e.g., Giora, 1997) may be at the root of this difference. However, the study had several limitations. Based on the targets given in the examples, all figurative targets were abstract, while all literal targets were concrete objects (e.g., for the idiom *bury the hatchet*, literal target: AXE, figurative target: FORGIVE). Additionally, and most importantly, a native speaker control group was not present. Thus, the priming differences between literal and figurative targets cannot be clearly attributed to differential L1 and L2 processing and may be the result of other differences (e.g., abstractness).

Beck and Weber (2016a) followed up on these claims with another cross-modal priming study. In this study, both a native and a highly proficient non-native group of speakers (German L1) participated in the lexical decision task (again with figuratively and literally related targets), in which idioms were varied only systematically in their translatability from English to German (e.g., *lend one's ear* is translatable word-for-word into the same German idiom, while *kick the bucket* has an equivalent idiom using different lexical items). Targets were presented at the offset of the sentences, which coincided with the last idiom constituent (e.g., targets for *kick the bucket*: figurative: DIE, literal: PAIL, controls: ZOO, BOAT). The authors found significant priming effects in both types of targets

compared to their controls for both native and non-native speakers. Based on the similar behaviour of L1 and L2 participants, the authors suggest that highly proficient non-native processing is indeed similar to native processing, a result also confirmed in a later but similarly designed cross-modal priming study conducted by van Ginkel and Dijkstra (2019).

Additional studies have compared native and non-native processing using idioms as well as other formulaic language compared to novel phrases and have found more mixed results. Using a self-paced reading paradigm, Conklin and Schmitt (2008) found reading advantages for idioms compared to novel phrases in both L1 and L2 readers, suggesting that both groups are sensitive to the processing advantages of formulaic phrases. On the other hand, Underwood et al. (2004) as well as Siyanova-Chanturia, Conklin, and Schmitt (2011) investigated such phrasal advantages using eye-tracking in reading and found that while natives fixated less often and for shorter periods of time on final elements in idiomatic or formulaic sequences compared to novel phrases, non-natives did not reliably show the same effects. Rather, Siyanova-Chanturia et al. (2011) pointed out that idioms were processed in the same manner as novel phrases for L2 readers. These studies suggest that while access to meaning may be available online, L2 speakers may not be reliably sensitive to predictive advantages seen in L1 speakers.

In the current study, we asked if predictability can also impact figurative priming for non-native listeners as it has been shown to do so in native listeners (e.g., Cacciari & Tabossi, 1988; Tabossi & Zardon, 1993). In particular, when considering the Configuration Hypothesis and the idea of the idiomatic key, is figurative meaning available as early as the idiomatic key? If we expect that L2 processing occurs in the same manner as L1 processing, the predictability of an idiom and therefore the placement of the idiomatic key should impact the availability of figurative meaning in online processing. However, if L2 processing differs from L1 processing (i.e., in speed of access to figurative meaning), we do not expect an idiom's predictability to reliably impact meaning access. We conducted two cross-modal priming studies in order to investigate this question.

4.2 EXPERIMENT 1

4.2.1 Method

In a cross-modal priming study, German L2 listeners heard English sentences with short, non-biasing contexts containing idioms. In a lexical decision task, response times to target words related to the figurative meaning or a literal constituent meaning in the idiom were compared to matched unrelated control words in order to measure lexical priming. The idioms varied in predictability, or the placement of the idiomatic key (early-key idioms and late-key idioms), and presentation of the target words was presented at the offset of the idiomatic key (i.e., during the sentence for early-key idioms, and at the

end of the sentence for late-key idioms). If L2 processing is sensitive to predictability in the same manner as L1 processing, we expected figurative priming for both types of idioms following the idiomatic key. If L2 priming is not sensitive to predictability, we expected to find priming only in the presentation of the idiomatic key at the end of the sentence (following late-key idioms), and not earlier. We expected literal priming to occur regardless of idiom-type.

4.2.1.1 Participants

A total of 40 non-native speakers of English (German L1) took part in the study, and 38 were included in the analysis. One participant was excluded from the analysis for not meeting the participation requirements (L1 only German), and 1 was excluded based on data loss. All participants (30 female, average age of 23.71, *SD* = 3.32) received financial compensation for participation and were recruited at the University of Tübingen, primarily from the department of English. Participants were all advanced users of English and reported an average of 10 years of formal English instruction and rated their English skills at 5.8 on a 7-point scale (1 corresponds to very poor and 7 native-like). A total of 5 participants were left-handed, and none reported any visual or hearing impairments.

4.2.1.2 Materials

The experiment consisted of 36 target and 104 filler trials. Target trials consisted of short, neutral sentences ending in idioms (e.g., I think she *got up on the wrong side of the bed*.). The majority (29) of the idioms were taken from Beck and Weber (2016a), and 7 additional idioms (to represent both low and high-predictable idioms comparably) were taken from the German Database of Idiom Norms (**Chapter 2**). Idioms were selected that were rated highly for familiarity of encounter (mean = 5.2, *SD* = 0.81, on a scale from 1 to 7) and familiarity with the meaning (mean = 5.8, *SD* = 0.56, on a scale from 1 to 7). Idioms with a literality in the mid-range were selected in order not to bias the availability of literal interpretations one way or another.

Idioms varied systematically in their predictability and therefore the placement of the idiomatic key. Idioms were divided into two categories with either an early idiomatic key or a late idiomatic key based on the cloze-probability testing done in the DIN database. Highly predictable idioms that were completed by at least 70% of participants correctly at any word before the final constituent word were classified as early-key idioms, and idioms with less than 70% correct completions before the final constituents were classified as late-key idioms. Early-key idioms had an average correct completion rate of 85% (*SD* = 10) at the point identified as the idiomatic key, whereas late-key idioms had an average of only 23% (*SD* = 20), and groups significantly differed from one another (t = -11.56, df = 25.62, $p < .001$). For an example of each idiom type, see Table 4-1. Notably, idioms classified as early-key idioms align with highly predictable idioms, and those classified as late-key idioms as unpredictable. However, as the idioms were never correctly identified as idioms in the cloze-

probability testing, there is a small possibility that these unpredictable idioms are not recognized as idioms. In the current study, the use of familiar idioms is meant to guard against this possibility.

Table 4-1. Example items

Key	Idiom	Correct Completions	Targets	Controls
Early	get up on the wrong | side of the bed	86 %	MISTAKE (L) MOODY (F)	PLANET BOTTLE
Late	Pop the question |	18 %	ANSWER (L) PROPOSE (F)	ORANGE PROBLEM

Note: The Symbol | is placed after the idiomatic key. (F) is the figurative target, and (L) is the literal target.

Four target words associated with each idiom were either literally related to the idiomatic key (or last content word before the idiomatic key), figuratively related to the idiom, or a respective control. For the late-key idioms taken from Beck and Weber (2016a), the same targets were used. For all other idioms and targets, literal targets were taken from established associative norms (Nelson, McEvoy, & Schreiber, 1998), and figurative targets were developed by two native speakers as there is no similar database for figuratively related meanings known to these authors. For the early-key idiom *get up on the wrong side of the bed*, the literal target MISTAKE was chosen as it corresponds to the key "wrong" and the control target PLANET was matched for orthographic complexity and frequency. Likewise, the figurative target MOODY relates to the overall idiomatic meaning, to have a bad start to the day or wake up in a bad mood, and the matched control was BOTTLE. For a late-key idiom such as *pop the question*, the literal target ANSWER corresponded to the final constituent "question" and the figurative target PROPOSE still corresponded to the overall meaning of the idiom, to propose marriage, both with matching controls ORANGE and PROBLEM, respectively. While the literal and figurative targets tend to vary in abstractness (i.e., MOODY vs. BOTTLE), this limitation did not prevent figurative priming in the original study (for a more detailed discussion of these limitations, see Beck and Weber, 2016a, **Chapter 3**). Both idioms with their corresponding targets can also be found in Table 4-1.

In order to make as few changes as possible to the data from Beck and Weber (2016a), only targets for new idioms or literal targets that were related to words earlier than the final constituent were changed, and there were slight differences in list-frequencies based on the $SUBTLEX_{US}$ corpus (Brysbaert & New, 2009). Individual t-tests confirmed that figuratively related targets were overall more frequent than their controls (means= 164.472, 131.611 respectively, $t = 2.716$, $p < 0.01$) and the literal control targets (mean= 130.722, $t = 2.879$, $p < 0.01$). Since lexical frequency might therefore disadvantage figurative targets, frequency counts were included in the final models in the statistical analysis.

Half (18) of the target trials consisted of early-key idioms, and the other half of late-key idioms. The 104 filler trials were the same ones used in Beck and Weber (2016a), consisting of 20 additional idiomatic trials. These filler idiomatic trials were followed by non-words, helping ensure a balance of word and non-word targets for all trial types. Overall, half of the targets were words, and half non-words, and of the word-targets some filler targets also had meanings related to the literal meanings of a constituent word in the filler sentence. The experiment was divided into four counter-balanced lists so that each participant heard each idiom only once. Only target items varied between lists.

The sentences used in the experimental trials were recorded by a male speaker of American English in the recording studio of the LingTüLab (in the English Department of The University of Tübingen). Each participant heard the same sentences, as only the visual targets varied per list. The experiment began with four practice trials followed by the 140 randomized experimental trials. Participants were allowed a self-directed concentration break after trial 70.

4.2.1.3 Procedure

Participants were tested individually in single rooms in the LingTüLab (University of Tübingen). Individuals were instructed that they would be hearing English sentences and would be presented with either an English word or non-word during or at the end of the auditory sentence. Participants were told to listen carefully and respond as quickly, but accurately as possible by pushing the red or green button on a Cedrus Button-box. The green "YES" button was always pushed with the dominant hand, while the red "NO" button was pushed by the non-dominant hand. Participants were advised that careful listening was important, even while making lexical decisions, as they would be asked questions about what was heard following the lexical decision task.

Once it was clear that participants understood the task, the experiment began. Auditory stimuli were presented at a clear volume on closed headphones. Target words appeared on a computer screen with a black background and white text in size 20 font. Audio files played to completion for each trial, but targets were presented at different times during the trials. The targets were presented immediately at the offset of the idiomatic key (up to 4 words before the end of the idiom) and varied for filler trials. For late-key idioms, the presentation of the target coincided with the final word of the idiom; for early-key idioms, the presentation of the target appeared immediately at the offset of the word identified as the idiomatic key. If a button was pressed still during the presentation of the auditory sentence, then the next trial began 2000ms after the end of the auditory stimulus. If the button press occurred after the end of the auditory stimulus, then the next trial began 2000ms later. If no button was pressed, the next stimulus began after 3000ms (ensuring at least 2000ms of silence).

Following the lexical decision task, participants completed a short comprehension task and a language background questionnaire. Both tasks were completed on forms using Adobe Professional.

For the comprehension task, participants saw a list of 60 sentences, and had to decide whether they heard the sentences listed by checking "YES" or "NO." Half of the sentences on the test were heard in the experiment, and half were new (see Beck & Weber, 2016a; Cieślicka, 2006). The language background questionnaire collected basic information from the participants about their language skills and confirmed their status as L1 German and L2 English speakers.

4.2.1.4 Results

The analysis of reaction times included only trials with correct responses to targets. On average, participants responded correctly to 90% of all items and 98% of all target items. Based on these performance levels, high proficiency in English can be assumed. Additionally, outliers were identified and removed from the analysis if they were either outside of 2 standard deviations from the mean reaction time per participant considering the main experimental factors of *figurativeness*, *relatedness*, and *key-timing* or they exceeded 1500ms (a total of 2.6% of the data).

Table 4-2. Mean reaction times (ms)

	Early Key				**Late Key**			
	Figurative		**Literal**		**Figurative**		**Literal**	
	RT	**inverse RT**	**RT**	**inverse RT**	**RT**	**inverse RT**	**RT**	**inverse RT**
Related	660	-1,641	716	-1,546	592	-1,834	584	-1,861
Unrelated	708	-1,568	692	-1,571	595	-1,784	576	-1,855

R (R Core Team, 2013) and lme4 (Bates, Maechler, Bolker, & Walker S., 2015) were used to perform a linear mixed effects analysis of the relationships between the factors of *figurativeness*, *relatedness*, and *key-timing* on the inverse reaction times. The inverse reaction times were used, as this transformation created the most normal distribution based on the Box-Cox transformation test and a visual confirmation of the data (Venables & Ripley, 2002). The mean reaction times and inverse reaction times, measured from the offset of the idiomatic key, for each condition are displayed in Table 4-2. The mean reaction times are shown in Figure 4-1, with bars representing the standard error of the mean.

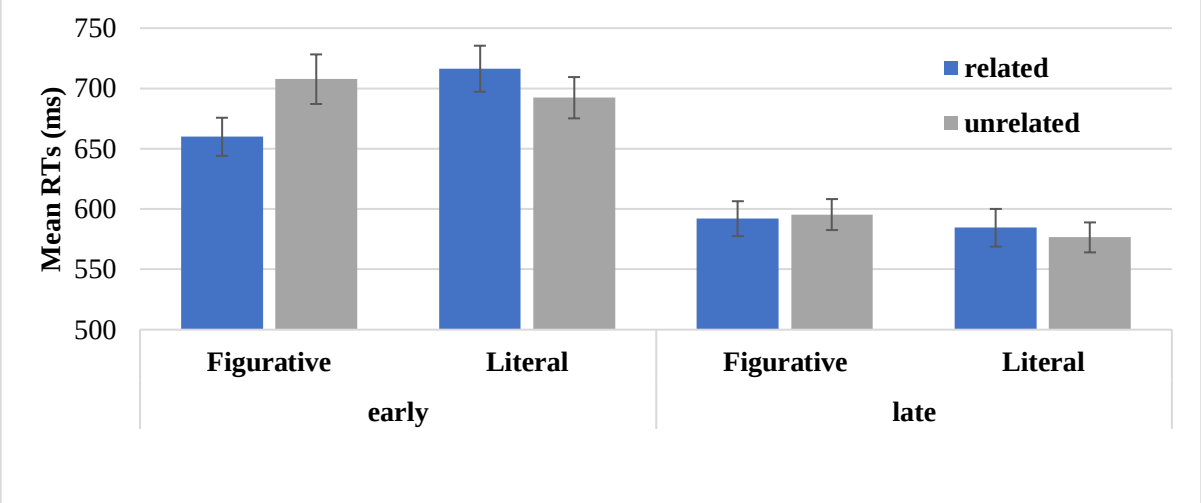


Figure 4-1. Experiment 1 mean RTs (ms) with standard error bars

Note: Grouped by *key-position* (early vs. late from left to right), *figurativeness* (figurative vs. literal from left to right), and *relatedness* (related vs. unrelated).

LMER models with inverse RTs as the dependent variable and fixed factors, centered around 0, were coded and included as follows: *figurativeness* (figurative: 0.5 and literal: -0.5), *relatedness* (related: 0.5 and unrelated: -0.5), and *key-timing* (late: 0.5 and early: -0.5). Additional numeric fixed factors included after centering were *order* (order of experimental trial), *frequency* (target frequency per million), *length* (target length in letters), *L2 meaningfulness* (ratings on the meaningfulness of the idiom, on a scale from 1-7), *L2 familiarity* (ratings on the subjective frequency of encounter of the idiom), *L2 literality* (ratings on the literal interpretability of the idiom), and *proficiency* (self-scored proficiency ratings on reading, writing, speaking, and listening, on a scale from 1 (not at all proficient) to 5 (native-like proficiency) averaged across all four ratings). The idiomatic ratings included were taken from the DIN database as they have been shown to affect access to meaning in cross-modal priming experiments (e.g., S. D. Beck & Weber, 2019; Titone & Connine, 1994a; Titone & Libben, 2014). Subjects and items were included in the models as random factors with random slopes. A maximally justified random effects structure was determined by step-wise selection and model comparison (see e.g., Bates, Kliegl, Vasishth, & Baayen, 2015) using RePsychLing (Baayen, Bates, Kliegl, & Vasishth, 2015) that included random slopes for *figurativeness*, *relatedness*, and *key-timing*. Backward step-wise selection was used to eliminate the additional factors not relevant to the theoretical questions of this study where model comparison showed that they did not contribute to a better fit of the model. The values of the final model are shown in Table 4-3.

Table 4-3. Experiment 1 final LMER model

Fixed Effects and Controls	Effect Size	SE	*t*-Value	Pr(<|*t*|)
Intercept	-1.690	0.0495	-34.165	<2e-16***
Figurativeness (coded)	0.006	0.0271	0.218	0.828
Relatedness (coded)	-0.007	0.0276	-0.237	0.813
Key-Timing (coded)	-0.265	0.0355	-7.465	<1.3e-10***
Frequency	-0.070	0.0144	-4.848	3.63e-6***
L2 Meaningfulness	-0.097	0.0352	-2.760	0.007**
L2 Familiarity	0.082	0.0333	2.467	0.015*
L2 Literality	0.028	0.0159	1.732	0.086.
Proficiency	-0.115	0.0488	-2.364	0.024*
Figurativeness* Relatedness (coded)	-0.066	0.0532	-1.235	0.219
Figurativeness* Key-timing (coded)	0.056	0.0536	1.054	0.294
Relatedness* Key-timing (coded)	0.013	0.0534	0.251	0.802
Figurativeness*Relatedness* Key-timing (coded)	0.072	0.1068	0.672	0.503

Note: *.p<.10 *p<.05 **p<.01 ***p<.001*

Of the three factors under investigation in this study, only *key-timing* shows a significant main effect (ß = -0.26, *t*= -7.46, *p*< .001); all targets appearing in the late-key position (i.e., sentence offset) were responded to more quickly than those in the early-key position. Additional fixed factors also affected reaction times. *Frequency* (ß = -0.07, *t*= 4.84, *p*< .001) affected reaction times in that more frequent targets were reacted to more quickly. The idiomatic factors *L2 familiarity* (ß = 0.08, *t*= 2.46, *p*< .05), *L2 meaningfulness* (ß = -0.09, *t*= -2.76, *p*< .01), and, marginally, *L2 literality* (ß = 0.02, *t*= 1.73, *p*= .08) all improved the model fit, but did not impact the interpretation of the main results. Familiarity with the meaning of the idiom decreased reaction times, while subjective frequency of encounter increased reaction times. While the latter result may seem surprising, the ratings were collected from another group of non-native speakers and reflect only how often L2 speakers thought they had come across the idiom. The literal interpretability of an idiom also slowed down reaction times where an idiom was more literally interpretable. Though, this result is only marginal, it reflects other results commonly found in idiomatic processing literature (e.g., Titone & Connine, 1994a). Finally, a significant effect of *proficiency* (ß = 0.08, *t*= 0.03, *p*< .05) confirms that an increased self-rated English proficiency decreased reaction times.

4.2.2 Discussion

The analysis shows that facilitatory priming was neither observed for literally or figuratively related targets nor was there an interaction of these factors with the timing of the idiomatic key. While it was

unclear whether or not priming would be found for figuratively related targets in the early-key condition, it is surprising that no priming was found in the late-key condition. Both Beck and Weber (2016a) as well as van Ginkel and Dijkstra (2019) found figurative and literal priming following the offset of the final idiomatic constituent, and while Cieślicka (2006) did not find figurative priming in this position, literal priming was also present in her study. Although different items were used from the latter two studies, the current study re-used many of the experimental items from Beck and Weber (2016a), so it would be surprising that these idioms and targets already shown to reproduce similar results (e.g., Beck & Weber, 2016b) would be the cause of the null result in the current study.

One possible reason for the lack of figurative priming in the current study is that previous studies did not consider the factor of predictability. While this has been shown to affect L1 processing in a number of studies (e.g., Titone & Connine, 1994a), it has largely been ignored in many L2 studies. A closer look at the items from Beck and Weber (2016a) shows that the idioms included were both predictable and non-predictable idioms. Additionally, neither van Ginkel and Dijkstra (2019) nor Cieślicka (2006) included such norms in their data, and it appears that here, too, there is possible variation. While all of the studies mentioned here tested either after the offset or the penultimate word in the idiom, results indicated the presence of priming that should be comparable to the late-key idioms in the current study. However, one result of such predictability variability is that the high-predictability of some idioms may cause an increased priming for figuratively related targets at the offset (or even penultimate) position, and overall priming for figuratively related targets may be found though there is variation between individual idioms. However, this factor should only affect the figurative targets, and we should still have expected to see literal priming in the late-key targets.

Another possible explanation for the lack of facilitatory priming not just in figuratively related targets, but overall, is that the task demands were too great on the L2 listeners. Unlike in previous experiments, the current experiment asked the participants to make fast, online decisions both during listening and immediately at the offset to the lexical items presented at targets. Additionally, in the instructions, the participants were asked to listen carefully, as there would be questions about the sentences heard at the end of the experiment. While the goal of such instructions was to ensure that participants actively listened rather than reacting only to the visual targets, the result is that the demands of listening may have impeded their ability to respond to the targets or overshadowed priming effects (e.g., Cutler & Clifton, 1999). Slow reaction times displayed by L2 listeners when directly compared to L1 listeners in previous and similar cross-modal priming studies (e.g., Beck & Weber, 2016a; van Ginkel & Dijkstra, 2019) suggests that these listening demands are exaggerated for L2 listeners in a cross-modal priming task.

In order to test the latter theory, a second experiment was conducted in which task demands were decreased for listeners. First, the possibility of semantic activation for both literally and

figuratively related targets was increased by displaying targets 400ms after the offset of the idiomatic key (see e.g., Titone & Libben, 2014) rather than directly at the offset. Second, rather than playing all sentences to completion, listeners only heard the sentences until the offset of the idiomatic key. While this results in sentence fragments for the early-key idioms (and some filler sentences), participants' attention should be equally and wholly focused on the lexical decision task for both idiom types. If, after these changes, facilitatory priming can be found for the late-key idioms as in previous experiments, we assume that the demands of the experimental task influenced the results, and the results for both early- and late-key idioms can better be interpreted in terms of the placement of the idiomatic key.

4.3 EXPERIMENT 2

4.3.1 Method

Experiment 2 followed the same method as Experiment 1 with the exception of the timing of the target presentation and the cut-off of the auditory stimuli immediately after the offset of the idiomatic key.

4.3.1.1 Participants

A total of 41 L2 speakers of English (German L1) took part in the study who did not participate in Experiment 1, and 40 were included in the analysis. One participant was excluded from the analysis for not meeting the participation requirements. All participants (27 female, average age of 23.56, SD= 3.43) received financial compensation for participation and were recruited as in Experiment 1. All were advanced users of English and reported an average of 11 years of formal English instruction and rated their English skills at 5.7 on a 7-point scale (1 corresponds to very poor and 7 native-like). One participant was left-handed, and none reported any visual or hearing impairments.

4.3.1.2 Materials

The same materials were used as in Experiment 1.

4.3.1.3 Procedure

The same laboratory space, task, and order of procedure was used as in Experiment 1. Crucially, there were two differences from Experiment 1:

1. The presentation of all targets occurred 400ms after the offset of the idiomatic key rather than directly at the offset.
2. The presentation of the auditory stimuli sentence stopped after the offset of the idiomatic key. For late-key idioms, there was no change from Experiment 1 during listening as the key occurred at the end of the sentence. For early-key idioms, the full idiom was not presented in the stimuli.

As in Experiment 1, following a button-press and 2000ms or a maximum of 3000ms with no response, the next trial began.

4.3.1.4 Results

The analysis followed the same order and procedure as in Experiment 1. Again, analyzed reaction times included only trials with correct responses to targets. On average, participants responded correctly to 91% of all items and 99% of all target items. Additionally, outliers were removed from the analysis if they were either outside of 2 standard deviations from the mean reaction time of the sample considering the main experimental factors of *figurativeness*, *relatedness*, and *key-timing* or they were faster than 300ms or exceeded 1500ms (1.8% of the data).

Table 4-4. Mean reaction times (ms) and inverse reaction times

	Early				**Late**			
	Figurative		**Literal**		**Figurative**		**Literal**	
	RT	**inverse RT**	**RT**	**inverse RT**	**RT**	**inverse RT**	**RT**	**inverse RT**
Related	641	-1.657	590	-1.782	611	-1.722	593	-1.771
Unrelated	623	-1.696	607	-1.728	635	-1.660	613	-1.716

A linear mixed effects analysis of the relationships between the factors of *figurativeness*, *relatedness*, and *key-timing* was performed on the transformed reaction times, again the inverse reaction times as determined by the Box-Cox transformation test and a visual confirmation of the data (Venables & Ripley, 2002). The mean reaction times and inverse reaction times, measured from the offset of the idiomatic key, for each condition are displayed in Table 4-4. The mean reaction times are graphed in Figure 4-2, with bars representing the standard error.

LMER models with inverse RTs as the dependent variable and fixed factors, centered around 0, were coded and included as follows: *figurativeness* (figurative: 0.5 and literal: -0.5), *relatedness* (related: 0.5 and unrelated: -0.5), and *key-timing* (late: 0.5 and early: -0.5). Additional numeric fixed factors included after centering were *order*, *frequency*, *length*, *L2 meaningfulness*, *L2 familiarity*, *L2 literality*, and *proficiency*. Subjects and items were included in the models as random factors with random slopes. A maximally justified random effects structure was determined by step-wise selection and model comparison using RePsychLing (Baayen et al., 2015) that included a random slopes for *relatedness*. Backward step-wise selection was used to eliminate the additional factors not relevant to the theoretical questions of this study where model comparison showed that they did not contribute to a better fit of the model. The values of the final model are shown in Table 4-5.

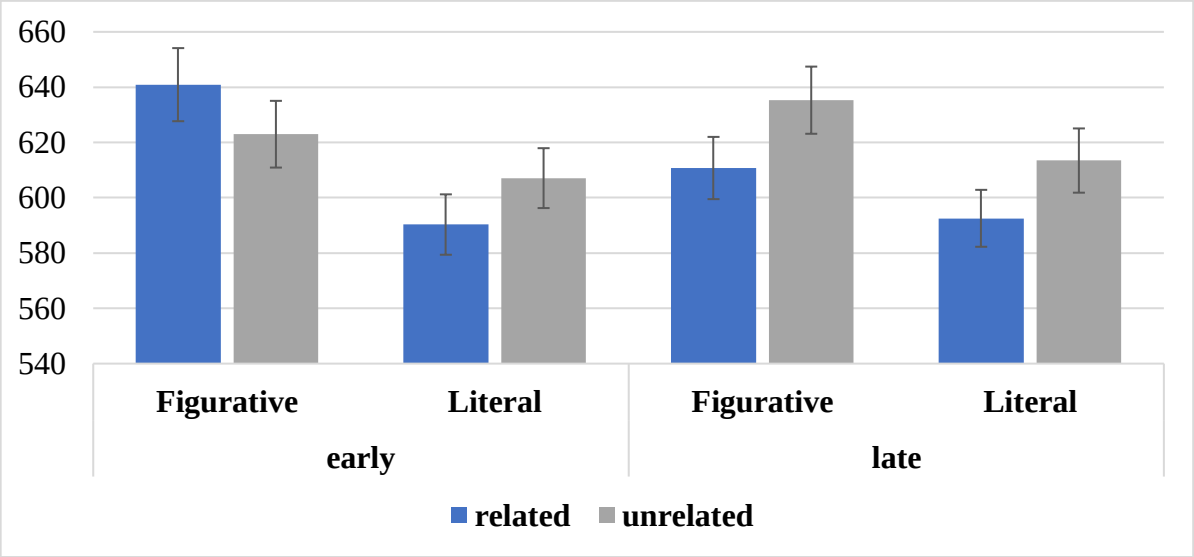


Figure 4-2. Experiment 2 mean RTs (ms) with standard error bars

Note: Grouped by *key-position* (early vs. late from left to right), *figurativeness* (figurative vs. literal from left to right), and *relatedness* (related vs. unrelated).

The factor figurativeness showed a main effect (*ß* = 0.04, *t*= 2.47, *p*< .05) suggesting that all figurative targets were responded to more slowly than literal targets (regardless of relatedness). Additionally, *frequency* (*ß* = -0.03, *t*= -3.91, *p*< .001) showed the effect of faster reaction times for more frequent targets. No other factors or interactions reached significance. However, Figure 4-2 does align visually with predicted patterns.

Table 4-5. Experiment 2 final LMER model

Fixed Effects and Controls	**Effect Size**	**SE**	***t*-Value**	**Pr(<|*t*|)**
Intercept	6.394	0.021	305.049	2e-16***
Figurativeness (coded)	0.040	0.016	2.476	0.015*
Relatedness (coded)	-0.009	0.015	-0.611	0.542
Key-Timing (coded)	0.005	0.015	0.342	0.733
Frequency	-003.1	0.008	-3.915	1.45e-04***
Figurativeness* Relatedness (coded)	0.027	0.031	0.883	0.379
Figurativeness*Key-timing (coded)	-0.009	0.031	-0.283	0.777
Relatedness*Key-timing (coded)	-0.022	0.031	-0.698	0.486
Figurativeness*Relatedness* Key-timing (coded)	-0.056	0.062	-0.903	0.368

Note: *.p<.10 *p<.05 **p<.01 ***p<.001*

4.3.2 Discussion

As in Experiment 1, no facilitatory priming was found for literally or figuratively related targets in either the early-key or late-key positions. Thus, in addressing the concerns of task demands in

Experiment 1, Experiment 2 does not deliver a clear answer. Though, the effect of *key-timing* is no longer significant in Experiment 2. While the reaction times graphed in Figure 4-2 do present the expected patterns as well as those seen in previous experiments (e.g., Beck & Weber, 2016a; Cieślicka, 2006; van Ginkel & Dijkstra, 2019), they fail to reach significance. Since facilitatory priming is neither reliably found in early-key nor late-key positions, the latter of which has been consistently found for L2 listeners, the results should not be interpreted as negative evidence for prediction. Possible issues will be discussed in the general discussion.

4.4 GENERAL DISCUSSION AND CONCLUSION

In addressing the question of whether predictability can impact L2 listeners' access to meaning, particularly considering the placement of the idiomatic key, both experiments provide inconclusive results. While the mean reaction times in Experiment 2 generally reflect expected patterns of responses, unlike in Experiment 1, these patterns are not reliably corroborated in statistical analyses. Experiment 2 is therefore also not conclusive in whether task demands may have impacted reaction times in Experiment 1. Thus, a closer look at possible issues with the experimental methods and/or materials must be taken into consideration.

One possible issue is the (un)reliability of native speaker predictability measures. The idioms for the current study were taken from the DIN database, a database with both L1 and L2 ratings on over 300 idioms. While the results are generally similar to other databases that have been released (e.g., Titone & Connine, 1994b), marked differences in the same idioms between studies suggest that participants in the ratings experiments may not be accurate or unbiased in their judgements. Inaccuracies in ratings are particularly crucial in determining the placement of the idiomatic key. For example, the idiom *kick the bucket* was rated as highly predictable (70% correct completions after "the"), a result that is surprising given the broad and general nature of "kick" as a verb. These results may be skewed by the bias of a large number of idioms in the ratings tasks. Where a large number of idioms are present both L1 (e.g., Bobrow & Bell, 1973) and L2 (e.g., Beck & Weber, 2016b) studies have shown that figurative completions or interpretations become more frequent or the strength of these activations can increase. Thus, the reliability of such cloze-probability tasks with a large number of idioms cannot be taken for granted.

Another issue with the predictability scores is that they were taken from L1 speaker judgements. Cieślicka (2006) also addressed the issue of using L1 predictability for L2 listeners, noting that a best practice for such studies would be to include predictability measures from non-native speakers rather than native speakers. However, in order to accurately represent the knowledge variability in an L2 group in a particular experiment, the same speaker group should be used to determine predictability from the experiment, presenting an unrealistic experimental situation in which

either participants must be exposed to the idioms before the experiment and risk a change in behaviour, or the participants must see the idioms afterwards, which would impact their responses. A more plausible compromise for further studies would be cloze-probability testing in a similar group of L2 speakers as the one tested, much like the L2 familiarity measures used in **Chapter 2**. However, the same concerns remain, that subjective familiarity with the idiom will greatly impact such responses. Furthermore, the tendency for L2 speakers to use literal language preferentially in offline and production tasks may prevent any idioms from appearing as predictable by L2 listeners (Irujo, 1993).

Word frequency also had a marked impact on the data. In both experiments, frequency was a highly significant factor for reaction times. The word lists were originally controlled for frequency based on Beck and Weber (2016a) using the CELEX database (Max Planck Institute for Psycholinguistics, 2001) for consistency. However, due to a large number of missing frequencies, the words were later re-calculated based on the $SUBTLEX_{US}$ corpus (Brysbaert & New, 2009) for inclusion in the LMER models. The differences in lists were discussed in the materials section, and, importantly, there were list differences between the figuratively related and -unrelated control word lists. In particular, figuratively related targets were more frequent than the two control target groups. While the inclusion of this measure in the regression model should help identify the effects of the factors under consideration, the strength of frequency as a predictor suggests that the materials were not adequately controlled for considering the speaker group at hand, and this may provide some indication as to why figurative priming that appears in a visual inspection of the late-key data do not reach significance.

Critically, however, the concept of the idiomatic key and its role in idiomatic processing should be looked at with more scrutiny. With consideration of the criticisms of the identification of the idiomatic key are taken to heart, the challenge of correctly identifying the key remains unanswered. Whether or not such a task (i.e., identifying a true idiomatic key) is possible suggests that the concept of the idiomatic key may need reconsideration, as should studies attempting to identify predictability in idioms. Namely, previous authors have defined this as the point at which an idiom can be identified as such (e.g., Cacciari & Tabossi, 1988; Tabossi & Zardon, 1993), though, more recent research has suggested that access to idiomatic meaning is incremental (see Cacciari, 2014 for an overview) and subject to a variable number of factors during processing (e.g., Libben & Titone, 2008). For example, the presence of context is one factor that can also aid in the speed of idiomatic recognition (e.g., Colombo, 1993, 1998; see also **Chapter 5**). If the idiomatic key represents the point at which enough information accumulates for recognition, then we should recognize the idiomatic key as a concept more fluid than a single word within an idiom. The idiomatic key may vary for an individual in the same idiom given a variety of contexts, but also it may vary greatly between individuals. Whereas an L2 speaker with a great deal of English experience and knowledge of L1 idioms may recognize an idiom

very early during listening, another less-experienced or simply otherwise-experienced L2 listener may not recognize the idiom until its completion, or even at all. While this does not negate the importance of studies using predictability as a measure in idiom processing studies, as it has been shown to affect processing (e.g., Cacciari & Tabossi, 1988; Tabossi & Zardon, 1993; Titone & Connine, 1994b), it's important to consider predictability and the idea of the idiomatic key in combination with other factors that may influence idiom processing and access to figurative meaning.

Overall, no evidence was found for early access to idiomatic meaning in L2 listeners based on a high-predictability or an early idiomatic key. However, the inconclusive results also should not be used as negative evidence. Rather, further investigation using highly-predictable idioms based on the target listener group and careful experimental controls may help further investigate the issue. In particular, later presentation of the target (i.e., one word later in early-key idioms) may be another experimental variation worth exploring. Considering L2 ratings for predictability may also add worthwhile information during item selection. Finally, the use of more sensitive measures such as eye-tracking (see e.g., Underwood et al., 2004) or EEG (e.g., Rommers et al., 2013) may help to re-examine this effect where reaction times to a lexical decision task may be insufficient.

4B L1 AND L2 LEARNERS KEEP THEIR EYES ON THE PRIZE: EYE-TRACKING EVIDENCE DURING IDIOM RECOGNITION[5]

Ruth Keßler & Sara Beck

Abstract

Idioms appear to be both compositional and non-compositional at the same time. For an idiom like *to let the cat out of the bag* (figuratively, to reveal a secret), it is possible to access the figurative meaning of an idiom without making use of the literal constituent words, though the literal meaning is also well-formed. While adult native speakers activate literal meaning even after accessing figurative meaning in some cases (e.g., Keßler, Weber & Friedrich, submitted; **Chapter 3**), there may be some differences in the processing strategies used by language learners (e.g., Cieślicka, 2006). In this study, we asked to what extent language learners, namely L1 children (aged 13) and L2 adults, anticipate idioms' final constituents and activate their literal meanings. We tested these questions using a visual world eye-tracking paradigm on children with German L1 and adults with proficient German L2 (English L1). Participants listened to neutral sentences containing familiar, predictable idioms without their final word (*Hannes let the cat out of the …*) and had to choose the correct idiom completion from one of four displayed words by giving a verbal response. Displayed words were (a) correct completions (*BAG*), (b) distractors semantically related to the correct completion (*BASKET*), and (c) two unrelated distractors (*ARM, STOMACH*). Like Keßler, Weber and Friedrich (submitted), we found that both L1 children and L2 adults anticipate final constituent words and activate their literal meanings. However, there are differences in the time-course and overall fixation patterns between the groups tested. Namely, children are faster than L2 adults, but slower than L1 adults, in prediction and literal activation. Additionally, children show a literal activation pattern similar to adults, while L2 adults show overall more uncertainty and stronger literal activation. These results are discussed in terms of possible processing differences based on language learned as well as proficiency.

[5] This chapter is adapted from an article also to appear in the dissertation of Ruth Keßler (in preparation).

4.1 INTRODUCTION

Idioms are both formulaic sequences as well as figurative expressions, identities which both present their own implications and possible challenges for language processing. Although idiomatic meaning appears to challenge compositional theories of language processing, for adult native speakers of a language, the use and comprehension of idioms is a breeze. In fact, research shows that not only are idioms like *to let the cat out of the bag* and *to have butterflies in one's stomach* understood figuratively by native speakers online (e.g., Beck & Weber, 2016a; Cacciari & Tabossi, 1988), but there may even be an advantage in speed of comprehension when compared with similar novel expressions (e.g., Gibbs, 1980; McGlone, Glucksberg, & Cacciari, 1994; Swinney & Cutler, 1979). This advantage might stem from predictive properties based on their formulaic nature, in which idiomatic strings are recognized early during comprehension. However, the same speed advantages may also stem from a direct mapping of the form to the idiomatic or figurative meaning, bypassing compositional processes and associated word-meaning activation processes. The figurative meaning of the idiom (to reveal a secret) is thus associated with a fixed, word-like form rather than building the meaning of an idiom like *to let the cat out of the bag* online from its individual constituents (i.e., CAT and BAG). In spite of this lack of apparent contribution of the literal words to the idiomatic meaning, it seems that, in many instances, native speakers still activate individual literal constituents (e.g., Beck & Weber, 2016a; Cacciari & Tabossi, 1988; Titone & Connine, 1994) and even literal phrasal meaning (e.g., Holsinger & Kaiser, 2013; see also **Chapter 5B**) without hindering the speed of access to the idiomatic meaning.

However, these processing advantages in idioms are not to be underestimated, particularly when comparing first (L1) and second (L2) language learners. For a language learner to understand an idiom, they must first recognize the phrase as an idiom, have been exposed to the figurative meaning of the idiom, and must ignore the possible literal interpretation achieved through traditional meaning computation. Additionally, they must do so quickly if predictive processes are to offer the same processing advantages as are evidenced in native speakers. However, language learners have less experience with idiomatic expressions, which might negatively influence idiom recognition and speed of processing. Furthermore, the role of literal meaning might differ both between adult native speakers as well as between L1 and L2 learners of a language. Language learners may activate literal meaning differently than experienced users of a language based on language experience. Both learner groups may also differ because of different underlying mechanisms of acquisition (Arnon & Christiansen, 2017). For instance, children may acquire formulaic sequences as units, and literal word meanings might be less prominent in processing of those sequences, while adult L2 learners are more likely to combine words within formulaic sequences and rely more on literal word meaning.

The current study addresses both the issue of prediction and literal activation. Specifically, the current study asks to what extent adult L1 predictive processes associated with formulaic sequences also apply to language learners, and what processes occur following idiom recognition, particularly concerning literal meaning activation. In order to address these questions, we will first review these processes in native speakers and then attend to the current research in language learners, specifically proficient non-native (L2) speakers and child (L1) speakers.

4.1.1 Predictive processing in idioms

The processing advantages of idioms compared to similar novel phrases are well-documented (e.g., Conklin & Schmitt, 2008; Gibbs, 1980; McGlone et al., 1994; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Swinney & Cutler, 1979). Some theories of idiomatic processing account for this processing advantage by suggesting that this holistic access to figurative meaning is either partially, in the case of hybrid theories (e.g., Cacciari & Tabossi, 1988; Sprenger et al., 2006), or even fully responsible for fast idiomatic processing (e.g., Swinney & Cutler, 1979). Alternatively, processing advantages are often also attributed to familiarity with and the fixed nature of idioms (e.g., Tabossi, Fanari, & Wolf, 2009). Providing evidence for the latter instance, reading studies using eye-tracking techniques have consistently shown processing benefits for idioms as well as other formulaic sequences. Underwood, Schmitt, and Galpin (2004) used a reading paradigm that measured fixations on individual words within formulaic sequences, including idioms embedded in unambiguous contexts. They found that participants fixated less while reading all types of formulaic sequences compared to novel phrases. They also found fewer fixations on terminal words in the phrases, which indicates prediction of these words based on the previous elements of the sequence. This result was confirmed by Siyanova-Chanturia et al. (2011), who found that idioms, used in both literal and figurative contexts, incited fewer and shorter fixations than similar novel phrases (e.g., *at the end of the day* vs. *at the end of the war*). However, they did not find any differences between idiomatic and literal readings of the idioms. This difference provides even further evidence that the processing benefit associated with idioms may indeed relate to the fixed nature of the phrase.

Additionally, idioms pose an interesting case because of their distinct idiomatic and phrasal meanings, and these meaning differences are a common testing ground for predictive processes or anticipatory online processing that occurs during listening. There is evidence of qualitative differences in idiomatic processing compared to other types of formulaic language. Carrol and Conklin (2019) compared processing in idioms, collocations and binomials and found that specific idiomatic properties such as phrasal frequency, familiarity, and decomposability play an important role in the speed of processing and that familiarity with the phrase might explain shorter reading times associated with idioms, unlike other types of formulaic language. Additionally, ERP studies conducted by Vespignani, Canal, Molinaro, Fonda, and Cacciari (2009) and by Canal, Pesciarelli, Vespignani, Molinaro, and

Cacciari (2017) provide further evidence for such differences by comparing literal and figurative processing. The former study found that prediction within idioms is qualitatively different from predictive processing in novel, literal language based on the timing and distribution of N400 and P300 effects. The latter study added to this finding with evidence of effects in biased idiomatic sentences as early as the first word of the idiom, suggesting that prediction and possible reintegration processes occur immediately at recognition of the idiom, which may be sooner than in other types of formulaic language. Thus, the reasons for idiomatic advantages are still in need of further investigation.

While these well-established general speed and processing advantages in idioms indirectly indicate predictive mechanisms for final words as a result of their familiar and conventionalized form, only few studies have looked directly at prediction within idioms. Cloze-probability tests show that native speakers can produce predictions for constituent words within idioms. In fact, most native speakers predict the final word *BAG*, when encountering the phrase *let the cat out of the …* (e.g.,Van Lancker Sidtis et al. 2015; see also **Chapter 2**), and some idioms allow for prediction very early in the phrase. An online study by Kessler, Weber, and Friedrich (submitted) confirm these processes and found evidence for predictive processing both with eye-tracking and ERPs. In their eye-tracking study, native speakers listened to a non-biasing sentence ending with a predictable idiomatic phrase missing its final constituent. Participants saw the correct final constituent of the idiom, a literal distractor, and two unrelated distractors. Fixation rates on the correct completion were significantly higher than to all other options early during listening. In their ERP data, in line with the results of Rommers, Dijkstra, and Bastiaansen (2013), Kessler and colleagues found that when the final constituents of these predictable idioms were replaced with incorrect and unexpected words (e.g., *let the cat out of the* BASKET/STOMACH), N400 effects occurred, yielding higher amplitudes for incorrect completions compared to the correct completion. These results are in line with previous work showing the sensitivity of the N400 to predictive properties, such as lexical violations like unexpected words (Federmeier & Kutas, 1999). Thus, regardless of the root of idiomatic processing advantages, adult native speakers consistently show evidence of prediction during idiomatic processing.

4.1.1.1 Predictive idiomatic processing in language learners

Most comparable research on predictive processing of idioms in language learners has been conducted with proficient non-native speakers. However, the results of these studies appear to be mixed. In several studies, the same advantages for formulaic sequences or idiomatic phrases compared to novel phrases, such as faster reading times and fewer fixations during the phrase and on terminal words, have also been present when compared directly to native speakers (e.g. Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Underwood et al., 2004). Conklin and Schmitt (2008) tested both figurative and literal uses of idioms in a self-paced reading study on L1 and L2 readers and found a speed advantage for formulaic phrases compared to novel phrases in both reader groups. In direct

contrast to these findings, Siyanova-Chanturia, Conklin, and Schmitt (2011) found differences between L1 and L2 readers in eye-movements depending on the phrase-type; unlike in their L1 reader group, idiomatic readings did not produce any speed or fixation differences from novel phrases. Furthermore, idiomatic readings produced a processing cost for L2 readers not present in L1 readers. However, the results of the study conducted by Siyanova-Chanturia, Conklin, and van Heuven (2011) suggest that the inconsistencies found between studies may be due to experience with the language. Namely, in their study on frequency in binomial processing (e.g., *bride and groom* vs. *groom and bride*), phrasal frequency interacted with proficiency in a manner suggesting that only known, memorized forms can be subject to phrasal frequency effects and this experience is necessary for processing effects to be visible in L2 readers.

Beyond the limited L2 research on the possible advantages of idiomatic and formulaic phrases, there is little research directly on L2 prediction in idiomatic processing. It's clear, however that L2 learners are also able to predict during processing, though these processes may be more limited than in L1 speakers (see Kaan, 2014 for an overview). For example, there is growing evidence that L2 speakers use contextual information to form predictions of upcoming words differently from L1 speakers in literal language (e.g., Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013; Ito, Martin, & Nieuwland, 2017). For instance, Ito, Martin, and Nieuwland (2017) conducted an ERP study to test prediction of form and meaning in highly predictable sentence-endings (e.g., *The student is going to borrow a…book.*) on Spanish-English bilinguals. While the authors found reduced N400 amplitudes for words semantically related to predicted words, this reduction was not dependent on predictability. However, as also suggested by the reading and eye-tracking studies discussed previously, proficiency may modulate prediction abilities rather than distinguishing them from native processing. More studies looking at prediction in idioms are needed in order to compare L1 and L2 predictive processes in idiom comprehension. To our knowledge, there are not yet comparable eye-tracking or ERP studies on prediction within idioms in the L2.

For L1 language learners, there is some evidence that children show a processing advantage for idioms over novel expressions. In a study by Qualls, Treaster, Blood, and Hammer (2003), 10-year-old children read correct idioms with high, moderate, or low familiarity. These were compared to control phrases that were manipulated idiom forms in which either the initial or the final word was replaced with an unrelated word (e.g., *<u>put/stomp</u> their heads together; go around in <u>circles/trouble</u>*). For each sentence, children had to judge whether it was an idiom or not. Children showed shorter response latencies for idiomatic than for control expressions but only when the final word was substituted. Furthermore, children showed shorter response times for highly familiar idioms reflecting the L1 adult data showing that higher familiarity leads to faster processing (Carrol & Conklin, 2019). Thus, similar to second language learners, children might show a stronger processing advantage if they

are familiar with the canonical form of the idiom. Taken together, idioms also seem to have a privileged status for children, but direct evidence for predictive processing within idioms is missing.

L1 language learners are also able to perform predictions about upcoming words in literal language (e.g., Mani & Huettig, 2012, 2014). Mani & Huettig (2014) conducted a visual world experiment in which children listened to sentences like *the boy eats the big cake* while at the same time seeing pictures of eatable and non-eatable objects on the screen. Children showed more fixations towards eatable objects shortly after hearing the verb *eat* (Mani & Huettig, 2014). In another experiment, Mahler and Chenery (2019) looked at naming latency. Children were presented auditorily with high and low cloze sentences (*The dog buried the bone/stick*) and had to repeat the final word. They were faster at repeating words in high (*bone*) compared to low cloze words (*stick*) suggesting a benefit for predictable items. Still, more evidence is needed to explain predictive processing mechanisms in children, but overall research supports general predictive abilities.

4.1.2 Literal activation in idiomatic constituents

When idioms are highly predictable, and an idiomatic meaning is activated, ongoing literal computation and therefore literal word activation seems unnecessary. However, evidence has repeatedly shown that online literal activation occurs in adult native speakers. In addition to consistent semantic priming effects for literal constituents of idioms even in the presence of figurative meaning activation (e.g. *kick the bucket* primes PAIL, Cacciari & Tabossi, 1988; Beck & Weber, 2016a), eye-tracking also provides further evidence of such activation. For example, Holsinger (2013) tested literal word activation in a visual world paradigm. While listening to idioms (*kick the bucket*), participants saw four words on the screen: one related to the figurative meaning of the idiom (DEATH), one literally related to an idiom constituent (FOOT) and two unrelated distractors (TRIANGLE, ANIMAL). Despite increased fixations towards the figuratively related word, participants also showed more fixations towards literally related words compared to unrelated distractors. Thus, eye-tracking also shows evidence of literal constituent meaning activation.

With regard to literal word activation during processing of highly predictable idioms, evidence is more mixed. In a production study, Sprenger, Levelt, and Kempen (2006) found that speakers preparing to produce an idiom-final word (e.g., English translation: *Jan walked against the lamp*) produce words semantically related to the respective idiom-final word (*candle*) faster than unrelated words. In contrast, there is some evidence that literal constituents are not always activated during idiom processing. Rommers, Dijkstra, and Bastiaansen (2013) conducted a semantic expectancy paradigm using ERP in which the final word of Dutch idioms in biasing contexts was highly predictable, but the idiom was presented either with its correct completion or replaced by either a word semantically related to the correct completion or an unrelated word (e.g., English translation: *After many transactions the*

careless scammer eventually walked against the LAMP/CANDLE/FISH yesterday.). Neither the N400 nor the P600 effect showed sensitivity towards semantic relatedness, which suggested that literal activation of the final constituent did not occur. In contrast, a recent study of native German speakers using a similar paradigm by Kessler et al. (submitted) challenges this finding when testing auditorily instead of visually presented idioms (*Hannes let the cat out of the BAG/BASKET/ARM*). The authors found evidence of semantic expectancy within idioms as indexed by reduced N400 amplitudes for related compared to unrelated words in neutral contexts during listening. Thus, context as well as presentation mode (auditory vs. visual) may have played a role in the results. In a second experiment, Kessler et al. (submitted) adapted the paradigm used by Holsinger (2013) and tested semantic expectancy in a visual world paradigm. Participants listened to idioms missing the final, highly predictable constituent word while viewing the four possible words adapted from earlier studies on the screen. Increased fixations towards the correct completion around 460 ms prior to the offset of the acoustic stimulus indicated prediction of the correct idiom form. Simultaneously, there were more fixations to related than to unrelated distractors. Therefore, in spite of somewhat mixed evidence, literal word meaning is often co-activated during predictive processing after recognition of the idiom.

4.1.2.1 Literal constituent activation in language learners

For language learners, it is possible that literal meaning has a different status than figurative meaning than for adult speakers of a language. And because of possibly different acquisition mechanisms involved (Arnon & Christiansen, 2017), L1 and L2 language learners might also contrast in their reliance on individual constituents and thus, their literal word meaning activation. Although research with children has focused on the understanding and interpretation of figurative expressions (e.g., Bernicot, Laval, & Chaminaud, 2007), the topic of literal meaning activation has not yet been explored. However, this topic has been widely addressed in L2 learners.

For non-native speakers, a possible difference between L1 and L2 in literal activation may be its status as highly salient in comparison to figurative language as a whole (e.g., Cieślicka, 2006; Giora, 1997). This line of research suggests that not only does literal activation occur even when a figurative meaning is likely (i.e., in a biasing context), but activation of literal word and even phrasal meaning is obligatory for idioms as a general processing strategy (i.e., for *kick the bucket*, both the word BUCKET and the literal phrasal meaning of striking a pail with one's foot are activated automatically). In L2 cross-modal priming studies, both Cieślicka (2006) and later Beck and Weber (2016a) found that the literal meaning of idiom-final words was activated (e.g., *kick the bucket* activated PAIL), and these priming effects appeared to be stronger than the priming of the figurative meaning (e.g., DIE). However, Cieślicka (2006) did not include native speakers as a control group in her study. In contrast, Beck and Weber (2016a), and an additional study also using both speaker groups by van Ginkel and Dijkstra (2019) confirmed that there were not differences between L1 and L2 activation of literal

meaning. The studies in **Chapter 5B** and **5C** also look into the special status of literal meaning in L1 and L2 speakers on a phrasal level in two self-paced reading studies and do not find clear evidence that literal meaning is obligatorily computed in L2 speakers compared to L1 speakers. Thus, while L2 speakers do activate literal constituent meaning during processing, it's unclear whether literal constituent words play more importance in L2 compared to L1 processing.

4.1.3 Research questions

We conducted two visual world eye-tracking experiments using L1 and L2 language learners to address the questions:

(1) Do L1 and L2 learners predict words within an idiom during listening?

(2) Are literal words activated during this process, and if so, when?

By looking at both L1 and L2 language learners of German, we aim to fill in largely unexplored gaps on idiom processing in the progression of language development and look at processing differences that might result from different acquisition mechanisms.

In the current study, we use the visual world paradigm to answer our research questions. The visual paradigm is suitable in answering both research questions because it is sensitive towards predictive mechanisms (Altmann & Kamide, 1999, 2007; Kamide, Altmann, & Haywood, 2003)a s well as semantic competition (Huettig & Altmann, 2005). Furthermore, this paradigm has been previously used to study lexical access in idioms (Holsinger, 2013; Kessler, Weber & Friedrich, submitted). Literal word activation can be measured by looking at co-activation of semantically related words. Here, we used the experimental design of Kessler, Weber, and Friedrich (submitted): participants listened to neutral sentences containing idioms without their final word (e.g., German: *Hannes ließ die Katze aus dem …*/English translation: Hannes let the cat out of the…). We chose idioms in which the final word was highly predictable. At the same time participants viewed words displayed on the screen (German/English translation) that were (a) correct completions of the idiom (*SACK*/BAG), (b) distractors semantically related to the correct completion (*KORB*/BASKET), and (c) two unrelated distractors (*ARM*/ARM, *BAUCH*/STOMACH).

In answering question (1), we hypothesized that prediction will be seen in many and early fixations towards the correct completion that is not presented auditorily. If evidence of an idiomatic processing advantage for both L1 and L2 listeners extends to predictive processes, then we should find looks to the correct completion before or around the offset of the final auditorily presented word. Concerning question (2), increased fixations towards related compared to unrelated distractors will index effects of semantic competition and thus, literal word activation. We expect the results of previous priming experiments to extend to the current experiment, and L2 participants are expected to

look more towards related compared to unrelated distractors. If there is a literal priority for L2 listeners, the amount of these fixations may be greater than in the L1 listeners groups examined in Keßler and colleagues (submitted). As there is no evidence to the contrary for L1 language learners, we expect that they, too, will activate the literal meaning of idioms. Though, if only L2 learners show activation of the literal meaning, this would imply that L1 and L2 learners store and access idioms differently.

4.2 EXPERIMENT 1: L1 LANGUAGE LEARNERS

4.2.1 Methods

4.2.1.1 Participants

Twenty-six 7th graders (17 female; mean age = 12;8 years) recruited from local schools participated in the experiment. L1 learners from this age group were chosen based on (1) the high probability of exposure to the idioms included in the study while (2) still early enough in language development to still be considered learners. Parents gave informed consent prior to the experiment. All participants were native, monolingual speakers of German and had no history of hearing disorders with normal or corrected to normal vision. As compensation, the children received a voucher for a local toy store. The design of the study was approved by the local ethical committee (reference number: 2016/1027/22).

4.2.1.2 Materials

Experimental items from Kessler, Weber & Friedrich (submitted) were used in the current study. These items consisted of twenty well-known German idioms (see Appendix C) embedded in sentences using the following structure (see Table 4-6): (i) a person carrying out the action of the sentence, (ii) a sentence body that originated from a German idiom and (iii) the final target word of the idiom (which was not presented auditorily). Sentences were recorded using a native speaker of German, and the final target word was cut from the recording. Participants heard each idiomatic sentence once, during which four visual words were presented on a computer screen. The words presented visually were categorized as: (1) Correct Completion: correct completion of the idiomatic phrase, (2) Related Distractor:a word semantically related to the correct completion, (3&4) Distractors Unrelated 1 and Unrelated 2: words semantically unrelated to the correct completion. Unrelated 1 and Unrelated 2 words were matched word pairs from Correct Completions and Related Distractors used from other experimental sentences in the experiment (avoiding phonological overlap). All presented words matched the grammatical gender expected from the preceding sentence context.

Table 4-6. German example sentence with English equivalent

(i) Person	**(ii) Sentence Body**	**(iii) Target words**			
		(1) Correct	**(2) Related**	**(3) Unrelated 1**	**(4) Unrelated 2**
Hannes	ließ die Katze aus dem	Sack	Korb	Bauch	Arm
Hannes	*let the cat out of the*	*bag*	*basket*	*stomach*	*arm*

Semantic relatedness between correct and related words was confirmed by comparing semantic spaces using the R package LSAfun (Günther, Dudschig, & Kaup, 2015). The semantic similarity between correct and related words was significantly higher than between correct words and both unrelated words (Wilcoxon signed rank test: unrelated1 $Z = 189$, $p < .001$; unrelated2 $Z = 185$, $p = .002$). There was no difference in semantic similarity between correct words and both unrelated words ($Z = 75$, $p = .28$). Target words were visually presented in white font (Arial, font size 28) on a grey background (see Figure 4-4), and the position of displayed words was counterbalanced across items and participants. The inter-trail interval was 1500 ms, and the presentation of a fixation cross for 500 ms marked the beginning of a new trial. Next, the set of four words were displayed on the screen, and the presentation of the audio stimuli began 2150 ms later. The order of the trials was randomized.

Figure 4-3. Example of visual display and auditory stimuli used in the experiment

4.2.1.3 Procedure

The experiment was conducted either at local schools or in the lab of the University in a single session. Participants were given both written and oral instructions for the experimental task. They were instructed to sit in a relaxed position and fixate on the screen during the task. The auditory stimuli were presented via closed headphones. Participants were asked to listen to each sentence and decide which of the visually presented words was the best completion for the idiom by saying their choice out loud. Participant responses were noted by the experimenter. After their oral response, participants were instructed to press a button in order to continue to the next trial. Prior to the experimental task, a 5-

point eye calibration was carried out for each participant followed by a practice block consisting of five trials. Fixations were recorded using a portable Tobii eye-tracker with a sampling size of 60 Hz. The eye-tracking experiment took around 20 minutes in total including instructions, eye-calibration and the experimental task; which lasted about 10 minutes on its own.

4.2.1.4 Analysis

In order to analyze fixations on the four visually presented words, we divided the screen into four areas of interest (AOIs). The time window of the analysis was aligned to the offset of each audio stimulus (offset = 0 ms). Only items with correct responses were included in the analysis (i.e. trials in which the participants chose the correct completion of the idiom). We assume that correct answers indicate knowledge of the idiom forms. Values of the two unrelated distractors were averaged. We used the fixation data to conduct two analyses to determine 1) the point at which participants reliably looked to the correct response, and 2) the time-course of fixations to each response type. In our first analysis, we conducted running t-tests ($p < .01$) comparing fixations towards correct completions and unrelated distractors at succeeding measurement points (every 16.67 ms resulting from 60 Hz sampling rate of the eye-tracker). These tests were used to determine whether anticipation occurred by identifying a point of divergence between fixations on correct completions compared to unrelated distractors. In the second analysis, we conducted a Growth Curve Analysis (GCA) with orthogonal polynomials (Mirman, Dixon, & Magnuson, 2008) to compare the amount and time-course of fixations towards Distractor Types (related and unrelated). As the starting point of the GCA time window, we chose the start of the anticipation for a duration of 1200 ms.

4.2.2 Results

In 90.72 % of the trials, participants responded correctly. In our first analysis, a running t-test showed that fixations on correct completions and unrelated distractors diverged around 128 ms before the stimulus offset and therefore indicate anticipation. The overall fixation pattern is displayed in Figure 2, Panel (A). The vertical, dashed line visualizes the start of the anticipation. In our second analysis, we modeled fixation patterns for related and unrelated distractors with third-order orthogonal polynomials. The modeled data are also depicted in Figure 2, Panel (B). Estimated parameter terms of Distractor Type are summarized in Table 4-7.

Table 4-7. Parameter estimates for the model including distractor type (Related vs. Unrelated)

Term	Estimate	Standard Error	*t*	*p*
Intercept	-0.02	0.00	-4.78	.00
Linear	0.02	0.03	0.80	.42
Quadratic	0.03	0.00	6.35	.00
Cubic	-0.01	0.00	-2.93	.00

A significant effect can be found on the intercept that indicates more fixations on related than unrelated distractors. The quadratic term indicates that the curves for both distractors differ in their central inflection. Thus, the peak for unrelated distractors is more negative than for related distractors. The effect on the cubic term suggests that towards the extremities of the model, inflections for both distractors differ, resulting in convergence of both curves. Together, these curves indicate that while participants generally fixate more on related distractors overall, this difference is largest over central parts of the model.

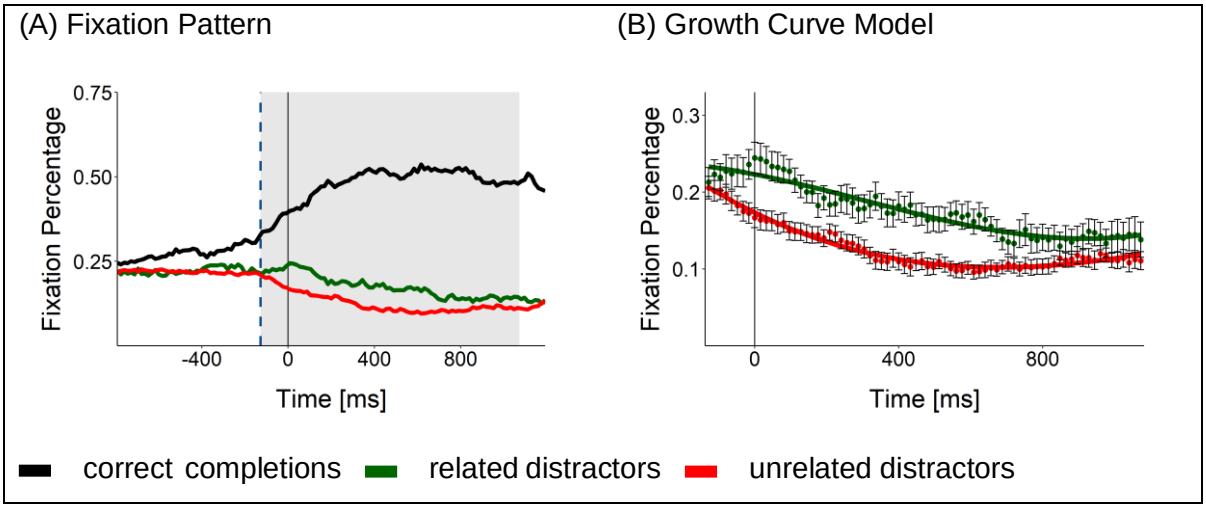


Figure 4-4. Experiment 1 fixation patterns

Note: (A) Fixation patterns for L1 Children: Fixation Percentage for Correct Completions (black), Related Distractors (green) and Mean of Unrelated Distractors (red); eye fixations are aligned to the offset of the acoustic stimulus (0 ms); start of anticipation (dashed, blue); time window for GCA (grey). (B) Fixation Percentage for Related and Unrelated Distractors (points = mean; error bars = standard error) with fit of growth curve model (line).

4.2.3 Discussion

In Experiment 1, we tested idiom processing in first language learners (7th graders). They showed relatively good performance on idiom recognition as displayed by correct responses in more than 90 % of the trials. This recognition of the idiom final word was also observable in predictive fixations of the correct completion. Thus, children not only show a processing benefit for the recognition of highly familiar idioms (Qualls et al., 2003) but also fast online prediction of idiom final constituents. This is in line with other studies showing the abilities of children to predict upcoming words in sentences (e.g., Mani & Huettig, 2012, 2014). However, unlike previous studies, the current study examines highly predictable idioms rather than using semantic cues for prediction (e.g., EAT increases looks to CAKE), and indicates an additional type of prediction in L1 children. We interpret these results as positive evidence for our first research question concerning prediction during listening, as seen via early anticipation of the idiom-final word.

Regarding lexical activation, children indeed show more fixations on related than unrelated distractors, and we interpret this as positive evidence for literal activation in L1 child listeners. Lexical activation seems to be strongest shortly after anticipation of the correct idiom final word.

4.3 EXPERIMENT 2: L2 LANGUAGE LEARNERS

4.3.1 Methods

4.3.1.1 Participants

Thirty-three non-native speakers of German (English L1, 12 female; mean age = 25.7 years) were recruited from the University of Tübingen for the experiment. Participants were living in Germany at the time of the experiment and were asked only to participate if they had at least a B2 level of proficiency ("Common European Framework of Reference for Languages," 2001). Of the 33 participants, only 26 were included in the analysis because five participants fell below the criteria for appropriate language proficiency based on task performance (correct responses < 40 %), LexTale score (scores < 50/100; see Lemhöfer & Broersma, 2012), and the language background questionnaire (years of study < 2, average self-rated proficiency < 3 on a scale of 1-7). The 26 participants included in the analysis had an average of 5.05 (*SD* = 2.34) years of German instruction, an average LexTale score of 68.62 out of 100 (*SD* = 9.39), and an average self-reported proficiency (averaged across reading, writing, speaking and listening) of 5.01 (*SD* = 0.94) on a scale of 1-7. Participants had no history of hearing disorders and normal or corrected to normal vision.

4.3.1.2 Materials

The same materials from Experiment 1 were used in Experiment 2.

4.3.1.3 Procedure

The procedure was the same as in Experiment 1 with the addition of a LexTale vocabulary test (Lemhöfer & Broersma, 2012) prior to the eye-tracking task and a language background questionnaire at the end of the experimental session. These additional tasks were used as measures of language proficiency. The experiment lasted about 30 minutes.

4.3.1.4 Analysis

The analyses were conducted following the same procedure as Experiment 1. Again, only items with correct responses were included in the analysis, and fixations were used to determine 1) the point at which participants reliably looked to the correct response via t-tests, and 2) the time-course of fixations to each response type via Growth Curve Analysis.

4.3.2 Results

On average, participants responded in 64.81 % of the trials with the correct completion of the idiom. The average responses to the related competitor were 17.88 %. In our first analysis, a running t-test revealed a reliable difference in fixations towards correct completions and unrelated distractors starting around 48 ms prior to the stimulus offset. The overall fixation pattern is displayed in Figure 3, Panel (A). The vertical, dashed line visualizes the start of the anticipation. In our second analysis, differences between fixation patterns for related and unrelated distractors were again modeled using a Growth Curve Analysis (GCA) (Mirman et al., 2008) with third-order orthogonal polynomials because by visual inspection of the time-course, two bendswere observable in the curve. The modeled data are also depicted in Figure 4-6, Panel (B). Estimated parameter terms of Distractor Type are summarized in Table 4-8.

Table 4-8. Parameter estimates for the model including distractor type (Related vs. Unrelated)

Term	Estimate	Standard Error	t	*p*
Intercept	-0.02	0.00	-4.16	0.00
Linear	0.04	0.03	1.06	0.29
Quadratic	0.05	0.01	8.39	0.00
Cubic	-0.07	0.01	-11.37	0.00

There was a significant effect of Distractor Type on all terms but the linear term (see Table 4-8). The main effect of Distractor Type indicates overall more fixations towards related than unrelated distractors. The quadratic term reflects that the curves for both distractors differ in their inflection around the center. Together with the depiction of the model in Figure 4-6, this result indicates that the

curve for related distractors peaks higher than for unrelated distractors. Significant effects on the cubic term reflect differences at the extremities of the curve.

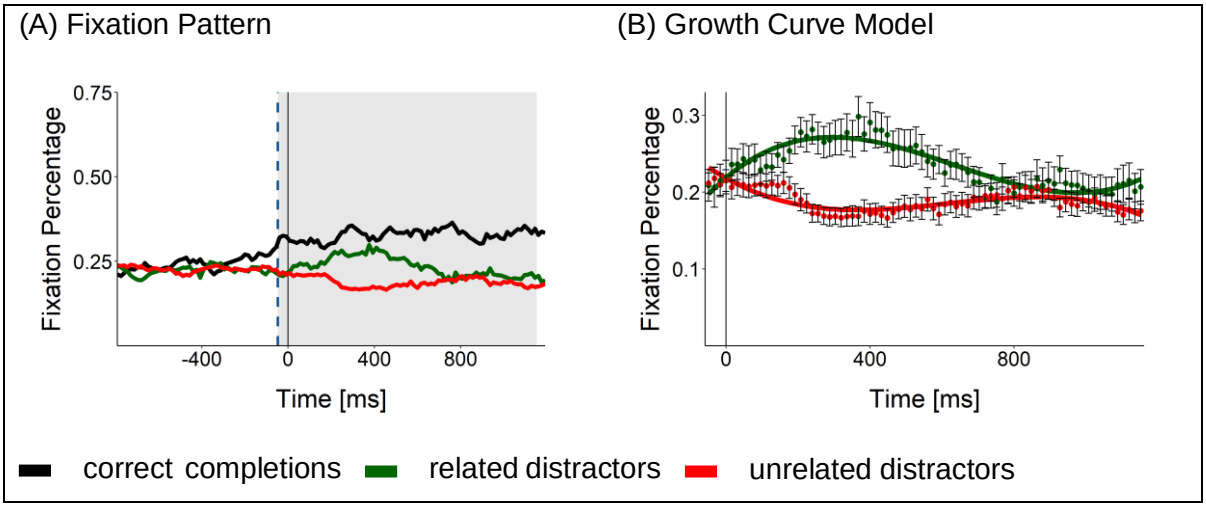


Figure 4-5. Experiment 2 fixation patterns

Note: (A) Fixation patterns for L2 Adults: Fixation Percentage for Correct Completions (black), Related Distractors (green) and Mean of Unrelated Distractors (red); eye fixations are aligned to the offset of the acoustic stimulus (0 ms); start of anticipation (dashed, blue); time window for GCA (grey). (B) Fixation Percentage for Related and Unrelated Distractors (points = mean; error bars = standard error) with fit of growth curve model (line).

4.3.3 Discussion

In Experiment 2, we tested idiom processing in second language learners of German (English L1). Their performance, averaging about 65 %, reflected moderate figurative language proficiency, which was lower than the proficiency of the L1 learner group in Experiment 1. However, when idioms were recognized, there was also evidence of prediction. Although looks to the correct completions only occurred 48 ms before the offset of the phrase, the programming of a saccade following a critical word takes approximately 200 ms (Saslow, 1967), a delay which suggests that recognition occurred prior to the anticipation point of 48 ms identified here. These results are in line with studies suggesting that there may be processing advantages associated with idioms, even for non-native learners (e.g., Conklin & Schmitt, 2008; Underwood et al., 2004). We interpret these results as positive evidence for our first research question, namely that L2 learners exhibit predictive processing of idioms during listening.

Concerning literal activation, more fixations on related distractors compared to unrelated distractors suggest that, like L1 learners, L2 learners activated the literal meaning of constituent words. This pattern of results is consistent with literature suggesting strong activation of literal constituents in L2 idiom processing (e.g., Beck & Weber, 2016a, 2016b; Cieślicka, 2006), especially considering that this literal activation continues to increase even after correct recognition of the final constituent. While

this does not provide evidence of an L2 literal priority directly (e.g., Cieślicka, 2006), it is further confirmation of the importance of literal word meaning even in figurative phrases. It should also be noted that activation occurs even without the auditory cue of the final word, suggesting that this activation is also predictive in nature and strengthens our conclusions on prediction in L2 listening. Overall, these results indicate that L2 learners also display patterns of predictive processes and semantic activation early on in idiom processing.

4.4 GENERAL DISCUSSION

Overall, the results from Experiments 1 and 2 show evidence that both groups of L1 and L2 language learners are able to predict idiom-final constituents during listening in addition to activating the literal meaning of these constituents, yet there were differences related to language proficiency when comparing both groups to native adult speakers. Regarding the task itself (i.e. identifying the correct completion of each idiom), we obtained graded effects in accuracy and timing of eye movements. In a previous study, adult native speakers determined idioms' final word at ceiling level with 99 percent correct responses (Keßler et al., submitted). In the present study using the same materials and procedures, L1 language learners determined 91 percent of the idiom completions correctly, while L2 language learners only determined 65 percent correctly. Similarly, overall looks to the correct response were higher for the L1 language learners than for the L2 language learners (see Figures 4-5 and 4-6 Panel A). Furthermore, adult native speakers directed their gazes earlier to the correct completion (ca. 460 ms before its onset) than L1 language learners (ca. 130 ms) and L2 language learners (ca. 50 ms), respectively. Graded accuracy in fixating and recognizing the correct idiom completions might simply reflect that adult native speakers had the most experience with the language, and L1 children had more experience than our L2 learners did. Differences in proficiency and therefore experience and familiarity with respective idioms, may impact stored frequency information (e.g., Lee, Hsin-Yi Lu, & Garnsey, 2013; Qualls et al., 2003; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tabossi et al., 2009) resulting in differences in the efficiency of predictions based on the same phrases.

Like native adults (Keßler et al., submitted), language learners appeared to activate literal constituents of idioms for predictive processing. All three groups showed more fixations towards related distractors than compared to unrelated distractors. While predicting the correct completion *bag* of the idiom onset *to let the cat out of the …,* participants' fixations were biased towards *BASKET* (being a semantic associate of *bag).* They appeared to predict single word constituents, which made spreading activation to semantic associates of these single constituents possible. Also similar to the timing obtained for fixations to the correct completions, the timing of fixations towards related distractors (relative to the anticipation of the correct completion) was graded between the three groups.

Again, this outcome might relate to differences in proficiency. It might suggest less rapid spreading activation in language learners compared to proficient native adults (e.g., Gernsbacher & Faust, 1991).

Looks towards related distractors also revealed qualitative differences in processing between L1 and L2 speakers of German. While looks to semantically related competitors linearly decreased in time in L1 learners (see Figure 4-5, Panel B) and in L1 adults (Keßler et al., submitted), they initially increased for L2 learners (see Figure 4-6, Panel B). In terms of qualitative differences, it has been suggested that literal meaning play a more important role in L2 idiom processing compared to L1 idiom processing (e.g., Cieślicka, 2006; Giora, 1997), though evidence has been mixed before this study (e.g., Beck & Weber, 2016a; van Ginkel and Dijkstra, 2019). Responsible for the current result may indeed be an underlying learning mechanism that is more additive in nature for L2 learners compared to L1 learners (Arnon & Christiansen, 2017), which might account for a stronger reliance on constituent meanings. According to this line of argumentation, increased looks to semantically related competitors in L2 might reflect obligatory activation of literal meaning rather than a more automatic or shallow activation that may be displayed in both L1 groups (e.g., Peterson, Burgess, Dell, & Eberhard, 2001). Following this line of argumentation, single word meanings play a different role in online processing.

Alternatively, however, this outcome might also still be a result of graded language proficiency. Participants of the least proficient group (L2 learners) might either be less confident in their choice of the idiom-final word and/or weaker in inhibiting automatic activation of semantic associates. Furthermore, increased looks to the related distractors may be a result of second-guessing their given responses as these choices sometimes reflect a likely response (e.g., a cat may be let out of a BAG/*SACK* or a BASKET/*KORB*). In the case of the latter possibility, the visual presence of both the correct final constituent as well as its semantic associate may cause a greater challenge for non-native speakers, who are less skilled at inhibiting related but inappropriate information (e.g., Gernsbacher & Faust, 1991). As the final idiomatic constituent as well as its related distractor were available on the screen during the listening process, the current study is limited in its ability to clearly tease apart those alternative possibilities.

In conclusion, our data shows evidence that L1 and L2 language learners make use of predivtive processes during listening and activate the meaning of literal constituents. Whereas prediction indicates that both sets of learners interact with the idiom in some manner as a unit, literal constituent activation indicates that the parts of the unit still remain available. Differences in timing between these groups and adult native speakers (Keßler et al, submitted) suggest that proficiency may play a role in both the timing of anticipation during listening as well as the pattern of literal activation. Additionally, in line with growing evidence that L1 and L2 idiomatic processing are similar (e.g., Beck & Weber, 2016a, 2019; Conklin & Schmitt, 2008; Tabossi et al., 2009; van Ginkel & Dijkstra, 2019), we propose that the differences found between our L1 and L2 learners here may be an interaction

between proficiency and L2-specific challenges rather than fundamental differences in storage and processing. Future studies must include more varied proficiencies in learner groups in order to better understand how prediction and activation evolves and use additional methods such as ERP (e.g., Keßler et al., submitted) in which literal activation can be more clearly separated from other possibilities such as spreading activation by excluding target words from the experimental trials.

5 CONTEXT AND ACCESS TO FIGURATIVE AND LITERAL MEANING

Outside of experimental research on idiom processing, idioms occur in rich contexts including situational, pragmatic, and linguistic, and meaning a debate between a possible literal and figurative meaning is unnecessary (see e.g., Gibbs & Colston, 2012). In psycholinguistic research, such as the current work, however, this context is limited for the purpose of experiments, and is often left out of the picture. This chapter explores the flexibility of access to figurative and literal meaning, with a broader focus on context. Context is used in a general sense to include both experimental or global context, as in 5A, as well as linguistic context that biases interpretations as in 5B and 5C. In addition to the impact of context, 5B and 5C consider an additional aspect, literality, that may interact with context in a manner that could affect meaning access and how language users make use of biasing context.

5A L2 IDIOM PROCESSING: FIGURATIVE ATTUNEMENT IN HIGHLY IDIOMATIC CONTEXTS[6]

Sara D. Beck & Andrea Weber

Abstract

Using cross-modal priming, we investigated the processing of idioms in non-native listeners in varying experimental contexts. As idiomatic processing models have presented evidence for an idiomatic mode of processing that can be activated for non-native speakers in highly figurative contexts (Bobrow & Bell, 1973), this experiment revisits those claims while also examining access to figurative meaning in addition to the literal meaning of individual words within an idiom. This experiment showed increased priming for visual targets related to the figurative meaning of an idiom when the experimental list contained a large proportion of idiomatic sentences compared to when the list contained only a small proportion of idiomatic sentences. Non-native speakers not only showed online access to figurative meaning but were also sensitive to highly idiomatic contexts; though, responses to the targets related to literal meaning of the final word of the idiom were faster in all instances than figuratively related targets.

[6] This chapter is adapted from the published version of the following article: Beck, Sara D.; Weber, Andrea (2016): L2 Idiom Processing: Figurative Attunement in Highly Idiomatic Contexts. In A. Papafragou, D. Grodner, D. Mirman, J. C. & Trueswell (Eds.): *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

5.1 INTRODUCTION

While understanding idioms is *a piece of cake* for native speakers of English, non-native (L2) speakers often struggle to recognize and understand them. Not only is figurative language extremely prevalent in everyday English use, but idioms are among the most frequent figurative expressions used by native speakers and an integral part of non-native language competence (see e.g., Cieślicka, 2006; 2013b). In defining idioms, researchers generally agree that 1) the meaning of an idiom often differs from the literal meaning of the words comprising the phrase, 2) idioms have fixed formulations or structures with limited variation, and 3) idioms are multi-word expressions (Liu, (2008). Although the challenges that these expressions pose for L2 learners are well-documented (see e.g., Cooper, (1999)), the underlying processes are still in need of research. While we know that highly proficient L2 listeners can have access to figurative meaning in some instances, it may not be the case for all idioms. The current experiment tested online processing of figurative and literal meaning for L2 listeners in two different contexts: a highly figurative and a less figurative context.

5.1.1 Native idiom processing

Idiom processing has been the subject of many L1 studies for quite some time. Idioms are a particularly interesting linguistic phenomenon since many allow for both figurative and literal interpretations. For instance, *a piece of cake* can refer literally to a slice of cake or it can be figurative and mean "easy." This aspect of idioms has been the basis of one of the most studied questions of idiomatic processing: How does the processing of figurative meaning compare to that of literal meaning? A number of models have been proposed to address this issue.

One of the first models of processing developed specifically to address idiomatic processing was proposed by Bobrow and Bell (1973). Like standard pragmatic models, the Idiom List Hypothesis assumes that figurative and literal meaning undergo separate processes and, in normal contexts, literal meaning has processing priority over figurative meaning. In the idiomatic mode of processing, idiomatic meanings are retrieved from a list and do not undergo the same composition that literal language does. Following contradictory psycholinguistic evidence, Swinney and Cutler's (1979) Lexical Representation Hypothesis proposed that simultaneous processing occurs; however, figurative meaning is accessed first due to lower processing costs. Both theories assume that idioms are stored as one unit and processed as long words; thus, figurative meaning need only be retrieved, while a literal phrase involves both retrieval and composition. An alternative proposed by (Cacciari & Tabossi, 1988), the Configuration Hypothesis, assumes that literal word meaning is processed until an idiomatic key in the idiom is reached. At this point, the configuration of words is recognized as an idiom, and the figurative meaning is accessed, giving priority at that point to figurative processing.

Literal-first models have been widely discounted based on the mounting psycholinguistic evidence that figurative meaning is often faster than not only literal meaning in a non-biased sentential context (e.g., Cacciari & Tabossi, 1988; Swinney & Culter, 1979) but also comparative novel phrases (e.g., Conklin & Schmitt, 2008; Tabossi, Fanari & Wolf, 2009). However, it is not clear whether processing occurs simultaneously as two separate processes, one process that differentiates at a recognition point, or even one process influenced by other individual idiomatic properties such as decomposability (see e.g., Gibbs et al., 1989), literal saliency (e.g., Cieślicka, 2006), frequency (e.g., Tabossi, Fanari, & Wolf, 2009), or many of these properties at once (e.g., Libben & Titone, 2008).

5.1.2 Non-native idiom processing

Much research on non-native processing of idioms focuses on both the comparison of access to figurative and literal meaning in addition to the comparison of idiomatic processing to novel language. Unlike L1 research, there is more variation in the access to figurative meaning and most L2 results focus on supporting or refuting the existing L1 models of processing. Researchers such as Conklin and Schmitt (2008), Underwood et al. Galpin (2004), and Siyanova-Chanturia, Conklin, and Schmitt (2011) investigated access to figurative meaning in comparison to novel phrases using eye-tracking methods during reading. Conklin and Schmitt (2008) found that idioms, examined in their research as a subset of formulaic language, were read more quickly than comparable novel phrases whether used figuratively or literally. Underwood et al. (2004) found a similar advantage based on the number of fixations, but not for total fixation length, suggesting a more complex picture of the processing of idioms and other formulaic language that also accounts for L2 disadvantages. However, neither make any claims about figurative access in comparison to literal access. Siyanova-Chanturia, Conklin, and Schmitt (2011), on the other hand, found no advantage for idioms compared to novel phrases in proficient L2 users and also found that the figurative meanings of idioms required more time to retrieve than the literal interpretation, which seems to support a literal-first model of processing for L2 users. Though, Siyanova-Chanturia, Conklin, and van Heuven (2011) suggest that these results may vary due to factors such as frequency and exposure.

Some L2-specific idiom processing models have been proposed in addition to the L1 models. The Idiom Diffusion Model of Second Languages (Liontas, 2002; 2015) proposes a two-stage comprehension model. The first stage involves prediction, eased by idioms which are the same in a learner's L1 and L2; the second stage consists of confirmation or replacement and/or reconstruction when necessary. Though a comprehension model, it suggests that processing is eased for translatable idioms, supported by an offline study from Irujo (1986) and a timed production task from Liontas (2002). In her Model of Dual Idiom Representation, Abel (2003a) proposes that the important factor for L2 processing is decomposability, or the relation of the individual constituents to the idiomatic meaning. This model assumes that non-decomposable and frequently encountered idioms are

represented by idiom entries in the mental lexicon, as in the Idiom List Hypothesis, while non-decomposable idioms are represented by lexical entries of the individual constituent words. Abel's model, however, is based solely on offline ratings. Finally, see Cieślicka's (2006) Literal Salience Model directly addresses idiom processing and is based on online data. The model suggests that literal meanings remain most salient for L2 users, even for well-known idioms, as they are more likely to be used and encountered by learners. Based on ideas presented by Giora (1997), salient meanings are accessed more quickly than non-salient meanings. The model is builds on findings from a cross-modal priming experiment that showed that access to literal meaning occurs prior to figurative meaning for L2 listeners. Using non-biasing sentences followed by literally and figuratively related targets, reaction times to literal targets were faster than figurative ones when compared to matched controls. However, see **Chapter 3** for more discussion on this model.

Research on L2 idiom processing is less developed than and lacks the quantity that L1 research has been afforded. While some evidence supports fast access to figurative meaning, the speed of access in comparison to novel or literal language is still inconclusive. And, like L1 research, it is unclear whether or not figurative language has its own mode of processing.

5.1.3 The current experiment

While the L1 Idiom List Hypothesis based on research by Bobrow and Bell (1973) has been dismissed, among other reasons (see e.g., Cacciari, 2014; Cacciari & Tabossi, 1988) based on false assumptions about slow access to figurative meaning in the absence of a biasing context, this model is not necessarily refuted by the literature for L2 users and still reflects some intuitions about the way we comprehend idioms. As Swinney and Cutler (1979) also observed, when an individual becomes aware of a highly idiomatic context, L1 listeners often become more attuned to figurative meaning occurring in natural communicative situations and might even fail to see the literal meaning of an idiom. The experiment from Bobrow and Bell (1973) presented highly literal idioms with the possibility of both literal and figurative interpretations following a biasing context containing several sentences with either literal-only or figurative-only interpretations. Participants were asked to note which interpretation—literal or figurative—they first perceived. Based on increased literal interpretations first following the literal-only contexts and figurative interpretations first following figurative-only contexts, Bobrow and Bell argued for two separate modes of processing. Following a highly figurative context, the figurative mode of processing becomes active and leads to deviation from a standard or normal literal-first mode of processing. What Bobrow and Bell did not consider, and what Swinney and Cutler's observations suggest, is that the adjustment observed might be due to a contextual adaptation rather than a mode of processing unique to idiom or figurative language processing.

The ability of L2 listeners to adapt to their environment in language comprehension is well-documented, and there is evidence for rapid attunement to varying linguistic situations. Listeners can adapt, for example, to speech rate, surrounding noise and idiosyncrasies of a speaker both in their L1 and their L2 (see e.g., Sebastián-Galĺes, Dupoux, Costa, & Mehler, 2000). L2 listeners are also able to quickly attune to changes in their environment such as surrounding noise (McQueen & Huetting, 2012) and foreign accents (e.g., Weber, Di Betta, & McQueen, 2014). In addition, sequential effects of item presentation can influence the listening process. Perea and Carreiras (2003) found that listeners are able to shift their response criterion on a trial-by-trial basis to adjust to the lexical status and frequency of a previous trial. While an offline, even conscious, adaptation to the presence of figurative meaning is a common intuition, this phenomenon has not been examined in on online setting, and it warrants further research to determine if a context-dependent shift based on the presence of figurative language occurs in L2 listeners—be it via a figurative mode of processing or a contextual figurative attunement.

The current experiment will revisit the idea of an idiomatic mode of processing for non-native listeners activated in a biasing experimental context using current psycholinguistic methods. We address the questions still left open in light of current research on idiom processing: Will a highly literal or idiomatic experimental context affect online processing of figurative meaning in L2 listeners as it appears to do in L1 listeners? If so, do the results support a unique mode of idiomatic processing or attunement to the figurative context?

5.2 METHOD

In the present English cross-modal priming study, based on the experimental design from **Chapter 3**, we presented idioms with a medium degree of literal and figurative interpretation (neither highly literal nor figurative based on L2 ratings) in one of two experimental contexts to German learners of English. Participants either encountered target idioms embedded in sentences among more sentences containing idioms and very few literal sentences or among only literal sentences.

Listeners were presented an auditory prime (the sentence containing the idiom) followed by a visual target. In a lexical decision task, German participants had to decide whether or not the visual target was a real word of English or not—reaction times (RTs) to targets are known to be faster when prime and target are semantically related compared to when they are unrelated. Facilitatory priming provides information about the processing of the auditory prime, and faster targets compared to their unrelated controls indicate the activation of meaning. For the current experiment, we were interested in both facilitatory priming for targets related to the figurative meaning of the idioms and targets related to the literal meaning of constituent words compared to matched controls. If the same kind of contextual attunement found by Bobrow and Bell (1973) applies to non-native listeners, then we would expect increased priming effects for figurative targets in an idiomatic experimental context compared

to the non-idiomatic experimental context. Additionally, access to literal constituent words can give us more insight into non-native idiom processing.

5.2.1 Participants

Eighty-one native speakers of German were paid a small fee to participate in the experiment. Participants were University of Tübingen students who identified themselves as skilled speakers of English. One participant was excluded as she reported that she was unable to hear the stimuli.

5.2.2 Materials

The auditory materials and sentences for this experiment were taken from **Chapter 3** (S. D. Beck & Weber, 2016a). 25 target idioms were embedded at the end of non-biasing short sentences and presented in one of two varying experimental contexts containing 100 trials. The two experimental contexts differed only in 75 filler items, specifically the amount of sentences with idioms in the fillers (explained below). The target idioms all had a VP syntactic structure and were controlled for familiarity, meaningfulness, literality, and translatability (English to German) as rated by L1 and L2 users in the DIN database (**Chapter 2**). Each trial consisted of an auditory sentence prime followed by a visual target. All sentence primes were paired with four different targets.

The four targets included literal or figurative targets and their respective unrelated control targets. Targets semantically related to the literal meaning were based on the last content word of the sentence and chosen from the Nelson et al. (1998) association norms database. For the idiom *to kick the bucket (*primed in the sentence *His uncle kicked the bucket.)*, the literal target PAIL was chosen. The unrelated control word was matched for orthographic complexity and length (BOAT as a control for PAIL). Targets related to the figurative meaning of the idiom were chosen based on relation to the overall meaning of the idiom. For *to kick the bucket*, the target DIE was chosen, as the overall meaning is "to die." Similarly, figurative control targets were also controlled for orthographic complexity and length (ZOO as a control for DIE). The four lists of targets were also controlled for lexical frequency. See Table 5-1 for reference. The auditory sentence primes remained the same for all four lists. Targets were equally distributed across 8 lists (all four targets in two experimental contexts) and presented 400ms after the offset of the final word in each sentence.

Table 5-1. Sample of experimental items

	Literal		**Figurative**	
Stimuli	Target	Control	Target	Control
His uncle *kicked the bucket*.	PAIL	BOAT	DIE	ZOO

The experiment was performed using Presentation® software (Version 17.2, www.neurobs.com). Experimental sentences were recorded by a female speaker of American English

(first author) in an experimental lab setting. The eight lists were randomly distributed among participants. Each list began with four practice trials followed by 25 experimental and 75 filler trials (a total of 100 experimental trials) and was presented to an equal number of participants. Conditions were evenly distributed across lists.

5.2.2.1 High-idiomatic context

This variation included a highly figurative context by increasing the amount of idiomatic trials to a total of 75. In addition to the 25 critical trials, 50 filler sentences also embedded idioms into the end of neutral sentences. Only 25 filler trials did not include idioms. Fifty of the filler trials contained non-word targets, and the other half contained word targets. The four lists in this variation differed only in experimental target words.

5.2.2.2 Low-idiomatic context

This variation kept the idiomatic context to the 25 experimental trials by including only literal filler items. Half of the trials contained non-word targets, and the other half contained word targets. The four lists in this variation differed only in experimental target words.

5.2.3 Procedure

Participants were tested individually in a quiet room. First, participants were given instructions in English on the lexical decision task. Participants were instructed that they would hear sentences directly followed by the appearance of a word or non-word on the computer screen in front of them. Participants were told to listen to sentences and then decide whether the string of letters on the screen was a word or not. Subjects were asked to make their decision by pressing a green button with their dominant hand for 'YES' and a red button with the other hand for 'NO' as quickly and accurately as possible. Participants were also instructed that it was important both to listen and to respond to the visual targets, as they would be asked about what they heard after experiment.

Once participants understood the instructions and answered the instruction questions correctly, they could participate in the priming study. The participants listened over closed headphones, and the visual targets were presented on a laptop screen. The targets appeared on the screen 400ms after the offset of the auditorily presented sentence. The next trial began 1000ms later. Reaction times were measured from the onset of the presentation of the visual stimuli. The experiment concluded with a short yes/no comprehension test on items and a language background questionnaire (see also Cieślicka, 2006). The entire experiment took about 15 minutes.

5.2.4 Analysis

Ten correct responses with RTs longer than 2000ms (0.5% of the total data) were considered outliers and were removed from these analyses. Additionally, three targets (*CHIME, SHRUB, YARN*) were

answered correctly by participants only 50% or less of the total trials and were excluded from the results (29 responses or 1.5% of total data). In total, 2.0% of the data were not included in these analyses.

Table 5-2. Mean reaction times (in ms)

	Variation 1 (75% Idioms)		**Variation 2** (25% Idioms)	
	Figurative	**Literal**	**Figurative**	**Literal**
Related	749	709	737	674
Unrelated	813	739	740	711

Analyses of Variance (ANOVAs) were conducted on correct responses of the remaining RTs across participants (F_1) and across items (F_2) to examine the within-subject effects of *figurativeness* (with two levels *figurative* and *literal*) and *relatedness* (with two levels *related* and *unrelated*) and the between-subject effect of *experimental context* (with two levels *high-idiomatic* and *low-idiomatic*) between-participants and between-items. Table 5-2 reports the mean RTs measured from target onset for each condition, and the corresponding priming effects for RTs are shown in Figure 5-1.

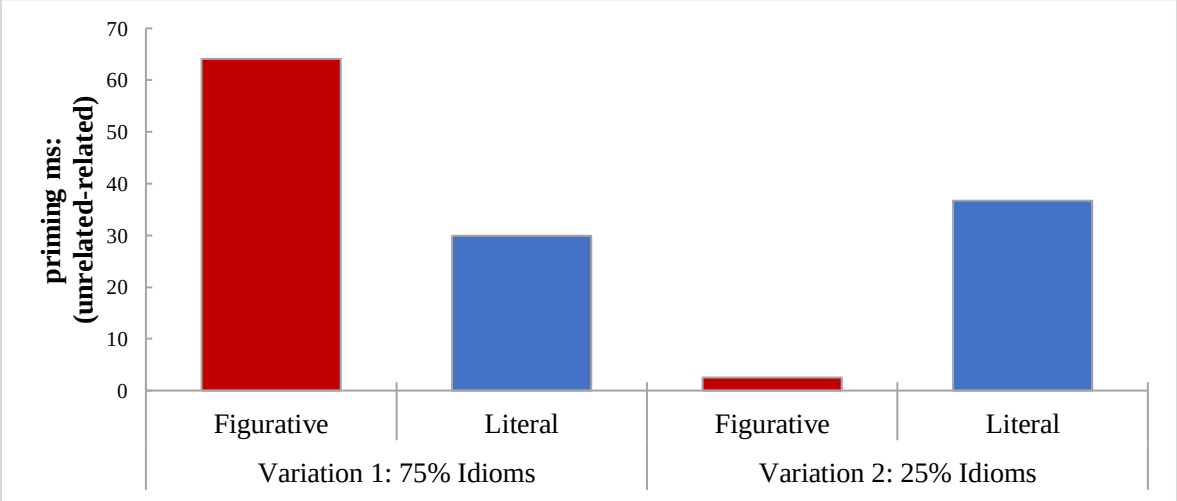


Figure 5-1. Priming effects (ms)

Note: The priming effect, graphed here for figuratively related targets (red) and literally related targets (blue) is calculated by subtracting the average mean reaction times for each condition from the unrelated target words.

Overall ANOVAS on RTs showed main effects of *figurativeness* (F_1[1,78]=20.29, p<.001; F_2[1,42]=11.41, p=.002), *relatedness* (F_1[1,78]=9.89, p<.01; F_2[1,42]=2.80, p<.05), and a weak

interaction between *figurativeness*, *relatedness*, and *experimental context* (F_1[1,78]=4.02, p<.05; F_2<1). No other interactions were significant.

To further explore the interaction in the RT analysis, separate analyses for *figurativeness* were conducted. For *literal* targets, there was a main effect of *relatedness* (F_1[1,78]=5.88, p<.05; F_2[1,42]=2.37, p>.05) and no significant interactions. RTs for literally related targets were faster than RTs for unrelated targets across both experimental contexts. Effects were more consistent across subjects than across items, suggesting variation between individual targets.

For *figurative* targets, there was a main effect of *relatedness* (F_1[1,78]=6.14, p<.05; F_2[1,48]=1.87, p>.05) and an interaction between *relatedness* and *experimental context* (F_1[1,78]=5.24, p<.05; F_2[1,48]=1.22, p>.05). Again, both effects are more consistent across subjects than items. RTs for targets figuratively related to the prime, as in the literal analysis, were faster than unrelated targets. Thus, we find facilitatory priming for relatedness in both figurative and literal targets.

In order to further explore the interaction in the *figurative* analysis, separate analyses were conducted for each *experimental context*. In the *high-idiomatic* context there was a main effect of *relatedness* (F_1[1,39]=10.58, p<.01; F_2[1,24]=2.75, p>.05). However, in the *low-idiomatic* context, no main effects were present. While facilitatory priming was present for figurative targets in the presence of the high-idiomatic context, this effect disappears in a more literal context for non-native listeners.

5.3 RESULTS

Though there was no main effect of experimental context, our results show that the experimental context significantly impacted facilitatory priming for figurative targets. As shown in Figure 51, the facilitatory priming for figurative targets (red bars) varies considerably from one variation to the other, while the literal targets (blue bars) show no significant changes. In the high-idiomatic context, a facilitatory priming effect for figurative targets of 64ms is observed compared to a non-significant 3ms in the low-idiomatic context. Figurative targets in this context represented the only condition in which relatedness was not a significant effect. These results provide evidence that even in a non-biasing sentence context, a highly figurative global environment can impact online processing of figurative meaning for idioms. This data is in line with the offline results from Bobrow and Bell for native participants (1973).

The priming effect for literally related targets showed less variation—an increase of only 7ms—suggesting that while processing for figurative targets was impacted, the processing of individual literal constituents was not significantly impacted. In the case of the low-idiomatic experimental context, we can also argue that where figurative meaning is not facilitated, literal meaning

is more dominant. This result also supports the L1 data collected by Bobrow and Bell (1973) as well as the L2 data from Cieślicka (2006) and Siyanova-Chanturia, Conklin, and van Heuven (2011).

Considering the impact of figurativeness, overall, literally related targets were faster than figuratively related targets. While this reflects the same results found by Cieślicka (2006), it does not necessarily imply a processing priority for literal meaning over figurative meaning. While the figuratively related targets correspond to the overall meaning of the idiom, the literally related targets correspond only to the literal meaning of the final constituent of the idiomatic phrase. It does, however, indicate that literal processing of constituents is present even when figurative meaning is processed. And, when figurative meaning is not facilitated by context, literal meaning appears to be dominant.

Our results are further supported by those gathered by Sprenger and colleagues (2006), who also found strong literal constituent activation in production tasks as well as the hybrid model for idiom production proposed in Sprenger et al. (2006). Our results differ, though, from Rommers, Dijkstra and Bastiaansen (2013). Rommers, and colleagues collected EEG data from a top-down procedure which indicated that related literal targets were not activated in highly predictable idioms when the final word of an idiom was replaced with the related target. Thus, while hearing a word might activate the literal meaning of that word, the functionality of these literal words may be limited or even switched off in some cases. However, it should be noted that this study used biasing contexts and highly predictable idioms, neither of which were part of the present study.

The consistency of effects—generally stronger across subjects than across items—supports general knowledge of variation across idioms (see e.g., Titone & Connine, 1994a) and possibly varying association strength of our targets, as these were selected by this author and confirmed only by ratings (see **Chapter 3**) rather than a database. While our idioms were controlled for much of this variation, the presence of remaining differences cannot be excluded.

5.4 DISCUSSION

Based on the results of this study, we will briefly consider the fit of our data with L1 and L2 models of idiom processing.

5.4.1 L1 Processing models

Our results are not compatible with the assumptions of stage models of processing. The Idiom List Hypothesis (literal-first) is problematic as our participants activated literal constituent meaning in addition to figurative meaning of the idiom in a high-idiomatic context. Additionally, while priming of figurative meaning increased in the high-idiomatic context, literal constituent meaning still precluded figurative meaning. This hypothesis predicts, however, that the literal meaning should not be activated at all in a high-idiomatic context. While our data does not rule out a second, idiomatic

mechanism that might be primed by a global figurative context, we argue that attunement or adaptation to this contextual setting is a more suitable explanation for this phenomenon. The Lexical Representation Hypothesis, assuming that figurative meaning is retrieved faster than compositional processing of literal meaning, is likewise not supported by participants' RTs in our experiment. However, the tenant of this model suggesting that composition and idiom retrieval can occur simultaneously cannot and should not be dismissed.

Furthermore, our data cannot make strong claims for or against the Configuration Hypothesis as we included only highly familiar idioms, and this model focuses on predictable idioms based on the recognition point of the idiom. Though familiarity generally correlates with predictability (Titone & Connine, 1994a), any further interpretation of compatibility would be far-reaching.

5.4.2 L2 Processing models

We can make limited claims about the previously discussed L2 processing models, but it seems that our results showing online access to figurative meaning are not compatible with the challenges many of these models present. As we used only non-translatable idioms, we would expect processing difficulties for our participants based on the Idiom Diffusion Model of Second Languages. Although we cannot compare our results with translatable idioms to make a stronger claim, the access to figurative meaning makes an overall case against it.

Our data is generally compatible with the Literal Salience Model. Our participants responded faster to literally related targets than to figuratively related targets as predicted by this model. However, the results solely from this observation are not compelling enough to warrant a strict interpretation that literal meaning has a priority over figurative meaning. Sprenger, Levelt and Kempen (2003) also found literal constituent priming in production tasks and claimed that, rather than a processing priority, activation of constituent word lemmas in addition to a superlemma, a phrasal representation of the idiom on a lexical-syntactic processing level, are activated. Though this production model does not clearly lay out a time course for processing, we assume that activation must spread to a superlemma from individual constituents, and it is possible that this activation occurs more slowly for L2 listeners than activation of an individual word lemma does. While intended to explain production, this model aligns well with the data presented here.

5.4.3 Figurative attunement

Our findings can generally be interpreted as compatible with current ideas on L2 listening adaptation. Based on the strong differences in processing of figurative language between the experimental variations, we argue that listeners are able to quickly adapt to a figurative context on the processing level.

5.5 CONCLUSION

Though our data does not provide strong evidence for or against idiom processing as its own mode of processing, separate from literal processing, it does provide strong evidence that proficient L2 listeners can have online access to figurative meaning. Additionally, the experiment supports the idea that figurative attunement is possible even in a very short amount of time, furthering evidence that L2 listeners can detect and shift their response-criterion in the presence of a highly figurative context.

5B CONTEXT AND LITERALITY IN L1 IDIOM PROCESSING: EVIDENCE FROM SELF-PACED READING[7]

Sara D. Beck & Andrea Weber

Abstract

In a self-paced reading study, we investigated how effects of biasing contexts in idiom processing interact with effects of idiom literality. Specifically, we tested if idioms with a high potential for literal interpretation (e.g., *break the ice*) are processed differently in figuratively and literally biasing contexts than idioms with a low potential (e.g., *lose one's cool*). Participants read sentences that biased towards a figurative or literal reading of idioms and continued with resolutions that were congruent or incongruent with these biases (e.g., [The new schoolboy/the chilly Eskimo] just wanted to *break the ice* [with his peers/on the lake] ...). While interpretations of high-literality idioms were strengthened by supporting contexts and showed costs for incongruent resolutions, low-literality idioms did not show this effect. Rather, interpreting low-literality idioms in a literal manner showed a cost regardless of context. We conclude that biasing contexts are used in a flexible process of real-time idiom processing and meaning constitution, but this effect is mediated by idiom literality.

[7] This chapter is adapted from the published version of the following article: Beck, Sara D.; Weber, Andrea (2020). Context and Literality in L1 Idiom Processing: Evidence from Self-Paced Reading. *J Psycholingusitic Res.* https://doi.org/10.1007/s10936-020-09719-2

5.1 INTRODUCTION

Idioms challenge standard notions of meaning composition as they are, by definition, multi-word strings with a figurative meaning that differs from the sum of its parts. For example, the idiom *to take the bull by the horns* means, figuratively, to take charge of a difficult situation, but a literal interpretation, denoting an event in which a male cow is grabbed by its horns, is also available following standard meaning composition. In recent years, a number of studies have shown that the availability of these two meanings can be influenced by linguistic context, and while there is a general consensus that context can facilitate access to the figurative meaning of an idiom (e.g., Fanari, Cacciari, & Tabossi, 2010), there is less agreement on the impact of context on the availability of the literal interpretation (e.g., Rommers et al., 2013; Holsinger & Kaiser, 2013). The current study adds to the body of literature examining contextual facilitation, by specifically considering the mediating effect of idiom literality in native speakers.

Idiom literality refers to the variation between idioms in their potential to be interpreted literally. While *to take the bull by the horns* can be used figuratively just as well as literally (i.e. has high literality), an idiom like *to be on cloud nine* has a low potential for literal interpretation in the absence of any contextual information (i.e. has low literality). Figuratively, it means to be very happy, and literally, a context is needed in order to create a situation in which a person might find themselves "on" a cloud called "nine." This variance in literality could mediate contextual effects on literal and figurative interpretations and particularly help to explain the varying effects of context on literal interpretations in previous studies. For high-literality idioms, supporting contexts could strengthen both the figurative and the literal interpretation, since both interpretations have a high potential to start with. For low-literality idioms, supporting contexts could still strengthen the figurative interpretation, but fail to affect the literal interpretation.

These assumptions were tested in a self-paced reading study on native speakers of English with idioms of varying literality. Following up on the methods used in a self-paced reading study conducted on phrasal verbs (Holsinger & Kaiser, 2013), our study examined idioms which were embedded into sentences containing literally or figuratively-biasing contexts and followed by resolutions that were congruent or incongruent with the bias (e.g., [The new schoolboy/the chilly Eskimo] just wanted to *break the ice* [with his peers/on the lake] ....). By comparing the costs and benefits of processing these expected and unexpected variations in sentences containing idioms of varying literality, our study looked closely at how literal and figurative meanings are integrated into sentential meaning in online processing.

While this study does not directly intend to test individual models of idiom processing, some of the various psycholinguistic models of idiom processing and their implications on the present

investigation will first be outlined in order to place it within this field of research before we discuss the relevant literature for the contributing factors: context and literality.

5.1.1 Idiom processing

Idiomatic models of processing have offered various accounts of the duality of meaning presented by idioms, and the role of literal meaning in idiom processing varies starkly among these models. Early models treated idioms as complex words for which the figurative meaning is directly retrieved from the lexicon. The consequence of such models is that literal meaning composition plays no direct role in accessing the figurative meaning of an idiom. The order of this composition and retrieval approach could be considered as either literal-first, as in the Lexical Access model (Bobrow & Bell, 1973); figurative-first, as in the Direct Access Hypothesis (Gibbs, 1980a); or parallel in both literal composition and figurative retrieval, as in the Lexical Representation Hypothesis (e.g., Swinney & Cutler, 1979). Of these models, only the Direct Access Hypothesis implies that literal composition need not occur. According to this hypothesis, idiomatic retrieval can begin immediately without considering literal composition. In the other two cases, literal composition occurs obligatorily, though the order, and thus predictions for speed of access, differs.

The stepwise nature of meaning access implied by such lexical theories of idiom retrieval, however, is not substantiated by psycholinguistic research. Figurative meaning has consistently been shown to be available online (e.g., Beck & Weber, 2016a; Cacciari & Tabossi, 1988; Libben & Titone, 2008) and sometimes even earlier than literal meaning (e.g., Cacciari & Corradini, 2015), but access to one interpretation does not necessarily preclude access to the other. Cacciari and Tabossi (1988) and, later, Beck and Weber (2016a) demonstrated that, in cross-modal priming experiments, the activation of individual literal constituents occurs quickly and in addition to the activation of figurative meaning when presented in non-biasing sentence contexts (e.g., *kick the bucket* actives DIE and PAIL). Though few studies have looked at both literal and figurative interpretations on a phrasal level, particularly without the impact of biasing context, Ortony, Schallert, Reynolds, and Antos (1978) found that literal interpretations were faster than figurative ones where only short, non-biasing, contexts were present, while longer contexts showed the reverse. Using conventional metaphors, McElree and Nordlie (1999), found no differences in the timing for access to both interpretations using a speed-accuracy trade-off procedure, providing even more evidence against simple stepwise models of idiomatic processing.

An alternative to such models can be found in so-called hybrid approaches (e.g., Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Sprenger, Levelt, & Kempen, 2006). These hybrid models attempt to integrate the processing of idioms into the framework of existing processing and representational accounts. The Configuration Hypothesis (Cacciari & Tabossi, 1988), for example,

assumes that literal composition occurs necessarily until enough information has accrued for the idiom to be recognized as such. Thus, processing is literal by default until signaled to proceed otherwise, and the literal meaning plays no clear role after recognition. Libben and Titone (2008) added to this idea in their Constraint-based Model of Idiom Processing by suggesting that idiomatic properties, such as compositionality or literality, as well as linguistic context can affect access to meanings at different times during processing (see also Titone & Libben, 2014), and figurative meaning also appears to accrue cumulatively.

The Hybrid Model of Idiom Representation, a model of idiom production put forward by Sprenger et al. (2006), links the meaning of literal constituents more clearly to the figurative meaning of an idiom on a representational level (see also Cutting & Bock, 1997). Individual lemmas and *superlemmas*, or representations of idioms on the lexical-syntactic level, are simultaneously activated starting with the first word of the idiom, unlike in the Configuration Hypothesis. Activation of both the literal and figurative meanings compete with one another as a sentence unfolds, and access to the figurative meaning follows a spreading procedure that increases as more information becomes available, much like the literal composition process. As in the Configuration Hypothesis, a clear prediction for literal composition after recognition is not made, though the employment of meaning competition, as opposed to a building recognition in the Configuration Hypothesis and Constraint-based Model, suggests that literal composition may continue on some level. However, in all of these hybrid models, figurative activation proceeds cumulatively and has the potential to be impacted by local and global factors such as literality and context.

5.1.2 Contextual effects

A number of studies on preceding linguistic context provide evidence that figuratively biasing contexts can ease access to figurative meaning in idioms (e.g., Colombo, 1993; Fanari, Cacciari, & Tabossi, 2010; Holsinger & Kaiser, 2013; Ortony, Schallert, Reynolds, & Antos, 1978). Studies using preceding linguistic biasing contexts have found faster paraphrasing of figurative interpretations (e.g., Gibbs 1980), better monitoring of matching, rhyming, and related words (e.g., Estill & Kemper, 1982), larger priming effects in cross-modal priming tasks (e.g., Colombo 1993), as well as faster reading times (e.g., Colombo 1993, 1998). Some studies have even suggested that figuratively biasing contexts might be a necessary condition for early activation of figurative meanings, particularly in the case of unpredictable or short idioms, and idioms with high literality (e.g., Cacciari, Padovani, & Corradini, 2007; Colombo, 1993; Fanari et al., 2010).

Access to the literal meaning of individual constituent words or phrasal meaning has not been shown to be as consistently affected by context in psycholinguistic studies. While cross-modal priming studies have shown literal constituent activation in idioms in neutral contexts (e.g., Beck & Weber,

2016a; Cacciari & Tabossi, 1988), EEG evidence found in a study by Rommers, Dijkstra, and Bastiaansen (2013) showed that the literal meaning of final words in predictable idioms (e.g., *lamp* in *liep tegen de lamp*, 'walked against the lamp', meaning to get caught) was not activated when the words were presented in idiomatic sentences. When the same final words were presented in literal sentences (e.g., *draaide het peertje in de lamp*, 'screwed the light bulb into the lamp'), their literal word meaning was activated. That is, an N400 effect for semantically related words (e.g., *kaars*, 'candle') was found in literal sentences, but no N400 effect was found in idiomatic sentences. This result suggests that literal interpretation was rendered unnecessary in this context. In another EEG study, Canal, Pesciarelli, Vespignani, Molinaro, and Cacciari (2017) similarly found evidence suggesting facilitation for figurative and literal meanings following biasing contexts, this time for the entire phrase rather than only constituent words. In line with Rommers, Dijkstra, and Bastiaansen (2013), evidence for qualitative changes in processing after idiom recognition suggests that literal composition mechanisms are suppressed or discontinued following a figurative context. Thus, context has an impact on both literal and figurative interpretations during processing.

Alternatively, there is also some evidence that the literal meaning of idioms is not improved by preceding contexts to the same extent that figurative bias improves access to idiomatic interpretations, if at all (e.g., Gibbs, 1980). Like idioms, phrasal verbs can have a literal and figurative interpretation, and Holsinger and Kaiser (2013) investigated the costs for recovery of meaning in phrasal verbs where the resolution of the verb was either congruent or incongruent with expected literal or figurative interpretations (e.g., The daring archaeologist/The hungry waitress *dug into* the tomb/the sandwich). Their self-paced reading study showed that figurative interpretations benefitted from the presence of biasing contexts, but literal interpretations did not show differences regardless of whether the biasing context was figurative or literal. Furthermore, they found that the greatest processing costs occurred when a figurative interpretation followed a literal biasing context. Thus, unlike the previous studies discussed above (Canal, Pesciarelli, Vespignani, Molinaro, & Cacciari, 2017; Rommers et al., 2013) Holsinger and Kaiser conclude that literal composition occurs necessarily, regardless of context, whereas context is required for achieving early access to figurative meaning. Notably, while the reading study by Holsinger and Kaiser used phrasal verbs rather than idioms and a mismatch design focusing on meaning integration (i.e., biasing contexts matched or mismatched with a later resolution in the sentences), the EEG studies already mentioned compared the effects of contexts in figurative sentences with idioms to comparable literal sentence processing (Canal et al., 2017; Rommers et al., 2013). These key differences leave unanswered questions about the impact of context on literal interpretations and the obligatory nature of literal composition in idiom processing.

5.1.3 Idiom literality

Idioms vary starkly in their potential to be used in a literal sense (e.g., Popiel & McRae, 1988), and this potential for literal interpretation, referred to here as literality, has been shown to impact idiom processing. Literality is commonly based on ratings across a scale determining the plausibility of a literal interpretation of an idiom (see e.g., Titone & Connine, 1994a; b; DIN database, **Chapter 2**). This property is closely related to other idiomatic properties such as the ambiguity of an idiom's meaning and is sometimes even referred to interchangeably so (see e.g., Briner & Virtue, 2014; Cacciari & Tabossi. 1988). This overlap in definitions between such properties as literality and ambiguity may be because, in ambiguous idioms, the likelihood of figurative and literal interpretations is comparable, as may be the case of some idioms with a high literality. However, a highly-literal idiom may not necessarily be highly ambiguous if the figurative meaning is more likely (e.g., an idiom such as kick the bucket, with a high literal potential may not be used literally often, and is thus not highly ambiguous). The current study will focus on literality, as defined above, and will target idioms at both ends of the scale: high- or low-literality idioms.

While it might seem feasible that the potential for both literal and figurative interpretations in the meaning of high-literality idioms (e.g., *to take the bull by the horns*) slows down processing, several studies have found the opposite pattern. Cronk and Schweigert (1992), for example, observed faster reading times for high-literality idioms, and Mueller and Gibbs (1987) found that high-literality idioms were classified more quickly than low-literality idioms. Both attributed these findings to multiple entries in the lexicon (i.e. one figurative and one literal), thereby increasing the likelihood of encountering the phrase in either usage and contributing to overall increased speed in meaning access.

Literality has also been found to interact with other idiomatic properties to affect access to figurative and literal meaning. Titone and Connine (1994b), for example, showed that literality affects access to literal constituent meaning in highly predictable idioms (i.e., idioms in which the final word can be predicted). In their cross-modal priming study, highly predictable, low-literality idioms in neutral contexts showed no activation of words semantically related to the final constituent words (e.g., *burn the midnight oil* and FUEL). However, for unpredictable idioms, literal constituent meaning was activated in both high- and low-literality idioms. Titone and Connine interpreted the results to mean that competition between figurative and literal meanings may stay active when relevant, for instance in high-literality idioms, but this competition can decay where it's not necessary (see also e.g., Giora, 1997: Graded Saliency Hypothesis).

What is lacking in the current state of research, however, is a look at context and literality together, in particular, an investigation of how context and literality might affect one another. For instance, in unpredictable idioms, where Titone and Conine (1994b) found literal activation regardless of literality, it is unclear whether the presence of biasing contexts might impact this literal activation.

A re-examination of both high- and low-literality idioms in the presence of context—both literal and figurative—may allow for a more precise picture of both the impact of literality on processing as well as the possible limitations of contextual effects where literality varies.

5.2 SELF-PACED READING EXPERIMENT

To this end, a self-paced reading study in English on native speakers was conducted to further determine the activation and role of literal and figurative meaning during the time-course of L1 idiomatic processing. Like Holsinger and Kaiser (2013), we aimed to determine how context can impact access to these meanings in both cases where contextual expectations are met and turn out to be false. Specifically, we intended to focus on the point in reading at which either the confirmation of congruent meanings or the recovery from incongruent expectations occurs. We tested both high-literality and low-literality idioms in speakers of American English to investigate whether the effects of context vary with the literality of an idiom.

The study had a total of three manipulations: idiom type (high- or low-literality idioms), biasing context (literal or figurative bias), and resolution (literal or figurative resolution). High-literality idioms, like *break the ice* (meaning figuratively to ease the nervousness in a social situation), followed contexts that either biased a literal or a figurative interpretation (e.g., literal: The cold Eskimo, who was eager to catch some fish; figurative: The new schoolboy, who didn't know anyone in his class) and preceded resolutions that either corresponded to a literal or a figurative interpretation (e.g., literal: on the lake…; figurative: with his peers…). Low-literality idioms like *lose one's cool* (meaning figuratively to lose control of one's emotions) also followed the same type of biasing contexts (e.g., literal: The freelance writer, who often started political debates; figurative: The sweaty runner, who was recovering under a tree) and preceded corresponding resolutions (e.g., literal: from the shade; figurative: out of anger).

We predicted that our results for figurative interpretations should replicate those found by Holsinger and Kaiser (2013). Namely, there should be a benefit of faster reading times when a context biases for a figurative interpretation and a figurative resolution occurs compared to when a context biases for a literal interpretation and a figurative resolution occurs. These predictions are also in line with other contextual studies showing the benefits of figurative contexts on figurative interpretations (e.g., Colombo, 1993). Since both high- and low-literality idioms have a high potential for a figurative interpretation (i.e., per definition all idioms have a figurative meaning), we predicted the facilitatory effect of figurative contexts to occur for both idiom types. For literal interpretations, however, we expected a difference in context effects for high- and low-literality idioms. Contrasting Holsinger and Kaiser (2013), however, we expected a benefit when a context biases for a literal interpretation and a literal resolution occurs compared to when a context biases for a figurative interpretation and a literal

resolution occurs. Although neither Rommers and colleagues (2013) nor Canal and colleagues (2017) used a mismatch design, such a result would be in line with their studies as it suggests an influence of context also on the literal interpretation. If idiom literality is a mediating factor, then, in contrast with high-literality idioms, we would expect no benefit of a biasing context for low-literality idioms in the present study. This prediction would be in line with evidence from Titone and Connine (1994b) showing that literal activation may not occur in some cases for low-literality idioms, something which may be attributed to a lack of saliency for the literal meaning in these idioms (see e.g., Giora, 1997). If literality, however, does not outweigh the general saliency of idioms' figurative interpretations, differences in contextual influence on literal interpretations may not be present or measurable in such an experiment.

5.2.1 Method

5.2.1.1 Participants. Fifty-two native speakers of American English (35 female; average age of 21.62, SD = 2.44) received financial compensation for taking part in the study. Participants were recruited at the University of Maryland from the Department of Linguistics subject pool and grew up in monolingual households. A total of 6 participants were left-handed, and all participants had normal hearing and normal to corrected vision. ***Materials*** The experiment consisted of 22 target and 78 filler trials. Sentences in target trials began with a biasing context (noun phrase + relative clause) followed by the infinitive form of the idiom, followed by a resolution (prepositional phrase) congruent or incongruent with contextual expectations, and ended with two additional short phrases shared across all conditions (see Appendix D for all items). Two example stimuli with all four conditions are provided in Table 5-3. As the experiment was conducted using a phrase-by-phrase reading paradigm, each column in the table represents the phrases seen by participants during the experiment. Phrases were controlled for letter length and average word frequencies, with only minimal differences between conditions (see analysis section for more information). All idioms had the same syntactic structure (to-infinitive verb + determiner + noun) and were short and unpredictable.

Table 5-3. Example stimuli

		Biasing Subject	Biasing Context 1	Biasing Context 2	Pre-idiom	Idiom	Resolution	Resolution+1	Wrap-up
High literality	a.	The new schoolboy	who didn't know	anyone in his class	just wanted to	*break the ice*	with his peers	sooner than later	Monday morning.
	b.	The new schoolboy	who didn't know	anyone in his class	just wanted to	*break the ice*	on the lake	sooner than later	Monday morning.
	c.	The chilly Eskimo	who was eager	to catch some fish	just wanted to	*break the ice*	on the lake	sooner than later	Monday morning.
	d.	The chilly Eskimo	who was eager	to catch some fish	just wanted to	*break the ice*	with his peers	sooner than later	Monday morning.
Low literality	a.	The emotional writer	who often started	political debates	didn't want to	*lose his cool*	in his anger	too quickly	that morning.
	b.	The emotional writer	who often started	political debates	didn't want to	*lose his cool*	from the shade	too quickly	that morning.
	c.	The overheated runner	who was resting	under a tree	didn't want to	*lose his cool*	from the shade	too quickly	that morning.
	d.	The overheated runner	who was resting	under a tree	didn't want to	*lose his cool*	in his anger	too quickly	that morning.

Note: The contexts in (a.) and (b.) bias for the figurative interpretation and (c.) and (d.) for the literal interpretation while the resolutions in (a.) and (d.) resolve figuratively and (b.) and (c.) resolve literally.

Sixty-eight sentences varying in length and structure were selected as filler trials, and ten of these sentences also contained idioms. Additionally, ten filler trials consisted of arithmetic questions (e.g., (30-28)*5= ?). One third of all trials, including all target trials, were followed by comprehension questions targeting differing types of sentential information. The comprehension and arithmetic questions were included to hold participants' attention and ensure concentration throughout the task, and correctness of response was considered in the analysis. All experimental items were divided into 4 lists, counterbalanced so that each idiom occurred only once per list, but idiom-type and context- and resolution-type conditions were balanced between the lists. The order was then randomized and reversed for each list so that there were 8 lists in total. All lists contained the same filler items.

Three regions of interest, also labeled in Table 5-3, were identified for our analysis: the Idiom, the Resolution, and the Resolution + 1 (the phrase following the resolution) regions. However, our main predictions concern the Resolution +1 region. Since previous studies have found that meaning is typically available later in short and unpredictable idioms than in longer, more predictable ones (e.g., Titone & Connine, 1994), we expected the effects of (mis-)matching contexts to show up in the region following the resolution rather than during the resolution itself. These expectations align with the late effects found by Holsinger and Kaiser (2013). Furthermore, because the Resolution +1 region is identical across conditions for each item (see Table 5-3), so observed differences in reading times based on context and resolution cannot be attributed to differences in lexical items or syntactic structure in this region.

Before deciding on the 22 target idioms, a total of 30 idioms were pre-selected from the 300 idioms in the English-German Database of Idiom Norms (Beck & Weber, 2016b) based on a number of factors that are known to affect idiomatic processing (see e.g., Titone & Libben, 2014). Only

unpredictable idioms (i.e. with a low predictability score) with a high subjective familiarity were pre-selected (ratings higher than 3.5 on a 7-point scale). Idioms shared a VP syntactic structure, beginning with a verb and ending with a noun (e.g., *break the ice*), but they differed in levels of literality (high literality had a minimum of 4 and low-literality idioms had a maximum of 3 on a 7-point scale). For all 30 pre-selected idioms, biasing sentences were constructed and subjected to norming studies.

5.2.1.3 Norming.

Two norming studies were conducted on the pre-selected 30 idioms and their biasing sentences in order to choose the items that were a) most appropriately biased for a literal or figurative reading of the idioms following the given contexts and b) most plausible (see e.g., Ratcliff, 1987). Based on the results from these norming studies (see below), a total of 22 idioms were selected for the reading study: 11 high-literality idioms (average of 5.0) and 11 low-literality idioms (average of 2.8; $t = 14.12$, $p < 0.001$). The selected 22 idioms had mean constituent frequencies of 3.25 per million and were rated as highly familiar by L1 speakers (average of 6.5).

5.2.1.3.1 Norming Experiment 1: Strength of context.

In a multiple-choice test, 40 American English participants were presented with the idioms preceded by either their figuratively or literally biasing contexts and were asked to choose the meaning of the highlighted phrase (the idiom) that most appropriately fits the provided sentential context. Each participant saw only one biasing context for each idiom, and four answer choices were available: the literal meaning, the correct figurative meaning, an incorrect figurative meaning, and "I don't know".

For the final selection of 11 high literality idioms, the percentages of completions consistent with the biasing context were 98% for figurative contexts and 52% for literal contexts respectively; for the final selection of 11 low literality idioms, the percentages of consistent completions were 97% for figurative contexts and 30% for literal contexts respectively. Linear models were conducted on the average correct completions based on context using *idiom literality* (high- and low-literality) and *biasing context* (figurative and literal) as fixed effects. Our model confirmed that figurative contexts were completed consistently more often than literal contexts ($ß = -20.4$, $t = -2.8$, $p < .01$) and that for literal contexts, the high-literality idioms were completed consistently more often than low-literality idioms ($ß = -46.4$, $t = -8.8$, $p < .001$). Thus, in a number of cases participants responded with figurative interpretations even when a literal context rendered them inappropriate, thereby in fact ignoring contextual information. If this preference for figurative interpretations in the norming study indeed implies that figurative contexts bias more than literal contexts, it is possible that we will not see effects of literal bias in our self-paced reading results. Nonetheless, in order to best capture the potential individual variation based on item, biasing strength, included as the by-item average of correct

completions based on context, was later included as a predictor in the regression models in our analysis of reading times where warranted.

5.2.1.3.2 Norming Experiment 2: Plausibility

Eighty American English participants were asked to rate the 30 pre-selected idioms with their four biasing sentence contexts (see Table I) on how plausible the situation described by the sentences was (on a 7-point scale). Each participant only saw each idiom in one of the four conditions.

For the final selection of 22 items, the mean plausibility rating was 4.15 (*sd*= 1.17), and item averages ranged from 2.16 to 6.85, suggesting that while no items were rated as completely implausible, some were rated as very plausible. We again used simple linear models to verify any differences between items based on our predictive variables. Using the mean plausibility rating as the dependent factor, *idiom literality* (high- and low-literality), *biasing context* (figurative and literal), and *resolution type* (figurative and literal) were included as fixed effects. Our model showed that ratings for literal endings were overall lower than those for figurative endings (*ß*= -1.9, *t*=-5.8, *p*<.001), and literal contexts were likewise rated lower than figurative contexts (*ß*= -2.2, *t*=-6.7, *p*<.001). Additionally, an interaction of resolution by context suggested that congruent endings (literal contexts and resolutions / figurative contexts and resolutions) had higher ratings than incongruent resolutions (*ß*= 3.5, *t*=-7.5, *p*<.001), and a three-way interaction suggests that low-literality idioms were rated worse than high-literality idioms in the condition with a literal bias and resolution (*ß*= -1.6, *t*=-2.2, *p*<.05), whereas low-literality idioms showed higher plausibility ratings in cases where endings were congruent with their contexts.

A further look at the comments section at the end of the rating study suggested that participants were likely responding to preferential use of idioms rather than the plausibility of the sentences as described in the instructions, and it is possible that this norming study did not fully capture plausibility as intended. The differences between both congruent and incongruent endings as well as the preference for low-literality idioms with a figurative resolution also aligns with these expectations, as both cases represent unlikely uses of idioms. In spite of the difficulties in capturing this effect, in order to account for possible item differences based on plausibility, as collected in this norming study, the average ratings were included as a predictor in the regression models of the reading data where justified.

5.2.1.4 Procedure

Participants gave informed consent following the ethical guidelines of The University of Maryland College Park and were tested individually in quiet rooms at the Language Science Center. The experiment was programmed and executed using E-Prime (Psychology Software Tools, 2013), and a subsequent idiom recognition test (described below) and a language background questionnaire were

completed using Adobe Acrobat. Participants wore noise-cancelling headphones to ensure that no noise distractions occurred during the reading study, and responses were recorded with the spacebar of a keyboard.

Participants began the reading study with four practice trials, and halfway through the experiment a self-timed break was offered. A standard moving-window phrase-by-phrase presentation was used in which each phrase was masked by hyphens corresponding in length to the phrase to be presented. Phrase-by-phrase rather than word-by-word was used in order to better mimic natural reading patterns and avoid a forced incremental processing pattern (e.g., Jegerski, 2014), which might directly affect the questions at hand. This method, while not precise enough to measure fine differences in constituent meaning activation as the sentence unfolds, provides a good balance between natural reading and experimental procedures that can capture the costs and benefits of integrating activated meanings into contexts. Participants were instructed to begin a trial by pressing the space bar, and to press the space bar again once they had read the phrase presented. After pressing the space bar, the phrase again became masked, and the new phrase was unmasked. 30 of the 90 reading trials were followed by multiple choice comprehension questions, and 10 by basic arithmetic problems. After the reading study, participants completed an idiom recognition test and a language background questionnaire. In the idiom recognition test, participants were presented with all 22 target idioms in the sentences from the condition including the figurative context and figurative resolution. The task was to select the figurative meaning of the idiom from a randomly ordered multiple choice selection including 1) the figurative meaning, 2) a possible literal meaning, 3) an incorrect meaning, and 4)"I don't know." if the idiom was unfamiliar. This was done to ensure that the idiomatic meaning was indeed familiar for the participants of the study. The results of the recognition task were used to inform the analysis of our data.

5.2.2 Results

We used R (R Core Team, 2017) and lme4 (Bates, Maechler, Bolker, & Walker, 2015) to perform linear mixed effects analyses of the relationships between *biasing context* (figurative, literal), *resolution type* (figurative, literal), and *idiom literality* (high-literality, low-literality) on normalized reaction times in each of the three previously identified areas of interest: *idiom*, *resolution*, and *resolution + 1* regions. Trials were only included in the analyses if answers to the directly following comprehension questions were answered correctly. One participant fell below a threshold of answering 85% of questions correctly (68% correct) and was therefore excluded from the analysis). Two idioms (*play the field* and *turn the tide*) were also excluded as the idiom recognition test following the reading study revealed that more than half of participants were not familiar with these idioms. Outliers were excluded based on a visual inspection of reading times in each section. First, theoretically irrelevant times were excluded, second, inspection of the mean and general distribution of overall reading times

were used to determine critical reading times for each region on an individual basis. Critical reading times were identified as 150ms-1500ms in the *idiom* section, 150ms-1800ms in the *resolution* section, and 150-1500ms in the *resolution+1* section, and 3.89%, 2.62%, and 4.85% of the data was excluded in each region respectively. The difference in size between regions is in line with the mean length, difficulty, and average reading times in each region. Table 5-4 reports the mean and log-transformed reading times for the phrases in each region across all factors.

Table 5-4. Raw reading times (ms and log) by region and main effects

		Figurative Context				**Literal Context**			
		high-literality		**low-literality**		**high-literality**		**low-literality**	
Region	***Resolution-Type***	**RTs**	**log RT**	**RTs**	**log RT**	**RTs**	**log RT**	**RTs**	**log RT**
Idiom	figurative	607.62	6.33	634.95	6.37	611.40	6.34	647.32	6.40
	literal	598.56	6.33	611.56	6.36	607.57	6.34	653.87	6.40
Resolution	figurative	639.59	6.36	718.83	6.42	680.36	6.48	756.30	6.53
	literal	617.58	6.33	675.48	6.37	641.68	6.43	711.87	6.47
Resolution +1	figurative	572.19	6.26	613.32	6.34	616.95	6.36	615.57	6.36
	literal	610.36	6.34	705.41	6.48	605.24	6.31	680.60	6.43

The analyses for each region of interest were done separately on normalized log-transformed values (see Baayen, 2008) following the same procedure each time. With the exception of the idiom section (see details below), independent measures and fixed factors, numerically centered around 0, were coded and included as follows: *biasing context* (figurative: 0.5, literal: -0.5), *resolution type* (figurative: 0.5, literal: -0.5), and *idiom literality* (high-literality: 0.5, low-literality: -0.5). All three theoretically relevant factors were kept in the models, and items and participants were included as random factors with random slopes. A maximally justified random effects structure was determined for each region by stepwise selection and model comparison (see e.g., Bates, Kliegl, Vasishth, & Baayen, 2015) consulting RePsychLing (Baayen, Bates, Kliegl, & Vasishth, 2015). To account for additional variation, fixed factors of *trial order*, *region length* (in number of letters), *average lexical frequency*, and norming values for *idiom familiarity*, *plausibility*, and *strength of context* were also included in full models, and were eliminated from the models if backward stepwise selection showed that they did not improve the model. While these additional factors did not contribute to the interpretation of the results, their inclusion in the final model was justified by model comparisons, and any additional effects found were more likely to be caused by our manipulations rather than these factors. After models were selected, extreme outliers with a distance of greater than 2.5 standard deviations from zero were removed in order to improve model fit (see e.g., Baayen, 2008), this process

excluded an additional 1.72%, 1.97%, and 2.40% of the data respectively. The analyses for each section are discussed in the following, with a main focus on the *resolution + 1* region as it is the only region which directly compares reading times of the same phrases in all four conditions, and it is a likely point in time for effects of the matching or mismatching resolutions to be observable.

5.2.2.1 Reading times in the idiom region.

The model for reading times for the idiom region did not include *resolution type* or the norming factor for *plausibility* as the ending has not yet been encountered at this point. The final model output is displayed in Table 5-5. Significant effects of *region length* (β = .038, t= 2.36, p < .05) and *trial order* (β = -.06, t= -7.08, p < 0.001) suggest that idioms with more letters took longer to read and that reading times improved over the course of the experiment. These effects are consistent throughout the successive regions but did not contribute to the overall findings other than to justify their inclusion in the models and will therefore be listed in the tables for models but not discussed further in the text.

Table 5-5. Idiom region LMER model output

Factor	β (SE)	t	p	
Intercept	6.368(.045)	140.324	2.00E-16	***
Biasing Context	-0.005(.022)	-0.256	0.8016	
Idiom Literality	-0.051(.033)	-1.549	0.1398	
Region Length	0.038(.016)	2.361	0.0309	*
Trial Order	-0.067(.009)	-7.088	3.74E-10	***
Context x Literality	0.037(.043)	0.865	0.4009	

Note: *p<.05 **p<.01 ***p<.001

The predicted log reading times for the idiom region, collapsed across literality, can be seen in Figure 5-2. As reflected in the graph, no significant interactions between factors emerged, and there was no significant effect of context at this point in reading.

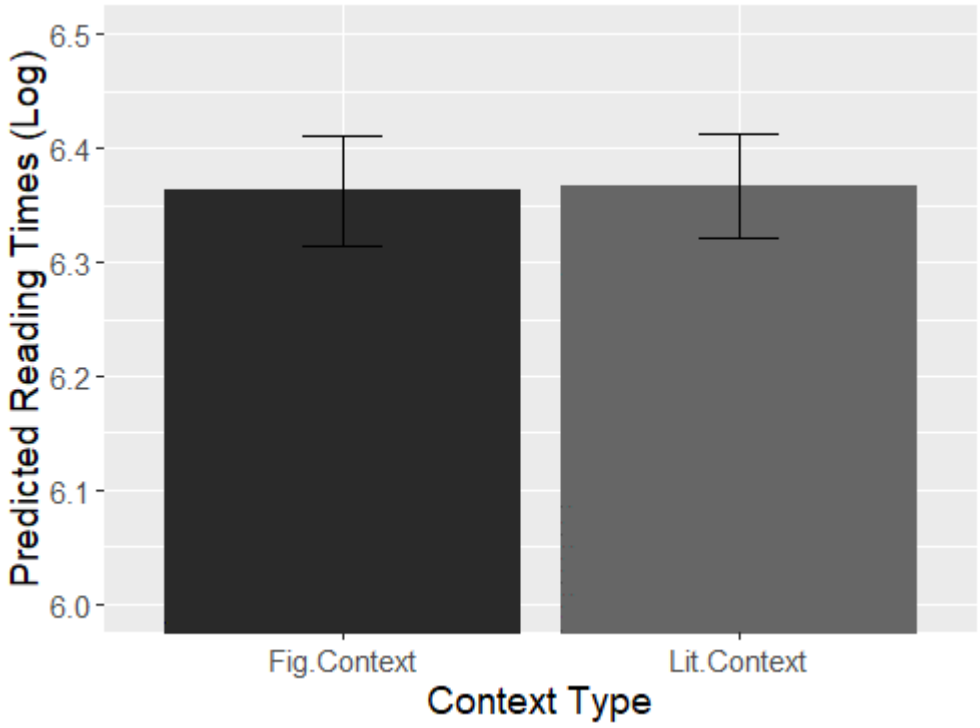


Figure 5-2. Model-predicted log reading times in the idiom section

Note: The predicted reading times are collapsed across literality. The black bar represents mean log reading times for idioms following a figurative context and the grey bar following a literal context, both with standard error bars.

5.2.2.2 Reading times in the resolution region.

For the resolution region, Table 5-6 displays the output of the best-fitting model after backward, stepwise selection and further removal of extreme residuals. Again, *region length* and, marginally, *trial order* effects were significant. Additionally, *idiom familiarity* improved overall reading times in this region (β = -.04, t= -2.26, $p < 0.05$), suggesting that reading times improve with increasing familiarity. *Resolution type* was also significant as a main effect (β = -.04, t= 1.96, $p < 0.05$) and as an interaction with *biasing context* (β = -.09, t= -2.00, $p < 0.05$).

Table 5-6. Resolution region LMER model output

Factor	*β* (SE)	*t*	*p*	
Intercept	6.409(046)	137.713	2.00E-16	***
Biasing Context	-0.028(.028)	-0.995	0.3341	
Idiom Literality	-0.066(.039)	-1.667	0.1144	
Resolution Type	0.043(.022)	1.969	0.0494	*
Region Length	0.030(.016)	1.852	0.0827	.
Trial Order	-0.089(.011)	-7.828	1.98E-14	***
Idiom Familiarity	-0.043(.019)	-2.260	0.0375	*
Context x Literality	-0.030(.056)	-0.532	0.6022	
Context x Resolution-Type	-0.090(.045)	-2.004	0.0455	*
Literality x Resolution-Type	0.015(.045)	0.353	0.7243	
Context x Literality x Resolution-Type	-0.018(.089)	-0.209	0.8347	

Note: .p<0.1, *p<.05, **p<.01, ***p<.001

In order to better interpret this effect in light of the interaction, a post-hoc analysis was conducted using simple slopes (e.g., Aiken & West, 1991). By recoding *biasing context* (first as figurative context =0, literal context =1, subsequently reversed), its effect on *resolution type* can be examined more clearly. The same model used in the original analysis was run using the change in coding listed above, and the effect of *resolution type* is significant when contexts are literal (β = .08, t= 2.78, p < 0.01) and not figurative (β = .00, t= -0.03, p = 0.96). Namely, as displayed in Figure 5-3, where contexts are literal, reading times are significantly slower for figurative resolutions. This effect is in line with one of the main conclusions of Holsinger and Kaiser (2013), though, it should be noted that the comparisons in this region were between different lexical units, and this result will be placed in this context in the discussion section.

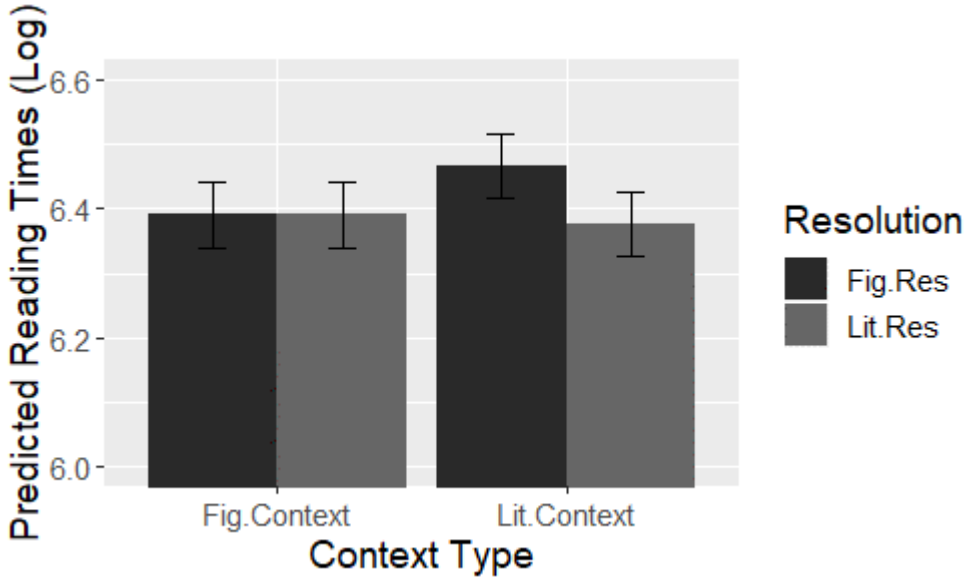


Figure 5-3. Model-predicted log reading times in the resolution section

Note: The predicted reading times are collapsed across literality. The black bars represent mean log reading times resolving figuratively (Fig.Res) and the grey bars resolving literally (Lit.Res), all with standard error bars.

5.2.2.3 Resolution + 1 region

The final model output for the *resolution + 1* region, derived following the same procedure as the other regions, is displayed in Table 5-7. The predicted log values from the final model can be seen in Figure 5-4. Unlike in the other two regions, the text in this region was identical in all four conditions for each idiom. Main effects of *resolution type* (β = -0.07, t = -3.86, $p < 0.001$) and, marginally, *idiom literality* (β = -0.07, t= -1.84, p = 0.07) were present. Additionally, the effects of *region length* and *trial order* were generally in line with those found in the previous regions. In addition to main effects, several interactions were also significant, setting our data apart from earlier studies. Both *biasing context* by *resolution type* (β = -0.11, t= -3.23, $p < .01$) and *idiom literality* by *resolution type* (β = 0.10, t = 2.75, $p < 0.01$) were significant, and the three-way interaction of *biasing context* by *resolution type* by *idiom literality* was marginally significant (β = -0.13, t= -1.77, p = 0.07).

Table 5-7. Resolution + 1 region LMER model output

Factor	*β* (SE)	*t*	*p*	
Intercept	6.354(.04)	142.065	2.00E-16	***
Biasing Context	-0.021(.01)	-1.168	0.2431	
Idiom Literality	-0.072(.03)	-1.842	0.0766	.
Resolution Type	-0.070(.01)	-3.869	0.0001	***
Region Length	0.049(.01)	2.945	0.0091	**
Trial Order	-0.091(.01)	-7.561	1.89E-11	***
Context x Literality	-0.031(.03)	-0.858	0.3911	
Context x Resolution-Type	-0.118(.03)	-3.238	0.0012	**
Literality x Resolution-Type	0.100(.03)	2.754	0.0060	**
Context x Literality x Resolution-Type	-0.130(.07)	-1.777	0.0759	.

Note: .p<0.1, *p<.05, **p<.01, ***p<.001

In order to better interpret the interactions, we performed post-hoc analyses using simple slopes on the relevant variables individually. First, the effect of *idiom literality* on *biasing context* as well as its involvement in the three-way interaction with *resolution type* was investigated by recoding *idiom literality* (first as high-literality =0, low-literality =1, subsequently reversed) and re-running the original model. Results showed that *resolution type* was significant in the case of low-literality idioms (β = -0.12, t= -4.44, p < 0.001) but not in high-literality idioms (β = -0.02, t= -0.82, p = 0.40). This confirms the pattern in low-literality idioms that figurative resolutions were significantly better than literal ones (see Figure 5-4). However, the results also show that the interaction between *biasing context* and *resolution type* was significant for the high-literality idioms (β = -0.18, t= -3.81, p < 0.001), and not for the low-literality idioms (β = -0.05, t= -0.97, p = 0.32).

To further examine the nature of this interaction in high-literality idioms, *biasing context* was also recoded (first as figurative context =0, literal context =1, subsequently reversed), and the same model was used again, keeping high-literality idioms as the intercept. The results display main effects of *resolution type* for both literal (β = 0.07, t= 2.04, p < 0.05) and figurative contexts (β = -0.11, t= -3.36, p < 0.001) in opposite directions. This finding suggests that, for high-literality idioms, congruently resolving endings were read more quickly than incongruently resolving endings (e.g., "The new schoolboy… wanted to *break the ice* with his peers…" and "The chilly Eskimo…wanted to *break the ice* on the lake…" vs. "The new schoolboy… wanted to *break the ice* on the lake …" and "The chilly Eskimo…wanted to *break the ice* with his peers …"). While this is a reflection of the marginally significant three-way interaction, these results confirm the trends displayed in the high-literality idioms (right) in Figure 5-4.

In order to investigate any possible effects not already summarized by the recoding of *idiom literality* and then *biasing context*, *biasing context* was recoded (first as figurative context =0, literal context =1, subsequently reversed) and applied to the original model with no other changes. Where

contexts were figurative, both the effects of *idiom literality* (β = -0.08, t= -2.03, $p < 0.05$) and *resolution type* (β = -0.13, t= -5.05, $p < 0.001$) were significant. This result suggests that figurative contexts cause significant benefits and costs where resolutions do not match these expectations and that figurative contexts improve readings for high-literality idioms more than low-literality idioms. In the case of literal contexts, these factors were involved in a two-way interaction between these factors (β = -0.16, t= -3.172, $p < 0.01$), which was already explored in the first post-hoc analysis.

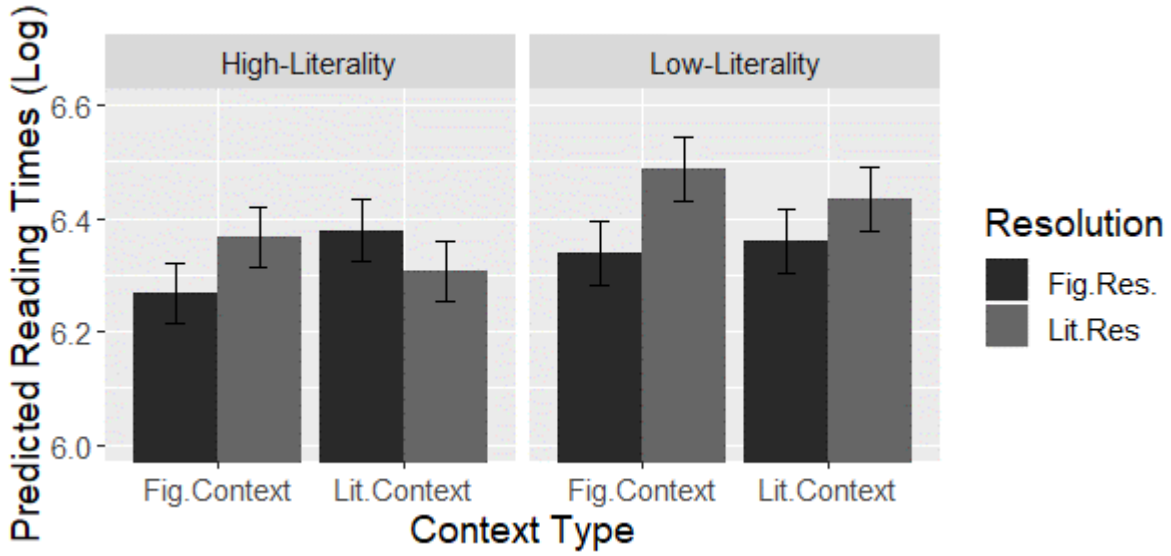


Figure 5-4. Model-predicted log reading times in the resolution+1 section

Note: Predicted log reading times are divided into low- (left) and high-literality (right) idioms by biasing context. The black bars represent mean log reading times following a figurative resolution (Fig.Res.) and the grey bars following a literal resolution (Lit.Res.), all with standard error bars.

5.3 DISCUSSION AND CONCLUSIONS

With this self-paced reading study, we aimed to further examine the activation and role of literal and figurative meaning during the time-course of idiomatic processing. Specifically, we asked how context impacts access to these interpretations and what limits the level of literality of an idiom can impose in both cases where contextual expectations are met and turn out to be false. By examining three regions of interest (the *idiom*, the *resolution*, and the *resolution + 1* regions), we were able to see effects of context and literality as both figurative and literal interpretations were integrated with contextual meaning. Our study fills some of the gaps in similar research conducted in the field and provides evidence that context has a limited role in meaning activation, and factors such as idiom literality can

supersede standard processing and integration mechanisms under certain conditions. We will briefly discuss the implications of the reading times in the first two regions before focusing on the final region.

Notably, our findings show some critical differences from those of previous studies (e.g., Holsinger & Kaiser, 2013) and also corroborate previous findings on meaning constitution that suggest a processing strategy that adapts to contextual and linguistic cues rather than unfolding in a uniform manner. We did not detect any early differences (i.e., *idiom* region) between idiom types based on context, which suggests similar processing during reading. The well-documented effect that idioms are read faster following a supporting context (e.g., Gibbs, 1980; Holsinger & Kaiser, 2013; Swinney & Cutler, 1979) was not reliably observed in our data. However, we presume that one reason this effect fails to reach significance could be due to the short length and unpredictability of the idioms in the present study (see e.g., Fanari et al., 2010) as well as due to the possibility that effects which appear only very briefly (i.e., on single words) are lost in a phrase-by-phrase design.

Our findings in the *resolution* region, combined with the differences between this and the following region, suggest that as sentence comprehension unfolds, processing strategies may also be adaptive. The effect found in the *resolution* region in which there was a cost for figurative interpretations following literal contexts is in line with the conclusions drawn by Holsinger and Kaiser (2013). However, unlike the region in which this effect was found in their study, there is not yet evidence for the benefit of context in this region in our data, so conclusions based on this effect alone must be drawn with care. It may be evidence that early processing strategies indeed involve necessary literal composition, and for that reason figurative interpretations are costly where immediately unexpected. However, it should also be considered that this region draws comparisons between reading times of differing phrases, and, even under controlled conditions, subtle phrasal differences may also play a role in the effect found here.

Most importantly, the results in the final region (*resolution + 1*) are both in line with earlier studies (e.g., Canal et al., 2017; Rommers et al., 2013) and show key differences in reading time patterns suggesting that processing strategies may divide based on idiom-literality. Like Holsinger and Kaiser (2013), in our high-literality idioms, the greatest facilitation was found for sentences with a figurative bias and a figurative resolution. Figurative meanings did benefit the most from context for these idioms. But unlike their results, ours show that facilitation was still present for sentences with a literal bias and a literal resolution and that costs were greatest when literal endings followed figurative biasing contests. Our results are in line with those from both Canal et al. (2017) and Rommers et al. (2013) identifying literal contextual effects. These results also follow suit with Titone and Connine (1994b) concerning high-literality idioms, as both studies show evidence of literal meaning activation. Notably, the contexts in the current study were much longer than those used by Holsinger and Kaiser,

and we expect that this may be one reason that effects not detected in the previous study were found in ours (see e.g., Ortony et al., 1978).

Another important result from the final region was that our low-literality idioms were not impacted by context. Rather, reading times were faster in the case that sentences resolved figuratively. These results were consistent with our predictions for this group of idioms, and generally with other studies examining low-literality idioms (e.g., Titone & Connine, 1994b). While Titone and Connine (1994b) only found such a result with predictable, low-literality idioms, they only looked at automatic activation on the word-level. Thus, while this result provides a difference to this earlier study, the results are not contradictory. The results from this region also further impress upon differences in contextual effects found in the Holsinger and Kaiser (2013) study as well as the results found by Canal et al. (2017) and Rommers et al. (2013b), in which literal contextual effects were identified as a whole.

Notably, the implications of these differences found in the *resolution +1* region implies that high- and low-literality idioms are not equally sensitive to contextual effects and may encourage differing processing strategies. In the case of high-literality idioms, processing may be more adaptive and take context into account. However, for low-literality idioms, this seems not to be the case; rather, a purely figurative processing strategy is preferred. Thus, these idioms provide evidence for one of the limitations of the impact of context on idiom processing. Unlike the resolution region itself, lexical items in this region were identical in all conditions and observed differences in reading times here can be attributed to the manipulations in our study.

Importantly, these results also impress upon the need for literal composition during idiom processing. The differences in effects between the two types of idioms suggest that, in contrast with the claims of Holsinger and Kaiser (2013); literal processing may be abandoned or suppressed at some point during processing if either appropriate supporting context is present or an idiom has a low-literality. While we did see an early cost for unexpected figurative interpretations, this may reflect the process of late idiom recognition before processing becomes adaptive. In line with research from Rommers et al. (2013) and Canal et al. (2017), our low-literality idioms show evidence for abandonment of literal computation regardless of context. In addition to a lack of facilitation of literal resolutions compared to figurative resolutions, the general cost for all literal resolutions suggests that the conventional nature of these idioms might contribute to this processing strategy (i.e., experience with these idioms, used primarily figuratively, renders literal meanings unlikely). Though, our lack of evidence for a contextual effect in this case is not necessarily evidence that it does not occur. Rather, the low literality of these idioms may simply be a stronger signal than context. For the high-literality idioms, the same process of abandonment of literal composition after recognition cannot entirely explain the data. Canal et al. (2017) found evidence in high-literality, predictable idioms that context mediated whether semantic analysis occurred. Their results also suggested that these effects occur very

early in recognition. While our effects were found later than the effects discussed in their EEG data, typical of reading studies, it may be a case of shifting to a differentiated or even shallower processing of literal constituents (see also Peterson, Burgess, Dell, & Eberhard, 2001). This more subtle processing shift would be consistent with the lacking facilitation effect and heightened recovery costs in incongruent conditions. Following this pattern, it is feasible that the same effect may simply occur earlier or more prominently in low-literality idioms. In this case, context and literality could mediate when shallow processing becomes suppressed literal computation. Of course, given the nature of the task in the current study, it should also be noted that the late effects may also reflect evaluative processes in reading (e.g., D. N. Rapp & Mensink, 2011).

While the goal of our study was not to determine a model of processing for idioms, our results can be aligned with general research in the field already outlined. Unlike earlier studies, our data cannot speak to support an early advantage for literal interpretations, nor a purely figurative-first approach as they make uniform predictions about meaning retrieval and integration, and therefore cannot account for differences in context integration based on idiom literality. However, both hybrid models can account for the effects present in the final region, though, they do not explain the differences between idiom literality equally well. While the Configuration Hypothesis (Cacciari & Tabossi, 1988) should support earlier or stronger activation of low-literality idioms (see e.g., Titone & Connine, 1994b), it is unclear about predictions of literal composition, and therefore silent on whether or not it is in line with the results seen for our high-literality idioms. A later (or weaker) recognition of high-literality idioms may account for the stronger use of context in processing, and therefore the facilitation effect evidenced by the advantage for matching biases and resolutions. The Hybrid Representation theory (Sprenger et al., 2006) can account for this behavior in a more direct manner. As this hypothesis allows for a stronger level of competition between the two interpretations, the model seems to predict that idioms with a greater capacity to be interpreted literally should have more competition between meanings than those without (i.e., low-literality idioms). Thus, it can explain why literal interpretations are abandoned in low-literality idioms in all contextual situations, while context more readily impacts the competition between meanings in high-literality idioms.

Importantly, the current study has added to the body of literature that supports more heterogenous accounts of idioms and their processing, but it faces limitations in its claims about exactly what causes these differences in processing strategies. Here, literality, as defined by the potential for literal interpretation and examined as a binary variable, was shown to interact with context in idiomatic processing in a critical manner: literality may cause processing to proceed either more or less flexibly (concerning high-literality and low-literality idioms respectively). However, as discussed briefly in the introduction, this factor overlaps critically with other idiomatic properties such as ambiguity (e.g., Cacciari & Tabossi, 1988), saliency (e.g., Giora, 1997), and even meaning dominance (e.g., Milburn

& Warren, 2019), and it is unclear precisely which of and to what extent these factors may have impacted the results of the current study. Although high- or low-literality idioms can greatly overlap with ambiguous and non-ambiguous idioms respectively, these two terms differ from one another in that literality accounts only for the potential of a literal interpretation; ambiguity should also account for the likelihood of such an interpretation, as an idiom is only ambiguous if both interpretations are actually likely. This likelihood for ambiguity is also complicated by subjective familiarity with an idiom's figurative and literal uses as well as the dominance of such uses (e.g., Cronk, Lima & Schweigert, 1993; Milburn & Warren, 2019), all factors which contribute to the salience of a particular meaning (see e.g., Giora, 1997). Though some idioms may have a high-literality, or have both plausible literal and figurative interpretations, these two interpretations may not be equally salient for users. In light of the differences between the two types of idioms examined in this study, it is also conceivable that similar results may be achieved when looking at ambiguity, saliency, or even meaning dominance as predicting factors, as low-literality idioms seem to also critically differ from high-literality idioms among these properties (i.e., low-literality idioms are typically less ambiguous; may be more salient in the figurative meaning; and are more often used in the figurative sense, as reflected in our norming studies). While the current study was not designed to tease apart these properties, future studies should look at a broader spectrum of such properties and may even employ them as a scale rather than a binary factor in order to better do so.

Overall, our findings fill in some of the gaps in current research on idiom processing, specifically in addressing questions of figurative and literal meaning constitution on a phrasal level. Our data support a processing strategy that is sensitive to context, but in which these contextual effects are mediated by idiom literality. We conclude that processing follows a single, adaptive pattern until a threshold of information can be reached that suppresses or qualitatively changes the nature of literal computation. This threshold can be mediated by context, pushing more ambiguous, high-literality idioms across this threshold only when context also supports such a strategy, and by literality, which can supersede context in strength if low-literality idioms cause the threshold to be reached earlier, for example. Future studies are still needed to provide more evidence on what happens earlier in processing concerning context and literality, among other properties, and what additional factors can impact a change in processing strategies.

5C CONTEXT MATTERS, FIGURATIVELY, FOR L2 READERS: EVIDENCE FROM SELF-PACED READING[8]

Sara D. Beck & Andrea Weber

Abstract

In a self-paced reading study, we investigated the extent to which non-native speakers use biasing context in idiom processing, and whether idiom literality limits these effects as it does in native speakers. Idioms with a high potential for literal interpretations (e.g., *break the ice*) and a low potential for literal interpretations (e.g., *lose one's cool*) were embedded into sentences biasing literal or figurative readings and followed by resolutions that were either congruent or incongruent with these expectations. Context significantly impacted figurative interpretations, but not literal ones. While the reading time patterns replicate those of native speakers, literality did not significantly interact with contextual effects as in L1 readers, particularly in how it impacts literal readings. The results highlight differences between L1 and L2 competencies.

[8] This chapter is adapted from the published version of the following article: Beck, Sara D.; Weber, Andrea: Context matters, figuratively, for L2 readers: Evidence from self-paced reading. In : *Proceedings of The 11th International Conference on the Mental Lexicon*. Edmonton, Canada, University of Alberta Libraries.

5.1 INTRODUCTION

Idioms require speakers to be able to interpret both a literal and a figurative interpretation with ease in real-time processing. Although some studies have provided evidence that non-native (L2) speakers can process idioms similarly to native (L1) speakers (e.g., S. D. Beck & Weber, 2016a; Conklin & Schmitt, 2008), others report findings that vary (e.g., Siyanova-Chanturia, Conklin, & Schmitt, 2011). Not only must L2 speakers be familiar with an idiom in order to interpret it figuratively (e.g., Titone & Libben, 2014), but speakers must also learn to deal with additional complexities such as linguistic context in order to decide whether a literal or figurative interpretation is appropriate. In addition to context, idioms vary greatly in their potential for a literal interpretation (literality). An idiom like *break the ice*, clearly interpretable in both a literal and a figurative (i.e., to relieve social tensions) sense, has high- literality compared to a low-literality idiom like *lose one's cool*, which has a less clear literal sense compared to its figurative one (to lose control of one's emotions). While L1 experience with the language and individual idioms has given these speakers the tools to integrate these cues into online processing strategies, it's unclear whether L2 speakers also have these same abilities.

In a self-paced reading study in **Chapter 5B**, the effects of biasing context in L1 idiom processing were investigated and it was asked whether idiom literality limited contextual effects. Following the basic design of a study on highly literal phrasal verbs by Holsinger and Kaiser (2013), high- and low-literality idioms were embedded into sentences with contexts biasing either a literal or a figurative interpretation, and followed by resolutions that were congruent or incongruent with these expectations (e.g., [The new schoolboy/the chilly Eskimo] just wanted to *break the ice* [with his peers/on the lake] ...). The results showed that reading times for figurative and literal interpretations were faster where context and resolutions were congruent in high-literality idioms. By contrast, low-literality idioms showed only a benefit for figurative interpretations compared to literal ones, regardless of context. Thus, for native speakers, context is an important cue for idioms with a higher potential for literal interpretations (i.e., high-literality), and a figurative interpretation is more salient for low-literality idioms. This outcome suggests that L1 experience deems whether or not context should be considered in meaning integration.

The current study is a replication of the study in **Chapter 5B** for highly-proficient L2 readers. In particular it asks: (1) how sensitive L2 readers are to linguistic context in idiomatic processing, and (2) whether idiom literality limits these effects. In answering these questions, this study briefly looks into possible differences in L1 and L2 idiom processing and their abilities to integrate linguistic and lexical cues (i.e., context and literality) in online processing.

5.1.1 Context and literality effects

As in L1 research, studies have shown that context is beneficial to L2 figurative meaning activation, (e.g. Cieślicka & Heredia, 2011). However, several studies have suggested that, even in the presence of biasing context, literal meaning is activated more quickly than figurative meaning (e.g. Cieślicka, 2006; Siyanova-Chanturia, Conklin, & Schmitt, 2011), and possibly prioritized due to its salience (e.g., Giora, 1997). In line with L1 research (e.g., Holsinger & Kaiser, 2013), Cieślicka, Heredia, and Olivares (2014) compared the effects of context in an eye-tracking study on dominant and non-dominant bilinguals in high-literality idioms. They found that biasing context impacted the fixations and reading times of figurative and literal interpretations, but they determined that non-dominant bilinguals did seem to show a literal preference, in contrast with dominant bilinguals. As the current study investigates high-proficiency L2 speakers, they may behave like non-dominant bilinguals and prefer literal readings regardless of context. This result would indicate a clear difference from our L1 speakers.

Literality has been investigated in only a limited number of L2 studies, (e.g. Cieślicka, 2006), and so far no differences between the idiom types based on literality have been found to impact L2 processing. However, the scarcity of such studies does not allow for clear predictions. Thus, following the L1 results from **Chapter 5B**, we expect that L2 readers are either a) too inexperienced with idioms and do not show any differences between the idiom types or b) able to take literality as a salient cue that limits the impact of context.

5.2 SELF-PACED READING STUDY

We conducted a phrase-by-phrase self-paced reading study following the same method presented in **Chapter 5B** with three manipulations: idiom *literality* (high- or low-literality idioms), *biasing context* (literal or figurative bias), and *resolution-type* (literal or figurative resolution). By comparing the results from the present study to the results from native speakers of English, the expectation is that the results will highlight possible differences in L1 and L2 abilities to integrate context and use idiom literality in their online processing strategies.

5.2.1 Method

5.2.1.1 Participants

Forty-seven native speakers of German (40 female; average age of 24.48, SD = 3.01) were paid to take part in the study. All were highly proficient in English (5 years minimum education), with an average self-reported proficiency of 6.05 (SD = 0.68) on a 7-point scale. 7 participants were left-handed.

5.2.1.2 Materials

The experiment consisted of 22 target and 78 filler trials. Target trials began with a biasing context (noun phrase + relative clause) followed by the infinitive form of the idiom, then by a resolution (prepositional phrase) congruent or incongruent with contextual expectations and ended with two additional short phrases shared across all conditions. Phrases were controlled for letter length and average word frequencies, with only minimal differences between conditions (see analysis section for more information). See example in Table 1.

Table 5-8. Example stimuli

Biasing Subject		Idiom	Resolution	Resolution+1	
[The new schoolboy]	just wanted to	*break the ice*	[with his peers]	*sooner than later*	...
[The chilly Eskimo]			[on the lake]		

Note: For more examples, see Table 5-3 in **Chapter 5B**.

Eleven low- and 11-high literality idioms were included. All idioms were taken from **Chapter 5B**; were pre-tested and normed (DIN database, **Chapter 2**) based on familiarity (L1 and L2), literality, length, and constituent frequency; had the same syntactic structure (to-infinitive verb + determiner + noun); and were short and unpredictable (i.e., the final word is necessary in order for idiom recognition where no context is present). Any minor differences were accounted for by including these variables in the analysis.

Sentences were also normed for strength of biasing context and plausibility of the sentences. While there were differences between items (see **Chapter 5B**), the norming values for each item were included as fixed variables in the final analysis, where warranted. The stimuli are listed in Appendix D.

Three regions of interest were identified for our analysis: the Idiom, the Resolution, and the following phrase (Resolution + 1), labeled in Table 5-8. Following [4], we expected the effects of (mis-) matching contexts to show up in the Resolution + 1 region. Since this region is identical across conditions for each idiom, observed differences in reading times cannot be attributed to differences in lexical items or syntactic structure.

5.2.1.3 Procedure

The experiment was programmed and executed using E-Prime (Psychology Software Tools, 2013), and a subsequent idiom recognition test (to ensure participants knew the idioms) and a language background questionnaire were completed using Adobe Acrobat. Participants sat in a lab setting and wore noise-cancelling headphones. Responses were recorded with the spacebar of a keyboard.

The study began with four practice trials and included a self-timed break at the half-way point. A standard moving-window phrase-by-phrase presentation was used in which each phrase was masked by hyphens corresponding in length to the phrase to be presented. Only one phrase was seen at a time. Phrase-by-phrase rather than word-by-word was used in order to better mimic natural reading patterns and avoid a forced incremental processing pattern, (e.g., Jegerski, 2014), which might directly affect the questions under investigation in the present study. 30 of the 90 reading trials (including all target trials) were followed by multiple choice comprehension questions, and 10 were followed by arithmetic problems.

5.2.1.4 Results

We used R (R Core Team, 2013) and lme4 (Bates, Maechler et al., 2015) for linear mixed effects analyses of the relationships between *biasing context* (figurative, literal*), resolution type* (figurative, literal), and *idiom literality* (high-, low-literality) on reading times in three areas of interest: *idiom*, *resolution*, and *resolution + 1* regions. Only trials with correctly answered comprehension questions were analyzed. One participant was excluded for poor performance on the comprehension questions (56%). Two idioms (*play the field* and *turn the tide*) were also excluded as they were unknown to more than half of participants. Based on a visual inspection of all reading times, outliers beyond overall minimum and maximums were excluded per section as reading times less than 200ms (all regions) and greater than 1600ms (*idiom*, 4.4%), 1800ms *(resolution*, 4.4%), and 1700ms (*resolution + 1*, 3.8%), respectively. The difference in time windows is justified based on the length and difficulty of the corresponding areas of interest. See Table 2 for means in each region.

Table 5-9. Mean reading times (ms), all regions

		Figurative Context		**Literal Context**	
Region	***Resolution type***	**high-lit. RTs**	**low-lit. RTs**	**high-lit. RTs**	**low-lit. RTs**
Idiom	-	770	740	771	765
Resolution	figurative	781	818	773	874
	literal	767	804	754	794
Resolution+1	figurative	685	738	752	770
	literal	763	827	716	802

Note: The idiom region does not distinguish between resolution-types as this section occurs prior to the resolution.

Each region was analyzed individually on normalized log-transformed values. With the exception of the idiom section (details below), independent measures and fixed factors, numerically centered around 0, were coded and included as follows: *biasing context* (figurative: 0.5, literal: -0.5*)*, *resolution type* (figurative: 0.5, literal: -0.5), and *idiom literality* (high-literality: 0.5, low-literality: -0.5). Theoretically relevant factors were kept in the models, and items and participants were included

as random factors with random slopes. A maximally justified random effects structure was determined for each region by step-wise selection and model comparison, (e.g. Bates, Kliegl et al., 2015) consulting RePsychLing (Baayen et al., 2015). Additionally, fixed factors of *trial order*, *region length* (number of letters), *average lexical frequency*, and norming values for *idiom familiarity*, *plausibility*, and *strength of context* were also included in full models, and they were eliminated from the models if backward step-wise selection showed that they did not improve the model. The analyses for each section are discussed in the following, with a focus on the resolution + 1.

5.2.2 Analysis and discussion

The model for reading times in the idiom region did not include *resolution type* or the norming factor for *plausibility* as the ending had not yet been encountered at this point. Significant effects of *region length* (β = .031, t= 2.33, p < .05) and *trial order* (β = -.06, t= -5.48, p < 0.001) suggest that longer idioms were read more slowly and reading times improved during the experiment. These effects are consistent throughout the successive regions but did not contribute to the overall findings other than to justify their inclusion in the models. They will therefore not be discussed further in the subsequent regions. The full model output can be seen below in Table 5-10.

Table 5-10. Idiom region full model output

Factor	β	t	p	
Intercept	6.582	140.289	2.00E-16	***
Biasing Context	-0.006	-1.144	0.262	
Idiom Literality	-0.031	-0.803	0.430	
Ave. Lexical Frequency	-0.029	-1.665	0.113	
Region Length	0.039	2.329	0.031	*
Trial Order	-0.057	-5.480	4.85E-07	***
Biasing Context x Literality	0.026	0.639	0.531	

Note: *p<.05 **p<.01 ***p<.001

In the *resolution* region, significant effects of *region length* (β = .028, t= 2.20, p < .05), *trial order* (β = -.05, t= -5.01, p < 0.001), and *literality* (β = -.07, t= -2.70, p < 0.05) were present. The effect of literality shows the pattern that resolutions following high-literality idioms were read more quickly than those following low-literality idioms. This is both consistent with L1 results and in line with research showing faster reading times for idioms with higher literality, (e.g. Cronk & Schweigert, 1992). The full model output can be seen below in Table 5-11.

Table 5-11. Resolution region full model output

Factor	*β*	*t*	*p*	
Intercept	6.615	137.485	2.00E-16	***
Biasing Context	-0.010	-0.475	0.638	
Literality	-0.074	-2.695	0.014	*
Resolution Type	0.052	1.366	0.187	
Region Length	0.028	2.196	0.041	*
Trial Order	-0.054	-5.013	4.33E-06	***
B. Context x Literality	0.006	0.160	0.873	
B. Context x Resolution Type	-0.021	-0.549	0.583	
Literality x Resolution Type	-0.053	-0.707	0.488	
B. Context x Literality x Resolution Type	0.061	0.804	0.422	

Note: *p<.05 **p<.01 ***p<.001

Finally, in the *resolution+1* region, main effects were found again for *region length* (β = .055, t= 3.08, p < .01), *trial order* (β = -.05, t= -4.99, p < 0.001), and *literality* (β = -.08, t= -2.33, p < 0.05). Additionally, a marginal effect of *resolution-type* (β = -.04, t= -1.89, p = 0.07) and a significant interaction of *biasing context* by *resolution-type* (β = -.11, t= -2.65, p < 0.05) were found. In order to better interpret the data in the presence of this interaction, the data were split by biasing context for further analysis. The full model output is displayed in Table 5-12.

Factor	*β*	*t*	*p*	
Intercept	6.568	140.841	2.00E-16	***
Biasing Context	-0.013	-0.548	0.590	
Literality	-0.082	-2.330	0.030	*
Resolution Type	-0.045	-1.885	0.073	.
Region Length	0.055	3.082	0.006	**
Trial Order	-0.045	-4.938	9.96E-07	***
B. Context x Literality	0.005	0.100	0.921	
B. Context x Resolution Type	-0.107	-2.645	0.016	*
Literality x Resolution Type	0.024	0.510	0.616	
B. Context x Literality x Resolution Type	-0.075	-0.924	0.366	

Note: *p<.05 **p<.01 ***p<.001

The additional analyses revealed that, for figurative contexts, figurative endings were significantly faster than literal ones (β = -.10, t= -4.07, p < 0.001), but this effect is not present for literal contexts and endings. The result for figurative contexts reflects the L1 data in **Chapter 5B**, but the literal context analysis does not. A close look at Figure 5-5, however, indicates that the same pattern exists, in which congruently resolving phrases for high-literality idioms appear faster than incongruent, whereas for low-literality idioms, figurative resolutions are preferred overall. While this result is not

significant in the L2 data here, the smaller effect size is likely due to a lower proficiency, a conclusion in line with the differences shown in Cieślicka, Heredia, and Olivares (2014).

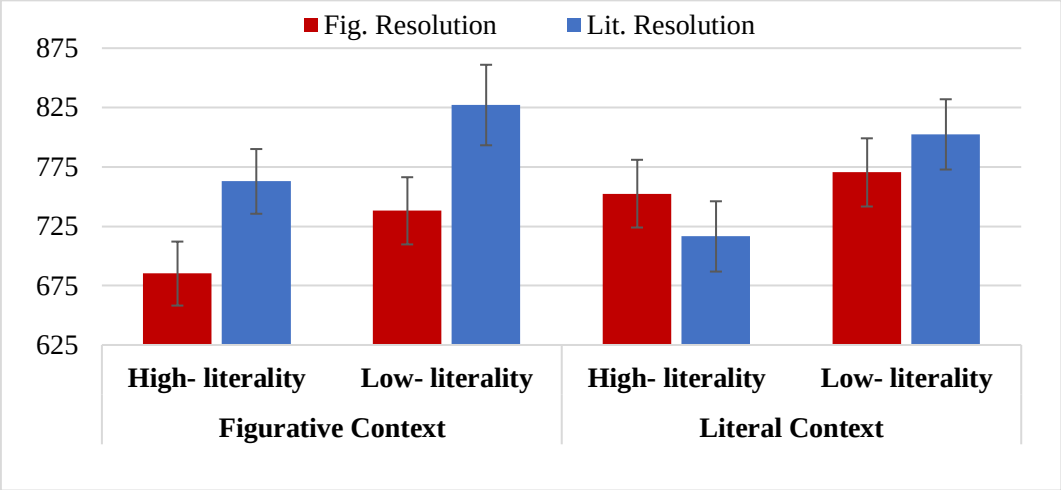


Figure 5-5. Average RTs in the Resolution+1 Region

Note: Averaged reading times are divided into low- (left) and high-literality (right) idioms by biasing context. The red bars represent mean reading times following a figurative resolution (Fig.Resolution) and the blue bars following a literal resolution (Lit.Resolution), all with standard error bars.

5.3 CONCLUSION

The current study asked (1) how sensitive L2 readers are to linguistic context in idiomatic processing, and (2) whether idiom literality limits these effects. We conclude that L2 readers do integrate linguistic context in idiomatic processing, particularly for figurative interpretations. For literal interpretations, our results are inconclusive, nor did we find significant evidence that literality impacts this process. However, based on the patterns present in low-literality idioms, and their similarities to the previous L1 data presented in **Chapter 5B**, we conclude that this lack of finding may be due to proficiency. Furthermore, we would expect that as proficiency increases, so too will language experience, both which should be reflected in processing patterns. Thus, looking at varying proficiencies in future studies may help confirm whether this pattern continues.

These results are in line with previous L2 research that indicates that context matters for figurative interpretations (Siyanova-Chanturia, Conklin, & Schmitt, 2011), but we maintain that our results do not provide evidence for a literal priority that is distinct from L1 processing (e.g., Cieślicka, 2006). Rather, like L1 readers, experience with the language can indicate when context is more or less relevant in reaching an interpretation. Though, more studies with direct L1 and L2 comparisons are needed to investigate these claims.

6 PHRASAL LEARNING IS A HORSE APIECE: NO MEMORY ADVANTAGES FOR IDIOMS IN L1 AND L2 ADULT LEARNERS[9]

Sara D. Beck & Andrea Weber

Abstract

Native (L1) and to some extent non-native (L2) speakers have shown processing advantages for idioms compared to novel literal phrases (e.g., Conklin & Schmitt, 2012). This study tested whether these advantages generalize to memory. When encountering new idioms, L2 learners must cope with both figurative meaning and unfamiliar vocabulary, and therefore possible competition between phrasal and word learning, particularly since word meaning need not contribute to idiomatic meaning. This study employed a learning paradigm to test whether there is a memory advantage for idioms compared to literal phrases in adult L1 and L2 learners. Additionally, we asked whether the presence of new words competes with phrasal learning. In Experiment 1, L2 learners showed equal recall for literal and idiomatic phrases in which either all words were highly familiar, or one word was unfamiliar. Though, unfamiliar words decreased overall recognition and were also remembered better in literal compared to idiomatic phrases. In Experiment 2, L1 learners showed no recall differences between phrase types and an increase in recognition in the presence of unfamiliar words. We conclude that there is no inherent memory advantage for idioms based on figurativeness alone, and word- and phrasal meaning interact differently in learner groups.

[9] This chapter is adapted from the submitted version of the following article: Beck, Sara D.; Weber, Andrea (submitted). Phrasal learning is a horse apiece: No memory advantages for idioms in L1 and L2 adult learners.

6.1 INTRODUCTION

Idioms challenge language learners as they represent a figurative meaning that is not achieved overtly through language composition. *To shoot the breeze*, for example, means to have an informal conversation, and neither the meaning of "shoot" nor "breeze" contributes straightforwardly to this meaning as is the case in compositional literal language. Rather, in order to understand the idiom, one must simply learn the definition, and this holistic meaning may be associated with the idiomatic phrase or unit in a word-like manner (e.g., Wray, 2002). Not only are adult native (L1) and highly proficient non-native (L2) users of a language able to interpret idioms quickly and without difficulty (e.g., Beck & Weber, 2016, **Chapter 3**), but they also may be able to do so more quickly than with comparable novel phrases (e.g., Underwood, Schmitt, & Galpin, 2004). Additionally, these processing advantages extend to memory in native children (Reuterskiöld & van Lancker Sidtis, 2012). That is, newly learned idioms may be remembered better than comparable literal phrases.

In less proficient language users and learners, however, the same processes are decidedly more difficult, and processing advantages of idioms compared to novel phrases are less stable and sometimes even show slower processing when an idiom is used figuratively (e.g., Siyanova-Chanturia, Conklin, & Schmitt, 2011). It is also unclear whether there may be memory advantages for idioms in L2 learners, as this topic has not yet been investigated to our knowledge. Furthermore, in addition to storing and accessing the phrasal meaning signalled by the idiomatic unit, the individual words present additional challenges for language learners, particularly in cases where these words are less familiar (e.g., *breeze* in *shoot the breeze*). This competition may cause phrasal learning to be affected by the presence of such words in idioms, and, in turn, word-learning to be negatively affected in idiomatic phrases.

In the current study, we compared the learning of figurative and literal phrases. In contrast to previous processing studies which compared different phrases, identical phrases were used and either learned with an idiomatic or a literal meaning. This allowed us to prevent the possible influence of structural and semantic differences between phrases on the outcome. The study addressed three questions in two experiments: 1. Is there a memory advantage for newly learned idioms compared to literal phrases for L2 learners? 2. Does competition between unfamiliar words and phrases interact and affect memory? 3. Do adult native speakers show a comparable memory advantage for newly learned idioms and identical literal phrases?

6.1.1 Idiomatic processing and memory

Though there are few studies dealing directly with memory for idioms compared to similar novel phrases, there is a multitude of literature dealing with idiomatic processing and how it may compare to novel, literal language processing. Idiomatic processing theories vary widely but generally take basis

in the idea that idiomatic meaning is at least partly non-compositional. This unique feature of idioms has caused disagreement in how storage and processing occur. Namely, idioms may be stored holistically as units to which meaning must then be accessed similarly to individual words in a list-like manner either following the rejection of literal meaning or immediately upon encounter (e.g., Bobrow & Bell, 1973; Swinney & Cutler, 1979); they may be represented as associations of words from which meaning is accessed only after enough information has accumulated (e.g., Cacciari & Tabossi, 1988); or they may be represented on another conceptual level, for example a lexical-syntactic level as a "superlemma", where figurative meaning immediately competes with words and composition via spreading activation during processing (e.g., Sprenger, Levelt, & Kempen, 2006). Each theory of processing attempts to explain the presence of the so-called *idiom superiority effect*, which describes the fast nature of idiomatic processing compared to other types of phrases. While there is not a consensus on how this process occurs, recent psycholinguistic literature favors the latter two ideas along with other hybrid approaches to idiomatic processing. Such approaches are supported by consistent evidence that processing occurs quickly and without cost in advanced language users while also remaining flexible and subject to linguistic influence, for example from idiomatic properties and context (e.g., Beck & Weber, 2016; Titone & Connine, 1994; Titone & Libben, 2014; **Chapter 5B**).

When compared to similar novel (literal) phrases, idioms consistently show processing advantages in the sense of faster or more efficient reading (e.g., Conklin & Schmitt, 2008; Gibbs, 1980; Siyanova-Chanturia, Conklin, & Schmitt, 2011). Underwood et al. (2004) looked at eye-movements during the reading of passages containing embedded idioms compared to novel phrases. Native speakers fixated less on the final constituent word of an idiom compared to the same word in a non-formulaic context (e.g., "teeth" in *met the deadline by the skin of his teeth* vs. *the dentist looked at his teeth*), and the duration of fixations was also shorter. In a self-paced reading paradigm, Conklin and Schmitt (2008) also found that reading times were shorter for idiomatic phrases compared to more closely matched controls (e.g., *hit the nail on the head* vs. *hit his head on the nail*). Siyanova-Chanturia, Conklin, and Schmitt (2011) replicated these results with eye-movements and found that not only were fixations shorter on idiomatic phrases, but first-pass and total reading times were also shorter than matched control phrases. Interestingly, however, the latter two studies also compared literal and figurative readings of these familiar idioms and found that native speakers did not show any differences between the two readings. Thus, it appears that the processing advantages associated with speed of reading are not associated solely with figurative uses alone. Rather, these advantages may stem from familiarity with the phrase.

This processing advantage also appears to extend to memory for idioms in native speakers. Reuterskiöld and Van Lancker Sidtis (2012) looked at young girls' retention of idioms after a single exposure in conversation compared to novel literal phrases. They found that both age groups tested (8-

9 and 12-14 years) scored higher on recognition and comprehension of target idioms compared to literal phrases and non-target idioms. Authors concluded that idiomatic phrases are acquired holistically, and this allows for a more rapid retention compared to similar literal phrases. However, few participants were tested (only 6 per group) and previous knowledge of and exposure to the target idioms was measured only via questionnaires and cannot be entirely ruled out. Nonetheless, these results are generally supported by earlier memory studies conducted on idiomatic or unitized word pairs suggesting that idiomatic word pairs are represented in a holistic or unitized manner (e.g., Horowitz & Manelis, 1973; Schachter & McGlynn, 1989). Schachter and McGlynn (1989) compared implicit memory for such unitized (idiomatic) word pairs (e.g., sour grapes) to non-unitized word pairs (e.g., soft soap). In free association tasks, implicit recall of non-unitized word pairs was improved by all elaborative study measures (e.g., defining tasks and synonym naming) whereas such improvement was not seen in all elaborative conditions for the unitized pairs. The differences between these types of word pairs support the view that idioms are represented holistically, and processing and memory advantages may be a result of such storage. Unclear, however, is whether these processes are as immediate as suggested by the Reuterskiöld and Van Lancker Sidtis (2012) study, and whether this is related to the figurative meaning associated with idioms or rather their formulaic nature.

Unlike in native speakers, there is mixed evidence that processing idioms is faster than novel literal phrases in non-native speakers. L2 readers were examined in Underwood et al. (2004), Conklin and Schmitt (2008), and Siyanova-Chanturia, Conklin, and Schmitt (2011), and each found slightly different results. Underwood et al. (2004) found that while L2 readers fixated less on final constituent words in idioms compared to nonformulaic phrases, the duration of these fixations was not shorter as it was found to be in L1 readers. Considering reading times of such phrases, however, Conklin and Schmitt (2008) found consistent advantages for idioms compared to controlled novel phrases in both figurative and literal contexts, just as in L1 readers. However, Siyanova-Chanturia and colleagues (2011) failed to replicate these results using eye-tracking. Not only did L2 readers show no differences in the number of fixations, first-pass, and total reading times between idioms and controlled novel phrases, unlike their L1 counterparts, but figurative readings of idioms were actually read more slowly than literal readings. Crucially, each experiment used different methods to test previous knowledge of the idioms, and it is unclear whether all speakers were familiar with the idioms used. Thus, idiomatic units may not carry the same processing advantages for L2 speakers present for L1 speakers, and their figurative nature presents additional processing challenges.

While we are unaware of any existing L2 memory studies dealing with idioms compared to novel literal phrases, there is a plethora of literature focusing on how to increase learning of and memory for idioms in L2 learners. While the challenges of learning idioms are often discussed in research (e.g., Cooper, 1999), there are a number of techniques that increase L2 idiom learning.

Processes that improve recall of idioms and other multi-word units include, but are not limited to, noticing exercises (e.g., Boers et al., 2006), typological enhancement (e.g., Peters, 2012), the use of dictionaries and look-up tasks (e.g., Laufer, 2011), translation exercises (e.g., Laufer & Girsai, 2008), and focus on sound repetition (e.g., Boers & Lindstromberg, 2005). Additionally, idiom-specific differences such as imageability and transparency may also play a role in their retention. For example, Steinel, Hulstijn, and Steinel (2007) found that highly imageable idioms, or those easily pictured mentally (e.g., *keep a straight face*), have an advantage in receptive learning over those less easily pictured (e.g., *hang fire*). Additionally, they found that transparency, the overlap between figurative and literal meaning, also affected recognition (see also Tiv, 2016). Thus, while comparisons between idioms and similar control phrases in L2 memory are lacking, it is clear that elaborative processes are helpful, and some idioms have inherent memory advantages over others.

6.1.2 Phrasal- and word- competition

Unlike most other formulaic language, idioms have a figurative meaning that is not derived from the meaning of the constituent words, and meaning may be stored holistically (e.g., Wray, 2002). While a holistic representation may lead to the processing advantages described above, it also causes a disconnect and possible competition between activation of the meaning of the individual constituent words and the phrasal meaning. Where individual word meaning is unnecessary, there is evidence that native speakers may give literal constituents less attention. In a memory study on unitized word pairs, Horowitz and Manelis (1973) found evidence that recognition of individual words in idiomatic pairs was worse than recognition in non-unitized word pairs (e.g., *cold war* vs. *cold egg*). However, a study investigating the effect of word frequency on the phonetic duration of words in highly frequent multi-word units suggests that even in frequent units, such as idioms, the frequency of words continues to have effects on the entire phrase (Arnon & Cohen Priva, 2015). The authors suggest that such sequences need not be represented holistically, and that constituent words retain their ability to affect the phrase. Thus, where phrases with equal frequency are compared, whether literal or figurative, it may be the case that the competition between word- and phrasal-meaning remains. This idea is also substantiated by evidence that both L1 and L2 speakers still activate both literal constituent and figurative meanings in online processing (e.g., Beck & Weber, 2016a).

However, there is some debate as to whether L1 and L2 speakers diverge in their attention to constituent word meanings. L2 speakers may rely more on literal constituent meanings during language processing as a whole because of the saliency of literal word meaning in L2 language use and learning (see e.g., Cieślicka, 2006; Giora, 1997). Evidence for this tendency in L2 idiom processing was based on cross-modal priming results from Cieślicka (2006) in which priming was greater for literal constituents of an idiom compared to the figurative meaning of the phrase. Later studies using L1 controls, however, did not support this conclusion (Beck & Weber, 2016a; van Ginkel & Dijkstra,

2019). Thus, it is unclear whether the results from Horowitz and Manelis (1973) in which constituent words were recalled better in non-idiomatic compared to idiomatic word pairs would also apply to L2 speakers. Additionally, experimental methods looking at meaning activation in idiom processing have focused on highly familiar idioms and assumed vocabulary knowledge for all constituent words. For L2 speakers, however, it may be the case that an idiom is both unknown as a phrase and contains unfamiliar constituent vocabulary. This competition between word- and phrasal-meaning may have consequences both on the memory for the phrase as well as the memory for the individual words.

Though few studies have compared the learning of idioms or other formulaic phrases with familiar and unfamiliar vocabulary, collocations may offer some insight into the topic. Kasahara (2011) found that L2 English learners' recall for collocation meanings containing new words was better than recall for the individual words alone (i.e., the L1 Japanese equivalent of *delicious morsel* was better recalled than the Japanese word for *morsel* alone). By combining familiar and unfamiliar words in the target phrases, familiar words (e.g., delicious) serve as cues to better recall the meaning of the phrases (e.g., delicious morsel). However, individual words in collocations differ from idioms critically in their contribution to overall meaning: individual words in collocations often provide additional meaning information, unlike in idioms. Because of this difference, it may neither be the case that individual words are recalled better in idioms nor that the idioms containing unfamiliar words themselves are better remembered. Furthermore, it is also possible that word- and phrase-learning compete with one another when both the phrase and individual words are unfamiliar, and attention to individual words in comparable literal and figurative units may therefore differ between L1 and L2 speakers.

6.2 Experiment 1: Non-native speakers

The present study looked at memory for novel phrases that were either learned with a figurative meaning or a literal meaning by non-native speakers. Based on established memory and processing advantages for native speakers, we asked whether the advantage for idioms over literal phrases holds for L2 speakers. Additionally, we explored the competition between word- and phrasal-meaning by examining the effects of unfamiliar vocabulary on memory for both the phrase and the unfamiliar word. By conducting a training and testing paradigm in which the identical fabricated phrases were either used with an idiomatic meaning or a literal meaning, we could more accurately test differences that are based solely on figurativeness and the impact of familiarity of constituent vocabulary. Our paradigm used a learning task followed by recognition tasks on form and meaning in addition to a translation task for words. Phrases differed either in meaning (figurative or literal) or by the familiarity of a content noun to language learners (familiar or unfamiliar).

This experiment addresses the first two research questions. Considering the first question of whether there is evidence of memory advantages for idioms compared to identical literal phrases following learning, we predicted that if L2 learners acquire idioms rapidly as units, like L1 users do (e.g., Reuterskiöld & Van Lancker Sidtis, 2012), recognition of figurative phrases should be better than literal phrases. On the other hand, if figurative meaning poses a particular challenge for L2 learners (e.g., Siyanova-Chanturia, Conklin & Schmitt, 2011), and this disadvantage extends to memory and learning, literal phrases should be better recalled than figurative phrases. Considering the second question of whether competition between phrasal- and word-learning interact and affect memory for the phrases and/or unfamiliar words, we predicted that it does. However, it may do so in several ways. If the combination of familiar and unfamiliar words improves memory for figurative phrases as it does for collocations (e.g., Kasahara, 2011), then the presence of unfamiliar words should improve the recall of phrases overall as both literal and figurative phrases will have the same combination of familiar and unfamiliar words. Though, if this effect in collocations is a result of compositional meaning benefits, then this improvement may only be seen in literal phrases. Additionally, if the saliency of this literal word-meaning cue competes with phrasal meaning (e.g., Cieślicka, 2006), it may decrease recall for figurative phrases. Finally, considering the recall of the unfamiliar words, we predicted that if less attention is given to these constituents in figurative phrases, as is the case for L1 speakers (e.g., Horowitz & Manelis, 1973), there should be better recall for unfamiliar words in literal phrases compared to figurative phrases.

6.2.1 Methods

6.2.1.1 Participants

65 non-native speakers of English (German L1) participated in the study on two separate days. Participants (45 female, average age of 25.35, *SD*= 3.11) were given financial compensation for their time and were recruited from the University of Tübingen via university-wide emails. All participants identified as native speakers of German, though their English abilities varied. Scores on the LexTale lexical decision task in English (Lemhöfer & Broersma, 2012) varied (range: 45-100, mean 78.75, *SD*= 12.70), and participants' self-reported English ratings averaged across reading, writing, listening, and speaking (on a scale from 1-7, 1 as very poor and 7 as native-like) were 5.49 (*SD*= 0.69, range= 3.75-7). On average, participants reported 7.69 years (*SD*= 4.50) of formal English instruction. Five participants were left-handed.

6.2.1.2 Materials

A total of 30 target phrases were developed for the study, with four variations per phrase. Each phrase had a literal paraphrase and a figurative meaning and also had a variation in which one well-known

(familiar) word was replaced with an unfamiliar, less frequent word (e.g., *bell* and *bugle*). The figurative meaning was invented by native speakers to be a plausible idiomatic meaning for the phrase, though not transparent enough that the meaning is apparent without definition, and the literal meaning was a paraphrase using simple vocabulary (e.g., *the morning bell sounded*, literal: the church bell rang early in the day, figurative: a return from the distraction of day-dreaming). An example phrase with all four variations is displayed in Table 6-1. The phrases were generally short (mean phrase length in letters: 21.73, in words: 5.06) with no significant differences in length between any four variations. Phrase variations were divided into four counter-balanced lists for learning and testing between participants.

Table 6-1. Example phrase stimuli

Target Phrase	Meaning	Noun-type	Figurativeness
the morning ***bell*** sounded	a return from the distraction of daydreaming	familiar	figurative
the morning ***bell*** sounded	the church bell rang early in the day	familiar	literal
the morning ***bugle*** sounded	a return from the distraction of daydreaming	unfamiliar	figurative
the morning ***bugle*** sounded	the musical horn played early in the day	unfamiliar	Literal

6.2.1.2.1 Familiar/unfamiliar noun selection

The familiar and unfamiliar nouns were as similar as possible but differed systematically in their familiarity to the German L2 speaker group tested in this study (familiar or unfamiliar words). Unfamiliar target nouns were first chosen based on their conceptual and semantic comparability to the familiar words (e.g., *bell* and *bugle* are both types of instruments that make musical sounds). Additionally, since learners would later be tested on German translations of the nouns, we pre-tested whether native German students in several advanced courses of English studies at the University of Tübingen were able to translate the nouns.

A total of 143 nouns were tested on average by 24.86 students of English in four rounds. In the pre-test, students were given a list of up to 60 English nouns and were asked to translate as many of them as possible into German. Included in the list were a mixture of familiar and unfamiliar nouns. Only nouns were selected for Experiment 1 that all students either correctly translated (familiar) or none did (unfamiliar). The subjective familiarity differences identified between the two noun-types were also confirmed with differences in lexical frequencies. T-tests showed that known nouns differed systematically in frequency per million (familiar mean: 32.20, unfamiliar mean: 0.67, t= 3.912, p< 0.001) based on the $SUBTLEX_{US}$ corpus (Brysbaert & New, 2009). This frequency difference in combination with the pre-test ensured that the words presented as unfamiliar were indeed unfamiliar to the L2 participants.

6.2.1.2.2 Pre-study: Idiom material testing

In addition to pre-tests for the target nouns in the study, the idioms were also pre-tested as new idioms to ensure that their meanings were novel and not predictable based on their constituent words as well as for their levels of transparency and imageability, which have been shown to affect idiom-learning in native speaker adults (e.g., Steinel et al., 2007). Native speakers of English completed three short online tasks using "SurveyGizmo," to determine 1) whether the meaning of the idiom could be derived only from the parts, 2) a scalar rating for transparency, and 3) a scalar rating for imageability. The tasks were completed in two parts, first meaning and imageability were combined, and second, participants rated the idioms' transparency.

Twenty native speakers of English (15 female, average. age=31, SD= 8.54) participated in the 20-30-minute online study and were compensated by entry into a gift card lottery. 30 target idioms with familiar words and 20 filler idioms that were highly familiar for native speakers were included in the online study. Only targets with familiar words were included since the familiar and unfamiliar words are conceptually and visually similar and should therefore be equivalent in their contribution to the imageability and transparency on an idiom. The filler idioms were included to provide a spectrum of imageable to non-imageable as well as highly transparent to non-transparent idioms and to prevent frustration from interacting with only unfamiliar phrases.

First, participants were asked to give the meaning of the idiom and to rate its imageability. Participants were instructed to paraphrase the idiomatic meaning to the best of their abilities, if known, or to come up with a guess, if possible. Imageability was rated on a scale from 1-7 (1: TRUE, 7: FALSE) by rating the truth of the statement: "I could easily visualize this idiom." Then, in the second part of the survey, participants saw the idiom with a paraphrase of the idiomatic meaning. Transparency was defined as "how related the literal and figurative meanings of the idioms are" and participants rated the truth of the statement: "The figurative meaning of this idiom has a lot in common with its literal meaning." on a scale from 1-7 (1: TRUE, 7: FALSE).

None of the idioms' meanings were guessed by the participants. Additionally, idioms' imageability averages ranged from 2.90-5.95 with an average of 4.94 (SD= 0.74). The idioms were generally rather imageable which reflects the choice to include concrete nouns in each idiom. On the other hand, idioms were not very transparent with an average of 2.54 (SD= 1.02) and a range from 1.25-4.75. This also reflects of the choice to include only idioms with meanings not predictably derived from the literal words. While neither property includes a full range of values, the averages for each idiom are included in additional linear mixed effects on the experimental results for each task. Ratings for each idiom are included in Appendix E.

6.2.1.3 Procedure

6.2.1.3.1 Learning

The study was conducted in a sound-attenuated room in the LingTüLab at the University of Tübingen on two separate days. A learning task was completed on the first day and a short series of testing tasks on the second. On the first day, the LexTale vocabulary test (Lemhöfer & Broersma, 2012) was conducted using Presentation® Software, and the learning task followed using a PowerPoint presentation on a desktop computer. This session took about 20 minutes in total. The instructions for the learning task were self-guided slides, and the learning task itself was timed using automatically progressing slides. In the instructions, participants were informed that they would be learning both literal and figurative phrases, and an example of each was given and presented with the automatic timing to be used in the experiment. Participants were also informed that in the next session, they would complete three short tasks testing what they had learned, but no further information was given.

Each participant learned one of the four lists of phrases. The target phrases were presented on slides seen by each participant twice in separate rounds. Each slide contained the target phrase in black, German glosses of all nouns in the phrases in green (maximum of two), and a paraphrase (either literal or figurative) in blue (see Figure 6-1). Each slide was presented for a total of 8000ms. During the first 2000ms, only the phrase with German glosses was displayed, and the addition of the paraphrase followed and was presented for an additional 6000ms. Once a participant had seen each phrase, a 2-minute timed break was imposed, after which participants repeated the procedure again following a single mouse click. The order of phrase-presentation was reversed for half of all participants for each list.

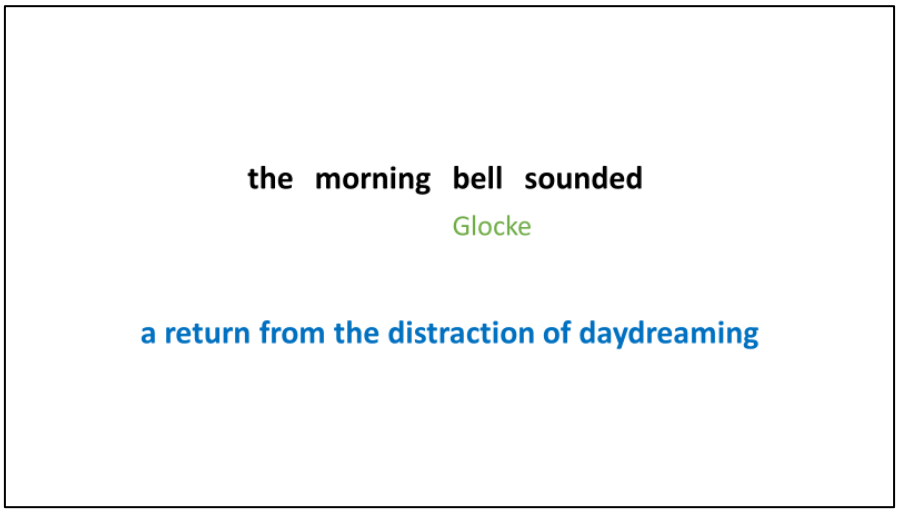


Figure 6-1. Example slide from learning task

Note: Depicted is the fully presented slide. The blue paraphrase appeared after 2000ms.

6.2.1.3.2 Testing

The testing session took place 3 days later, as a rule. However, in the case of illness or unforeseen scheduling conflicts, some participants came back 2 or 4 days later. This interval was chosen after pilot testing with learning and testing on 2 consecutive days showed ceiling performance and null effects. By testing 3 days after the learning session, performance remained high but was no longer at ceiling, providing more optimal conditions for hypothesis testing. Participants returned to the same room in which the learning task was completed. In this testing session, three tasks were conducted consecutively using jsPsych (Leeuw, 2015) to measure memory differences in: form recognition, meaning recognition, and a translation task. After the completion of the testing tasks, participants finished a language background questionnaire and were then informed more precisely about the phrases they learned.

6.2.1.3.2.1 Form recognition

The form recognition task was used to test recognition of the phrases learned during the learning phase. The test included all 30 phrases (15 idiomatic and 15 literal) presented to each person based on the list they learned in addition to 30 new fillers. All 60 phrases were presented in a randomized order. Fifteen easy fillers were phrases that were lexically different from the learned phrases (7 of them also contained unfamiliar words), and 15 difficult fillers were rather similar to the learned phrases; difficult fillers were in fact target phrases with the alternative word used in another list (i.e., *the morning bugle sounded* instead of *the morning bell sounded*). In this task, participants were instructed to decide quickly but accurately whether the phrase presented is one of the exact phrases learned in the first session. Responses were measured with a keyboard press marked with green for the dominant hand and red for the non-dominant hand. Participant responses based on correctness were coded for analysis.

6.2.1.3.2.2 Meaning recognition

The meaning recognition task tested recognition of the learned meaning of the idiomatic phrases. Participants were presented individually with all 15 idiomatic phrases learned in session 1 and given three multiple choice options. The multiple-choice options included 1) the correct answer, 2) the meaning of another learned idiom, and 3) an idiomatic meaning from an idiom not included in the experiment. For options 2) and 3), only plausible meanings were presented with each idiom. Additionally, a sliding scale was present at the bottom of the screen asking participants how sure they were of their choice (from not at all to very sure). The scale was not numbered, but yielded values from 1-100 in the output. Measures of correctness and subjective ratings of sureness were collected for analysis.

6.2.1.3.2.3 Translation

The translation task tested learning of the designated target word (familiar or unfamiliar) in the learned phrases. Participants were given all 30 learned phrases individually and asked to translate the bolded and underlined word back into German. They were asked to recall the word given in the first session and instructed to take a guess if unsure or to skip the word if nothing came to mind. Open-ended translations for each target word were recorded and later scored by judges (see Analysis section for details).

6.2.1.4 Analysis

R (R Core Team, 2013) was used to analyze the results of each task individually using mixed effects regression models on the tasks for *Form*, *Meaning*, and *Translation*. While the analyses differed slightly between tasks, the same procedure was used. *Correctness* (1=correct, 0=incorrect) was used as the dependent variable in each case, and, depending on the task, *Phrase-Type* (Figurative or Literal, coded as .5 and -.5 respectively) and *Word-Type* (familiar or unfamiliar, coded as .5 and -.5 respectively) were included as fixed effects. Items and participants were also included as random factors with random slopes, where justified. The random effects structure was tested by consulting RePsychLing (Baayen, Bates, Kliegl, & Vasishth, 2015), ensuring that models were sufficiently recognized. In order to account for further variation, factors such as *Trial*, *LexTale Score*, and *Testing Interval*, numerically centred around zero, were also included in the full models, and backward stepwise selection was used to eliminate these non-theoretically relevant factors in the case that model fit was not improved. Additional analyses were conducted to determine whether the factors *Imageability* and *Transparency* impacted the results. The results of each task will be discussed below individually.

6.2.2 Results

6.2.2.1 Form results

Overall, participants performed well on the task, and total correct responses averaged 82.59% (SD= 9.11) across the entire task. Examining target items only, performance was still at 72.82% (SD= 15.68), and participants displayed a wider range of performance averages (30-100%). *Correctness* as a function of *Word-Type* and *Phrase-Type* is graphed in Figure 6-2.

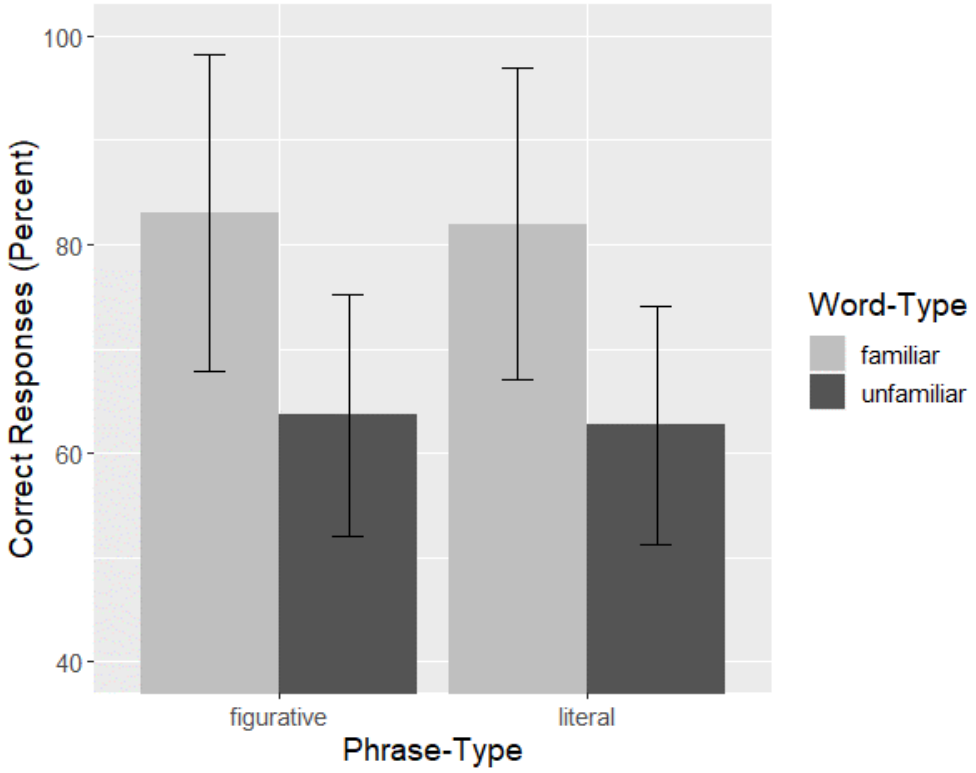


Figure 6-2. L2 Form accuracy

Note: The overall correctness of participant responses to form recognition are graphed as a function of *Phrase-Type* by *Word-Type*. Familiar words are represented by light grey and unfamiliar words by dark grey, and bars are the standard error of the mean.

The estimates from the final model can be seen in Table 6-2. Of the main factors under investigation, only *Word-Type* was significant (*ß*= 0.19, *t*= 10.49, *p*< 0.001), and no interaction was present. The effect of *Word-Type* confirms the pattern in Figure 6-2, namely, that the presence of an unfamiliar word significantly decreased accuracy, whereas the type of phrase (figurative or literal) had no effect. The effect of *Word-Type* suggests that phrasal- and word-meaning competition may be occurring (e.g., Cieślicka, 2006), and the lack of a *Phrase-Type* effect is novel, as this effect appears to be robust in L1 speaker groups (e.g., Reuterskiöld & Van Lancker Sidtis, 2012). Though, a lack of interaction does not conform to our expectations.

Table 6-2. L2 Form accuracy LMER output

Fixed Effects	***β***	***(SE)***	***t***	***Pr(>|t|)***	
Intercept	0.7283	(0.022)	33.097	2.00E-16	***
Phrase-Type	0.0038	(0.018)	0.206	0.837	
Word-Type	0.1933	(0.018)	10.496	2.00E-16	***
Trial	-0.0400	(0.009)	-4.285	1.92E-05	***
Phrase-Type x Word-Type	-0.0018	(0.037)	-0.048	0.962	
Random Effects	**Variance**	***SD***			
Subject	0.0190	0.138			
Item	0.0032	0.057			

Note: *p<.05 **p<.01 ***p<.001

An additional post-hoc analysis was conducted on the figurative phrases in order to ensure that the transparency and imageability of the idioms included did not play a role in our results (see e.g., Steinel et al., 2007). In order to do so, the same LMER model, excluding *Phrase-Type* as a factor, was used on the figurative phrases. *Imageability* and *Transparency*, centered around zero, were added to the models, and neither was significant (*Transparency*: ß= 0.00, t= -0.05, p= 0.963, *Imageability*: ß= -0.01, t= -0.38, p= 0.704). Furthermore, backward stepwise selection confirmed that they also did not improve model fit. Thus, we conclude that idiomatic differences based on these factors did not affect the results significantly.

6.2.2.2 Meaning results

The results from the meaning recognition task were recorded in two steps: first, the accuracy of the multiple-choice response (*Correctness*) and second, the ratings on how sure participants were of their choice (*Sureness*). In both cases, responses concerned only the figurative phrases, and therefore only *Word-Type* and not *Phrase-Type* was used as an independent variable in the analyses. While both the correctness and confidence of choice are reported below, the focus will remain on correctness.

6.2.2.2.1 Correctness

Overall performance on the task was high, and total correct responses averaged 77.13% (SD= 17.75) across the entire task. Participants varied widely in their average correctness ranging from 20-100%. Accuracy on the task as a function of *Word-Type* is graphed in Figure 6-3.

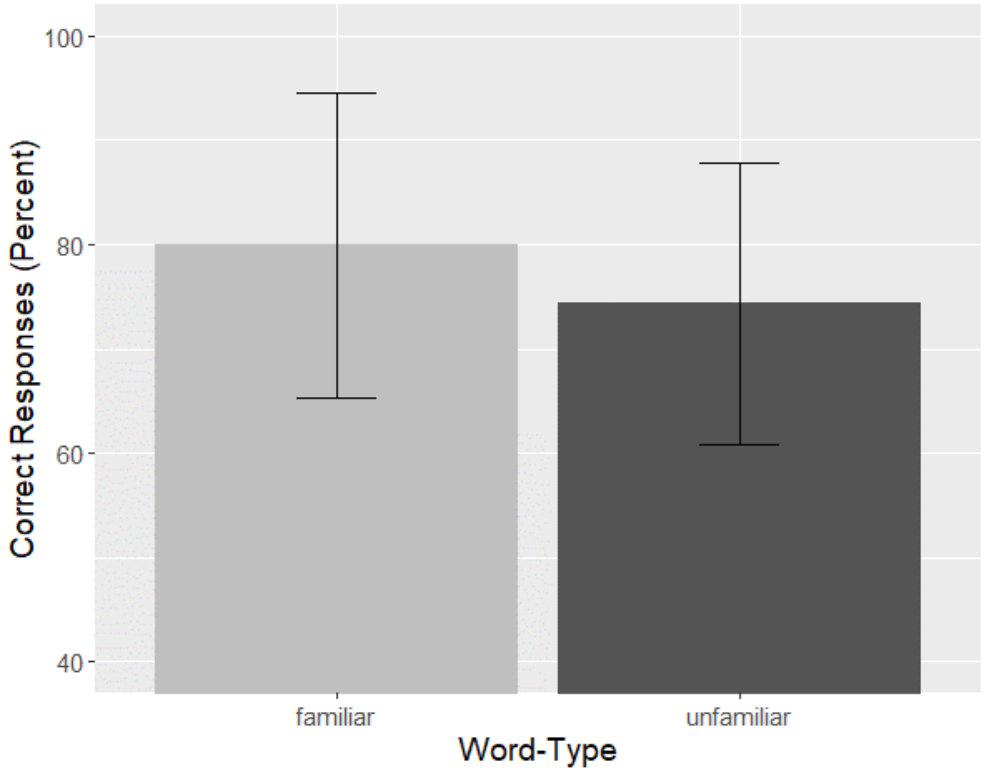


Figure 6-3. L2 Meaning accuracy

Note: The overall correctness of participant responses to the meaning of the phrase are graphed as a function of *Word-Type*. Familiar words are represented by light grey and unfamiliar words by dark grey, and bars are the standard error of the mean.

The analysis revealed a significant effect of *Word-Type* (*ß*= 0.19, *t*= 10.49, *p*< 0.001). No other factors were significant, nor improved model fit, and therefore, the final model included only this factor in addition to items and participants as random factors and slopes. The output of this model can be seen in Table 6-3. This result replicates those of form recognition in that unfamiliar words decreased performance in meaning recognition.

Table 6-3. L2 Meaning Accuracy LMER Output

Fixed Effects	***β***	***(SE)***	***t***	***Pr(>|t|)***	
Intercept	0.7713	(0.029)	26.902	2.00E-16	***
Word-Type	0.0563	(0.024)	2.305	0.0214	*
Random Effects	**Variance**	***SD***			
Subject	0.0221	0.149			
Item	0.0100	0.100			

Note: *p<.05 **p<.01 ***p<.001

Additionally, the factors *Imageability* and *Transparency* were neither significant, nor did they improve model fit (*Transparency*: *ß*= 0.02, *t*= 1.08, *p*= 0.286, *Imageability*: *ß*= 0.01, *t*= 0.71, *p*= 0.481). Thus, these factors did not play a significant role in the participants performance on meaning recognition, as in form recognition.

6.2.2.3 Confidence of choice

Confidence of choice (*Sureness*) was measured on a sliding scale (1-100) and was selected together with the multiple-choice answer used in the above analysis. Participants were generally confident in their choices, and the average sureness rating was 69.99% (SD= 30.92), though there was wide variability in this choice (range: 0-100).

While this measure was not the focus of our investigation, we also conducted linear mixed effects models on this score to determine if confidence of choice simply aligns with correctness, or whether the word-type also impacted this choice independently. In our model, we used *Word-Type* and *Correctness* (coded as -0.5 for incorrect, 0.5 for correct) as our independent predictors. The output of this model can be seen in Table 6-4. The analysis confirms the results reported on accuracy: both *Word-Type* (*ß*= 9.07, *t*= 4.85, *p*< 0.001) and *Correctness* (*ß*= 24.17, *t*= 11.72, *p*< 0.001) were significant predictors, though the effect size for *Correctness* was greater than *Word-Type*. No interaction was present. These results suggest that participants were overall more confident in their choices when they were correct but also when there were only familiar words present; unfamiliar words decreased overall sureness independent of the correctness of their choice.

Table 6-4. L2 Sureness LMER output

Fixed Effects	***β***	***(SE)***	***t***	***Pr(>|t|)***	
Intercept	63.389	(2.248)	28.201	2.00E-16	***
Word-Type	9.073	(1.868)	4.858	1.4E-06	***
Correctness	24.170	(2.061)	11.727	2.00E-16	***
Correctness x Word-Type	3.401	(3.825)	0.889	0.374	
Random Effects	**Variance**	***SD***			
Subject	146.930	12.121			
Item	56.340	7.506			

Note: *p<.05 **p<.01 ***p<.001

6.2.3 Translation results

In the translation task, participants were asked to provide a translation from English into German for one word in each phrase, and to reproduce the gloss shown in the learning phrase, if possible. The responses were first scored as correct or incorrect by hand, and *Correctness* was subsequently analyzed for the translations of unfamiliar words based on *Phrase-Type*. The process of scoring and then the analysis is described below.

6.2.3.1 Scoring

All responses were scored, including translations of both familiar and unfamiliar words. Answers were categorized as correct (1) or incorrect (0). The judges were two native speakers of German with highly proficient English skills. Where the two scorers disagreed, a third judge made the decision. Answers were scored as correct if they were one of the following: a direct match with the translation given in the learning task; a conceptually correct, close synonym with the word given in the learning task (e.g., *Armreif* and *Armband*, English: bangle); a partial match that was underspecified (e.g., *Fischschwarm* and *Schwarm*, English: shoal); or misspelled or abbreviated versions of the correct word (e.g., *Pullover* and *Pulli*, English: sweater). All other responses were marked as incorrect.

6.2.3.2 Correctness

Overall, participants did well on the task in spite of not being given any direct instructions on or indication of a word-learning task, and they had an average of 78% (SD= 17.88) correct translations including all target words. Additionally, performance on the familiar (mean: 97.02%, SD= 5.89) and unfamiliar words (mean: 33.33%, SD= 22.32) was significantly different (t= 39.67, p< 0.001), a result in line with pre-tests suggesting that the familiar words were indeed highly familiar compared to the unfamiliar words.

The analysis again used *Correctness* as the dependent variable, and *Phrase-Type* as an independent factor. In addition to the factors of *Trial*, *LexTale Score*, and *Testing Interval*, *Word Frequency* (based on word from frequency from Brysbaert & New, 2009, also centered around zero) was also used in the full model to account for any additional variation based on frequency within this group of unfamiliar words. The output of the final model can be seen in Table 6-5. Only *Phrase-Type* (ß= -0.05, t= -2.20, p< 0.05) and *LexTale Score* (ß= 0.08, t= 3.31, p< 0.01) remained in the final model, and both factors were significant. As shown in the graph in Figure 6-4, unfamiliar words were more accurately translated when they were encountered as part of a literal phrase compared to figurative phrases. Additionally, higher LexTale scores predicted better performance on the translation task.

Table 6-5. L2 Translation LMER output

Fixed Effects	***ß***	***(SE)***	***t***	***Pr(>|t|)***	
Intercept	0.333	(0.041)	8.212	8.36E-11	***
Phrase-Type	-0.056	(0.025)	-2.207	0.0276	*
LexTale Score	0.085	(0.026)	3.316	0.0015	**
Random Effects	**Variance**	***SD***			
Subject	0.032	0.179			
Item	0.030	0.173			

Note: *p<.05 **p<.01 ***p<.001

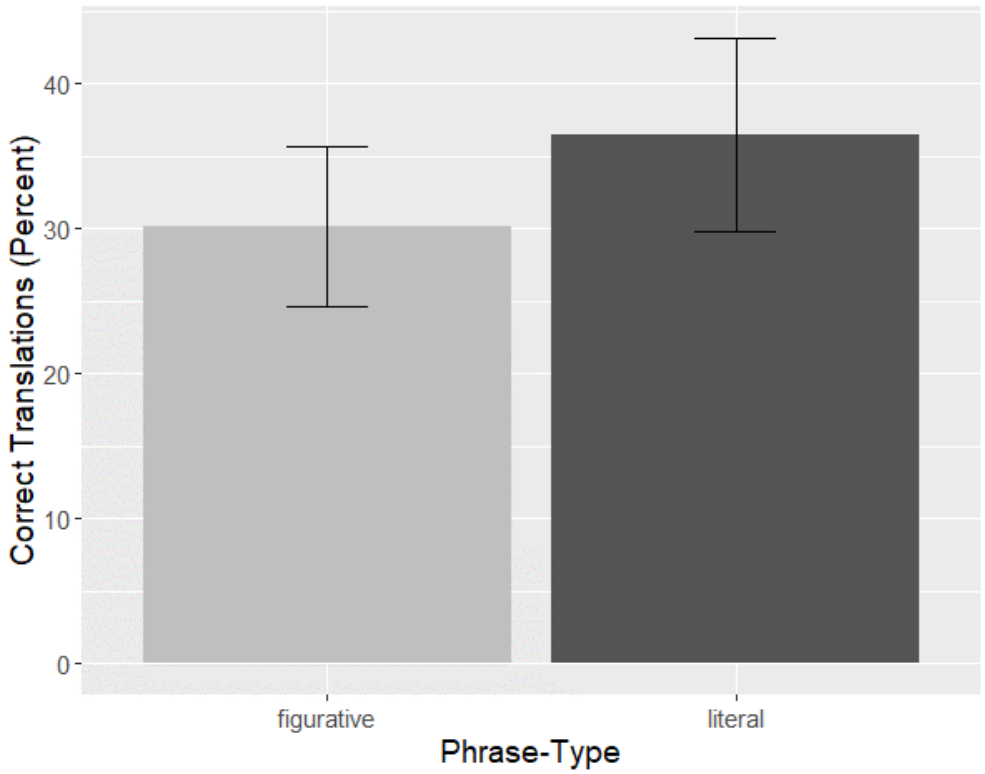


Figure 6-4. L2 Translation accuracy

Note: Overall correctness of participant translations of unfamiliar words are graphed as a function of *Phrase-Type*. Translations occurring in figurative phrases are represented by light grey and literal phrases by dark grey, and bars are the standard error of the mean.

Additionally, the factors *Imageability* and *Transparency* were applied only to the idiomatic targets, to see whether these factors may have impacted the results of the figurative phrases. They neither improved model fit nor were significant (*Transparency*: *ß*= 0.02, *t*= 0.66, *p*= 0.514, *Imageability*: *ß*= 0.01, *t*= 0.35, *p*= 0.724). Thus, we confirm that these factors did not play a role in participants' performance on the translation task.

6.2.4 Interim discussion

In the first experiment, we compared figurative and literal interpretations of identical phrases learned with equal exposure. We did not find evidence of memory differences based on the figurativeness of a phrase. These results may either suggest that L2 learners, in contrast to L1 learners, do not acquire figurative language rapidly in a manner different than literal language, or they suggest that there is generally no memory advantage when identical phrases are compared. In the case of the latter, L1 speakers should also show no memory advantage for the materials of Experiment 1. Note though that Reuterskiöld and Van Lancker Sidtis (2012) used low-frequency idioms and surveyed the parents of

the children involved on their knowledge and use of the idioms in question, and it may be the case that the children were exposed to the idioms at some point before the experiment. Additionally, as exposure happened in a naturalistic setting, phrases may have also presented differently (i.e., with differing prosodic cues) in a manner that could not be controlled for (see e.g., Van Lancker Sidtis, 2003). Thus, additional evidence is needed in order to determine whether this affect applies to both learner groups.

Considering the competition between phrasal and word meaning, we found that unfamiliar words impacted phrasal learning. Critically, this impact did not differ between figurative and literal phrases. We interpret the decrease in performance where unfamiliar words were present to a more divided attention on word- and phrasal-meaning. Although participants were asked to "learn the meaning of the phrase," the presence of an unfamiliar word still served as a salient cue, and more attention may have been given to the German glosses while these phrases were being presented. We also found that individual words were learned better when they were part of a literal phrase as shown by the translation task. This effect is generally in line with the L1 recognition results found by Horowitz and Manelis (1973) in which similar effects were found with unitized compared to non-unitized word pairs. Thus, while figurativeness did not impact the recall of the phrases, it did impact how well translations of unfamiliar words were recalled.

Overall, the negative result concerning memory differences between learned phrases based on figurativeness alone leaves questions unanswered. In particular, whether this lack of effect is unique to L2 learners or whether it also holds for adult L1 learners where identical phrases and learning environments are used. In order to determine whether this is the case, we conducted a follow-up experiment using the same phrases and procedure on adult native speakers.

6.3 EXPERIMENT 2: NATIVE SPEAKERS

Experiment 2 aims to replicate Experiment 1 for native speakers, and it asks whether idiomatic advantages for adult native speakers extend to memory when phrases are controlled for vocabulary and exposure. If L1 speakers do show advantages for idioms compared to literal phrases, then we expect better performance for figurative compared to literal phrases in the recognition tasks. Such a result would indicate a difference in storage and processing between L1 and L2 speakers. If, however, again no memory advantages are found, then this may suggest a more general effect of the experimental item control and learning conditions of the experiment.

Although not the focus of Experiment 2, we will also examine how the presence of unfamiliar words affects phrasal learning in native speakers. While it may be the case that the presence of less familiar words negatively impacts L1 phrasal learning similarly to L2 phrasal learning, it may also be the case that unfamiliar words are more unexpected for L1 speakers, and these words serve as highly

salient cues. In this case, we would expect an improvement of recall for either all phrases with such words, or an interaction following our last prediction.

6.3.1 Methods

6.3.1.1 Participants

Twenty-five native speakers of English participated in the study on two separate days. Participants (14 female, average age of 31.76, *SD*= 11.04) were given financial compensation for their time and were recruited from the University of Tübingen via university-wide emails. At the time of the study, participants had lived in Germany for an average of 5.83 years (*SD*= 5.87) but still reported using English an average of 74% of the time in their day-to-day lives (SD= 21.27). Two participants were left-handed.

6.3.1.2 Materials

The materials used were the same as in Experiment 1.

6.3.1.3 Procedure

6.3.1.3.1 Learning

The study was conducted, as in Experiment 1, in a sound-attenuated room in the LingTüLab Lab at the University of Tübingen on two separate days including a learning task on the first day and a testing task on the second. Unlike Experiment 1, participants were told that the figurative phrases were examples of uncommon idioms taken from other languages or varieties of English. This would ensure that these participants also believed themselves to be learning existing idiomatic phrases. Participants were also informed that in the next session, they would complete two short tasks testing what they learned, but no further information was given.

As in Experiment 1, participants learned one of the four lists of phrases. The target phrases were presented on slides seen by each participant twice. Each slide contained the target phrase in black and a paraphrase (either literal or figurative) in blue. The only visual difference from Experiment 1 was the exclusion of German glosses for nouns. All timing remained the same.

6.3.1.3.2 Testing

The testing session took place 3 days later for all participants. Participants returned to the same room in which the vocabulary was learned for testing. In this session, only two of the three tasks from Experiment 1 were conducted consecutively using jsPsych to measure learning: form recognition and meaning recognition. The translation task was not included, as no translations were presented during

the learning experiment. After the completion of the testing tasks, participants completed a language background questionnaire and were then informed more precisely about the phrases they learned.

Both form and recognition tasks were the same as in Experiment 1.

6.3.1.4 Analysis

R (R Core Team, 2013) was used, as in Experiment 1, to analyze the results of each task individually using mixed effects regression models on the tasks for *Form* and *Meaning*. Coding for the dependant variable *Correctness* (1=correct, 0=incorrect) as well as the factors of *Phrase-Type* (Figurative or Literal, coded as .5 and -.5 respectively), *Word-Type* (familiar or unfamiliar, coded as .5 and -.5 respectively), and *Trial (*numerically centered around zero) were identical to Experiment 1. The results are discussed by task below.

6.3.2 Results

6.3.2.1 Form results

Overall, total correct responses averaged 85.60% (SD= 8.19) across the entire task, slightly higher than L2 performance in Experiment 1 (82.59%). On target items, performance was still 78.13% correct (SD= 15.95), and participants again displayed a wide range of performance averages (33.33-96.66%). Overall correctness as a function of *Word-Type* and *Phrase-Type* is graphed in Figure 6-5.

The final LMER model estimates can be seen in Table 6-6. As in Experiment 1, *Word-Type* showed a main effect, though only marginally (*ß*= -0.06, *t*= -1.74, *p*= 0.09), and neither an effect of *Phrase-Type*, nor an interaction was present. The effect of *Word-Type* displays the trend seen in Figure 6-5, namely, that the presence of an unfamiliar word increased accuracy, whereas the type of phrase (figurative or literal) had no effect. While the effect of *Word-Type* showed the opposite pattern displayed by L2 learners in Experiment 1, the lack of effect of *Phrase-Type* replicates the pattern in Experiment 1.

Table 6-6. L1 Form accuracy LMER output

Fixed Effects	***ß***	***(SE)***	***t***	***Pr(>|t|)***	
Intercept	0.7809	(0.033)	23.572	2.00E-16	***
Phrase-Type	-0.0096	(0.028)	-0.348	0.7280	
Word-Type	-0.0650	(0.037)	-1.749	0.0935	.
Trial	-0.0506	(0.014)	-3.585	3.60E-04	***
Phrase-Type x Word-Type	0.0536	(0.055)	0.971	0.33	
Random Effects	**Variance**	***SD***			
Subject	0.0026	0.051			
Item	0.0205	0.143			
Word-Type	0.1251	0.570			

Note: *p<.05 **p<.01 ***p<.001

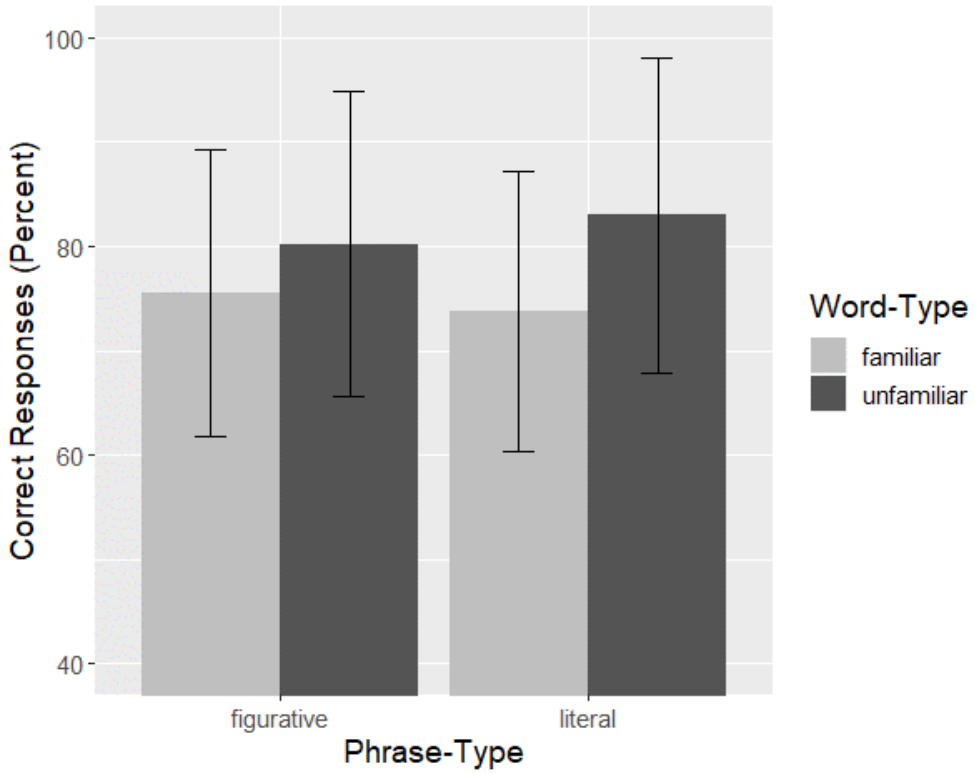


Figure 6-5. L1 Form accuracy

Note: The overall correctness of participant responses to form recognition are graphed as a function of *Phrase-Type* by *Word-Type*. Familiar words are represented by light grey and unfamiliar words by dark grey, and bars are the standard error of the mean.

As in Experiment 1, we conducted an additional post-hoc analysis on the figurative phrases in order to ensure that the transparency and imageability of the idioms included did not influence the results significantly (see e.g., Steinel et al., 2007). Again, the same LMER model, excluding *Phrase-Type* as a factor, was used on the figurative phrases. *Imageability* and *Transparency*, centered around zero, were added to the models, and neither were significant, nor did they change the findings in our full models (*Transparency*: ß= 0.00, *t*= 0.17, *p*= 0.864, *Imageability*: ß= 0.01, *t*= 0.92, *p*= 0.361). Furthermore, backward stepwise selection confirmed that they also did not improve model fit. Thus, we conclude that idiomatic differences based on these factors did not affect the results.

6.3.2.2 Meaning results

Following the procedure of Experiment 1, both accuracy of the multiple-choice response (*Correctness*) and ratings on how sure participants were of their choice (*Sureness*) were measured and will be discussed below. Only *Word-Type* was used as an independent variable in the analyses as the task

concerned only figurative phrases. Like Experiment 1, both accuracy and confidence of choice were measured, but the focus will remain on accuracy.

6.3.2.2.1 Correctness

Correct responses averaged 80.26% (SD= 18.40) across the entire task. Participants varied widely in their average correctness ranging from 33.33-100%. Accuracy on the task as a function of *Word-Type* is graphed in Figure 6-6.

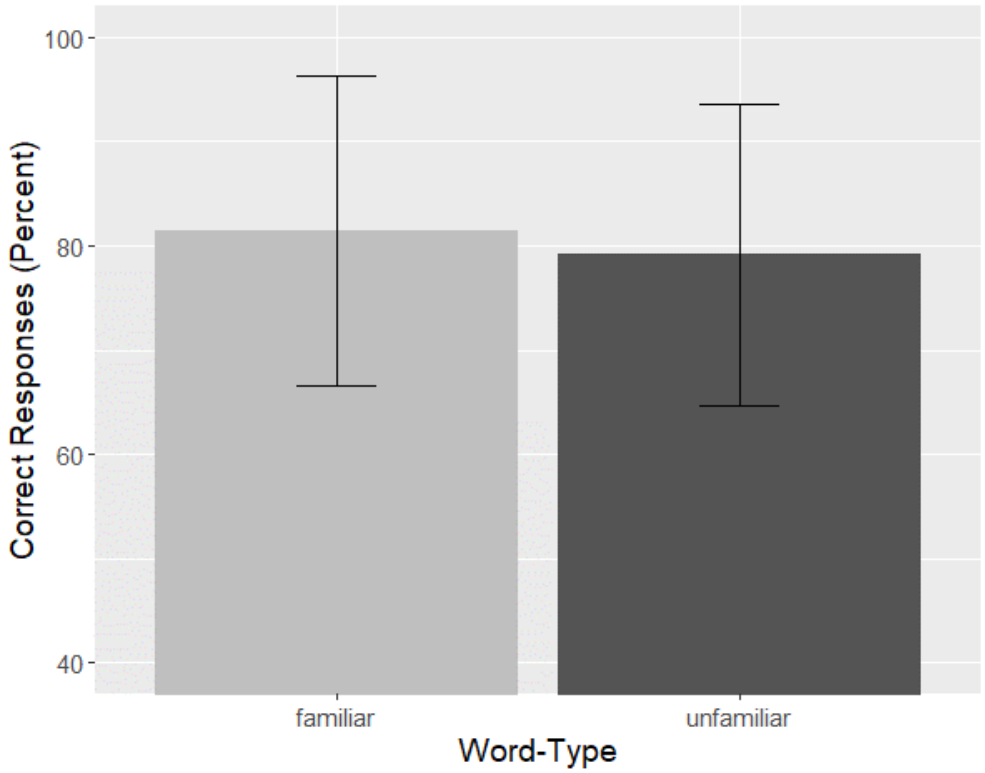


Figure 6-6. L1 Meaning accuracy

Note: The overall correctness of native speaker participant responses to the meaning of the phrase are graphed as a function of *Word-Type*. Familiar words are represented by light grey and unfamiliar words by dark grey, and bars are the standard error of the mean.

The analysis revealed no significant effects, though a marginal effect of *LexTale Score* was present. Unlike the results of the Form recognition task and the L2 results, *Word-Type* was not a significant factor. The output of the final model can be seen in Table 6-7.

Table 6-7. L1 Meaning LMER output

Fixed Effects	***ß***	***(SE)***	***t***	***Pr(>|t|)***	
Intercept	0.8028	(0.042)	8.912	2.00E-16	***
Word-Type	0.0277	(0.037)	0.743	0.4660	
Random Effects	**Variance**	***SD***			
Subject	0.0256	0.1600			
Item	0.0137	0.1170			
Word-Type	0.0485	-0.6900			

Note: *p<.05 **p<.01 ***p<.001

Additionally, the factors *Imageability* and *Transparency* were added and checked with model comparison, as in Experiment 1. Neither factor was significant, nor improved model fit (*Transparency*: *ß*= 0.03, *t*= 1.36, *p*= 0.184, *Imageability*: *ß*= 0.01, *t*= 0.69, *p*= 0.495). Thus, these factors did not play a role in the L1 and L2 participants' performances on meaning recognition, as in form recognition.

6.3.2.2.2 Confidence of choice

L1 participants were also generally confident in their choices, and the average sureness rating was 75.91% (SD= 27.42), and participants used the entire range of the scale (range: 0-100).

We conducted LMER models on this score using *Word-Type* and *Correctness* (coded as -0.5 for incorrect, 0.5 for correct) as our independent predictors. The output of this model can be seen in Table 8. The analysis confirmed the results reported on accuracy: only *Correctness* (*ß*= 25.93, *t*= 8.68, *p*< 0.001) was a significant predictor, not *Word-Type*. No interaction was present. These results suggest that participants were overall more confident of their choices when they were correct. As in the L1 accuracy results, and unlike the L1 results for meaning recall, *Word-Type* does not seem to affect the accuracy of meaning recall for native speakers.

Table 6-8. L1 Meaning LMER Output

Fixed Effects	***ß***	***(SE)***	***t***	***Pr(>|t|)***	
Intercept	68.0517	(3.042)	22.369	<2.00E-16	***
Word-Type	0.6897	(2.618)	0.263	0.7920	
Correctness	25.9341	(2.987)	8.683	<2.00E-16	***
Word-Type x Correctness	2.0271	(5.418)	0.374	0.71	
Random Effects	**Variance**	***SD***			
Subject	136.16	11.669			
Item	58.48	7.647			

Note: *p<.05 **p<.01 ***p<.001

6.4 GENERAL DISCUSSION

As in Experiment 1, Experiment 2 did not show evidence of memory differences in phrasal recall based on the figurativeness of a phrase. This repeat result from Experiment 1 suggests that the lack of difference between the two phrase-types is not unique to L2 learners. Although Reuterskiöld and Van Lancker Sidtis (2012) compared memory for literal and figurative phrases in L1 children following a single exposure (e.g., *cross swords with someone* and *in a new school*) and found better recall of the new idioms, the possibility of previous encounters with the idiomatic phrases cannot be entirely discounted. Thus, a "rapid uptake" (pp.221) of idiomatic phrases cannot be based on the figurativeness of the idiom alone. One possible reason for the difference in results may be based on the method of exposure to the idioms. Whereas the current study used visual presentation with a focus on phrasal learning, exposure to the phrases in the previous study occurred auditorily in a naturalistic setting in which experimenters used the phrases as part of normal communication during an activity. As the authors point out, prosody may be a crucial component of retaining formulaic language such as idioms, and such cues may have played a role in signalling the idiomatic phrases more prominently (see e.g., Van Lancker, Canter, & Terbeek, 1981; Van Lancker Sidtis, 2003).

Another possible reason that the current study did not show evidence of recognition differences between literal and figurative phrases of equal nature is that they were both learned as phrases. Previous studies which found differences in processing for L1 speakers comparing idiomatic phrases to novel literal phrases typically compared similar, though not equal phrases (e.g. *at the end of the day* vs. *at the end of the war*) or phrases with slight changes in word order (e.g., *hit the nail on the head* vs. *hit his head on the nail*). These novel phrases may have been encountered as phrases for the first time or as one of very few exposures (e.g., Conklin & Schmitt, 2008, Siyanova-Chanturia, Conklin, & Schmitt, 2011). In contrast, the same studies did not find differences between literal and figurative readings of the same idiomatic phrases. This result is also in line with research from Tabossi, Fanari, and Wolf (2009) comparing the processing of idioms and known clichés to matched controls. In a recognition task, participants showed differences between matched controls and both types of formulaic language, but idioms did not differ from clichés. The authors concluded that idioms have processing advantages because of their familiarity and not due to a holistic storage or their figurative nature. As both types of phrases in the current study were equally familiar based on exposure in the learning phase, our results would also be consistent with this characterization of idiomatic processing. Of course, it should also not be discounted that advantages present in processing may not transfer to memory.

Experiment 2 also showed differences in recall based on the presence of unfamiliar words, though the direction of this effect was in opposition to the L2 findings in Experiment 1. Namely, whereas L2 learners showed poorer recall of the phrase where unfamiliar words were present, L1

learners showed better recall when unfamiliar (and less frequent) words were present. Though, this effect did not significantly impact the recall of the meaning of the phrase for L1 learners as it did for L2 learners. This effect and lack of interaction, like Experiment 1 in nature though not direction, suggests that for idiomatic and literal phrases alike, properties of the individual words do affect the processing of the entire phrase. This overall result is in line with findings from Arnon and Cohen Priva (2015), in which word frequency impacted the phrasal duration of multi-word sequences, even when these phrases were highly frequent. Concerning the direction of this effect, there may be different forces at work for each learner group. For the L2 learners in Experiment 1, the saliency of literal word meanings during learning may have created competition between phrasal- and word-learning, regardless of figurativeness (e.g., Giora, 1997). Additionally, L2 learners may rely on individual words during learning overall and have more difficulty building multi-word units (e.g., Arnon & Christiansen, 2017). For the L1 learners in Experiment 2, we expect that while the unfamiliar words were infrequent, they were not entirely unknown. Rather than having the challenge of learning both phrasal- and word-meanings (the latter of which was not provided for L1 learners), the L1 learners could focus more on the phrase as a whole, and the surprise of unfamiliar L1 words may have caused the phrases containing such words to be more salient than other phrases. Rather than competition between word- and phrase-learning, as we suggest took place in the L2 learners, the presence of such words increased recall as L1 learners noticed but did not need to learn these words.

Finally, though there was no evidence of memory advantages for figurative phrases in the recall tasks in Experiments 1 and 2, the results of the translation task in Experiment 1 suggest that there may still be differences in the recall of word meaning based on the figurativeness of the phrase. L2 learners better recalled the German translations of unfamiliar words when they were part of literal phrases. While this effect seems in line with the L1 recognition results found by Horowitz and Manelis (1973) in which individual words in non-unitized (cold egg) compared to unitized word pairs (cold war) were recalled better, we do not suspect that this memory effect is due to a difference in storage or unitization between literal and figurative phrases. Had this been the case, we would have also expected differences in recall on this basis, as originally predicted. The effect found by Horowitz and Manelis might also be explained based on differences in exposure to the phrases used, which is not the case in the current study. Rather, the result from our L2 learners may indeed be more in line with the collocation results suggested by Kasahara (2011). The individual words in literal phrases from the current study were learned as part of (literal) phrases not unlike many collocations. In this manner, the unfamiliar words were part of a phrase containing familiar words in both figurative and literal phrases, but the unfamiliar words in literal phrases may have been better recalled due to this compositional meaning effect (i.e., all words contributed to the phrasal meaning). For the unfamiliar words in figurative phrases, as the words did not contribute to the phrasal meaning, such an effect is not present.

Thus, there is some evidence that figurativeness impacts the learning of individual words within phrases.

6.5 CONCLUSION

Motivated by previous research suggesting a difference in the acquisition and storage of idiomatic phrases compared to literal phrases (e.g., Reuterskiöld & Van Lancker Sidtis, 2012), the current study set out to examine the presence of this difference in adult native and non-native learners. Additionally, we aimed to examine the competition of constituent words and phrases in these learners, focusing on L2 learners. We found that recall for idiomatic phrases does not differ from literal phrases when these phrases are identical, and this holds for both native and non-native learners. Additionally, phrasal meaning does interact with the presence of unfamiliar constituent words in both types of phrases, but it does so differently for L1 and L2 learners. While the presence of unfamiliar words boosts recall for figurative and literal phrases in native speakers, it decreases overall recall for non-native speakers. Additionally, non-native speakers show more learning of these unfamiliar words in literal compared to figurative phrases. We also found that individual words have an impact on phrasal recall and processing, though, this impact differs between native and non-native learners. Unfamiliar words boost overall saliency to the phrase in L1 learners, whereas it increases competition between the phrase and the highly salient constituent words in L2 learners.

We conclude that idioms are not stored differently from literal phrases after a single learning exposure session. If holistic storage of idioms is systematically different from literal language and this difference is based on figurativeness alone, a single learning session with visual stimuli is insufficient in producing measurable differences. Future studies should continue to investigate memory differences between phrases more equal in vocabulary and exposure in order to better evaluate whether figurativeness or unitization may account for differences in findings.

7 L1 AND L2 IDIOM COMPREHENSION: AN FMRI STUDY[10]

Sara D. Beck & Alexander M. Rapp

Abstract

Idioms, multi-word units with figurative meanings, challenge theories of language processing for native (L1) and non-native (L2) speakers. Unclear is whether literal and figurative meaning activation involve similar or different processes, and even less clear is whether meaning constitution is the same for L1 and L2 speakers. While some authors propose differences in L1 and L2 idiom comprehension (e.g., Cieślicka, 2006), with L2 processing following a more literal, compositional strategy, others have not found evidence of such differences (e.g., Beck & Weber, 2016). Current L1 functional magnetic resonance imaging (fMRI) data indicates that there may be different neurophysiological activation in literal and figurative phrases (e.g., Rapp, 2019), but there is not yet comparable L2 data. If L1 and L2 processes differ, we also expect language-based differences in brain activation. Though, if proficiency plays a role, it may be the case that such differences decrease with increasing proficiency. Using fMRI, we investigated these assumptions by comparing meaning activation between literal and idiomatic sentences in and between L1 and L2 speakers. Thirty healthy, right-handed native English (13) and highly proficient non-native German (17) participants read short, non-biasing sentences containing English idioms or equivalent literal phrases (e.g., She just *went to pieces*/*London*.) in an fMRI scanner. While the results showed only minimal differences between literal and figurative processing in general, robust differences were found between L1 and L2 readers, and proficiency negatively correlated with various regions of the brain in L2 speakers during idiom comprehension. The results suggest generally similar processing in L1 and L2 speaker groups, but that L1 and L2 speakers do activate some different regions of the brain during comprehension. Furthermore, differences in activation may decrease over time.

[10] This study is the result of a collaboration with Dr. Priv. Doz. Alexander Rapp from the medical faculty at the University of Tübingen Hospital.

7.1 INTRODUCTION

Non-native (L2) competence includes mastering not only simple vocabulary and literal phrases but also figurative phrases such as idioms. Idioms are formulaic phrases with a figurative meaning. If, for example, John *spilled the beans,* it should be interpreted that John revealed a secret or a surprise and not that John made a mess of the legumes. This contrast between the compositional literal meaning and the intended figurative meaning of an idiom, which is, unlike metaphorical meaning, fixed in nature has made idioms an interesting topic for linguistic study. These properties also make idioms particularly challenging for non-native speakers. L2 speakers must both recognize the phrase as an idiom and be familiar with its figurative meaning in order to successfully comprehend its meaning. Although the challenges for L2 speakers in idiom comprehension are not new (e.g., Cooper, 1999), whether L1 and L2 speakers comprehend idioms differently and how proficiency plays a role in these processes is still unclear.

Studies on idiomatic processing have focused on differences between figurative and literal language and differences between idioms and novel expressions. Both avenues of research have found idioms to show superiority effects (e.g., Conklin & Schmitt, 2008; Gibbs, 1980; McGlone et al., 1994; Swinney & Cutler, 1979; Tabossi et al., 2009). Additional research has investigated whether idioms also show processing differences on a brain function level (e.g., Cacciari & Papagno, 2012; Citron, Cacciari, Funcke, Hsu, & Jacobs, 2019; Kana, Murdaugh, Wolfe, & Kumar, 2012; Mashal et al., 2008; A. M. Rapp, 2018; Yang et al., 2016; Zempleni, Haverkort, Renken, & A Stowe, 2007). Studies are available in English (Kana et al. 2012, Desai et al. 2013), Italian (Romero Lauro et al. 2012), Dutch (schuil et al. 2013) and Mandarin (Yang et al. 2013). However, all studies investigated native speakers. Studies in L2 are not yet available.

There is reason to assume that neural correlates of idiom processing in a second language might differ from first language correlates. Studies comparing brain activation in L1 and L2 processing show evidence that non-native speakers show increased activation compared to native speakers in a variety of language tasks (e.g., Hernandez & Meschyan, 2006; Shirley-Ann Rüschemeyer, Fiebach, Kempe, & Friederici, 2005; Shirley-Ann Rüschemeyer, Zysset, & Friederici, 2006). Though, there is also some evidence that such differences may diminish as proficiency increases (e.g., Perani & Abutalebi, 2005). Additionally, though this line of research is still limited in its conclusions, a picture of idiomatic processing has emerged that suggests that native (L1) speakers process and activate idiomatic meaning quickly and efficiently (e.g., Siyanova-Chanturia, Conklin, & Schmitt, 2011; Tabossi et al., 2009) using both hemispheres of the brain (see e.g., Cacciari & Papagno, 2012 for an overview), though there is less clarity for non-native (L2) speakers. Superiority effects in L2 processing are considerably less consistent (e.g., Cieślicka, 2006; Conklin & Schmitt, 2008; Siyanova-Chanturia,

Conklin, & Schmitt, 2011), and neuroimaging studies with L2 idiom processing are non-existent. Though possible differences in L1 and L2 idiomatic processing have been addressed in behavioural studies (e.g., S. D. Beck & Weber, 2016a; Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011), it's unclear as to whether differences are proficiency- or language-based. The current study addresses this research gap by using functional magnetic resonance imaging (fMRI) to examine possible differences in brain activation during idiomatic processing between L1 and L2 readers and adds to the body of research on brain activation during idiom comprehension. In doing so, we will first address the status quo on idiomatic processing research in L1 and L2 speakers first in behavioural research and then in neuroimaging.

7.1.1 Idiom processing in native and non-native speakers

Research on idiom processing in L1 and L2 speakers addresses the crucial contrast between an idiom's figurative meaning and possible literal meaning, the two of which need not be related. This difference likens them to other forms of figurative language such as metaphors and irony, but their conventional, fixed nature also makes them unique. This rather strict relationship between the idiomatic form and a word-like, conventional meaning suggests that idiomatic meaning may be stored and learned arbitrarily and retrieved directly in online processing rather than built in a compositional manner in L1 speakers (e.g., Bobrow & Bell, 1973; Swinney & Cutler, 1979). Whether or not this process is similar in L2 speakers remains to be seen, as research has so far provided conflicting evidence (e.g., S. D. Beck & Weber, 2016a; Cieślicka, 2006; Siyanova-Chanturia, Conklin, & Schmitt, 2011).

Research on L1 speakers has shown consistent processing advantages when idioms are compared with novel literal language, and such advantages have been shown, for example, in the form of faster phrasal recognition (e.g., Swinney & Cutler, 1979), better performance in monitoring tasks (e.g., Estill & Kemper, 1982), fast recognition of idiomatic meaning (e.g., Cacciari & Tabossi, 1988), and overall speed of reading as well as number of fixations in eye-tracking (Siyanova-Chanturia, Conklin, & Schmitt, 2011). These advantages may stem from a direct mapping of idiomatic meaning that is activated and retrieved quickly during online processing (e.g., Libben & Titone, 2008; Sprenger et al., 2006) or it may be due to the fixed nature of these phrases, as idioms show similar processing advantages to non-figurative formulaic phrases such as clichés and binomials (e.g., Carrol & Conklin, 2019; Wray, 2002). In any case, models of idiom processing generally address the difference between literal and figurative meaning and describe access as 1) literal-first, 2) figurative-first, or 3) a hybrid, interactive relationship between both meanings. While early research in L1 speakers favoured step-wise models of access demonstrating either literal- (e.g., Bobrow & Bell, 1973) or figurative-first (e.g., Swinney & Cutler, 1979) access, recent research has highlighted the complexity of this process in favor of hybrid models (e.g., Cacciari & Tabossi, 1988; Libben & Titone, 2008; Sprenger et al., 2006). Hybrid models are best able to explain the superiority effect that appears to occur in idiomatic

processing without discounting the complexities that arise through individual differences in idioms and linguistic contexts surrounding idiom comprehension.

While there has been a surge in L2 idiom processing research in recent years (e.g., S. D. Beck & Weber, 2016a; Cieślicka, 2006; Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011; van Ginkel & Dijkstra, 2019), there is still not a consensus on whether the same processing advantages found in native speakers apply to non-native speakers. While most studies identify similarities in L1 and L2 idiom processing (e.g., S. D. Beck & Weber, 2016a; Conklin & Schmitt, 2008; van Ginkel & Dijkstra, 2019), there is still some evidence suggesting that the figurative meaning of idioms may pose special challenges for L2 speakers (e.g., Cieślicka, 2006; Siyanova-Chanturia, Conklin, & Schmitt, 2011). In particular, literal language may place a more prominent role in L2 processing than L1 processing due to its saliency in learning (e.g., Cieślicka, 2006; Giora, 1997). This suggestion is also in line with researchers looking at differences in the acquisition of other types of formulaic language. Following the frameworks outlined by Wray (2002), Arnon and Christiansen (2017) suggest that L2 learners, already aware of individual word meanings, decompose chunks into their individual words, while native speakers need not do so. Whether a product of saliency or a learning strategy, it may be the case that non-native speakers focus more on literal meanings during language processing, and this may explain slower processing of idiomatic phrases (e.g., Siyanova-Chanturia, Conklin, & Schmitt, 2011) or increased activation of literal word-meanings during idiom-processing (e.g., Cieślicka, 2006).

7.1.2 Brain activation during idiom processing

Research on brain activation during L1 idiom processing has also made significant progress in recent years, while L2 research in this area has yet to be investigated. One critical line of investigation has been whether the right hemisphere is responsible for figurative language comprehension, including idioms and, for example, metaphors. An early study on metaphor comprehension looked at possible differences in hemispheres by comparing left- and right-brain damaged aphasic patients (Winner & Gardner, 1977). In a picture-metaphor matching task, the authors found that patients with left-hemisphere damage were able to correctly match metaphors to images, while those with right-hemisphere damage were not. Van Lancker and Kempler (1987) conducted a similar task using idioms, and found similar results. Both results suggested that figurative language processing occurs primarily in the right hemisphere (RH), unlike other types of lexical activation. Jung-Beeman (2005) later refined this idea in the Coarse Semantic Coding Theory, which suggests that each hemisphere is divided into fine- and course-coding (left and right hemispheres [LH and RH], respectively). In this case, figurative language, using more distant semantic associations and conceptual relationships than literal language, belongs to the course-coding that may be dominantly located in the right hemisphere. However, recent research has taken issue with the RH theory (e.g., A. M. Rapp, 2013; A. M. Rapp, Leube, Erb, Grodd,

& Kircher, 2007). Critically, the differences found between patients in Winner and Gardner (1977) may apply more to difficulties with image identification rather than idiom comprehension as both types of patients were able to correctly describe the meaning of the metaphors. Difficulty with images rather than figurative language itself may also be attributed to the right hemisphere, an issue that extends to the idiom study from Van Lancker and Kempler (1987). Furthermore, while course-coding theory may apply straight-forwardly to metaphorical meaning, idioms are associated with very specific, word-like meanings, and may indeed be more representative of long words (Cacciari & Papagno, 2012). Giora (1997) addresses this issue with a further distinction in language between salient (or familiar, fixed expressions, i.e., idioms) and non-salient (unfamiliar non-literal language, i.e., novel metaphors) language. According to this theory, the right hemisphere is primarily responsible for processing only novel non-literal expressions. Thus, not all non-literal language may be RH dominant (see e.g., Cieślicka & Heredia, 2011; A. M. Rapp, 2018).

However, rather than showing clear divisions in hemispheric locality of activation between literal and figurative language, L1 idiom research shows a complex picture of idiom comprehension involving both hemispheres of the brain. Several metanalyses (A. M. Rapp, 2018; A. M. Rapp, Mutschler, & Erb, 2012) have relooked at the body of non-literal language processing studies and determined that results are less lateralized than previously suggested. In the first metanalysis, A. M. Rapp et al. (2012) identified 38 fMRI studies on non-literal language, including 7 idiom studies, and found that of only 33% of the activated foci were in the right hemisphere. In particular, the authors identified the parahyppocampal gyrus, possibly dealing with disambiguation (e.g., Hoenig & Scheef, 2005) and emotional connotation (e.g., Proverbio, Crotti, Zani, & Adorni, 2009); the medial prefrontal cortex, critical in perspective taking and Theory of Mind (e.g., Amodio & Frith, 2006), or suppression of alternative interpretations in idiom comprehension (e.g., Papagno & Romero-Lauro, 2010); the left precentral gyrus, reflecting motor semantics (e.g., Pulvermüller & Fadiga, 2010); the LH thalamus, identifying attributive categories in metaphor comprehension (e.g., Stringaris et al., 2006) or indicating increased cognitive demand (e.g., Saalmann & Kastner, 2011); and the right cerebellum, which may play a role in comprehension (e.g., Cook, Murdoch, Cahill, & Whelan, 2004) but is also an extremely connected area of the brain. In an updated metanalysis, A. M. Rapp (2018) looked at metaphor and idiom comprehension separately, and 11 studies on idioms (and 1 on proverbs) were included. Authors found an equal contribution in the number of loci reported per hemisphere, though the right hemisphere was more spatially distinct. While some of the studies showed distinct patterns when comparing figurative and literal interpretations (e.g., right precuneus, right middle front gyrus, and right posterior middle temporal gyrus in Mashal et al., 2008 and Lauro, Tettamanti, Cappa, & Papagno, 2008), others did not find differences using this contrast (e.g., Kana et al., 2012; Raposo, Moss, Stamatakis, & Tyler, 2009). Overall, these studies suggest that idiom processing is an extremely complex process which

seems to incorporate many of the skills needed for figurative language processing. Furthermore, native speakers make use of many different parts of the brain in doing so.

There is little research including L2 idiom comprehension and brain activation. However, evidence is available divided visual field research. Cieślicka and Heredia (2011) investigated the hemispheric contributions in L1 and L2 idiom processing using a divided visual field paradigm and lexical priming. Context and salience were in focus for this study as both have been predicted to cause asymmetries in activation between hemispheres. According the Giora's Graded Salience Hypothesis, the critical distinction separating hemispheric activation and processing mechanisms is not literal vs. figurative language but rather salient vs. non-salient language (e.g., Giora, 1997). Non-salient, rather than non-literal language, may cause more activation in the RH, in line with the Fine-Course Coding Theory (Jung-Beeman, 2005) and may be at the root of the inconsistencies in findings in L1 non-literal language processing studies (e.g.,A. M. Rapp, 2018). Increased RH activation in L2 (non-salient) compared to L1 (salient) idiom processing was predicted following these theories.

Additionally, contextual studies have suggested that the RH is less subject to contextual influence than the LH (e.g., Cieślicka & Heredia, 2011), and only the LH would show differences between figurative and literal interpretations of idioms. The authors did not find significant differences between hemispheres regarding context, but the language being presented did play a role in literal and figurative as well as hemispheric activation. Overall priming effects for literal language were much greater in L2 processing compared to L1 processing, and figurative meanings in the L2 showed more priming in the RH compared to the LH. Increased LH activation of literal meanings in the L2 condition are in line with research by Fogliata et al. (2007) suggesting that these salient meanings require more inhibitory control to suppress competing literal meanings during figurative activation. These results are also in agreement with saliency-based theories in L2 idiom processing research suggesting that literal meanings have a special status in L2 processing (here equated with native language; e.g., Cieślicka, 2006). Notably, in a follow-up study using the same paradigm focusing on idiom decomposability (Cieślicka, 2013a), these hemispheric differences were only partially substantiated. Citron et al. (2020) investigated L1 and L2 processing of metaphors in a recent fMRI study and did not find evidence of hemispheric differences. Instead, their data suggests that L2 readers were less affected by figurativeness than L1 readers during comprehension. With such mixed results, further investigation using idioms and fMRI is necessary to adequately look at this complex topic. To our knowledge, there are no such fMRI studies looking at L1 and L2 idiom comprehension simultaneously to date.

In addition to hemispheric differences in idiom processing based on language, L2 readers face additional challenges for language processing that may be impacted by proficiency. Studies have found increased activation in tasks for L2 compared to L1 processing such as picture naming (e.g., Hernandez & Meschyan, 2006), sentence-reading (e.g., Shirley-Ann Rüschemeyer et al., 2006), listening (e.g.,

Shirley-Ann Rüschemeyer et al., 2005), and even mathematical calculations (e.g., Wang, Lin, Kuhl, & Hirsch, 2007). While early accounts of L2 neural organization relied on the idea that different systems are responsible for L1 and L2 language processing and therefore account for such differences, recent studies show evidence that not only do L1 and L2 grammar and neural activations overlap early in acquisition (see e.g., Luk, Green, Abutalebi, & Grady, 2011), but proficiency modulates the similarity of L1 and L2 processing (see e.g., Perani & Abutalebi, 2005). Green, Crinion, and Price (2006) even go so far as to propose that qualitative differences between L1 and L2 processing in neural activation decrease as proficiency increases. This idea is based on evidence of increased effort in production as shown by brain activity in low-proficient L2 participants (e.g., Briellmann et al., 2004; De Bleser et al., 2003) compared to highly-proficient L2 participants, which also applies to comprehension (e.g., Perani et al., 1998). Thus, while the idiom study done by Cieślicka and Heredia (2011) may point to neural differences in idiom processing based on language, these differences may become less pronounced or even disappear with increasing proficiency. Though more research is necessary in order to determine whether these such differences apply to activation of literal and figurative meaning in idiom processing, L2 idiom processing likely presents similarly or even more complex processes than those needed in L1 idiom processing.

7.2 THE CURRENT STUDY

In this study, we aimed to answer three questions:

1) Does idiom comprehension differ between native and non-native readers?
2) Does proficiency correlate with activation patterns?
3) Does idiomatic and literal sentence processing differ within and between L1 and L2 readers?

Based on the existing body of literature, we predict that regarding Question 1, idiom comprehension will show differences in activation between L1 and L2 readers. We also expect that, if L2 idiom processing becomes more native-like as proficiency increases, activation patterns may show correlations with proficiency in areas of the brain associated with idiom comprehension (see e.g., A. M. Rapp, 2018). Finally, regarding Question 3, we predict that there may be differences in activation based on the figurativeness of the sentence. However, we do not expect these differences to show isolated hemispheric involvement (i.e., right-lateralization for figurative language).

7.2.1 Materials and methods

7.2.1.1 Participants

A total of 30 participants (17 female, 19-51 years old) were recruited via email from the University of Tübingen. All participants were healthy, right-handed, had at least college-entry level education, and did not report any issues with hearing or vision at the time of the experiment. Basic participant

information is summarized in Table 7-1. L1 and L2 participants did not significantly differ in age (t= -0.593, p= 0.558). L2 participants were non-native speakers of English (German L1) with a high proficiency in English. Their LexTale scores (Lemhöfer & Broersma, 2012) averaged 77.71 out of 100 (SD= 13.04), and their average self-rated proficiency was 5.79 (SD= 0.75) on a scale from 1-7. 13 native speakers of (Predominantly American) English all grew up monolingually and had at least low proficiency in German (level A2, see e.g., "Common European Framework of Reference for Languages," 2001). The study was approved by the Ethical Committee of the University Clinic of Tübingen, Germany, and informed consent was given following an experimental description prior to taking part in the study.

Table 7-1. Participant information by language group

Participant Group	*n*	*n* female	Age	Student participants	SPQ Score	LexTale Score	Ave. Self-rated Proficiency (1-7)
L1 (native) participants	13	6	26.5	10	50.2	93.5	7.0
L2 (non-native) participants	17	11	28.0	13	41.5	77.7	5.5

7.2.1.2 Stimuli

Forty idioms were selected from the German-English Database of Idiom Norms (**Chapter 2**), which is a publicly available database of 300 idioms with norming values (e.g., familiarity with the form and meaning, predictability, and literality). Idioms used in the present study were selected based on their high familiarity to L1 and (German) L2 English speakers (average rating of 6.09 and 5.14 from L1 and L2 speakers, respectively on a scale from 1-7). Idioms also had the same syntactic form (verb + article + noun), and the final word was a concrete noun. Idioms with this form were chosen since they would be comparable in properties such as predictability and literality that have been shown to impact the timing of access to figurative and literal meaning during idiom processing (e.g., Titone & Connine, 1994b). Each idiom was embedded at the end of short, neutral sentences in its original idiomatic form (e.g., I told him to *get a grip*.) as well as with the final word replaced by another concrete noun to create a comparable literal version of the sentence (e.g., I told him to *get a drink*.) The idiom-final words and their literal variations neither differed in length (mean letter length idiom: 4.97, literal: 4.79, p= 0.574) nor in frequency (mean cob/ml idiom: 124.36, literal: 124.39, p= 0.499, Max Planck Institute for Psycholinguistics, 2001). In addition to these 80 sentences, 20 non-idiomatic filler sentences in German were included. All sentences were divided into two parts for presentation: the beginning of the sentence and the final verb phrase containing the idiom or the literal version of the idiom (see Table 7-2). Additionally, 20 sets of non-linguistic symbols were included as controls. A complete list of stimuli is available in APPENDIX F.

Table 7-2.Example experimental items

Item	Figurativeness	First Half	Second Half
1a	Idiom	I think that John	spilled the beans.
1b	Literal	I think that John	spilled the milk.
2a	Idiom	I told him to	get a grip.
2b	Literal	I told him to	get a drink.

7.2.1.3 Procedure

During the fMRI scanning procedure, participants lay supine in the scanner with their heads secured by foam to minimize movement. Stimuli were presented visually on a translucent screen viewed by the subjects via a mirror. Sentences were displayed 1500ms apart in two parts, each for 1500ms. Participants were instructed to read sentences silently for comprehension, and to press a button with their right hand on the second half of sentences which were presented in German. This procedure was exercised in a training session prior to the experiment using sentences not included in the experiment. In order to ensure reading comprehension, participants were reminded that they would be asked questions about the English sentences at the end of the experiment. Items were divided into two balanced experimental blocks so that each sentence occurred in its idiomatic version in one block and its literal version in the other, and the order of the two blocks was pseudorandomized between participants. Stimuli in each block consisted of the 40 target sentences (20 idiomatic sentences and 20 literal sentences), 10 German sentences, matched on length and grammatical complexity, and 10 low-level visual baseline "sentences" (i.e., literal sentences presented in ''SPSS marker set" characters/letter strings). Stimulus sequence was unpredictable and optimised using optseq software.

In addition to the fMRI task, all participants completed the LexTale vocabulary test prior to entering the scanner. Following the fMRI scanning procedure, they also completed a comprehension check, a language background questionnaire, and the schizotypal personality questionnaire (Raine, 1991). In addition to self-reported measures of proficiency, the LexTale vocabulary test (Lemhöfer & Broersma, 2012) serve as an objective measure of participants' vocabulary and general English proficiency. The comprehension check was used as a measure to be sure that participants were reading the items carefully, as they were instructed that they would be asked questions about the English read during the fMRI task.

7.2.1.4 Imaging methods

Imaging was performed on a 3-T Scanner (Siemens, TIM TRIO). Functional images were acquired with an echoplanar image sequence which is sensitive to blood-oxygen-level dependent contrasts

(BOLD responses; TR 1500 ms, 72 slices). Two runs consisting of 205 volumes each were acquired during the experiment.

7.2.2 Analysis

For image processing and all statistical analysis, SPM12 (The Wellcome Centre for Human Neuroimaging) was used. The functional images of each subject were slice time-corrected to the middle slice and were corrected for motion and realigned by using the first scan of the block as a reference. Three participants (1 L1, 2 L2) were excluded from further analysis due to movement artifacts. T1 anatomical images were co-registered to the mean of the functional scans and spatial normalized to the MNI space by the combined segmentation bias correction and spatial normalization tool in SPM12. The calculated nonlinear transformation was applied to all functional images. Finally, the functional images were smoothed with an 8-mm full-width, half-maximum (FWHM) Gaussian filter.

7.2.2.1 Contrasts between conditions

A general linear model (GLM) was constructed for each participant to analyze the hemodynamic response function. In each GLM, regressors were generated by convolving the hemodynamic function with a box car function. Separate regressors were used to model the hemodynamic responses during presentation of the sentence-beginning, idiomatic sentence-ending, literal sentence-ending, German sentences, and the visual baseline condition. Moreover, a high-pass filter (1/128 Hz) was applied to remove low-frequency drifts. For each subject, T-Test contrasts for the idiomatic sentence-ending (*idiom*) and literal sentence ending (*literal*) conditions versus low level baseline were calculated. To directly compare brain activation between these conditions, subtraction analysis of *idiom* versus *literal* and vice versa were calculated. Random effects analyses on group level were calculated for each of these contrasts. Then, between-group comparisons were calculated for each of these contrasts using two sample T-Tests Activation for all the differential contrasts is reported if it exceeds a significance of $p < 0.001$ (uncorrected) and an extent threshold of 10 voxels.

7.2.2.2 Language proficiency correlations

To investigate the influence of language proficiency on brain activation, a simple regression analysis of SPM data was used. In this type of analysis, each single voxel in the brain is individually examined with respect to whether the size of the BOLD response is correlated with a variable over subjects. The result of this analysis is a brain map, which depicts the voxels with which there is a significant correlation between the variables. Because of the exploratory character of this study, we chose a liberal threshold of $p < 0.001$ (uncorrected) and an extent threshold of 10 voxels for analysis.

Separate tests were performed to detect positive correlation (i.e., the higher the score of the individual subject in the questionnaire, the stronger the BOLD response) and negative correlation (i.e., the higher the score of the individual subject in the LEXTALE total score, the weaker the BOLD response)

associated with language proficiency. Moreover, separate tests were performed for all participants as well as for the subgroups of native speakers (L1) and non-native speakers (L2).

7.3 RESULTS

Main effects for reading idiomatic (*idiom*), literal (*literal*) and German sentences against low level baseline (*baseline*) showed robust activations in a predominantly left lateralized network including visual cortices, temporal lobe and prefrontal cortex. The results concerning each question (concerning only the conditions *idiom*, *literal*, and *baseline*) are displayed below and placed into context in the discussion section.

7.3.1 L1 and L2 differences in idiom activation

Differential contrasts between native speakers (L1) and second language (L2) group are listed in Table 7-3 and displayed in Figure 7-1. Contrasts between these groups included comparisons between literal and figurative sentences, in both directions (i.e., *literal*>*idiom* and *idiom*>*literal*)[11], idioms against low level baseline figures, and literal sentences against low-level baseline figures. Robust differences were found in several areas of the brain not limited to a single hemisphere, though these regions are not those particularly associated with sentence processing.

Table 7-3. Differential contrasts between native (L1) and non-native (L2) groups during processing of idioms and literal sentences

	L1 > L2					**L2 > L1**				
	region	**hemisphere**	**extent**	**MNI coordinates**	**z**	**region**	**hemisphere**	**extent**	**MNI coordinates**	**z**
Idiom > literal	Posterior Cingulate	LH	12	-12 -50 20	4.530	Cerebellum	RH	5	36 -50 -48	3.760
	Insula	LH	10	-30 -4 20	4.020	temporal lobe	LH	33	-36 -58 -2	4.910
	Cingulate Gyrus	LH	22	-14 20 36	4.860					
literal > idiom	Sub-Gyral	LH	33	-36 -58 -2	4.912	Posterior Cingulate	LH	12	-12 -50 20	4.530
						White Matter	LH	10	-30 -4 20	4.021
						Cingulate Gyrus	LH	22	-14 20 36	4.860
idiom > baseline	Parahippocampal Gyrus	RH	10	22 -16 -26	3.780		*no activated voxels*			
	precuneus	LH	18	-12 -46 10	4.043					
literal > baseline	posterior cingulate	LH	45	-10 -44 10	4.457	precentral gyrus	LH	16	-40 2 30	-3.843

Note: Contrasts displayed between speaker groups are displayed from top to bottom on the left (L1>L2) and the right (L2>L1) for each comparison listed on the left side of the table.

[11] Contrasts are denoted directionally using > to indicate the difference in activation in the BOLD response (i.e., idiom>literal displays the difference in activation during idiom processing when compared to literal processing).

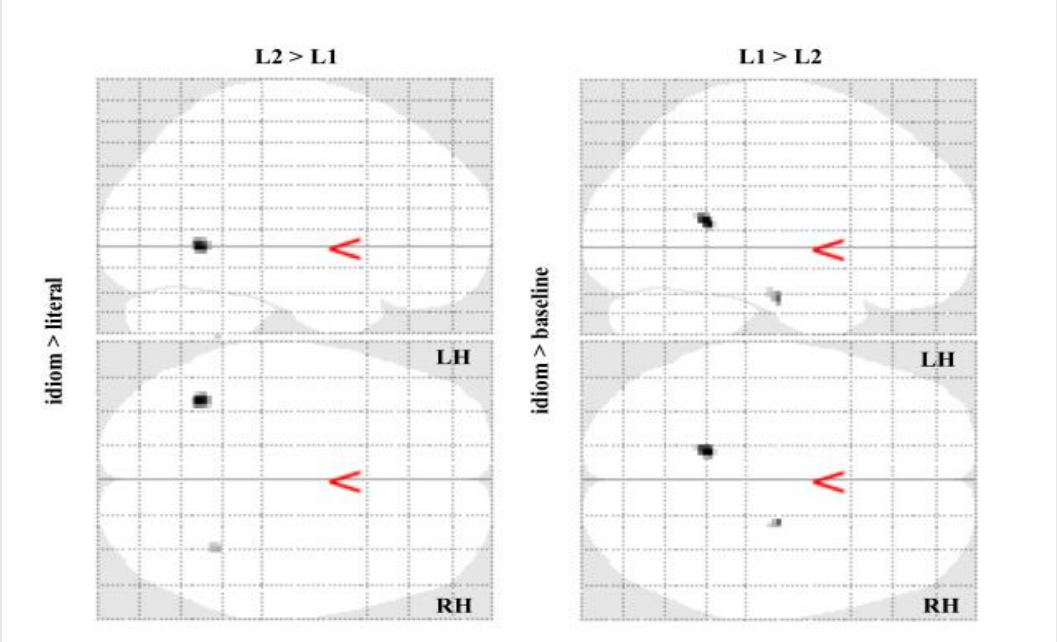


Figure 7-1. Glass brain comparisons between L1 and L2 speaker groups for 2 contrasts

Note: Top images show a right-view of the brain, and bottom images display a top-view. Right and left hemispheres are labeled RH and LH, respectively.

7.3.2 Proficiency (LexTale score) in idiom activation

Areas showing a BOLD response which is significantly correlated with language proficiency (as measured by the LexTale score) in simple regression analysis are listed in Tables 7-4 and 7-5. Both scores from all subjects combined and L1 speakers individually are shown in Table 7-4 and L2 participants individually are in 7-5.

Table 7-4. LexTale correlations with the BOLD response for all participants and L1 participants only

		Positive correlation					Negative correlation				
		region	hemisphere	extent	MNI coordinates	z	region	hemisphere	extent	MNI coordinates	z
All participants	***idiom* > *literal***	*no significant clusters*					*no significant clusters*				
	literal* > *idiom	*no significant clusters*					*no significant clusters*				
	idiom* > *baseline	*no significant clusters*					inferior frontal gyrus	LH	120	-48 26 8	4.107
							Inferior Frontal Gyrus	LH	14	-48 6 30	3.736
		region	**hemisphere**	**extent**	**MNI coordinates**	**z**	**region**	**hemisphere**	**extent**	**MNI coordinates**	**z**
L1 Speakers	***idiom* > *literal***	supplementary motor area	RH	14	10 -6 52	5.41	inferior frontal gyrus	LH	51	-46 36 -4	6.10
	idiom* > *baseline	sub gyral	RH	11	28 -22 10	5.858	*no significant clusters*				

Note: Positive and negative LexTale correlations are displayed from top to bottom on the left (positive) and the right (negative) for each contrast listed on the left side of the table. All subjects are displayed on the top rows and only L1 speakers on the bottom rows.

Table 7-5. LexTale correlations with the BOLD response for L2 participants

		Positive correlation	Negative correlation				
			region	hemisphere	extent	MNI coordinates	z
L2 speakers	***idiom > literal***	*no significant clusters*	Pons	LH	18	-18 -34 -40	6.146
			cerebellum	LH	14	-14 -62 -34	4.829
			cerebellum	RH	12	24 -36 -30	4.520
			cerebellum	RH	31	4 -54 -28	5.259
			inferior temporal gyrus	LH	15	-40 -40 -14	5.600
			hippocampus	RH	92	20 -26 -8	5.725
			midbrain		24	-6 -16 -6	5.223
			cerebellum	LH	16	-6 -40 -8	4.262
			inferior frontal gyrus	LH	26	-48 18 10	4.367
			sub-lobar	RH	15	4 -6 6	5.015
			sub-lobar	RH	84	24 -16 18	6.863
			Sub-lobar	LH	11	-24 2 16	4.122
			middle temporal gyrus	LH	11	-48 -48 20	4.106
			Medial Frontal Gyrus	LH	19	-12 34 34	4.727
			superior frontal gyrus	LH	58	-6 4 60	5.155
	idiom > baseline	*no significant clusters*	Cerebellum	RH	31	14 -54 -32	4.835
			white matter	LH	13	-42 -44 -8	4.617
			Calcarine	RH	18	6 -88 2	4.898
			Middle Occipital Gyrus	RH	10	38 -86 12	4.095

Note: Positive and negative LexTale correlations are displayed from top to bottom on the left (positive) and the right (negative) for each comparison listed on the left side of the table.

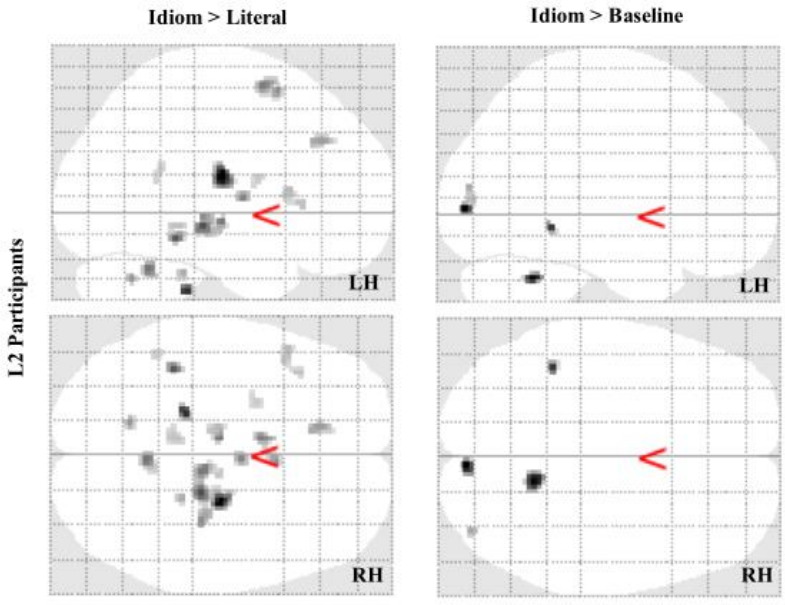


Figure 7-2. Glass brain visuals of activation of L2 participants showing negative correlation with proficiency

Note: Top images show a right-view of the brain, and bottom images display a top-view. Right and left hemispheres are labeled RH and LH, respectively.

Although all language groups were included in contrasts, most contrasts were present in the L2 group, particularly during literal language processing (Table 7-5). The L2 areas showing correlations for both contrasts are shown in Figure 7-2.

7.3.3 Differences in idiom and literal activation in both groups

Differential contrasts between idiomatic and literal sentences (*idiom>literal*) for all subjects together showed activation in only one cluster in the left hemisphere sub-lobar white matter. The opposite contrast (*literal>idiom*) showed activation in a bilateral posterior network including limbic and fusiform brain regions. These contrasts are listed in Table 7-6.

Table 7-6. Differential contrasts between idioms and literal sentences for all participants combined

Idiom > Literal					***Literal > Idiom***				
region	**hemisphere**	**extent**	**MNI coordinates**	**z**	**region**	**hemisphere**	**extent**	**MNI coordinates**	**z**
sub lobar white matter	LH	64	-20 -36 22	4.240	cerebellum	LH	12	-30 -42 -38	3.775
					parahippocampal gyrus	RH	35	26 -24 -28	3.912
					left inferior temporal gyrus	LH	138	-48 -56 -16	4.277
					fusiform gyrus	LH	36	-34 -22 -22	4.353
					fusiform gyrus	RH	45	34 -50 -16	3.893
					fusiform gyrus	LH	58	-26 -48 -10	4.091
					fusiform gyrus	RH	32	26 -86 -14	3.814
					Lingual Gyrus	LH	152	20 -82 -4	4.629
					posterior cingulate	RH	30	10 -44 20	5.226
					precuneus	RH	21	4 -62 22	3.972
					white matter	RH	57	28 -48 28	4.340
					Superior Parietal Lobule	LH	10	-28 -62 46	3.695
					Medial Frontal Gyrus	LH	10	-12 -18 56	3.896
					precentral gyrus	RH	18	24 -14 58	4.308

Note: Directional contrasts between idiom and literal conditions are displayed from top to bottom on the left (idiom>literal) and the right (literal>idiom).

Sub-analyses for the native speakers (L1) and second language (L2) group showed differences in their patterns of activation: whereas the L2 group showed one cluster of activation similar to the overall group in the sub-lobar white matter, in the L1 group, the left inferior frontal trigonum was active. These individual analyses are shown in Table 7-7 and the contrast *literal > idiom* is displayed in Figure 7-3.

Table 7-7. Differential contrasts between idioms and literal sentences for all participants combined

	Idiom > Literal					***Literal > Idiom***				
	region	**hemisphere**	**extent**	**MNI coordinates**	**z**	**region**	**hemisphere**	**extent**	**MNI coordinates**	**z**
L1 speakers	inferior frontal trigonum	LH	**9**	-40, 32 40	3.990	Fusiform Gyrus	LH	33	-22 -50 -12	3.79
						Inferior Occipital Gyrus	RH	7	30 -86 -14	3.52
						Inferior Occipital Gyrus	LH	36	-34 -92 -6	4.02
						Calcarine	RH	16	16 -98 -6	3.61
						Medial Frontal Gyrus		5	-12 -16 58	3.62
L2 speakers	Sub-lobar	LH	79	-22 -36 18	6.009	parahippocampal gyrus	RH	10	26 -24 -30	3.64
						inferior temporal gyrus	LH	35	-48 -44 -12	3.90
						lingual gyrus	LH	19	-12 -29 -6	3.38
						lingual gyrus	RH	54	22 -84 -2	3.75
						Sub-lobar	LH	24	-30 -4 20	4.01
						Precuneus	LH	11	-8 -66 22	3.26
						Posterior Cingulate	RH	19	4 -60 24	3.38
						Sub-lobar	RH	14	32 -30 26	3.64
						parietal lobe	RH	20	26 -46 30	3.63

Note: Directional contrasts between idiom and literal conditions are displayed from top to bottom on the left (idiom>literal) and the right (literal>idiom) for speaker group individually listed on the left side of the table.

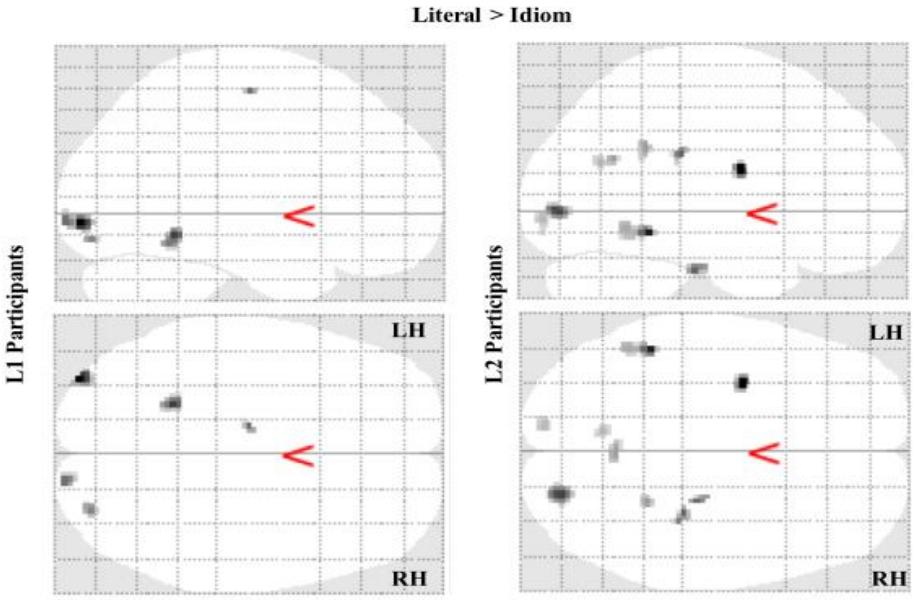


Figure 7-3. Glass brain contrast between literal and idiom processing for L1 and L2 participants individually

Note: Top images show a right-view of the brain, and bottom images display a top-view. Right and left hemispheres are labeled RH and LH, respectively.

7.4 DISCUSSION

We used functional magnetic resonance imaging to detect brain activation during processing of English idioms in native (L1) and second language (L2) speakers. To our knowledge, this is the first study that uses fMRI to investigate idiom comprehension in L2 speakers. The main finding of this study is that the L1 and L2 involved brain networks differ from one another in spatial activation during idiom comprehension, though not in quantitative activation. Furthermore, some differences in activation are modulated by language proficiency as measured by the English LexTale score.

7.4.1 Idiom comprehension between L1 and L2 readers

The first research question in the current study aimed to investigate differences in brain activations during idiom comprehension between L1 and L2 speakers, and particularly asked whether they differed—both qualitatively or even quantitatively in effort. In our study using salient English idioms, in line with our expectations, robust differences were found between the two groups (see Table 7-3), however, few of these differences were present in the classic language regions on the sentence level. Furthermore, differences were found bidirectionally.

L2 group are shown in Table 7-3 and Figure 7-1. Differential activation was found in both directions, meaning both for the L1 > L2 as well as the L2 > L1 contrast. This is compatible with a recent large meta-analysis of literal language processing (Sulpizio, Del Maschio, Fedeli, & Abutalebi, 2020), which also indicates stable differences in both directions. We calculated differences between groups both for the idioms > literal as well as against low level visual baseline contrasts. For the L2 > L1 contrast, differences were present in the cerebellum und left temporal lobe. In line with the recent meta-analysis (Sulpizio et al., 2020), processing of literal sentences activated the left precentral gyrus more strongly. This brain region is known to be involved both in the processing of figurative language (A. M. Rapp et al., 2012) including idioms (e.g., Boulenger, Hauk, & Pulvermüller, 2009; Yang et al., 2016) and –equally important – in L1 > L2 contrasts. Differential contrasts strongly activated in native versus non-native speakers predominantly included the medial brain regions including the anterior and posterior cingulate cortex. Again this activation is in line with findings in literal language on a sentence level (Sulpizio et al., 2020). Possible functions include language switching and semantic processing. However, these brain regions are also known to be involved in the comprehension of idioms (Zubicaray & Schiller, 2018). Cacciari and Papagno (2012) proposed a model according to which this brain area is responsible for suppressing irrelevant (literal) meanings during idiom comprehension (e.g., Fogliata et al., 2007), whereas the same region is also involved in suppressing foreign language meanings (Luk et al., 2011).

Finally, our result is compatible with the only other fMRI study examining figurative language in the L2 (Citron, Michaelis, & Goldberg, 2020). In our study, native (L1) speakers showed more

activation during non-literal comprehension compared to literal sentences in the language switching network including the anterior and posterior cingulate and LH insula. In line with Citron et al. (2020), it may be that native speakers are more engaged in figurative language processing than L2 speakers. Along these lines, this may indicate that the L1 speakers in our study possibly activate the figurative meanings of idioms more strongly compared to the L2 group, and thus an increase in overall activation during idiom processing is detected here, a result that may be in line with later activation in L2 speakers, particularly considering the high literality and low predictability of the idioms (e.g., Titone & Connine, 1994b).

Overall, there were no robust differences in the classical network for language and figurative language between speaker groups, although these findings are in line with recent research. Rather, the differences may fit in well with language switching networks as a whole (e.g., Cacciari & Papagno, 2012; Luk et al., 2011). This fits with language processing theories that neither suggest critical differences between figurative and literal language processing not between L1 and L2 idiom processing. Indeed, differences in the L1 and L2 language networks activated during idiom processing need not indicate different processing mechanisms. Rather, differences may be attributed to language differences in general. It should also be noted that these differences also cannot be attributed to common differences found in learning networks (i.e., hippocampus or inferior frontal gyrus). Thus, critically, we interpret these results to indicate that L1 and L2 processing of idioms do differ somewhat qualitatively. Though, these differences are not hemispheric in nature as predicted by previous literature (e.g., Cieślicka & Heredia, 2011). However, also critically, with language differences indicated bi-directionally, they do not suggest quantitative differences in effort during processing.

7.4.2 Proficiency and activation patterns

Considering the second question of whether proficiency correlates with brain activation, we find that the results positively support our expectations. We found a robust and significant negative correlation between language proficiency and brain activation in a large number of brain regions in L2 speakers, and L1 speakers also showed a positive correlation. Our study is therefore in contrast to the only other fMRI study on figurative language from Citron et al. (2010), which did not find correlations. A negative correlation indicates that a smaller BOLD response is associated with higher language proficiency. Positive correlations were only found within native speakers (L1). However, it should be noted that the LexTale task was originally designed to look at proficiency in non-native speakers and this positive interpretation shouldn't be interpreted on its own (see e.g., Cargnelutti, Tomasino, & Fabbro, 2019).

In the L2 speaker group, the network prominently activated during idiom comprehension included the cerebellum. Therefore, our study adds to the growing number of studies indicating that the cerebellum plays an important role in processing both L2 language (e.g., Cargnelutti et al., 2019)

as well as in a more general sense literal (e.g., Mariën et al., 2014) and figurative language (e.g., A. M. Rapp et al., 2012). Particularly, the cerebellum is known to be involved in the processing of idioms (e.g., Boulenger et al., 2009; Lauro et al., 2008; Yang et al., 2016; Zempleni et al., 2007). Cerebellar activity related to language proficiency has also been reported (e.g., Cargnelutti et al., 2019), aligning directly with the findings here. While the particular role of the cerebellum is unclear because of its broad range of activation across many different language-related and non-language-related tasks as well as its connection to many other areas of the brain, its increased usage in L2 compared to L1 language users is well-known. Another area similarly involved in a significant negative correlation is the hippocampus, which has been associated with learning processes and figurative language processing (e.g., A. M. Rapp et al., 2012). Following these lines, this area could also indicate a more active process of meaning searching where there is unclarity or ambiguity in language, in line with expectations of decreasing activation with increasing proficiency.

Another important finding concerning proficiency is a significant negative correlation between language proficiency and the activation of the left inferior frontal gyrus. This relationship is reputed during literal language processing. In a metanalysis, Cargnelutti and colleagues (2019) also found a robust negative correlation with L2 language proficiency in the same region. The left inferior frontal gyrus is a region that is often placed in conncetion with the integration of meaning in context (e.g., Binder, Desai, Graves, & Conant, 2009), and it has also been shown to play a roll in the comprehension of figurative language (Rapp et al., 2012). For instance, in the comprehension of novel metaphors, this region has been connected with the semantic mapping of meaning between two previously unconnected entities (e.g., A. M. Rapp, 2012; A. M. Rapp, Leube, Erb, Grodd, & Kircher, 2004).

Thus, these negative L2 correlations in combination with the positive L1 correlations suggest that L2 processing generally may become less effortful as proficiency increases. Though, it may be the case that the areas discussed here simply compensate less as proficiency increases. Instead, as suggested already by a lack of overall quantitative differences found in the previous research question, these L2-specific networks become less responsible during processing, and networks also associated with the L1 processing of this language take over. In line with previous literature (e.g., Green et al., 2006), L2 and L1 processing differences may be somewhat modulated by proficiency.

7.4.3 Comparing literal and figurative language

Finally, our study aimed to look at the question broadly asked in the field of figurative language research and brain activation, namely whether differences in literal and figurative language activation are present. In contrast with our expectations, the BOLD response between idioms and carefully matched literal sentences differed only minimally. When including all participants, L1 and highly proficient L2 speakers, differences (*idiom* > *literal*) were found only in a single cluster in the left

hemisphere. This finding contrasts previous fMRI studies, many of which found differences in brain activation, for example in the left frontal inferior gyrus when comprehending idioms (e.g., Desai, Conant, Binder, Park, & Seidenberg, 2013; Lauro et al., 2008; A. M. Rapp, 2018). Zempleni et al. (2007) found brain activation differences in this region during the processing of Dutch idioms. Analogous to other forms of non-literal language (e.g., A. M. Rapp, 2018), this region could be responsible for the integration of alternative figurative meanings in sentential contexts (see also A. M. Rapp et al., 2004; A. M. Rapp et al., 2012).

However, our results are still compatible with other previous fMRI literature as the findings up to now considering the differences between literal and idiomatic sentences have indeed been mixed (e.g., Kana et al., 2012; Raposo et al., 2009). For instance, Raposo and colleagues (2009) investigated idiomatic action sentences in a study in the United Kingdom. Although they included 56 stimuli and 22 participants, no contrast for idiomatic > literal sentences was detectable. Additionally, in Kana et al. (2012), no activation was found in the same comparison. Instead, the opposite contrast showed several activation clusters, mainly in the right hemisphere.

Our study also did not find any differences in the L1 group for idiomatic compared to literal sentences (*idiom>literal*). One possible explanation is that the highly salient idioms used in our study were processed as single units (e.g., Cacciari, 2014; Vespignani et al., 2009), putting similar demands on the brain and the prefrontal cortex as sentence meaning may be easily integrated in both literal and figurative sentences. Though, it may also likely indicate the general similarity of figurative and literal language processing. This interpretation is compatible with studies showing that there need not be differences in cognitive effort between figurative and literal readings of sentences, particularly considering behavioral data (e.g., Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011; also see **Chapters 3**, **4** and **6**) and brain activation (e.g., A. M. Rapp et al., 2007; A. M. Rapp et al., 2012; A. M. Rapp & Steinhäuser, 2013).

A further, yet very speculative interpretation is that the task applied in this study might have masked brain differences between figurative and literal meanings. In our study, the task for the subjects was to indicate by button press if a presented sentence is English or German. It is known that this task –like figurative-literal distinctions (e.g., A. M. Rapp, 2013)- requires the use of the language switching network (e.g., Citron et al., 2020; Luk et al., 2011; Sulpizio et al., 2020) and that these networks overlap. Of note, the only other neuroimaging study on figurative language in an L2 showed a negative contrast for metaphoric compared to literal language as well (Citron et al., 2020). Thus, while the reasons for the generally negative result in this question are still unclear, we did not find strong evidence of differential figurative and literal language networks, and notable such activation was not divided hemispherically.

7.5 CONCLUSIONS

In conclusion, the findings here added critically missing data comparing L1 and L2 language users during idiom comprehension. The study compared both language groups in terms of literal and figurative language comprehension during reading and used the LexTale score as a measure of proficiency to further investigate differences. While there were robust differences between L1 and L2 comprehension, these differences in activation were not found in areas most typically associated with figurative language processing or sentence processing. Rather, the differences seem to be more general in nature, and may be task- or more generally language-related. Differences in activation did, however correlate negatively with language proficiency in L2 readers. Though, the differences were overall more qualitative than quantitative. We interpret these results to suggest that as proficiency increases, language networks associated with more effortful L2 language processing may be less needed, and L2 language processing may look more like more L1 language processing with increasing proficiency. Finally, we found only minimal differences overall when investigating figurative and literal language processing altogether, and there was no evidence of a hemispheric division between these two types of sentence comprehension. Rather, as in L1- and L2-based differences, it seems that figurative and literal language processing may also be generally similar in nature on a neurological level.

GENERAL CONCLUSION

8 GENERAL CONCLUSION

This dissertation set out to look closely at idiom processing in native (L1) and non-native (L2) speakers. In particular, how and when meaning is accessed during comprehension has been in focus of this investigation. In examining this meaning and access in these speaker groups, the overarching question simply asked: How do L1 and L2 processing compare? In order to answer this general question in a broad manner, this dissertation set three goals:

I. Establish a large working database of English idioms containing L1 and L2 (L1 German) norming data to draw from.

II. Assess where critical differences between L1 and L2 idiom processing may lie and either collect experimental data from both groups where there is a paucity of data or collect experimental data from L2 speakers for comparison with existing L1 data.

III. Determine whether L1 and L2 speakers use different processing strategies and/or mechanisms during idiom comprehension.

The first goal of this thesis was met straight-forwardly with the database described in **Chapter 2**. The database was established early in research and used to supply normed idioms for all studies undertaken in the English language using existing idioms (all except the studies in **Chapters 4B** and **6**, which used German idioms and fictional idioms, respectively). The database is available online for future research as part of the collaborative research group (SFB 833), and hopefully other researchers using native speakers of (American) English or non-native speakers of English with a German L1.

The second goal of the thesis was met partially in the sub-questions identified in Section 1.3. These individual questions identified the possible differences in processing between languages and areas where either both L1 and L2 data was necessary or L2 data was needed to fill in research gaps. The collection of this data was undertaken in the subsequent chapters via experimental methods including offline ratings, cross-modal priming, self-paced reading, eye-tracking, and fMRI. In the summary of findings below (Section 8.1), these sub-questions will be answered directly, with the aim of giving a broad picture of where L1 and L2 processing show similarities and differences, based on the focal points given in Chapter 1.

The third goal of this thesis has been met only indirectly via the experiments in **Chapters 3-7**, and this goal will be discussed directly in Section 8.2. The focus will be on synthesizing the data collected throughout this dissertation into a clearer picture of how L1 and L2 idiom processing may occur. The most common existing models of idiom processing addressed by current literature for L1 and L2 speakers will be addressed in light of the sum of the current data, and the question of how and

whether these two speaker groups diverge in processing mechanisms and/or strategies will be discussed.

Finally, since this dissertation has directly addressed the L2 comprehension of idioms, implications for the teaching and learning of idioms for non-native speakers will be discussed briefly in Section 8.3. While raising this issue was not a primary goal of the current dissertation, theories of idiom processing directly deal with storage of idioms in the mental lexicon, and some of the pedagogical implications of how this process may occur should not be ignored.

8.1 SUMMARY OF FINDINGS

The findings here will be summarized based on the sub-points outlined in section 1.3 rather than chapter-by chapter, as each chapter contributes to several questions addressed in this dissertation.

8.1.1 Access to literal and figurative meaning in controlled environments

The first point of comparison between L1 and L2 processing is the access to figurative and literal meaning in controlled (i.e., experimental) environments. Aspects of this were addressed in **Chapters 3**, **4B**, **5B** and **5C**. Particularly, the question of whether both native and non-native speakers have access to the figurative meaning of an idiom during online processing was answered in the baseline cross-modal priming experiments in **Chapter 3**. In two cross-modal priming experiments, both native and non-native listeners responded to figuratively related targets presented at the offset of idioms presented auditorily at the end of non-biasing sentences. Compared to unrelated targets, both participant groups showed significant facilitatory priming, and therefore evidence of online access to the figurative meaning of the idioms. Likewise, this chapter answered the second question of whether these groups have online access to the meaning of literal constituents during online processing. Here, too, both groups showed facilitatory priming for literally related targets compared to unrelated controls. Thus, both groups showed similar capacities for meaning access to these two types of meaning on a basic level. While these findings alone are not novel, as previous research has shown similar results in native speakers (e.g., Cacciari & Tabossi, 1988; Titone & Connine, 1994b; Titone & Libben, 2014) and even non-native speakers (e.g., Cieślicka, 2006), the similarities in the behavior of both groups are. Though Cieślicka (2006) makes claims about a literal priority in processing due to the increased priming effects found in L2 listeners for literal targets, our study found that this effect also applies to native listeners, though the difference in priming was not significant. Possible reasons for differences in priming strength may be due to the abstractness of the figurative targets compared to more concrete literal targets (e.g., the figurative target DIE compared to the literal target PAIL for the idiom *kick the bucket*; see Brysbaert & Duyck, 2010). This lack of clear difference between groups was also corroborated and strengthened by van Ginkel and Dijkstra (2019), who did not find differences in figurative and literal

priming in both L1 and L2 listener groups in a later priming study. They even suggested that further differences may lie in cognate differences between target types.

However, the L2 experiment in **Chapter 4B** also addressed the issue of literal constituent activation during listening using an eye-tracking paradigm. In this study, L2 German participants (English L1) listened to incomplete German idioms and were asked to choose the correct final constituent word of the idiom from four possible words presented visually on the screen. The choices consisted of the correct idiom completion, a literally related word, and two unrelated distractor words. Like the L1 adults in Keßler et al. (submitted), looks to the literally related constituent were significantly higher than looks to unrelated distractors, suggesting that even when participants were knowingly interpreting idioms, the literal constituents were activated. This experiment, unlike the cross-modal priming studies in **Chapter 3**, did show qualitative differences between L1 (child and adult) and L2 behavior. Whereas looks to the literally related target decreased over time in L1 participants, L2 participants showed a late surge in these looks, indicating later and stronger literal activation. Thus, there is a possibility that L2 listeners do have different processing strategies than L1 listeners during idiom comprehension. However, there are also reasons to be wary in drawing quick conclusions. For instance, both the correct literal constituent and the related target were both present on the screen during listening. An increase in looks to this related target may also be due to a lack of inhibitory control of non-salient word activation (e.g., Gernsbacher et al., 1990; Gernsbacher & Faust, 1991). Additionally, the late nature of this increasing activation may also point to a lack of certainty in participants' choices of the correct response. As L2 participants performed overall worse than both L1 adults and children, and this is seen in fewer correct responses and overall looks to the correct response, proficiency may have played a role in this result. Thus, while there may be some evidence that literal constituent activation plays a more important role in L2 idiom processing compared to L1 processing, the reasons behind this difference may not warrant fundamental differences in processing.

Another aspect of meaning activation during online processing is whether both speaker groups compute the literal phrasal meaning of idioms during idiom comprehension. **Chapters 5B** and **5C** looked at this issue in native and non-native readers, respectively, in an indirect manner. In these self-paced reading experiments, contexts were biased to encourage either literal or figurative readings of idioms, which differed in their literality, and sentences ended in a manner than was either consistent or inconsistent with these expectations. In the case that idioms were biased figuratively, and even more so in the case where idioms also have a low capacity for a literal interpretation, the expectation was that a literal phrasal computation is unnecessarily costly and may not be undertaken at all. Indeed, native readers in **5B** and non-native readers in **5C** provided positive evidence for this expectation: following a figuratively biasing context, figurative readings of idioms were read more quickly than literal ones. However, for native speakers, literally biasing contexts produced similar results for literal

readings, where idioms had a high literality. Where idioms had a low literality, this was not the case. While non-native readers in **5C** repeated this pattern in the visual data, the result was not significant. Thus, for native speakers, it is neither the case that the literal meaning is always or never computed during processing, rather, this process is more flexible (see also Section 8.1.2). For non-native speakers, the data points to a similar result, though it's less clear that literal meaning is entirely abandoned where idioms have a low-literality.

Finally, the issue of the timeline for meaning access was addressed in the chapters already mentioned, as well as in **Chapter 4A** and **5A** (and this aspect will be addressed further in 8.1.2 and 8.1.3). What's consistently clear from the data from each experimental method (cross-modal priming in Chapters **3**, **4A**, and **5A**; eye-tracking in **4B**; and self-paced reading in **5B** and **5C**) is that L2 speakers show delayed access, both figurative and literal, compared to L1 speakers, as evidenced by slower reaction times, later looks, and slower reading times. This result is not particularly surprising, as L2 speakers are slower in their processing in general (see e.g., Cutler, 2012). However, this must be taken carefully into consideration when interpreting results. For instance, **Chapter 4A** used several variations of timing in two cross-modal priming studies in order to determine whether L2 listeners, like L1 listeners, can have access to the figurative meaning of an idiom once the idiomatic key has been presented (e.g., Cacciari & Tabossi, 1988; Tabossi & Zardon, 1993). Figuratively and literally related targets were presented at the offset of the idiomatic key rather than at the offset of the final idiom constituent word, as in **Chapter 3**. The experiments failed to produce figurative facilitatory priming effects, but it notably also failed to produce literal facilitatory priming effects, both of which were present in **Chapters 3** and **5A**. Without considering the generally slower reaction times and more difficult processing overall, a conclusion may be that L1 speakers have earlier access to figurative meaning than L2 speakers. While this indeed may be the case, as figurative meaning has been consistently slower in L2 listeners compared to L1 listeners (see discussion in Cieślicka, 2006), this conclusion can only be drawn when both groups are looked at together, and both types of meaning are compared. In **Chapter 3**, a full analysis with both speaker groups was unable to clearly corroborate this difference. The two language groups differed overall in speed, and the differences found between speaker groups based on figurativeness as a factor included both related and unrelated targets, suggesting that the figuratively related targets alone were not overall slower for L2 listeners compared to L1 listeners. Thus, L2 speakers are not necessarily disadvantaged in figurative meaning access, but rather, they show overall processing delays.

Also of note considering the timeline of access to meaning are the results from the eye-tracking experiment in Chapter **4B**. In addition to the evidence that overall recognition of idioms and activation of meaning (here, literal constituent meaning) is slower than in L1 listeners, the strength of the late activation of literal meaning may also give an indication that timelines for meaning access diverge. In

spite of the study's shortcomings, discussed above, the strength of the late surge in literal constituent activation may indeed point to ongoing literal meaning interpretation not present in L1 speakers. It may also highlight the efficiency present in native idiom processing. That is, whereas native speakers abandon irrelevant meaning activation as time goes on during processing, this may not be a strategy fully available to all L2 speakers. Indeed, L2 processing may be slightly less linear or efficient, where literal meanings are concerned, and suppressing such meanings or abandoning their processing may take place much later, if at all. This issue will be discussed further in 8.2.

8.1.2 Flexibility in meaning access and factors that may influence access to figurative and literal meaning

The first and second aspects of this point concern context: first, global context, and second situational context. **Chapter 5A** considered flexibility based on an aspect of global context: situational or experimental context. This experiment followed the same procedures and used the same materials from the cross-modal priming study in **Chapter 3** and asked whether a highly figurative or literal context can influence the availability of meaning in L2 participants. However, L2 participants saw one of two versions of the experiment. In the high-idiomatic context, a majority of experimental items (target sentences) were sentences containing idioms, and in the low-idiomatic context, only a minority of experimental items contained idioms. Similar to the offline L1 results from Bobrow and Bell (1973), L2 participants showed evidence of differing amounts of facilitatory priming for figuratively related targets depending on the context (i.e., high-idiomatic context increased figurative priming). However, there were no significant differences in the priming of literally related targets between the two contexts. While there is a possibility that literal priming does not show differences because it is indeed more dominant and less subject to flexibility in L2 listeners, it is also possible that its strength is less changed by context based on the associative nature of the literally related targets (e.g., Nelson et al., 1998) which is more direct than figurative target relationships.

Chapters **5B** and **5C** look at the impact of linguistic biasing context on literal and figurative readings of idioms. As mentioned in Section 8.1.1, these self-paced reading studies looked at both L1 and L2 readers and aimed to see whether context can impact access to meaning in terms of reading times. Supporting the results in **Chapter 5A**, figuratively resolving sentences were read faster following figurative contexts for both L1 and L2 readers. Additionally, in the case of L1 readers, literally resolving endings were also read faster following literally biasing contexts when idioms were highly literal. This result was not significant in L2 readers, though, the data visually replicates the L1 data. Again, these experiments support the finding that context may ease processing for figurative readings of idioms in both L1 and L2 speakers. Additionally, the somewhat novel result that literal contexts may also improve literal readings of idioms was also present for L1 speakers, though not for L2 speakers. Again, the evidence here, too, may point to a difference in literal access during idiom

processing for L2 readers, as literal readings seem not be influenced be context. It could be that either the effect was not strong enough to reach significance, indicating that no difference is present, or a general preference for or reliance on literal processing in L2 speakers may also be to blame, as a default would be less flexible and less likely to be impacted by contextual factors. However, the presence of another, idiomatic property may also be to blame for this lack of result, which will be discussed below.

The third aspect of flexibility in meaning access addressed here is the influence of idiom-specific properties. Chapter 2 identified many properties that may be responsible for differences in processing, and two of them examined in this thesis were literality and predictability. Predictability will be discussed below in 8.1.3, but the experiments in Chapters **5B** and **5C** looked at literality. As already mentioned, two types of idioms were included in the self-paced reading studies: high- and low-literality idioms. The literality of an idiom concerns the extent to which an idiom has the potential to be interpreted literally. Looked at in combination with context, this can shed light on whether some idiomatic cues, such as this one, have the potential to interact with or even override linguistic cues and impact meaning access. In line with previous findings (e.g., Titone & Connine, 1994b), literality did impact L1 meaning access in a critical manner: readers showed faster reading times for figuratively resolving sentences whenever low-literality idioms were present, regardless of context. On the other hand, readings of high-literality idioms were impacted by context in both directions (i.e., literal and figurative readings). While this result was not present in L2 readers, the pattern of reading times replicated those of L1 readers. Thus, we are again left with the difficulty of identifying why L1 access to meaning showed more flexibility than L2 access, particularly considering literal meaning, as here impacted by literality. Again, the options seem to be first, to consider that literal meaning is less flexible in nature in L2 speakers because of its importance in processing, or second, to consider that L2 speakers, lacking experience, may be overall less sensitive to such properties than L1 speakers. In the case of literality, identifying an idiom's potential for literal interpretation is just as possible for L2 speakers as L1 speakers (see **Chapter 2**). However, a lack of experience with these idioms may indicate less of a preference to do so in practice, meaning less sensitivity to this idiomatic property. Critically, there is evidence that exposure may also be an important driver in idiom processing (see e.g., Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tabossi et al., 2009), which would support this possible interpretation. This topic, however, will be discussed further in the following sections.

Finally, flexibility in access to meaning was also addressed based on L2-specific challenges, in particular, the impact of the L1 on L2 access to meaning. One aspect addressed frequently in offline studies (e.g., Irujo, 1986; Liontas, 2002, 2015) and more recently in online studies (e.g., Carrol et al., 2016; Carrol & Conklin, 2014) is the translatability of idioms between a speaker's L1 and L2. In **Chapter 3**, two types of idioms were included in the study: idioms were either translatable in a word-for-word manner with equivalent meanings between idioms or they did not overlap in words or in

meaning. Following expectations based on offline evidence, translatable idioms were expected to show larger figurative priming effects than non-translatable idioms in L2 listeners. However, no such evidence of this facilitation effect was found in our data. It may be the case that, for highly proficient L2 speakers, facilitation is not present as they are not using L1 knowledge or associations in their L2 idiom processing. A study published the same year, however, suggests that an idiom's L1 familiarity is just as important in this facilitation as an idiom's L2 familiarity (Carrol et al., 2016), an aspect not controlled for in our **Chapter 3** study. Thus, while some aspects discussed here show evidence of flexibility to meaning access in L1 and L2 speakers, others do not, particularly where L2 speakers are concerned.

8.1.3 Processes associated with the formulaic nature of idioms

Chapter 4 (**4A** and **4B**) deals with processing aspects that are often attributed to the formulaic or unitized nature of idioms (e.g., Wray, 2002). In contrast to solely looking at the figurative aspect of idiomatic processing, studies focusing on this aspect may also look at general processing advantages associated with idioms and compare these to novel language (e.g., Carrol & Conklin, 2019; Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tabossi et al., 2009). The first aspect of this issue is whether the idiomatic property predictability impacts access to figurative meaning. Specifically, this chapter looked at the role of the idiomatic key, after which point L1 listeners have been shown to have access to the figurative meaning of an idiom (e.g., Cacciari & Tabossi, 1988; Tabossi & Zardon, 1993). This property follows from the highly formulaic nature of idioms, which causes many of them to be highly predictable (see **Chapter 2**). Where idioms are highly predictable, so too are their figurative meanings for native speakers. In this set of experiments, L2 listeners were presented with literally and figuratively related targets at the offset of the idiomatic key, and in Experiment 1 the auditory stimuli continued, whereas in Experiment 2 the auditory stimuli stopped. Unlike the earlier L1 evidence of facilitatory priming for figurative meanings, there were no significant results for the L2 listeners in these experiments. Though, this lack of finding held for both literally and figuratively related targets. Again, only speculation is possible. A lack of finding may indeed show that L2 listeners are not impacted by an idiom's predictability as they may also not be as sensitive to idiomatic advantages based on their common form (e.g., Siyanova-Chanturia, Conklin, & Schmitt, 2011; Siyanova-Chanturia, Conklin, & van Heuven, 2011). However, considering the extra time needed in L2 processing (see 8.1.1), there may still be advantages present that were not detected given the timing involved in the task. Immediate target presentation after the key was unsuccessful in producing detectable effects, but presenting the targets one word later, while still before the end of the idiom in many cases, may have indeed shown priming effects not seen as early as the idiomatic key, which was based on L1 data. However, another possibility exists: language

learners may not be as skilled at the integration processes needed to access meaning based on upcoming linguistic cues.

The second aspect of this issue under investigation, directly following the negative results from **Chapter 4A**, is whether language learners can predict upcoming words within idioms, and how this impacts literal word activation. Consistently, L1 readers and listeners have been shown to predict upcoming words in idioms as evidenced by meaning activation of upcoming words in ERP studies (Keßler et al., submitted; Rommers et al., 2013) and eye-tracking (Keßler et al., submitted), but a research gap existed here for language learners. In the study in 4A, L1 and L2 language learners were tested using the eye-tracking methods already described in Section 8.1.1. Both learner groups showed a significant amount of looks to the correct idiom completion before the offset of the final word in the auditory cue (e.g., translated: before the offset of the final THE in *let the cat out of the*…). Thus, the final word was successfully predicted early during listening. Notably, though, this ability was graded between L1 adults (Keßler et al., submitted), who were the fastest, L1 children, and L2 adults, who were the slowest. Thus, while L2 learners indeed have the ability to predict these upcoming words, in line with other types of predictive processes (e.g., Kaan, 2014), they are not as skilled as native speakers in doing so. This result may also be in line with research suggesting that idiomatic processing advantages stem from experience, of which L2 speakers are lacking (e.g., Siyanova-Chanturia, Conklin, & van Heuven, 2011). Such an interpretation further supports weaker or non-significant results concerning flexibility in meaning access already discussed above, as experience may also dictate which cues are taken into account during idiomatic processing.

The impact of predictive processes on literal meaning access, already discussed in 8.1.1, also showed differences between L1 and L2 speakers. After recognition of the correct final constituent word, as evidenced by significant looks to this word, L1 speakers, both children and adults, showed generally decreasing looks to the literally related competitor word. In contrast, L2 to looks to this competitor word did not show such a simple pattern; rather, looks here displayed a late peak following recognition, before decreasing. Thus, where it seems that L1 literal activation can be inhibited once a highly predictable idiom is recognized (e.g., Keßler et al., submitted; Rommers et al., 2013; Titone & Connine, 1994b), this may not be the case for L2 literal activation.

8.1.4 Learning and memory for idioms

The studies in **Chapter 6** look at aspects of memory for idioms as another way of teasing apart possible differences in issues of processing between figurative and literal phrases. The first facet of this issue is whether idioms are recalled differently (i.e., better) than literal phrases. While evidence from L1 children (e.g., Reuterskiöld & Van Lancker Sidtis, 2012) and some L1 adult literature (e.g., Schachter & McGlynn, 1989) suggests that memory for and storage of idioms and similar literal phrases may be

fundamentally different, there was not yet evidence from L2 learners. In a learning and testing paradigm, the same phrases were presented with either a literal paraphrase or an idiomatic meaning to L2 learners of English. Two days later, participants were tested on recall of the form and meaning of these phrases. Against expectations, there were no differences in recall of literal and figurative phrases. As previous studies used similar but not equal phrases, these differences could neither clearly be associated with L2 differences nor with a more controlled testing paradigm. The same phrases were then presented to L1 participants in the same manner in a second experiment, and again no differences were found. Thus, these results strengthen the possibilities suggested in 8.1.3 that exposure rather than figurativeness may be a critical component in many of the advantages often associated with idiomatic processing. However, it may also be the case that idiomatic advantages found in auditory experiments have additional cues not found when tested only visually (e.g., Van Lancker et al., 1981; Van Lancker Sidtis, 2003), and other learning and testing measures are needed. It may also be the case that processing advantages do not apply straightforwardly to memory. Thus, while these results further impress upon the similarities in figurative and literal language processing, more evidence may be needed concerning memory specifically.

The second aspect of this issue concerns L2-specific challenges that may affect memory for and learning of idioms, in particular, competing unfamiliar cues on the word and phrasal level. In addition to varying the figurativeness of the phrases presented, each phrase was presented with either only frequent, familiar words or the same phrase with a single word replaced with an infrequent, unfamiliar word (e.g., morning and bell in *the morning <u>bell/bugle</u> sounded*). Unlike the figurativeness of the phrase, the type of word included had a clear impact on recall (both form and meaning): recall was worse where unfamiliar words were present for L2 learners, though L1 learners showed the opposite effect. In the case that competition negatively impacted recall in L2 learners, the saliency of unexpected, infrequent words may have been at play for L1 learners (e.g., Giora, 1997). Interestingly, however, a third translation task was included in the testing phase for L2 learners in which participants were asked to translate the key word (e.g., *bell* or *bugle* from the example above) back into the German gloss presented with the phrase in the learning session. In this task, translation scores were better overall for unfamiliar words presented in literal compared to figurative phrases. This result, however, does not directly point to recall differences based on figurativeness. Rather, this result is interpreted as in line with collocation evidence that individual words are learned better as part of collocational phrases rather than alone (e.g., Kasahara, 2011). In other words, figurative phrases do not enrich meaning processes via additional linguistic or contextual information in the way that literal phrases do. Thus, **Chapter 6** does not provide clear evidence for memory differences based on figurativeness alone, though, the constituent words present in these phrases do seem to play an important role in recall.

8.1.5 Brain activation during idiom comprehension

The final point of comparison between L1 and L2 idiom processing concerns brain activation during comprehension. The first point, again looking for differences between literal and figurative phrases, has been looked at prominently in native speakers, and has not before been looked at in non-native speakers. The study in **Chapter 7** compared reading of figurative and similar literal phrases (*kick the bucket/ball*) using fMRI. L1 research has suggested possible hemispheric differences in processing these two types of phrases based on figurativeness (i.e., figurative processing in the right hemisphere, see e.g., Winner & Gardner, 1977) and saliency (e.g., Cieślicka, 2013a), though more recent metanalyses of the relevant literature have suggested that this is likely not the case (e.g., Cacciari & Papagno, 2012; A. M. Rapp, 2018). Participants actively read the target sentences and responded to filler sentences with a button press. Looking at both L1 and L2 participants together, and separately, the results comparing figurative and literal activation showed only minimal differences in brain activation. Rather, both phrase-types activated wide areas in the brain in both hemispheres, adding more negative evidence to an already mixed field (e.g., Raposo et al., 2009; Rapp, 2018). Of course, it is still possible that differences indeed are present on this level, and these differences may have been masked by the task (e.g., Rapp et al., 2012) or by the type of idiom included in the study (e.g., Fanari et al., 2010; Titone & Connine, 1994b).

Looking at differences between L1 and L2 speakers during idiom comprehension, there were a number of differences in brain activations, though not in the areas classically associated with sentence-level language. Though, interestingly, L1 readers engaged more areas of the brain during idiom processing than L2 readers. While this result may seem surprising at first glance, as more activation may automatically be associated with more effort, this is indeed in line with other findings in the fMRI literature (e.g., Citron et al., 2020). This finding may suggest that L1 readers were more actively engaged during figurative language processing, whereas L2 readers were less able to do so. This may, again, be due to experience with the idioms, a more literal-based processing strategy, or other language-based differences not accounted for in our study. But overall, as differences were present in both directions between groups, qualitative rather than quantitative differences remain in focus.

Finally, the last point concerning this aspect of processing is whether proficiency impacts L2 idiom comprehension. The LexTale (Lemhöfer & Broersma, 2012) test was used as a measure of proficiency and was taken for all participants, including L1 speakers. Brain activation during idiom comprehension was considered using this measure as a marker, and L2 speakers showed a number of negative correlations, while L1 readers showed a positive one. Taken together with the results concerning L1 and L2 idiom comprehension above, this may be an indication that areas associated

with more effortful L2 processing recede in activation over time. Thus, L2 processing may become more native-like with increasing proficiency.

8.2 L1 AND L2 IDIOM PROCESSING: ONE MODEL FOR BOTH

Considering the findings summarized in Section 8.1, this section will address the overarching question, and final goal set for this dissertation: How do L1 and L2 processing compare? Specifically, the goal was set to answer this question by using the research already discussed to determine whether L1 and L2 speakers use different processing strategies and/or mechanisms during idiom comprehension.

In short, it seems that the mechanisms used during idiom processing for native and non-native speakers are likely one in the same. However, there are some critical differences which must also be considered. A comprehensive model of idiom processing for both speaker groups must allow for the following, based on the findings outlined above:

1. Fast, online access to figurative meaning

   Both speaker groups showed consistent access to figurative meaning using a variety of experimental methods throughout this work. This included the cross-modal priming studies in **Chapters 3** and **5A** as well as the self-paced reading studies in **Chapters 5B** and **5C**.

2. Fast, online access to literal constituent meaning

   Both speaker groups showed consistent activation of literal constituent meaning, again using a variety of experimental methods, even where figurative meaning was in focus. In addition to access in the cross-modal priming studies in **Chapters 3** and **5A**, this was also evident in the eye-tracking study in **Chapter 4B**.

3. Flexibility in access to both figurative and literal meaning based on idiomatic, linguistic, and global cues, considering in particular:

   a. Speed of access to figurative meaning

      Both global, figurative contexts as well as local linguistic contexts can increase priming effects as in the former (**Chapter 5A**) and speed up reading times as in the latter (**Chapters 5B and 5C**). Additionally, though not substantiated for L2 listeners (**Chapter 4A**), predictability can affect this access in L1 speakers.

   b. Availability of literal phrasal meaning

Linguistic biasing context can also impact the availability of the literal meaning in L1 readers as evidenced by faster reading times (**Chapter 5B**). Furthermore, this phrasal meaning availability is also affected by idiomatic properties and can be unavailable or extremely costly in the case of non-literal idioms (**Chapter 5B**). Again, though, these effects were not substantiated in L2 speaker results.

4. A close connection between figurative and literal processing that does not involve two separate processes

   Both the regular availability of literal and figurative meaning already listed as well as missing evidence of clear differences support this tenant. The lack of evidence suggesting differences that are based on figurativeness alone such as memory for idioms compared to literal phrases (**Chapter 6**) and brain activation (**Chapter 7**) suggests that literal and figurative comprehension may be best served with a single mechanism that takes both into account.

Additionally, a model that accounts for L1 and L2 idiomatic processing must also consider the differences found in the research presented here, some of which are clear in the previously listed requirements:

1. Overall slower processing of or access to figurative and literal meanings

   Evidence of this general speed disadvantage were present in studies that compared both L1 and L2 adults directly via reaction times from cross-modal priming experiments (**Chapters 3** and **5A**), recognition times calculated from the percentage of looks in eye-tracking (**Chapter 4B**), and reading times from self-paced reading (**Chapters 5B** and **5C**).

2. Less sensitivity to idiomatic properties

   i. Predictability

      While this has been shown to affect L1 access to figurative meaning in previous studies on which the study in **Chapter 4A** was based (e.g., Cacciari & Tabossi, 1988; Tabossi & Zardon, 1993), there was no evidence of sensitivity to this property in the current work. Though the possible reasons for this lack of finding are unclear (e.g., weaker or later activation), L2 speakers seem generally less sensitive to this property, at least in early processing.

ii. Literality

Based on this property, clear effects produced in L1 readers (**Chapter 5B**) were not present in L2 readers (**Chapter 5C**). However, the pattern of reading times in L2 results, mirroring the L1 patterns, suggest that here, too, it may be the case that effects based on literality were not strong enough to be measured in the task used. Thus, less sensitivity to this property should also be considered in this case.

3. Less flexibility in literal meaning access (constituent and phrasal)

Consistently, L2 speakers showed fewer effects that directly impacted the availability of literal meaning. In studies showing clear advantages to figurative interpretations, such an increase in priming (**Chapter 5A**) or faster reading times (**Chapter 5C**), literal constituent and phrasal meanings were unaffected. This is also reflected in the eye-tracking study (**Chapter 4B**) in which literal constituent activation does not decrease immediately after idiom recognition, unlike in L1 speakers. While the reasons for these findings may differ based on experimental design and method, the pattern that literal meaning access is not as flexible as figurative meaning access in L2 speakers should not be dismissed.

4. More sensitivity to competition of unfamiliar cues during learning

Unsurprisingly, L2 speakers also showed disadvantages in memory when an additional unfamiliar or salient cues were present (**Chapter 6**). While a model of processing need not address learning issues directly, it must, again, allow for weaknesses such as this one that may interfere with fast and eased processing.

Overall, it seems that indeed L1 and L2 processing harbor more similarities than differences, and many of the differences listed above may be a direct result of experience with the language. Experience may easily account for speed differences as well as varying sensitivities to the properties that are known to affect L1 processing. However, the issue of the lack of flexibility in access to literal meaning is indeed critical, as it is addressed frequently in the literature, and has been often addressed in this work. Whether this difference is fundamental in nature will be addressed below. First, however, the hybrid models presented in Section 1.2.2. will be addressed, ignoring step-wise models, as they are neither favored in the literature, nor supported by the findings presented here. Then, the possibility of a literal-based difference in L1 and L2 processing will be considered.

First, the Configuration Hypothesis presents of view of idiomatic processing in which literal processing proceeds necessarily until a point at which enough information has been accrued to retrieve

the figurative meaning. This point was identified early on as the idiomatic key (e.g., Cacciari & Tabossi, 1988), though later studies suggest that this point may be more flexible than single word (e.g., Tabossi & Zardon, 1993). The suggestion of fast, online access to figurative meaning, based on recognition that may happen early during comprehension is clearly in line with our L1 data. However, our L2 data are somewhat at odds with this characterization, particularly considering an idiomatic key. However, with the adjustment to the concept of the idiomatic key suggested in **Chapter 4A**, this hypothesis may still align with L2 data. Specifically, the idiomatic key, as a point of identification must be a more variable concept. As the frequency of occurrence or encounter with a variety of phrases, including idioms can increases the likelihood of fast recognition (e.g., Siyanova-Chanturia, Conklin, & van Heuven, 2011), a key may be more or less engrained in a person's mental lexicon. Just as L2 speakers are less sensitive to frequency effects, this seems also to apply to idioms. Furthermore, the identifying word may vary from speaker to speaker, particularly in the case of L2 speakers. While a native speaker may recognize the idiom *let the cat out of the bag* already on the word OUT, it may be that a non-native speaker only recognizes the idiom on the final THE, if at all. Thus, for access to figurative meaning, the research presented here may still be in line with this hypothesis. However, what is less clear, is the treatment of literal language access following recognition. Cacciari and Tabossi (1988) first suggested that literal processing and therefore activation no longer occurs after recognition, a tenant of the hypothesis not in line with the research in this work, Cacciari (2014) later adjusted this hypothesis to suggest that literal processing may continue on an automatic but shallow, syntactic-basis (see e.g., Peterson et al., 2001). See Figure 8-1 below. While this point cannot be directly supported or refuted by the cross-modal priming data presented throughout this dissertation, it does not seem entirely in line with the self-paced reading data presented in **Chapters 5B** and **5C**. While it may be that highly literal idioms are recognized later, and this literal processing may occur in the usual manner in these idioms compared to low-literality idioms, it seems to suggest that highly literal idioms in literally biasing sentences are not recognized as idioms or only recognized very late. While this may be the case for some idioms, it may not represent all idioms, and the behavior of literal processing remains a weak point for this hypothesis. Furthermore, all flexibility in meaning access must be attributed to recognition of the idiom, for both L1 and L2 speakers. In this case, all sensitivity issues follow from a lack of or strength in recognition. While this may be the case, it accounts for the amount of variability seen in L2 processing weakly. Thus, considering the variability of meaning access shown by L1 and L2 speakers in the current thesis, while not entirely refuted, the Configuration Hypothesis as it stands is also not entirely sufficient.

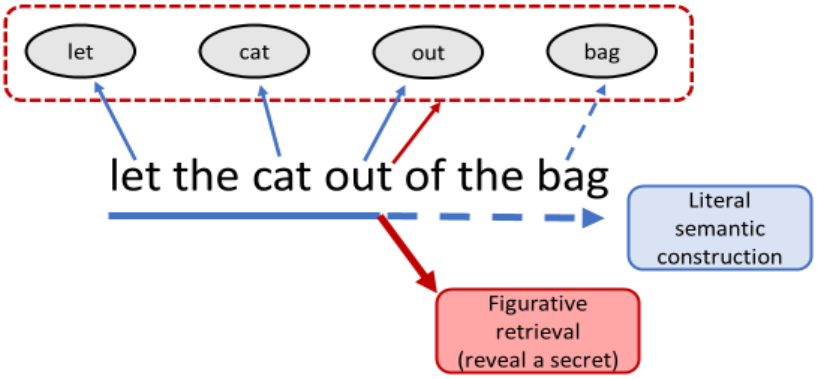


Figure 8-1. Illustration of the Configuration Hypothesis

Note: This illustration shows how processing may proceed temporally following the Configuration Hypothesis. Blue lines represent literal meaning access and red figurative meaning access. The timing of meaning access follows from left to right starting at the beginning of the phrase. The concepts for each word are in black ovals above the phrase, and the red outline represents the association of words.

The next model for consideration is the Hybrid Model of Idiom Representation (Sprenger et al., 2006). Notably, this model was devised for representation based on production and not processing. However, the claims made in this hypothesis are easily applied to processing, and it has since been considered widely in this manner. In this model, idioms have unitized entries on a lexical-syntactic level called superlemmas that are separate from individual word entries on the conceptual level and the lemma level. The superlemma then accounts for the missing connection between the two levels of concepts and words, and both the superlemmas as well as the individual concepts are co-actived, and the activation of one can spread to activate the other. In practice, this means that fast activation of the figurative meaning immediately affects activation of the individual literal constituents. It also suggests that activation can be cumulative, as individual concepts compete, in a sense, with the wholistic meaning of the idiom, and like the Configuration Hypothesis, meaning may build in strength over time. This model accounts immediately for the fast figurative activation as well as the availability of literal constituent meaning necessary in a successful model of processing. See Figure 8-2 for a depiction. Additionally, it suggests that processing idioms is not different than literal language, rather, it includes an additional activation of the superlemma. Flexibility can also be easily accounted for in this model, as the integration of linguistic and global cues during processing occurs in the same manner as it would during literal language processing. An increase in speed of access to one meaning or another, is then accounted for by a strengthening of the relevant superlemma (in the case of a figurative interpretation) or concepts (in the case of a literal interpretation). For this reason, literal processing should proceed as

long as the literal concepts remain in competition with the superlemma. It also follows that the differences found in L2 processing in the current work can also be explained as basic differences in language processing based on proficiency. In particular, the lack of flexibility of literal meaning activation may be due to stronger, more integrated literal concepts, and weaker supperlemmas, which increase greatly in strength and activation when pushed by additional cues such as context. One issue with this model not explored in the current work is the idea, unique to this model, that the figurative meaning of an idiom becomes active immediately. The research presented here cannot directly support this claim of the model, and the L2 data may even refute this tenant. However, an immediate activation in its weakest form may not be detectable via the methods used in the current work, and therefore it is also not inconsistent with the L1 and L2 data presented here.

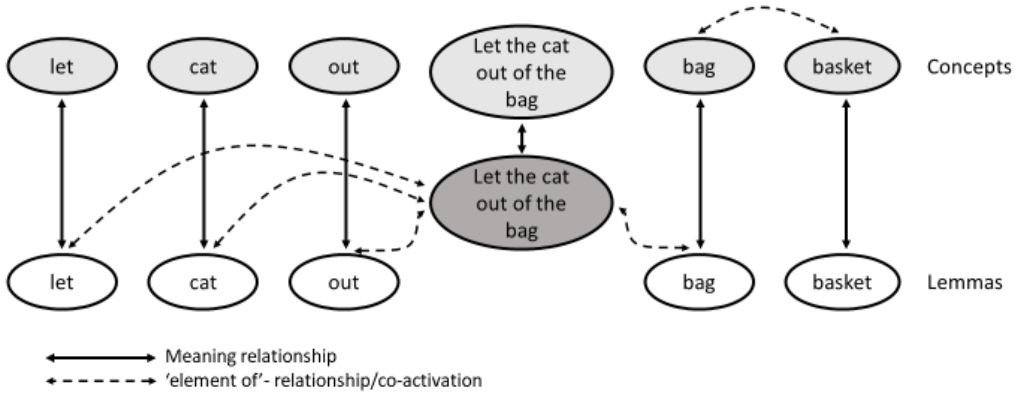


Figure 8-2.Illustration of the Hybrid Model of Idiom Representation for the idiom: let the cat out of the bag

Note: This is a recreation of the representation of idioms and constituent words recreated from Sprenger et al. (2006) for the idiom *let the cat out of the bag*. Processing levels are represented from top to bottom, and objects aligned are part of the same level.

Finally, the Constraint-based Model of Idiom Processing (Libben & Titone, 2008) should also be considered. Much like the Hybrid Model of Idiom Representation, individual words play a critical role in idiomatic processing. Following this model, idiomatic processing occurs in the same manner as literal processing, but familiar idiomatic phrases are represented and retrieved from the memory as units which can be recognized early in processing, and these meanings may interact with compositional phrasal (i.e., literal) analysis during processing. Following this model, just as in literal language processing, different aspects of linguistic environment and the phrase itself can interact with the time-course of meaning access throughout processing. The holistic storage and retrieval of idioms accounts well for the fast access to idiomatic meaning, and as literal processing continues, the activation of both the constituents and sometimes the phrase likewise does not pose a problem for this model. Additionally, it accounts for the flexibility of meaning, as factors such as predictability and literality of the idiom are easily integrated into the processing mechanism as they are recognized and

encountered, and it accounts for the variability already well-established in idioms that was not further explored in the current work (e.g., Nunberg et al., 1994). Furthermore, like the previous model, differences in L1 and L2 results can be accounted for based on experience, in particular because this model puts a high amount of importance on the familiarity of an idiom in its fast processing and early recognition. As seen in **Chapter 2**, familiarity does significantly differ between L1 and L2 speakers. However, this model still seems to consider idioms as separate entries in the lexicon, and the way these entries interact with ongoing processing is not entirely clear. In fact, it may be that this interaction is like the one presented by Sprenger et al. (2006). Thus, while also compatible with the data presented in the current work, the open questions about the temporal interaction between idiomatic representation and ongoing literal processing are less desirable than the answers provided more directly by the hybrid model considering superlemmas.

Finally, it is critical to look more closely at the Literal Saliency Model (Cieślicka, 2006) and whether the differences found in literal meaning activation between L1 and L2 warrant a differential approach in idiom processing between speaker groups. According to Cieślicka's model (see also Giora, 1997), literal processing has a necessary priority for non-native speakers as literal language is more salient than figurative language. For native speakers, who have encountered idioms in their figurative senses more often, figurative meanings may have a more salient status and thus be processed first. The general finding that literal meaning is less flexible and sensitive to change in L2 speakers (e.g., **Chapters 5A** and **5C**) as well as that literal meaning may be more strongly activated (e.g., **Chapter 4B**), seem to generally support such a claim. If literal language is the default, then it should be less sensitive to outside influence than a secondary processing route. However, each of these cases also had alternative explanations for the lack of findings (i.e., weaker effects, ceiling effects, or even a lack of inhibitory skills) which cannot be entirely discounted. Following this model, we also may not have expected to find the regularity of fast figurative access that was overall comparable to L1 speakers' access (e.g., **Chapters 3** and **5A**). Additionally, saliency may be integrated into the models of processing already presented without the necessity of a literal processing priority. For example, saliency is a concept that Giora (1997) explains is a cumulative function of factors such as familiarity, frequency, conventionality, or usage in a specific context. Each of these factors may impact the recognition point of an idiom (Configuration Hypothesis), the strength of associations and representations (Superlemma Hypothesis), or even how it is represented in the lexicon (Constraint-based Model). In each of these cases, literal meaning may indirectly be stronger or more active during online processing without the strict constraint that literal processing occurs before figurative processing. Because of this inflexibility inherent to this model, the current research is more in line with the hybrid theories presented above, which may also consider the saliency of an idiom's figurative

meaning, in both L1 and L2 processing. Additionally, more evidence supporting the similarities in L1 and L2 processing suggest that a singular processing mechanism for both groups suffices.

Thus, in light of the data presented in this work, and its compatibility with the models presented here, the Hybrid Model of Idiom Representation is best-suited to describe L1 and L2 idiom processing. Furthermore, while L1 and L2 idiom processing studies often produce varied results, the mechanisms appear to be the same. Differences can be accounted for based on general L2 processing challenges, familiarity with or exposure to idioms and their idiomatic meanings, and proficiency. Indeed, it follows that it may be the case that many differences disappear as speakers become more proficient and many of the constraints less-skilled comprehenders face decrease or even fade altogether. Though, future research must consider more varied proficiencies in order to confirm this suggestion.

8.3 IMPLICATIONS FOR LANGUAGE TEACHING AND LEARNING

In addition to a scientific contribution to the field of figurative language processing, the current research has implications for L2 teaching and learning. As learners become increasingly proficient in a language, they are expected to master not only literal expressions, but also idiomatic ones (e.g., Common European Framework of Reference for Languages, 2001). The challenges associated with this learning are well-established (e.g., Cooper, 1999), and there is still to-date little research in this field (Boers, 2001; Boers & Lindstromberg, 2005; Steinel et al., 2007). Though, such research is far from comprehensive and often lacking experimental controls.

Many of the results show promisingly, that L2 idiom processing behaves like L1 idiom processing, and learners should be given confidence in this finding—learners need not believe that they are at an indefinite disadvantage when it comes to idiomatic comprehension. However, a simple solution for overcoming the challenges of idiomatic processing presented here (e.g., a lower sensitivity to idiomatic properties and changes in literal processing) is still unclear. However, if many of the disadvantages associated with idiomatic processing are associated with familiarity and exposure to the idioms, then the simplest solution for learners, is to gain exposure, and the solution for L2 teachers, is to integrate idioms, and other formulaic expressions, into their L2 classrooms mor regularly.

With the goal of increasing exposure to learners of idioms, however, a few thoughtful applications should be taken into consideration. Critically, while the figurativeness of an idiom is an important part of what identifies idioms and makes them challenging for L2 learners, it may not be the key to learning them. Rather, identifying and reinforcing their formulaic nature may be the path forward. Formulaic phrases are an important building block in language (e.g., Wray, 2002), though this may be the case more for L1 learners than L2 learners (e.g., Arnon & Christiansen, 2017). Whereas L1 learners show evidence of building language in chunks, L2 learners are acutely aware of the

contribution of individual words in building meaning, and this is often the basis of learning in many L2 classrooms. It may be precisely this focus that encourages literal saliency (e.g., Cieślicka, 2006) and the strength of literal activations even where figurative meaning is encouraged (e.g., **Chapters 4B** and **5C**). In order to attempt to overcome this L2 tendency, classrooms must put more focus on larger units including, but not limited to, idioms in the teaching curriculum (e.g., Siyanova-Chanturia & Martinez, 2014). Although some additional lists of such phrases have already entered L2 classrooms (e.g., Ellis, Simpson-Vlach, & Maynard, 2008; Simpson-Vlach & Ellis, 2010), this effort must be broader and more sustained in order to impact more learners. Though, as **Chapter 6** suggests, even a few short exposures may be enough for learning in a productive learning environment.

Additionally, and critically for figurative language teaching, are the implications of the role of constituent words on learning from **Chapter 6**. While the study showed no differences in overall recall between figurative and literal phrases, two important differences were found. First, the presence of unfamiliar words was generally a hinderance to overall recall. This is critical in finding a balance between teaching more on the basis of phrases, as suggested above, and making sure that the individual words do not prove a distraction to the learning of these chunks. Thus, when learners need to learn both the words and the phrases at once, overall learning may suffer. Additionally, and relatedly, the second observation was that the constituent words were translated correctly more often when they were part of literal and not figurative phrases. This suggests that word- and phrasal-learning should not be combined when they do not aid in meaning (i.e., in figurative phrases where the words do not contribute to the meaning of the phrase). Thus, it is important that unfamiliar words present in idioms be learned first, and done so separately. Even better so if these words are learned in literal collocational phrases (e.g., Kasahara, 2011).

8.4 CONCLUSIONS

The present work has presented research on native and non-native idiom processing and has looked closely at how they compare with one another. As a whole, idiom processing does not seem to proceed via differing mechanisms between languages. Rather, a flexible model of processing must encompass the variability of meaning access (both literal and figurative) by allowing idiomatic processing to proceed in a manner connected with and fundamentally similar to literal processing. Both native and non-native speakers have online access to both figurative and literal constituent meaning, and literal phrasal meaning may also be constructed in some cases. While there were some differences established between listeners and readers of both language groups, these differences can be placed in the context of overarching differences that only partly stem from idiomatic properties, including figurativeness. Rather, differences are likely based on more general challenges that L2 learners face in language processing. This viewpoint, based on the various processing evidence from both native and non-native

comprehension presented here as well as a lack of clear differences in literal and figurative language in terms of memory and even brain activation, suggests that idioms may not be as special as they were once thought to be. Indeed, their conventionality and even popularity marks their ordinary nature, and the ease of processing that native speakers enjoy can also apply to proficient non-native speakers. The distinction between figurative and literal meaning is important, as it helps us recognize idioms and use them as effective communication tools, but it should not be used as a division in language processing mechanisms.

9 REFERENCES

Abel, B. (2003a). English idioms in the first language and second language lexicon: a dual representation approach. *Second Language Research*, *19*(4), 329–358. https://doi.org/10.1191/0267658303sr226oa

Abel, B. (2003b). *Sprecherurteile zur Dekomponierbarkeit englischer Idiome*. *Linguistische Arbeiten*. Tübingen: Max Niemeyer Verlag GmbH.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews. Neuroscience*, *7*(4), 268–277. https://doi.org/10.1038/nrn1884

Arnon, I., & Christiansen, M. H. (2017). The Role of Multiword Building Blocks in Explaining L1–L2 Differences. *Topics in Cognitive Science*, *9*(3), 621–636. https://doi.org/10.1111/tops.12271

Arnon, I., & Cohen Priva, U. (2015). Time and Again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, *9*(3), 377–400. https://doi.org/10.1075/ml.9.3.01arn

Ayto, J. (Ed.) (2010). *Oxford paperback reference*. *Oxford dictionary of English idioms* (3. ed.). Oxford: Oxford Univ. Press.

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

RePsychLing: Data sets from Psychology and Linguistics experiments (Version R package version 0.0.4) [Computer software] (2015).

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*, 255–278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models.

Lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-8) [Computer software] (2015): R Package. Retrieved from http://CRAN.R-project.org/package=lme4

Beck, S. D., & Weber, A. (2016a). Bilingual and monolingual idiom processing is cut from the same cloth: The Role of the L1 in literal and figurative meaning activation. *Frontiers in Psychology*, *7*, 1350. https://doi.org/10.3389/fpsyg.2016.01350

Beck, S. D., & Weber, A. (2016b). L2 Idiom processing: Figurative attunement in highly idiomatic contexts. In A. Papafragou, D. Grodner, D. Mirman, & J. C. & Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Beck, S. D., & Weber, A. (2019). Context matters, figuratively, for L2 readers: Evidence from self-paced reading. In *Proceedings of The 11th International Conference on the Mental Lexicon*. Edmonton, AB, Canada: University of Alberta Libraries. https://doi.org/10.7939/R3-MNCB-C270

Beck, S. [Sigrid], & Tiemann, S. (2016). Towards a model of incremental composition. In R. Truswell, C. Cummins, C. Heycock, B. Rabern, & H. Rohde (Chairs), *Sinn und Bedeutung 21*, Edinburgh.

Bernicot, J., Laval, V., & Chaminaud, S. (2007). Nonliteral language forms in children: In what order are they acquired in pragmatics and metapragmatics? *Journal of Pragmatics*, *39*(12), 2115–2132. https://doi.org/10.1016/j.pragma.2007.05.009

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A Critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*(2), 295–308. https://doi.org/10.1037/0278-7393.19.2.295

Bobrow, S. A., & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, *1*(3), 343–346. https://doi.org/10.3758/BF03198118

Boers, F. (2001). Measuring the impact of cross-cultural differences on learners' comprehension of imageable idioms. *ELT Journal*, *55*(3), 255–262. https://doi.org/10.1093/elt/55.3.255

References

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, *10*, 245–261. https://doi.org/10.1191/1362168806lr195oa

Boers, F., & Lindstromberg, S. (2005). Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. *System*, *33*(2), 225–238. https://doi.org/10.1016/j.system.2004.12.007

Bott, O., & Sternefeld, W. (2017). An event semantics with continuations for incremental interpretation. *Journal of Semantics*, *34*(2), 201–236. https://doi.org/10.1093/jos/ffw013

Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(8), 1905–1914. https://doi.org/10.1093/cercor/bhn217

Briellmann, R. S., Saling, M. M., Connell, A. B., Waites, A. B., Abbott, D. F., & Jackson, G. D. (2004). A high-field functional MRI study of quadri-lingual subjects. *Brain and Language*, *89*(3), 531–542. https://doi.org/10.1016/j.bandl.2004.01.008

Briner, S. W., & Virtue, S. (2014). Hemispheric processing of idioms: The influence of familiarity and ambiguity. *Journal of Neurolinguistics*, *28*, 1–18. https://doi.org/10.1016/j.jneuroling.2013.10.001

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. https://doi.org/10.1017/S1366728909990344

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Cacciari, C. (2014). Processing multiword idiomatic strings: Many words in one? *The Mental Lexicon*, *9*(2), 267–293. https://doi.org/10.1075/ml.9.2.05cac

Cacciari, C., & Corradini, P. (2015). Literal analysis and idiom retrieval in ambiguous idioms processing: A reading-time study. *Journal of Cognitive Psychology*, *27*(7), 797–811. https://doi.org/10.1080/20445911.2015.1049178

Cacciari, C., Padovani, R., & Corradini, P. (2007). Exploring the relationship between individuals' speed of processing and their comprehension of spoken idioms. *European Journal of Cognitive Psychology*, *19*, 417–445. https://doi.org/10.1080/09541440600763705

Cacciari, C., & Papagno, C. (2012). Neuropsychological and Neurophysiological Correlates of Idiom Understanding: How many Hemispheres are Involved? In M. Faust (Ed.), *The Handbook of the Neuropsychology of Language*. Oxford, UK: Wiley-Blackwell.

Cacciari, C., & Tabossi, P. (1988). The Comprehension of Idioms. *Journal of Memory and Language*, *27*, 668–683. https://psycnet.apa.org/doi/10.1016/0749-596X(88)90014-9

Canal, P., Pesciarelli, F., Vespignani, F., Molinaro, N., & Cacciari, C. (2017). Basic composition and enriched integration in idiom processing: An EEG study. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(6), 928–943. https://doi.org/10.1037/xlm0000351

Cargnelutti, E., Tomasino, B., & Fabbro, F. (2019). Language brain representation in bilinguals with different age of appropriation and proficiency of the second language: A Meta-analysis of functional imaging studies. *Frontiers in Human Neuroscience*, *13*, 154. https://doi.org/10.3389/fnhum.2019.00154

Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, *17*(04), 784–797. https://doi.org/10.1017/S1366728913000795

Carrol, G., & Conklin, K. (2019). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, 95-122. https://doi.org/10.1177/0023830918823230

Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in Translation: The influence of the L1 on the reading of idioms in a L2. *Studies in Second Language Acquisition*, *38*(3), 404–433. https://doi.org/10.1017/S0272263115000492

Carrol, G., Littlemore, J., & Gillon Dowens, M. (2017). Of false friends and familiar foes: Comparing native and non-native understanding of figurative phrases. *Lingua*, *204*, 21-44. https://doi.org/10.1016/j.lingua.2017.11.001

Charteris-Black, J. (2002). Second language figurative proficiency: A comparative study of Malay and English. *Applied Linguistics*, *23*(1), 104–133. https://doi.org/10.1093/applin/23.1.104

Chen, H.-C., & Ng, M.-L. (1989). Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory & Cognition*, *17*(4), 454–462. https://doi.org/10.3758/BF03202618

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. London: MIT Press. https://doi.org/10.2307/j.ctt17kk81z

Cieślicka, A. B. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, *22*(2), 115–144. https://doi.org/10.1191/0267658306sr263oa

References

Cieślicka, A. B. (2013a). 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*(6), 475–503. https://doi.org/10.1007/s10936-012-9232-4

Cieślicka, A. B. (2013b). Second Language Learners' Processing of Idiomatic Expressions: Does Compositionality Matter. In K. Droździał-Szelest & M. Pawlak (Eds.), *Second Language Learning and Teaching. Psycholinguistic and sociolinguistic perspectives on second language learning and teaching: Studies in honor of Waldemar Marton* (pp. 115–136). Berlin, New York: Springer.

Cieślicka, A. B., & Heredia, R. R. (2011). Hemispheric asymmetries in processing L1 and L2 idioms: Effects of salience and context. *Brain and Language*, *116*(3), 136–150. https://doi.org/10.1016/j.bandl.2010.09.007

Cieślicka, 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.), *Second Language Learning and Teaching. Essential topics in applied linguistics and multilingualism: Studies in honor of David Singleton* (pp. 21–41). Cham [u.a.]: Springer.

Citron, F. M. M., Cacciari, C., Funcke, J. M., Hsu, C.-T., & Jacobs, A. M. (2020). Idiomatic expressions evoke stronger emotional responses in the brain than literal sentences. *Neuropsychologia*, *131*, 233–248. https://doi.org/10.1016/j.neuropsychologia.2019.05.020

Citron, F. M. M., Cacciari, C., Kucharski, M., Beck, L., Conrad, M., & Jacobs, A. M. (2016). When emotions are expressed figuratively: Psycholinguistic and Affective Norms of 619 Idioms for German (PANIG). *Behavior Research Methods*, *48*(1), 91–111. https://doi.org/10.3758/s13428-015-0581-4

Citron, F. M. M., Michaelis, N., & Goldberg, A. E. (2020). Metaphorical language processing and amygdala activation in L1 and L2. *Neuropsychologia*, *140*, 107381. https://doi.org/10.1016/j.neuropsychologia.2020.107381

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, New Jersey: Lawrence Erlbaum Associates, Publishers.

Colombo, L. (1998). Role of context in the comprehension of ambiguous Italian idioms. In D. Hillert (Ed.), *Syntax and Semantics: Vol. 31. Sentence processing: A cross-linguistic perspective* (pp. 405–425). New York: Academic Press.

Colombo, L., & Williams, J. (1990). Effects of word- and sentence-level contexts upon word recognition. *Memory & Cognition*, *18*(2), 153–163. https://doi.org/10.3758/BF03197090

Colston, H. L. (2015). *Using figurative language*. New York, NY: Cambridge University Press.

Common European Framework of Reference for Languages: Learning, teaching, assessment (2001). Cambridge, U.K.: Press Syndicate of the University of Cambridge.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1), 72–89. https://doi.org/10.1093/applin/amm022

Conklin, K., & Schmitt, N. (2012). The Processing of formulaic language. *Annual Review of Applied Linguistics*, *32*, 45–61. https://doi.org/10.1017/S0267190512000074

Cook, M., Murdoch, B., Cahill, L., & Whelan, B.-M. (2004). Higher-level language deficits resulting from left primary cerebellar lesions. *Aphasiology*, *18*(9), 771–784. https://doi.org/10.1080/02687030444000291

Cooper, T. C. (1999). Processing of idioms by L2 learners of English. *TESOL Quarterly*, *33*(2), 233–304. https://doi.org/10.2307/3587719

Cronk, B. C., Lima, S. D., & Schweigert, W. A. (1993). Idioms in sentences: Effects of frequency, literalness, and familiarity. *Journal of Psycholinguistic Research*, *22*(1), 59–82. https://doi.org/10.1007/BF01068157

Cronk, B. C., & Schweigert, W. A. (1992). The comprehension of idioms: The effects of familiarity, literalness, and usage. *Applied Psycholinguistics*, *13*(2), 131–146. https://doi.org/10.1017/S0142716400005531

Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*: MIT Press.

Cutler, A., & Clifton, C. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). Oxford: Oxford Univ. Press.

Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blendsß. *Memory & Cognition*, *25*(1), 57–71. https://doi.org/10.3758/BF03197285

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

De Bleser, R., Dupont, P., Postler, J., Bormans, G., Speelman, D., Mortelmans, L., & Debrock, M. (2003). The organisation of the bilingual lexicon: a PET study. *Journal of Neurolinguistics*, *16*(4-5), 439–456. https://doi.org/10.1016/S0911-6044(03)00022-8

References

De Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, *30*, 90–123. https://doi.org/10.1016/0749-596X(91)90012-9

Desai, R. H., Conant, L. L., Binder, J. R., Park, H., & Seidenberg, M. S. (2013). A piece of the action: Modulation of sensory-motor regions by action idioms and metaphors. *NeuroImage*, *83*. https://doi.org/10.1016/j.neuroimage.2013.07.044

Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, *128*(3), 501–513. https://doi.org/10.1016/j.actpsy.2007.09.004

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*(2), 353–387. https://doi.org/10.1017/S0272263112000915

Duyck, W., & Brysbaert, M. (2004). Forward and backward number translation requires conceptual mediation in both balanced and unbalanced bilinguals. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(5), 889–906. https://doi.org/10.1037/0096-1523.30.5.889

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. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x

Elston-Güttler, K. E., Paulmann, S., & Kotz, S. A. (2005). Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience*, *17*(10), 1593–1610. https://doi.org/10.1162/089892905774597245

Estill, R. B., & Kemper, S. (1982). Interpreting idioms. *Journal of Psycholinguistic Research*, *11*(6), 559–568. https://doi.org/10.1007/BF01067612

Fanari, R., Cacciari, C., & Tabossi, P. (2010). The role of idiom length and context in spoken idiom comprehension. *European Journal of Cognitive Psychology*, *22*(3), 321–334. https://doi.org/10.1080/09541440902843866

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Fernando, C. (1996). *Idioms and idiomaticity*. *Describing English language*. Oxford: Oxford University Press.

Fogliata, A., Rizzo, S., Reati, F., Miniussi, C., Oliveri, M., & Papagno, C. (2007). The time course of idiom processing. *Neuropsychologia*, *45*(14), 3215–3222. https://doi.org/10.1016/j.neuropsychologia.2007.06.009

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(5), 1461–1469. https://doi.org/10.1037/a0036756

Francis, B. (Ed.) (2006). *Oxford idioms: Dictionary for learners of English* (2nd ed.). Oxford: Oxford University Press.

Gernsbacher, M. A., & Faust, M. E. (1991). The Mechanism of suppression: A Component of general comprehension skill. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *17*(2), 245–262. https://doi.org/10.1037/0278-7393.17.2.245

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 430–445. https://doi.org/10.1037/0278-7393.16.3.430

Gibbs, R. W. Jr. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, *8*(2), 149–156. https://doi.org/10.3758/BF03213418

Gibbs, R. W. Jr., & Colston, H. L. (2012). *Interpreting Figurative Meaning*. Cambridge: Cambridge University Press.

Gibbs, R. W. Jr., & Nayak, N. P. (1989). Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology*, *21*, 100–138. https://doi.org/10.1016/0010-0285(89)90004-2

Gibbs, R. W. Jr., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, *28*(5), 576–593. https://doi.org/10.1016/0749-596X(89)90014-4

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183–206. https://doi.org/10.1515/cogl.1997.8.3.183

Giora, R. (2002). Literal vs. figurative language: Different or equal? *Journal of Pragmatics*, *34*(4), 487–506. https://doi.org/10.1016/S0378-2166(01)00045-5

Glucksberg, S. (1989). Metaphors in conversation: How are they understood? Why are they used? *Metaphor and Symbolic Activity*, *4*(3), 125–143. https://doi.org/10.1207/s15327868ms0403_2

Green, D. W., Crinion, J., & Price, C. J. (2006). Convergence, degenerecy, and control. *Language Learning*, *56*, 99–125. https://doi.org/10.1111/j.1467-9922.2006.00357.x

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215. https://doi.org/10.1177/0267658312437990

Günther, F., Dudschig, C., & Kaup, B. (2015). Lsafun - An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*(4), 930–944. https://doi.org/10.3758/s13428-014-0529-0

Hernandez, A. E., & Meschyan, G. (2006). Executive function is necessary to enhance lexical processing in a less proficient L2: Evidence from fMRI during picture naming. *Bilingualism: Language and Cognition*, *9*(2), 177–188. https://doi.org/10.1017/S1366728906002525

Hoenig, K., & Scheef, L. (2005). Mediotemporal contributions to semantic processing: Fmri evidence from ambiguity processing during semantic context verification. *Hippocampus*, *15*(5), 597–609. https://doi.org/10.1002/hipo.20080

Holsinger, E. (2013). Representing idioms: Syntactic and contextual effects on idiom processing. *Language and Speech*, *56*(Pt 3), 373–394. https://doi.org/10.1177/0023830913484899

Holsinger, E., & Kaiser, E. (2013). Processing (non)compositional expressions: Mistakes and recovery. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(3), 866–878. https://doi.org/10.1037/a0030410

Horowitz, L. M., & Manelis, L. (1973). Recognition and cued recall of idioms and phrases. *Journal of Experimental Psychology: General*, *100*(2), 291–296. https://doi.org/10.1037/h0035468

Hubers, F., Cucchiarini, C., Strik, H., & Dijkstra, T. (2019). Normative data of Dutch idiomatic expressions: Subjective judgments you can bank on. *Frontiers in Psychology*, *10*, 1075. https://doi.org/10.3389/fpsyg.2019.01075

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23-32. https://doi.org/10.1016/j.cognition.2004.10.003

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Irujo, S. (1986). Don't put your leg in your mouth: Transfer in the acquisition of idioms in a second language. *TESOL Quarterly*, *20*(2), 287–304. https://doi.org/10.2307/3586545

Irujo, S. (1993). Steering Clear: Avoidance in the Production of Idioms. *IRAL*, *31*(3), 205–219. https://doi.org/10.1515/iral.1993.31.3.205

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 635–652. https://doi.org/10.1037/xlm0000315

Jegerski, J. (2014). Self-Paced Reading. In B. van Patten & J. Jegerski (Eds.), *Second language acquisition research series. Research methods in second language psycholinguistics* (pp. 20–49). New York: Routledge.

Johnson-Laird, P. N. (1993). Foreword. In C. Cacciari & P. Tabossi (Eds.), *Idioms: Processing, Structure, and Interpretation*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, *9*(11), 512–518. https://doi.org/10.1016/j.tics.2005.09.009

Kaan, E. (2014). Predictive sentence processing in L2 and L1. *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kana, R. K., Murdaugh, D. L., Wolfe, K. R., & Kumar, S. L. (2012). Brain responses mediating idiom comprehension: Gender and hemispheric differences. *Brain Research*, *1467*, 18–26. https://doi.org/10.1016/j.brainres.2012.05.031

Kasahara, K. (2011). The effect of known-and-unknown word combinations on intentional vocabulary learning. *System*, *39*(4), 491–499. https://doi.org/10.1016/j.system.2011.10.001

Kecskés, I. (2000). A cognitive-pragmatic approach to situation-bound utterances. *Journal of Pragmatics*, *32*(5), 605–625. https://doi.org/10.1016/S0378-2166(99)00063-6

Keßler, R., Weber, A., & Friedrich, C. (submitted). *Activation of literal word meanings in idioms: Evidence from Eye-tracking and ERP.*

References

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Kroll, J. F., van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism (Cambridge, England)*, *13*(3), 373–381. https://doi.org/10.1017/S136672891000009X

Kyriacou, M., Conklin, K., & Thompson, D. (2019). Passivizability of idioms: Has the wrong tree been barked up? *Language and Speech*, 23830919847691. https://doi.org/10.1177/0023830919847691

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Laufer, B. (2011). The Contribution of dictionary use to the production and retention of collocations in a second language. *Applied Linguistics*, *24*(1), 29–49. https://doi.org/10.1093/ijl/ecq039

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A Case for contrastive analysis and translation. *Applied Linguistics*, *29*(4), 694–716. https://doi.org/10.1093/applin/amn018

Lauro, L. J. R., Tettamanti, M., Cappa, S. F., & Papagno, C. (2008). Idiom comprehension: A prefrontal task? *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(1), 162–170. https://doi.org/10.1093/cercor/bhm042

Lee, E.-K., Hsin-Yi Lu, D., & Garnsey, S. M. (2013). L1 word order and sensitivity to verb bias in L2 processing. *Bilingualism: Language and Cognition*, *16*(4), 761–775. https://doi.org/10.1017/S1366728912000776

Leeuw, J. R. de (2015). Jspsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Levorato, M. C., & Cacciari, C. (1992). Children's comprehension and production of idioms: the role of context and familiarity. *Journal of Child Language*, *19*(2), 415–433. https://doi.org/10.1017/S0305000900011478

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, *36*(6), 1103–1121. https://doi.org/10.3758/MC.36.6.1103

Liontas, J. I. (2002). Context and idiom understanding in second languages. *EUROSLA Yearbook 2*, 155–185. https://doi.org/10.1075/eurosla.2.11lio

Liontas, J. I. (2015). Straight from the Horse's Mouth: Idiomaticity Revisited. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual figurative language processing* (pp. 301–340). Canbridge University Press.

Liu, D. (2008). *Idioms: Description, comprehension, acquisition, and pedagogy*. *ESL & applied linguistics professional series*. New York: Routledge.

Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2011). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes*, *27*(10), 1479–1488. https://doi.org/10.1080/01690965.2011.613209

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*(2), 157–201. https://doi.org/10.1080/01690969408402115

Mahler, N. A., & Chenery, H. J. (2019). A Developmental perspective on processing semantic context: Preliminary evidence from sentential auditory word repetition in school-aged children. *Journal of Psycholinguistic Research*, *48*(1), 81–105. https://doi.org/10.1007/s10936-018-9591-6

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake-but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. https://doi.org/10.1037/a0029284

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, *126*, 264–279. https://doi.org/10.1016/j.jecp.2014.05.004

Mariën, P., Ackermann, H., Adamaszek, M., Barwood, C. H. S., Beaton, A., Desmond, J., Ziegler, W. (2014). Consensus paper: Language and the cerebellum: An ongoing enigma. *Cerebellum*, *13*(3), 386–410. https://doi.org/10.1007/s12311-013-0540-5

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman, M. (2008). Hemispheric differences in processing the literal interpretation of idioms: Converging evidence from behavioral and fMRI studies. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *44*(7), 848–860. https://doi.org/10.1016/j.cortex.2007.04.004

Mashal, N., Vishne, T., & Laor, N. (2014). The role of the precuneus in metaphor comprehension: Evidence from an fMRI study in people with schizophrenia and healthy participants. *Frontiers in Human Neuroscience*, *8*, 818. https://doi.org/10.3389/fnhum.2014.00818

Matlock, T., & Heredia, R. R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. *Bilingual Sentence Processing*, *134*, 251–274. https://doi.org/10.1016/S0166-4115(02)80014-0

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

WebCelex [Computer software] (2001). Retrieved from http://celex.mpi.nl/

McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, *6*(3), 486–494. https://doi.org/10.3758/BF03210839

McGlone, M. S., Glucksberg, S., & Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Processes*, *17*(2), 167–190. https://doi.org/10.1080/01638539409544865

McQueen, J. M., & Huetting, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *Journal of the Acoustical Society of America*, *131*(1), 509–517. https://doi.org/10.1121/1.3664087

Milburn, E., & Warren, T. (2019). Idioms show effects of meaning relatedness and dominance similar to those seen for ambiguous words. *Psychonomic Bulletin & Review*, *26*(2), 591–598. https://doi.org/10.3758/s13423-019-01589-7

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. *Oxford studies in lexicography and lexicology*. Oxford: Oxford University Press.

Mueller, R. A. G., & Gibbs, R. W. Jr. (1987). Processing idioms with multiple meanings. *Journal of Psycholinguistic Research*, *16*(1), 63–81. https://doi.org/10.1007/BF01067751

The University of South Florida word association, rhyme, and word fragment norms [Computer software] (1998). Retrieved from http://www.usf.edu/FreeAssociation/

Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language*, *70*(3), 491–538. https://doi.org/10.1353/lan.1994.0007

OnExp [Computer software] (2011): Textstructures, University of Göttingen.

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, *17*(4), 465–477. https://doi.org/10.1016/S0022-5371(78)90283-9

Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, *121*(2), 221–236. https://doi.org/10.1016/j.lingua.2010.08.003

Papagno, C., & Romero-Lauro, L. (2010). The neural basis of idiom processing: Neuropsychological, neurophysical, and neuroimaging evidence. *Italian Journal of Linguistics*, *22*, 21–40.

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency. In J. Richards & R. Schmidt (Eds.), *Language and Communication* (pp. 191–225). London: Longman.

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, *15*(2), 202–206. https://doi.org/10.1016/j.conb.2005.03.007

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., Mehler, J. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain*, *121*(10), 1841–1852. https://doi.org/10.1093/brain/121.10.1841

Perea, M., & Carreiras, M. (2003). Sequential effects in the lexical decision task: The role of item frequency of the previous trial. *The Quarterly Journal of Experimental Psychology*, *56A*(3), 385–401. https://doi.org/10.1080/02724980244000387

Peters, E. (2012). Learning German Formulaic Sequences: The effect of two attention-drawing techniques. *Language Learning Journal*, *40*, 65–79. https://doi.org/10.1080/09571736.2012.658224

Peterson, R. R., Burgess, C., Dell, G. S., & Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1223–1237. https://doi.org/10.1037/0278-7393.27.5.1223

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pinker, S. (1991). Rules of language. *Science*, *253*(5019), 530–535. https://doi.org/10.1126/science.1857983

Pollio, H. R., Barlow, J., Fine, H., & Pollio, M. (1977). *Psychology and the poetics of growth: Figurative language in psychology, psychotherapy, and education*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Popiel, S. J., & McRae, K. (1988). The figurative and literal senses of idioms, or all idioms are not used equally. *Journal of Psycholinguistic Research*, *17*(6), 475–487. https://doi.org/10.1007/BF01067912

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15. https://doi.org/10.1016/S0001-6918(99)00050-5

Proverbio, A. M., Crotti, N., Zani, A., & Adorni, R. (2009). The role of left and right hemispheres in the comprehension of idiomatic language: An electrical neuroimaging study. *BMC Neuroscience*, *10*, 116. https://doi.org/10.1186/1471-2202-10-116

E-Prime (Version 2.0) [Computer software] (2013). Pittsburg, PA. Retrieved from http://www.pstnet.com

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience*, *11*(5), 351–360. https://doi.org/10.1038/nrn2811

Qualls, C. D., Treaster, B., Blood, G. W., & Hammer, C. S. (2003). Lexicalization of idioms in urban fifth graders: a reaction time study. *Journal of Communication Disorders*, *36*(4), 245–261. https://doi.org/10.1016/S0021-9924(03)00015-7

R: A language and environment for statistical computing (Version 3.4.3) [Computer software] (2013). Vienna, Austria: R Foundation for statistical computing. Retrieved from http://www.R-project.org/

Raine, A. (1991). The SPQ: A Scale for the assessment of Schizotypal Personality based on DSM-III-R criteria. *Schizophrenia Bulletin*, *17*(4), 555–564. https://doi.org/10.1093/schbul/17.4.555

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47*(2), 388–396. https://doi.org/10.1016/j.neuropsychologia.2008.09.017

Rapp, A. M. (2012). The brain behind nonliteral language: insights from brain imaging. In M. Faust (Ed.), *The Handbook of the Neuropsychology of Language* (pp. 406–424). Oxford, UK: Wiley-Blackwell.

Rapp, A. M. (2013). Where in the brain do metaphors become metaphoric? Research on the functional neuroanatomy of nonliteral language. In M. Połczyńska, Ł. P. Pakuła, & D. Jaworska (Eds.), *Young linguists' insights: Talking interdisciplinary approaches to the fore* (pp. 153–165). Poznań: Wydział Anglistyki Uniwersytetu im. Adama Mickiewicza.

Rapp, A. M. (2018). Comprehension of Metaphors and Idioms. In G. de Zubicaray & N. O. Schiller (Eds.), *The Oxford Handbook of Neurolinguistics* (pp. 709–735). New York NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190672027.013.28

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., & Kircher, T. T. J. (2004). Neural correlates of metaphor processing. *Brain Research. Cognitive Brain Research*, *20*(3), 395–402. https://doi.org/10.1016/j.cogbrainres.2004.03.017

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., & Kircher, T. T. J. (2007). Laterality in metaphor processing: Lack of evidence from functional magnetic resonance imaging for the right hemisphere theory. *Brain and Language*, *100*(2), 142–149. https://doi.org/10.1016/j.bandl.2006.04.004

Rapp, A. M., Mutschler, D. E., & Erb, M. (2012). Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. *NeuroImage*, *63*(1), 600–610. https://doi.org/10.1016/j.neuroimage.2012.06.022

Rapp, A. M., & Steinhäuser, A. E. (2013). Functional MRI of sentence-level language comprehension in schizophrenia: A coordinate-based analysis. *Schizophrenia Research*, *150*(1), 107–113. https://doi.org/10.1016/j.schres.2013.07.019

Rapp, D. N., & Mensink, M. C. (2011). Focusing effects from online and offline reading tasks. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 141–164). Charlotte, NC, US: IAP Information Age Publishing.

Reuterskiöld, C., & Van Lancker Sidtis, D. (2012). Retention of idioms following one-time exposure. *Child Language Teaching and Therapy*, *29*(2), 219–231. https://doi.org/10.1177/0265659012456859

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. https://doi.org/10.1162/jocn_a_00337

Rüschemeyer, S.-A. Zysset, S., & Friederici, A. D. (2006). Native and non-native reading of sentences: An fMRI experiment. *NeuroImage*, *31*(1), 354–365. https://doi.org/10.1016/j.neuroimage.2005.11.047

Rüschemeyer, S.-A., Fiebach, C. J., Kempe, V., & Friederici, A. D. (2005). Processing lexical semantic and syntactic information in first and second language: Fmri evidence from German and Russian. *Human Brain Mapping*, *25*(2), 266–286. https://doi.org/10.1002/hbm.20098

Saalmann, Y. B., & Kastner, S. (2011). Cognitive and perceptual functions of the visual thalamus. *Neuron*, *71*(2), 209–223. https://doi.org/10.1016/j.neuron.2011.06.027

Saslow, M. G. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*(8), 1024–1029. https://doi.org/10.1364/josa.57.001024

Schachter, D. L., & McGlynn, S. M. (1989). Implicit memory: Effects of elaboration depend on unitization. *The American Journal of Psychology*, *102*(2), 151–181. https://doi.org/10.2307/1422950

Schemann, H., & Knight, P. (2011). *Idiomatik Deutsch-Englisch: German-English dictionary of idioms* (2., durchges. Aufl.). *Englisch*. Hamburg: Buske.

Schweigert, W. A. (1986). The Comprehension of Familiar and Less Familiar Idioms. *Journal of Psycholinguistic Research*, *15*(1), 33–45. https://doi.org/10.2307/1422950

Schweigert, W. A., & Moates, D. R. (1988). Familiar Idiom Comprehension. *Journal of Psycholinguistic Research*, *17*(4), 281–296. https://doi.org/10.1007/BF01067198

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.

Sebastián-Galĺes, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception & Psychophysics*, *62*(4), 834–842. https://doi.org/10.3758/bf03206926

Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487–512. https://doi.org/10.1093/applin/amp058

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

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. https://doi.org/10.1075/ml.8.2.06siy

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*, *29*(2), 72–89. https://doi.org/10.1177/0267658310382068

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (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. https://doi.org/10.1037/a0022531

Siyanova-Chanturia, A., & Martinez, R. (2014). The Idiom Principle revisited. *Applied Linguistics*, amt054. https://doi.org/10.1093/applin/amt054

Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (Eds.) (2018). *Second language acquisition research series*. *Understanding formulaic language: A second language acquisition perspective* (1st). London: Routledge.

Sprenger, S., Levelt, W.J.M., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, *54*(2), 161–184. https://doi.org/10.1016/j.jml.2005.11.001

Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second language idiom learning in a paired associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition*, *29*(03), 199. https://doi.org/10.1017/S0272263107070271

Stringaris, A. K., Medford, N., Giora, R., Giampietro, V. C., Brammer, M. J., & David, A. S. (2006). How metaphors influence semantic relatedness judgments: The role of the right frontal cortex. *NeuroImage*, *33*(2), 784–793. https://doi.org/10.1016/j.neuroimage.2006.06.057

Sulpizio, S., Del Maschio, N., Fedeli, D., & Abutalebi, J. (2020). Bilingual language processing: A meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *108*, 834–853. https://doi.org/10.1016/j.neubiorev.2019.12.014

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition*, *28*(03), 387–422. https://doi.org/10.1017/S0272263106060177

SurveyGizmo [Computer software]. Boulder, Co., USA. Retrieved from www.surveygizmo.com

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645–659. https://doi.org/10.1016/S0022-5371(79)90355-4

Swinney, D. A., & Cutler, A. (1979). The Access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, *18*, 523–534. https://doi.org/10.1016/S0022-5371(79)90284-6

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language*, *27*(3), 324–340. https://doi.org/10.1016/0749-596X(88)90058-7

Tabossi, P., Arduino, L., & Fanari, R. (2011). Descriptive norms for 245 Italian idiomatic expressions. *Behavior Research Methods*, *43*(1), 110–123. https://doi.org/10.3758/s13428-010-0018-z

Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, *49*, 161–167. https://doi.org/10.1007/BF00308682

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. https://doi.org/10.1037/0278-7393.34.2.313

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, *37*(4), 529–540. https://doi.org/10.3758/MC.37.4.529

Tabossi, P., & Zardon, F. (1993). The Activation of Idiomatic Meaning in Spoken Language Comprehension. In C. Cacciari & P. Tabossi (Eds.), *Idioms: Processing, Structure, and Interpretation* (pp. 145–162). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Takashima, A., & Bakker, I. (2017). Memory consolidation. In H.-J. Schmid (Ed.), *Language and the Human Life Span (LHLS). Entrenchment, Memory and Automaticity: The Psychology of Linguistic Knowledge and Language Learning* (pp. 177–200). Berlin: de Gruyter Mouton. https://doi.org/10.1037/15969-009

Taylor, W. L. (1953). "Cloze Procedure": A New tool for measuring readability. *Journalism Quarterly*, *30*(4), 415–433. https://doi.org/10.1177/107769905303000401

Titone, D. A., Columbus, G., Whitford, V., Mercier, J., & Libben, M. R. (2015). Contrasting Bilingual and Monolingual Idiom Processing. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual figurative language processing* (pp. 171–207). Canbridge University Press.

Titone, D. A., & Connine, C. M. (1994a). Descriptive norms for 171 idiomatic expressions: Familiarity, composability, predictability, and literality. *Metaphor and Symbolic Activity*, *9*(4), 247–270. https://doi.org/10.1207/s15327868ms0904_1

Titone, D. A., & Connine, C. M. (1994b). Comprehension of idiomatic expressions: Effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1126–1138. https://doi.org/10.1037//0278-7393.20.5.1126

Titone, D. A., & Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics*, *31*, 1655–1674. https://doi.org/10.3758/MC.36.6.1103

Titone, D. A., & Libben, M. R. (2014). Time-dependent effects on decomposability, familiarity and literal plausibility on idiom meaning activation: A cross-modal priming investigation. *The Mental Lexicon*, *9*(3), 473–496. https://doi.org/10.1075/ml.9.3.05tit

Tiv, M. (2016). *Investigating the effects of transparency and ambiguity on idiom learning* (Undergratuate Thesis). University of Pittsburgh, Pittsburgh, PA. Retrieved from http://d-scholarship.pitt.edu/27803/

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

Trueswell, J. C., & Kim, A. E. (1998). How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. *Journal of Memory and Language*, *39*(1), 102–123. https://doi.org/10.1006/jmla.1998.2565

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. 155–172). Amsterdam: John Benjamins.

Van Ginkel, W., & Dijkstra, T. (2019). The tug of war between an idiom's figurative and literal meanings: Evidence from native and bilingual speakers. *Bilingualism: Language and Cognition*, *3*, 1–17. https://doi.org/10.1017/S1366728918001219

Van Lancker, D., Canter, G. J., & Terbeek, D. (1981). Disambiguation of ditropic sentences: Acoustic and phonetic cues. *Journal of Speech and Hearing Research*, *24*(3), 330–335. https://doi.org/10.1044/jshr.2403.330

Van Lancker, D., & Kempler, D. (1987). Comprehension of familiar phrases by left- but not right-hemisphere damaged patients. *Brain and Language*, *32*, 265–277. https://doi.org/10.1016/0093-934X(87)90128-3

Van Lancker Sidtis, D. (2003). Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. *Applied Psycholinguistics*, *24*(01). https://doi.org/10.1017/S0142716403000031

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). *Statistics and computing*. New York: Springer.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2009). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, *22*(8), 1682–1700. https://doi.org/10.1162/jocn.2009.21293

Wang, Y., Lin, L., Kuhl, P., & Hirsch, J. (2007). Mathematical and linguistic processing differs between native and second languages: An fMRI study. *Brain Imaging and Behavior*, *1*(3-4), 68–82. https://doi.org/10.1007/s11682-007-9007-y

Weber, A., Di Betta, A. M., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of Phonetics*, *46*, 34–51. https://doi.org/10.1016/j.wocn.2014.05.002

SPM12 [Computer software]. London, UK: UCL Queen Square Institute of Neurology.

Winner, E., & Gardner, H. (1977). The Comprehension of metaphor in brain-damaged patients. *Brain*, *100*, 717–729. https://doi.org/10.1093/brain/100.4.717

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, *32*(4), 430–449. https://doi.org/10.1093/applin/amr011

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*(4), 463–489. https://doi.org/10.1093/applin/21.4.463

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

Yang, J., Li, P., Fang, X., Shu, H., Liu, Y., & Chen, L. (2016). Hemispheric involvement in the processing of Chinese idioms: An fMRI study. *Neuropsychologia*, *87*, 12–24. https://doi.org/10.1016/j.neuropsychologia.2016.04.029

Zempleni, M.-Z., Haverkort, M., Renken, R., & A Stowe, L. (2007). Evidence for bilateral involvement in idiom comprehension: An fMRI study. *NeuroImage*, *34*(3), 1280–1291. https://doi.org/10.1016/j.neuroimage.2006.09.049

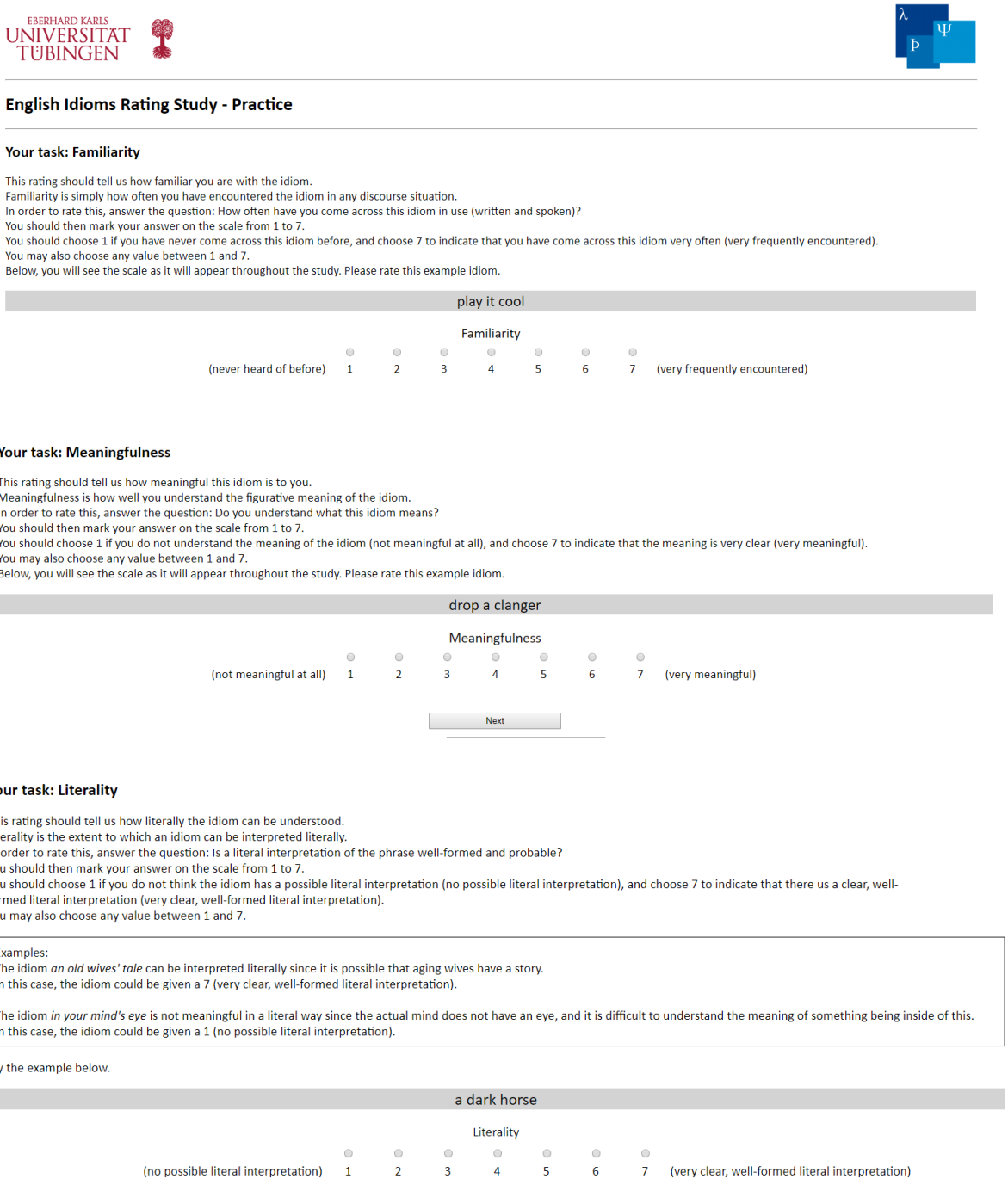
Zubicaray, G. de, & Schiller, N. O. (Eds.) (2018). *The Oxford Handbook of Neurolinguistics*. New York NY: Oxford University Press.

10 APPENDICES

10.1 APPENDIX A

Appendix A displays screen shots from the experiments in Chapter 2.

10.1.1 Experiment 1: Familiarity, meaningfulness, and literality



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

English Idioms Rating Study - Practice

Your task: Familiarity

This rating should tell us how familiar you are with the idiom.
Familiarity is simply how often you have encountered the idiom in any discourse situation.
In order to rate this, answer the question: How often have you come across this idiom in use (written and spoken)?
You should then mark your answer on the scale from 1 to 7.
You should choose 1 if you have never come across this idiom before, and choose 7 to indicate that you have come across this idiom very often (very frequently encountered).
You may also choose any value between 1 and 7.
Below, you will see the scale as it will appear throughout the study. Please rate this example idiom.

play it cool

Familiarity

(never heard of before) 1 2 3 4 5 6 7 (very frequently encountered)

Your task: Meaningfulness

This rating should tell us how meaningful this idiom is to you.
Meaningfulness is how well you understand the figurative meaning of the idiom.
In order to rate this, answer the question: Do you understand what this idiom means?
You should then mark your answer on the scale from 1 to 7.
You should choose 1 if you do not understand the meaning of the idiom (not meaningful at all), and choose 7 to indicate that the meaning is very clear (very meaningful).
You may also choose any value between 1 and 7.
Below, you will see the scale as it will appear throughout the study. Please rate this example idiom.

drop a clanger

Meaningfulness

(not meaningful at all) 1 2 3 4 5 6 7 (very meaningful)

Next

Your task: Literality

This rating should tell us how literally the idiom can be understood.
Literality is the extent to which an idiom can be interpreted literally.
In order to rate this, answer the question: Is a literal interpretation of the phrase well-formed and probable?
You should then mark your answer on the scale from 1 to 7.
You should choose 1 if you do not think the idiom has a possible literal interpretation (no possible literal interpretation), and choose 7 to indicate that there us a clear, well-formed literal interpretation (very clear, well-formed literal interpretation).
You may also choose any value between 1 and 7.

Examples:
The idiom *an old wives' tale* can be interpreted literally since it is possible that aging wives have a story.
In this case, the idiom could be given a 7 (very clear, well-formed literal interpretation).

The idiom *in your mind's eye* is not meaningful in a literal way since the actual mind does not have an eye, and it is difficult to understand the meaning of something being inside of this.
In this case, the idiom could be given a 1 (no possible literal interpretation).

Try the example below.

a dark horse

Literality

(no possible literal interpretation) 1 2 3 4 5 6 7 (very clear, well-formed literal interpretation)

Your task:

Read the idiom in the grey box.
Please rate your **familiarity** with this idiom (how often you have come across this item),
the **meaningfulness** of this idiom (how well you understand the meaning of the idiom),
and the **literality** of the idiom (to what extent a literal interpretation of the idiom is probable and well-formed).

twist someone's arm

Familiarity

(never heard of before) 1 2 3 4 5 6 7 (very frequently encountered)

Meaningfulness

(not meaningful at all) 1 2 3 4 5 6 7 (very meaningful)

Literality

(no possible literal interpretation) 1 2 3 4 5 6 7 (very clear, well-formed literal interpretation)

Next

Progress:

10.1.2 Experiment 2: Decomposability

Your task: Decomposability Judgement

Instructions:
For this task, you will be asked to make a judgement and rate the difficulty of your decision.
You will see the idiom and its paraphrase. The idiom will be on top in **bolded text** and its paraphrase will be visible below. First, you must decide whether or not an idiom is decomposable. If an idiom is decomposable, that means that the individual words contribute to the overall figurative meaning of the idiom.
If this is the case, you will answer "Yes."
If the individual parts do not contribute to the figurative meaning, you will answer "No."
Look at the following examples:

Examples:
The idiom **give someone a piece of one's mind** means *to tell someone what one thinks, especially in anger.*
This idiom is decomposable because *give someone* relates to the meaning *tell someone,* and the phrase *a piece of one's mind* figuratively relates to the idea *what one thinks.*
Therefore, you should respond with "Yes".

The idiom **go cold turkey** means *to suddenly or completely stop doing something, especially a bad habit.*
This idiom is NOT decomposable because the meaning of the individual words *go* and *cold turkey* do not directly relate to the overall meaning of the idiom.
Therefore you should respond with "No".

Once you decide if the idiom is decomposable, the next part of your task should be to rate the difficulty you had in deciding whether to choose "Yes" or "No".
Your answers can range from 1 (very easy) to 7 (very difficult). If you are able to decide quickly and easily, you should choose low values such as 1.
If you spend a lot of time deciding, and the decision is difficult, choose high values such as 7.
Remember that this is your own judgment, and there are not "right" and "wrong" answers.
Try the example below.

not move a muscle
to stay very still, without moving

Is the idiom **decomposable**?

Yes No

How **difficult** was this decision?

(Very easy) 1 2 3 4 5 6 7 (Very difficult)

Your task:

Is the following idiom **decomposable** (do the individual parts contribute to the overall figurative meaning)?
How **difficult** was this decision?

raise an eyebrow :
show surprise, e.g. by moving one's eyebrows

Is the idiom **decomposable**?

Yes No

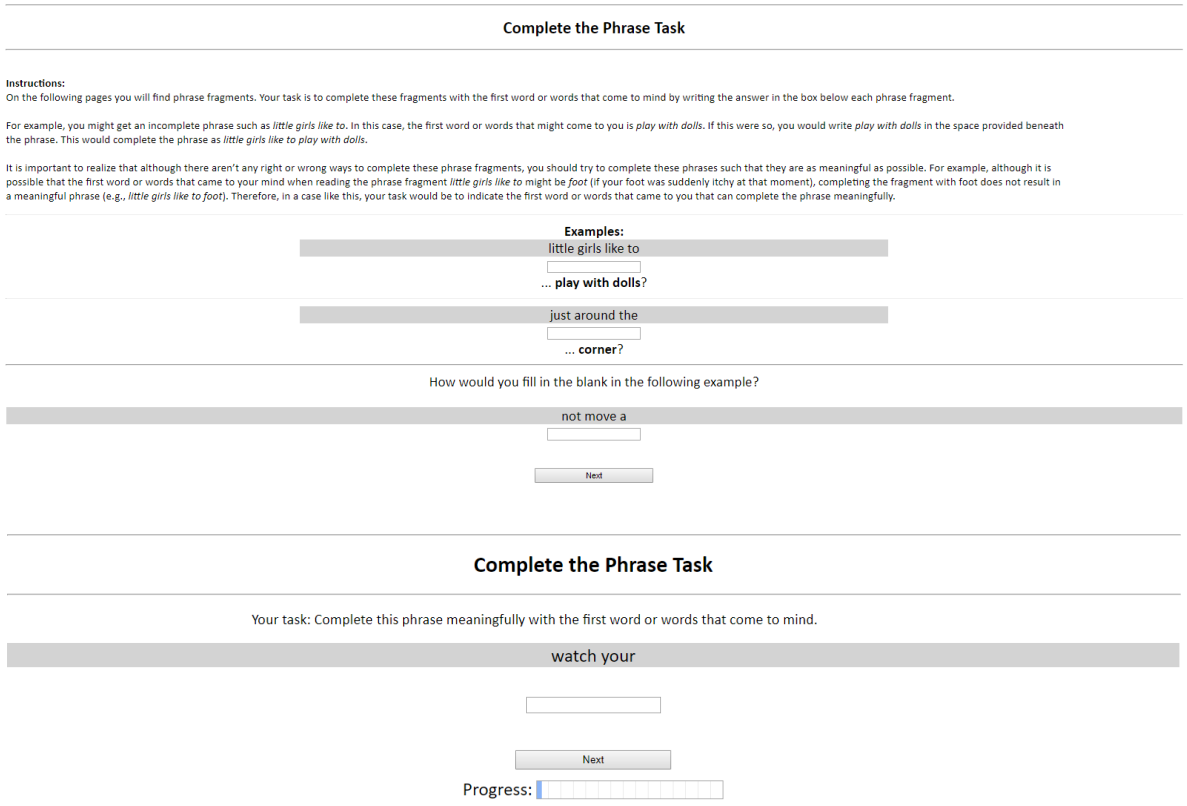
How **difficult** was this decision?

(Very easy) 1 2 3 4 5 6 7 (Very difficult)

Next

Progress:

10.1.3 Experiment 3: Predictability



10.2 APPENDIX B

Appendix B refers to the experimental items in Chapter 3.

			Figurative		Literal	
DIN Number	**Translatability**	**Sentence Context with *Idiom***	**Related**	**Control**	**Related**	**Control**
5	lexical	They are always *armed to the teeth*.	weapon	salad	dentist	sentence
10	lexical	For years she's been *beating the drum*.	support	laughter	band	song
18	lexical	I think he was *building castles in the air*.	daydream	coffee	cloud	frog
19	lexical	The teacher *bit that boy's head off*.	scold	trip	body	daisy
27	lexical	John was *born with a silver spoon*.	rich	thick	fork	bark
29	lexical	Let's *break the ice*.	cold	social	towel	slot
38	lexical	The boy *couldn't believe his ears*.	doubt	sheep	hear	mail
52	post-lexical	It's going to *cost an arm and a leg*.	expensive	attention	foot	soup
54	lexical	We need to *cover the territory*.	scope	white	area	radio
62	post-lexical	John will *deliver the goods*.	result	cement	product	pretend
63	lexical	Today I will *do the honors*.	announce	account	award	water
65	post-lexical	It didn't *do the trick*.	satisfy	gravity	magic	feature
70	lexical	Last night he *dropped a bombshell*.	surprise	lemon	explode	produce
74	post-lexical	I heard about her *fall from grace*.	disfavor	develop	prayer	cracker
75	lexical	It didn't *fall on deaf ears*.	ignore	hanger	nose	bar
89	post-lexical	I told him to *get a grip*.	control	shelter	hold	shirt
91	post-lexical	John *has the blues*.	sad	seal	music	nation
105	lexical	He *gave me the cold shoulder*.	ignore	autumn	arm	beet
106	lexical	I think he will *go the limit*.	endure	elbow	boundary	celery
108	post-lexical	The thing just *went to pieces*.	upset	album	parts	rent
113	post-lexical	He wanted to *have a go*.	try	spa	move	cheek
118	lexical	He might have *had cold feet*.	nervous	timber	toes	tax
120	lexical	He doesn't *have the heart*.	brave	trim	organ	agent
127	post-lexical	John *hit the road*.	leave	rate	highway	shadow
132	post-lexical	He tried to *hitch a ride*.	lift	raft	horse	yarn
145	post-lexical	His uncle *kicked the bucket*.	die	zoo	pail	boat
147	post-lexical	It's clear that he *knew the score*.	aware	apart	points	belts
153	lexical	I asked her to *lend me an ear*.	listen	logic	head	game
155	lexical	John *let the cat out of the bag*.	reveal	season	box	pin
162	lexical	He did it without *losing face*.	disgrace	purpose	eyes	eel
163	post-lexical	John worried he might *lose his touch*.	unable	animal	finger	meadow
164	lexical	At some point the man *lost his thread*.	stray	tree	needle	label
170	lexical	He does that to *make a living*.	earn	arch	dead	moan
172	post-lexical	John tried to *make a pass*.	flirt	nurse	throw	shrub
180	lexical	The man *missed the mark* .	fail	keep	spot	stack
200	lexical	I heard John *played a part*.	influence	objective	piece	gaze
201	post-lexical	They decided to *play it by ear*.	improvise	anything	wax	tab
204	post-lexical	The woman just tried to *play the game*.	behave	message	football	novel
207	post-lexical	He is going to *pop the question*.	propose	swallow	answer	orange

			Figurative		Literal	
209	lexical	Last night she *poured her heart out*.	confide	mandate	lung	bird
211	post-lexical	John likes to *pull my leg*.	joke	ship	walk	milk
212	lexical	He might have to *pull the plug*.	stop	clip	electricity	creativity
220	lexical	He asked us to *read between the lines*.	infer	envy	text	silk
223	post-lexical	He asked if it *rings a bell*.	remind	suspend	chime	thief
229	lexical	He *rules with an iron fist*.	reign	sink	knuckle	letter
234	lexical	He told me to *save my skin*.	rescue	moment	soft	mask
241	post-lexical	She finally *saw the light*.	accept	affair	dark	short
248	lexical	That's his way of *showing his flag*.	promote	circle	wave	hip
250	post-lexical	She spent a while *sitting on the fence*.	choice	shave	yard	wind
252	post-lexical	He accidentally *let it slip his mind*.	forget	shadow	brain	truck
255	post-lexical	I think John *spilled the beans*.	confess	level	rice	fox
257	lexical	Let's not *split hairs*.	nitpick	purchase	brush	speak
258	post-lexical	Let's *split the difference*.	compromise	talkative	similar	fabulous
263	post-lexical	It's supposed to *stir the blood*.	provoke	tribal	red	lock
275	lexical	He decided to *take the bull by the horns*.	confront	background	cow	mop
276	post-lexical	They sure *took the cake*.	rule	seam	birthday	journey
278	post-lexical	She will have to *take the stand*.	witness	curtain	sit	fan
281	lexical	She was just *throwing money out the window*.	waste	bench	glass	flame
283	post-lexical	She *tied the knot*.	marry	pillow	rope	mouse
288	post-lexical	Last night John *turned a corner*.	improve	attract	street	own
289	post-lexical	John is responsible for *turning the tables*.	advance	rhythm	chair	shot
296	post-lexical	Don't *waste your breath*.	ineffective	unattended	oxygen	element
297	lexical	I don't think he *wears the pants*.	command	debate	leg	moon
300	post-lexical	I think he would *give the world*.	crave	chat	universe	library

10.3 APPENDIX C

Appendix B refers to the experimental items in Chapter 4B. These items are in German as they were presented in the experiment.

Item	Sentence Stimuli	Correct	Related	Unrelated (1)	Unrelated (2)
1	Julia rutschte das Herz in die	Hose	Jacke	Welt	Venus
2	Annika war Balsam für die	Seele	Gefühle	Hose	Jacke
3	Marie stand Gabriel Rede und	Antwort	Frage	Seele	Gefühle
4	Jannis fiel die Decke auf den	Kopf	Bart	Fleck	Platz
5	Timo brachte den Stein ins	Rollen	Kugeln	Feuer	Holz
6	Sofia brachte die Aufgaben unter Dach und	Fach	Schrank	Ohren	Augen
7	Hannah schlug sich die Zeit um die	Ohren	Augen	Bewegung	Sprünge
8	Nora hatte ihr Herz am rechten	Fleck	Platz	Sack	Korb
9	Melina ließ die Kirche im	Dorf	Feld	Hals	Rücken
10	Isabell hatte Schmetterlinge im	Bauch	Arm	Wasser	Regen
11	Jasmin lebte wie die Made im	Speck	Käse	Kopf	Bart
12	Nico erblickte das Licht der	Welt	Venus	Antwort	Frage
13	Hannes ließ die Katze aus dem	Sack	Korb	Bauch	Arm
14	Linus verlor den Boden unter den	Füßen	Händen	Rollen	Kugeln
15	Lena setzte alle Hebel in	Bewegung	Sprünge	Löwen	Hasen
16	Julian hielt den Kopf über	Wasser	Regen	Füßen	Händen
17	Moritz begab sich in die Höhle des	Löwen	Hasen	Herzen	Magen
18	Helena fiel ein Stein vom	Herzen	Magen	Dorf	Feld
19	Amelie hatte einen Frosch im	Hals	Rücken	Speck	Käse
20	Robin legte für Natalie die Hand ins	Feuer	Holz	Fach	Schrank

10.4 APPENDIX D

Appendix D refers to the experimental items in Chapters 5B and 5C. Items 1-11 are high-literality idioms, and items 12-22 are low-literality idioms. For each example, a) has a figuratively biasing context and a figurative resolution, b) has a figuratively biasing context and a literal resolution, c) has a literally biasing context and a literal resolution, and d) has a literally biasing context and a figurative resolution.

	Subject	Context 1	Context 2	Pre-Idiom	Idiom	Resolution	Resolution+1	Wrap-up
1a	The new schoolboy,	who didn't know	anyone in his class,	just wanted to	break the ice	with his peers	as soon as possible	Monday morning.
1b	The new schoolboy,	who didn't know	anyone in his class,	just wanted to	break the ice	on the lake	as soon as possible	Monday morning.
1c	The cold Eskimo,	who was eager	to catch some fish,	just wanted to	break the ice	on the lake	as soon as possible	Monday morning.
1d	The cold Eskimo,	who was eager	to catch some fish,	just wanted to	break the ice	with his peers	as soon as possible	Monday morning.
2a	The fearless climber,	who was on a climb	alone in the mountains,	was ready to	play with fire	with any risk	if necessary	later on.
2b	The fearless climber,	who was on a climb	alone in the mountains,	was ready to	play with fire	from the grill	if necessary	later on.
2c	The young camper,	who was already bored	without any of his friends,	was ready to	play with fire	from the grill	if necessary	later on.
2d	The young camper,	who was already bored	without any of his friends,	was ready to	play with fire	with any risk	if necessary	later on.
3a	The trained therapist,	who just started	with a new patient,	really hoped to	scratch the surface	of the problem	without any delay	at all.
3b	The trained therapist,	who just started	with a new patient,	really hoped to	scratch the surface	off the ticket	without any delay	at all.
3c	The gambling addict,	who just wanted	to win the grand prize,	really hoped to	scratch the surface	off the ticket	without any delay	at all.
3d	The gambling addict,	who just wanted	to win the grand prize,	really hoped to	scratch the surface	of the problem	without any delay	at all.
4a	The new model,	who tried her best	to wear the latest fashion designs,	had chosen to	follow the crowd	with this trend	at her own risk	that day.
4b	The new model,	who tried her best	to wear the latest fashion designs,	had chosen to	follow the crowd	through the city	at her own risk	that day.
4c	The Italian tourist,	who was easily lost	when traveling to new places,	had chosen to	follow the crowd	through the city	at her own risk	that day.
4d	The Italian tourist,	who was easily lost	when traveling to new places,	had chosen to	follow the crowd	with this trend	at her own risk	that day.
5a	A tired housekeeper,	who just wanted	to go home and rest,	wasn't willing to	lift a finger	by offering help	even when asked	to do so.
5b	A tired housekeeper,	who just wanted	to go home and rest,	wasn't willing to	lift a finger	off the keys	even when asked	to do so.
5c	The office secretary,	who was always typing	during important meetings,	wasn't willing to	lift a finger	off the keys	even when asked	to do so.
5d	The office secretary,	who was always typing	during important meetings,	wasn't willing to	lift a finger	by offering help	even when asked	to do so.
6a	A bride-to-be,	who spent the night	thinking about her future,	hoped not to	have cold feet	about the wedding	the next morning	before breakfast.
6b	A bride-to-be,	who spent the night	thinking about her future,	hoped not to	have cold feet	from the snow	the next morning	before breakfast.
6c	The young hiker,	who spent the night	without fire in the woods,	hoped not to	have cold feet	from the snow	the next morning	before breakfast.
6d	The young hiker,	who spent the night	without fire in the woods,	hoped not to	have cold feet	about the wedding	the next morning	before breakfast.
7a	The popular teenager,	who liked to gossip	with his group of friends,	had managed to	spill the beans	about the surprise	even faster	than expected.

	Subject	Context 1	Context 2	Pre-Idiom	Idiom	Resolution	Resolution+1	Wrap-up
7b	The popular teenager,	who liked to gossip	with his group of friends,	had managed to	spill the beans	on the stove	even faster	than expected.
7c	The assistant chef,	who didn't seem	to do anything well,	had managed to	spill the beans	on the stove	even faster	than expected.
7d	The assistant chef,	who didn't seem	to do anything well,	had managed to	spill the beans	about the surprise	even faster	than expected.
8a	The working mother,	who always dreamed	of a long-lasting career,	was planning to	wear the pants	in the family	to the surprise	of many.
8b	The working mother,	who always dreamed	of a long-lasting career,	was planning to	wear the pants	with the pink dots	to the surprise	of many.
8c	The aspiring dancer,	who was worried	about her audition outfit,	was planning to	wear the pants	with the pink dots	to the surprise	of many.
8d	The aspiring dancer,	who was worried	about her audition outfit,	was planning to	wear the pants	in the family	to the surprise	of many.
9a	The financial manager,	who had to work	with a small budget,	would attempt to	pull the plug	on the project	after the discussion	that afternoon.
9b	The financial manager,	who had to work	with a small budget,	would attempt to	pull the plug	from the drain	after the discussion	that afternoon.
9c	The local plumber,	who was ready	to get right to work,	would attempt to	pull the plug	from the drain	after the discussion	that afternoon.
9d	The local plumber,	who was ready	to get right to work,	would attempt to	pull the plug	on the project	after the discussion	that afternoon.
10a	The gay couple,	who had been in love	for many years,	was able to	tie the knot	in a church	after much time	and effort.
10b	The gay couple,	who had been in love	for many years,	was able to	tie the knot	in the yarn	after much time	and effort.
10c	The expert knitter,	who was concentrated	on her newest scarf,	was able to	tie the knot	in the yarn	after much time	and effort.
10d	The expert knitter,	who was concentrated	on her newest scarf,	was able to	tie the knot	in a church	after much time	and effort.
11a	The substitute teacher,	who was in great need	of a break from the class,	was ready to	draw the line	at one interruption	right away	the next morning.
11b	The substitute teacher,	who was in great need	of a break from the class,	was ready to	draw the line	on the page	right away	the next morning.
11c	The new architect,	who had been sketching	the same picture for days,	was ready to	draw the line	on the page	right away	the next morning.
11d	The new architect,	who had been sketching	the same picture for days,	was ready to	draw the line	at one interruption	right away	the next morning.
12a	The teenage girl,	who was always speaking	without thinking,	just wanted to	eat her words	from the fight	late last night	in her room.
12b	The teenage girl,	who was always speaking	without thinking,	just wanted to	eat her words	from her spoon	late last night	in her room.
12c	The hungry student,	who was distractedly playing	with her alphabet soup,	just wanted to	eat her words	from her spoon	late last night	in her room.
12d	The hungry student,	who was distractedly playing	with her alphabet soup,	just wanted to	eat her words	from the fight	late last night	in her room.
13a	The freelance writer,	who often started	political debates,	didn't want to	lose his cool	out of anger	so quickly	that morning.
13b	The freelance writer,	who often started	political debates,	didn't want to	lose his cool	from the shade	so quickly	that morning.
13c	The sweaty runner,	who was recovering	under a tree,	didn't want to	lose his cool	from the shade	so quickly	that morning.
13d	The sweaty runner,	who was recovering	under a tree,	didn't want to	lose his cool	out of anger	so quickly	that morning.
14a	The young girl,	whose close friend	was openly upset,	would offer to	lend an ear	with some advice	as a gesture	of kindness.
14b	The young girl,	whose close friend	was openly upset,	would offer to	lend an ear	from her doll	as a gesture	of kindness.

	Subject	Context 1	Context 2	Pre-Idiom	Idiom	Resolution	Resolution+1	Wrap-up
14c	A little girl,	whose brother rebuilt	figurines from broken ones,	would offer to	lend an ear	from her doll	as a gesture	of kindness.
14d	A little girl,	whose brother rebuilt	figurines from broken ones,	would offer to	lend an ear	with some advice	as a gesture	of kindness.
15a	The jazz musician,	who had been trying	to get a record deal,	was willing to	give the world	for some fame	if it would help	his cause.
15b	The jazz musician,	who had been trying	to get a record deal,	was willing to	give the world	a new name	if it would help	his cause.
15c	The retired astronaut,	who had been tracking	a new planet in space,	was willing to	give the world	a new name	if it would help	his cause.
15d	The retired astronaut,	who had been tracking	a new planet in space,	was willing to	give the world	for some fame	if it would help	his cause.
16a	A car dealer,	who was able to recognize	customers' tastes in vehicles,	would try to	stretch a point	in any sales pitch	if necessary	for success.
16b	A car dealer,	who was able to recognize	customers' tastes in vehicles,	would try to	stetch a point	into a line	if necessary	for success.
16c	A sketch artist,	who was up late designing	her newest drawing,	would try to	stretch a point	into a line	if necessary	for success.
16d	A sketch artist,	who was up late designing	her newest drawing,	would try to	stretch a point	in any sales pitch	if necessary	for success.
17a	A young woman,	who had been fighting	with her boyfriend,	had decided to	face the music	with an apology	late last night	at the event.
17b	A young woman,	who had been fighting	with her boyfriend,	had decided to	face the music	from the speakers	late last night	at the event.
17c	The lead singer,	who had been struggling	with hearing problems,	had decided to	face the music	from the speakers	late last night	at the event.
17d	The lead singer,	who had been struggling	with hearing problems,	had decided to	face the music	with an apology	late last night	at the event.
18a	The cool kid,	who could pick	any date to the dance,	had decided to	play the field	instead of choosing	this time around	as expected.
18b	The cool kid,	who could pick	any date to the dance,	had decided to	play the field	behind the school	this time around	as expected.
18c	The soccer player,	whose team always	needed new challenges,	had decided to	play the field	behind the school	this time around	as expected.
18d	The soccer player,	whose team always	needed new challenges,	had decided to	play the field	instead of choosing	this time around	as expected.
19a	The crop farmer,	who was struggling	with a long drought,	wasn't able to	turn the tide	of his luck	as he hoped	he could.
19b	The crop farmer,	who was struggling	with a long drought,	wasn't able to	turn the tide	of the water	as he hoped	he could.
19c	The deep-sea diver,	who was swimming	much later than expected,	wasn't able to	turn the tide	of the water	as he hoped	he could.
19d	The deep-sea diver,	who was swimming	much later than expected,	wasn't able to	turn the tide	of his luck	as he hoped	he could.
20a	The famous songwriter,	who always used	his emotions in his songs,	wasn't hoping to	have the blues	from a heartbreak	when getting started	on the project.
20b	The famous songwriter,	who always used	his emotions in his songs,	wasn't hoping to	have the blues	on his paintbrush	when getting started	on the project.
20c	The art student,	who was often lazy	about cleaning his art supplies,	wasn't hoping to	have the blues	on his paintbrush	when getting started	on the project.
20d	The art student,	who was often lazy	about cleaning his art supplies,	wasn't hoping to	have the blues	from a heartbreak	when getting started	on the project.
21a	The powerful politician,	who was always	covering up a new scandal,	was going to	hit the headlines	of every newspaper	before going to work	Thursday morning.
21b	The powerful politician,	who was always	covering up a new scandal,	was going to	hit the headlines	with his fist	before going to work	Thursday morning.
21c	The old man,	who was always	upset while reading politics,	was going to	hit the headlines	with his fist	before going to work	Thursday morning.
21d	The old man,	who was always	upset while reading politics,	was going to	hit the headlines	of every newspaper	before going to work	Thursday morning.
22a	The new manager,	who didn't have	much working experience,	really needed to	fit the bill	for the job	without any delay	at work.
22b	The new manager,	who didn't have	much working experience,	really needed to	fit the bill	in the folder	without any delay	at work.
22c	The nervous waiter,	who was worried	about his observant boss,	really needed to	fit the bill	in the folder	without any delay	at work.
22d	The nervous waiter,	who was worried	about his observant boss,	really needed to	fit the bill	for the job	without any delay	at work.

10.5 APPENDIX E

Appendix E refers to the experimental items in Chapter 6. The underlined word in the target phrase is the word that varied between word-type (familiar and unfamiliar). The columns image. And transp. Stand for imageability and transparency from the ratings collected prior to the experiment.

Item	Phrase-Type	Word-Type	Target Phrase	Paraphrase	Image.	Transp.
1a	figurative	unfamiliar	hide in the back of the caboose	to be nervous of taking action	5.25	4.25
1b	literal	unfamiliar	hide in the back of the caboose	to take cover in the back of the final train car	5.25	4.25
1c	figurative	familiar	hide in the back of the bedroom	to be nervous of taking action	5.25	4.25
1d	literal	familiar	hide in the back of the bedroom	to take cover in the back of the bedroom	5.25	4.25
2a	figurative	unfamiliar	set up a new easel	to make a big change in one's life	5.4	4.75
2b	literal	unfamiliar	set up a new easel	to put together a new stand for painting	5.4	4.75
2c	figurative	familiar	set up a new desk	to make a big change in one's life	5.4	4.75
2d	literal	familiar	set up a new desk	to put together a desk for working	5.4	4.75
3a	figurative	unfamiliar	the morning bugle sounded	a return from the distraction of daydreaming	4.8	3.85
3b	literal	unfamiliar	the morning bugle sounded	the musical horn played early in the day	4.8	3.85
3c	figurative	familiar	the morning bell sounded	a return from the distraction of daydreaming	4.8	3.85
3d	literal	familiar	the morning bell sounded	the church bell rang early in the day	4.8	3.85
4a	figurative	unfamiliar	climb the tall blue trellis	to attempt an impossible task for the last time	5	2.15
4b	literal	unfamiliar	climb the tall blue trellis	to go up the blue frame used for a climbing plant	5	2.15
4c	figurative	familiar	climb the tall blue ladder	to attempt an impossible task for the last time	5	2.15
4d	literal	familiar	climb the tall blue ladder	to go up the blue equipment used for reaching high places	5	2.15
5a	figurative	unfamiliar	the little black coot near the road	a very clever person that surprises people	5.35	1.8
5b	literal	unfamiliar	the little black coot near the road	the small black aquatic bird close to the road	5.35	1.8
5c	figurative	familiar	the little black duck near the road	a very clever person that surprises people	5.35	1.8
5d	literal	familiar	the little black duck near the road	the small black aquatic bird close to the road	5.35	1.8
6a	figurative	unfamiliar	drive the long scow	to get in the way of something or someone	4.6	3.6
6b	literal	unfamiliar	drive the long scow	to drive the long vehicle	4.6	3.6
6c	figurative	familiar	drive the long truck	to get in the way of something or someone	4.6	3.6
6d	literal	familiar	drive the long truck	to drive the long vehicle	4.6	3.6
7a	figurative	unfamiliar	take it away from the shoat	to react harshly to something negative someone has done	5.4	2.6
7b	literal	unfamiliar	take it away from the shoat	to remove something from the young pig's grasp	5.4	2.6
7c	figurative	familiar	take it away from the puppy	to react harshly to something negative someone has done	5.4	2.6
7d	literal	familiar	take it away from the puppy	to remove something from the young dog's grasp	5.4	2.6
8a	figurative	unfamiliar	extra kibble in the kitchen	the best is yet to come	4.8	1.6
8b	literal	unfamiliar	extra kibble in the kitchen	the leftover dog food in the kitchen	4.8	1.6
8c	figurative	familiar	extra lentils in the kitchen	the best is yet to come	4.8	1.6
8d	literal	familiar	extra lentils in the kitchen	the leftover lentils in the kitchen	4.8	1.6
9a	figurative	unfamiliar	fresh tubers in the mix	too many unnecessary details (i.e. when telling a story)	5.35	2.1
9b	literal	unfamiliar	fresh tubers in the mix	the fresh root vegetable in the mixture	5.35	2.1
9c	figurative	familiar	fresh berries in the mix	too many unnecessary details (i.e. when telling a story)	5.35	2.1
9d	literal	familiar	fresh berries in the mix	the fresh fruits in the mixture	5.35	2.1
10a	figurative	unfamiliar	the lost pinion	another person's unneccessary effort	3	1.25
10b	literal	unfamiliar	the lost pinion	the missing bicycle gear	3	1.25
10c	figurative	familiar	the lost coin	another person's unneccessary effort	3	1.25
10d	literal	familiar	the lost coin	the missing money	3	1.25

Item	Phrase-Type	Word-Type	Target Phrase	Paraphrase	Image.	Transp.
11a	figurative	unfamiliar	the dirty smock on the floor	something everyone is talking about (gossip)	5.6	2.5
11b	literal	unfamiliar	the dirty smock on the floor	the unwashed clothing cover lying on the ground	5.6	2.5
11c	figurative	familiar	the dirty scarf on the floor	something everyone is talking about (gossip)	5.6	2.5
11d	literal	familiar	the dirty scarf on the floor	the unwashed piece of clothing lying on the ground	5.6	2.5
12a	figurative	unfamiliar	take one away from the shoal	to openly disagree with a popular opinion	2.9	4.1
12b	literal	unfamiliar	take one away from the shoal	to remove an individual from the group of fish	2.9	4.1
12c	figurative	familiar	take one away from the crowd	to openly disagree with a popular opinion	2.9	4.1
12d	literal	familiar	take one away from the crowd	to remove an individual from the larger group	2.9	4.1
13a	figurative	unfamiliar	the cleats are worn out	some ideas are not worth following	5.95	2.85
13b	literal	unfamiliar	the cleats are worn out	the sport shoes are not new	5.95	2.85
13c	figurative	familiar	the boots are worn out	some ideas are not worth following	5.95	2.85
13d	literal	familiar	the boots are worn out	the winter shoes are not new	5.95	2.85
14a	figurative	unfamiliar	an old gazebo made of wood	something which is not up to the current standards	5.05	4.35
14b	literal	unfamiliar	an old gazebo made of wood	an old freestanding structure built from wood	5.05	4.35
14c	figurative	familiar	an old playground made of wood	something which is not up to the current standards	5.05	4.35
14d	literal	familiar	an old playground made of wood	an old park for children built from wood	5.05	4.35
15a	figurative	unfamiliar	another bangle in the collection	someone who finds himself very attractive	5.25	1.7
15b	literal	unfamiliar	another bangle in the collection	an additional bracelet to add to the others	5.25	1.7
15c	figurative	familiar	another watch in the collection	someone who finds himself very attractive	5.25	1.7
15d	literal	familiar	another watch in the collection	an additional wristwatch to add to the others	5.25	1.7
16a	figurative	unfamiliar	the neatly folded lapel	a high standard of work that serves as an example to others	5.6	4.25
16b	literal	unfamiliar	the neatly folded lapel	a carefully folded and pressed jacket collar	5.6	4.25
16c	figurative	familiar	the neatly folded sweater	a high standard of work that serves as an example to others	5.6	4.25
16d	literal	familiar	the neatly folded sweater	a carefully folded and pressed warm top	5.6	4.25
17a	figurative	unfamiliar	under the oversized awning	getting better and becoming more successful	5.9	1.9
17b	literal	unfamiliar	under the oversized awning	beneath the very large sun protection	5.9	1.9
17c	figurative	familiar	under the oversized umbrella	getting better and becoming more successful	5.9	1.9
17d	literal	familiar	under the oversized umbrella	beneath the very large rain protection	5.9	1.9
18a	figurative	unfamiliar	wheel over the large gurney	to learn about an unpleasant truth	4.65	1.5
18b	literal	unfamiliar	wheel over the large gurney	to roll the large wagon somewhere	4.65	1.5
18c	figurative	familiar	wheel over the large suitcase	to learn about an unpleasant truth	4.65	1.5
18d	literal	familiar	wheel over the large suitcase	to roll the large luggage somewhere	4.65	1.5
19a	figurative	unfamiliar	the barrette that went missing	a difficult or unpleasant example of something	4.25	2.1
19b	literal	unfamiliar	the barrette that went missing	to find the lost hair clip	4.25	2.1
19c	figurative	familiar	the glasses that went missing	a difficult or unpleasant example of something	4.25	2.1
19d	literal	familiar	the glasses that went missing	to find the lost piece of eyewear	4.25	2.1
20a	figurative	unfamiliar	on the other side of the wide moat	the result following a heartbreak	5.3	2.65
20b	literal	unfamiliar	on the other side of the wide moat	on the opposite side of the canal	5.3	2.65

Item	Phrase-Type	Word-Type	Target Phrase	Paraphrase	Image.	Transp.
20c	figurative	familiar	on the other side of the wide bridge	the result following a heartbreak	5.3	2.65
20d	literal	familiar	on the other side of the wide bridge	on the opposite end of the overpass	5.3	2.65
21a	figurative	unfamiliar	one's golden tassel	something less powerful than it appears	5.7	1.55
21b	literal	unfamiliar	one's golden tassel	a bundle of golden hanging threads	5.7	1.55
21c	figurative	familiar	one's golden necklace	something less powerful than it appears	5.7	1.55
21d	literal	familiar	one's golden necklace	a golden piece of hanging jewelry	5.7	1.55
22a	figurative	unfamiliar	don't forget the fancy garter	do what you can to get your life on track	3.85	1.5
22b	literal	unfamiliar	don't forget the fancy garter	remember the nice silk undergarment	3.85	1.5
22c	figurative	familiar	don't forget the fancy belt	do what you can to get your life on track	3.85	1.5
22d	literal	familiar	don't forget the fancy belt	remember the nice leather waistband	3.85	1.5
23a	figurative	unfamiliar	don't catch the perch	beware of asking about something	3.9	1.4
23b	literal	unfamiliar	don't catch the perch	do not hook this specific type of fish	3.9	1.4
23c	figurative	familiar	don't catch the salmon	beware of asking about something	3.9	1.4
23d	literal	familiar	don't catch the salmon	do not hook this specific type of fish	3.9	1.4
24a	figurative	unfamiliar	decorate the place with tinsel	to cover up one's mistakes	5.7	2.8
24b	literal	unfamiliar	decorate the place with tinsel	to make a room look nice with shiny material	5.7	2.8
24c	figurative	familiar	decorate the place with candles	to cover up one's mistakes	5.7	2.8
24d	literal	familiar	decorate the place with candles	to make a room look nice with candles	5.7	2.8
25a	figurative	unfamiliar	your grandmother's doily	a useless activity that no one wants to do	5.25	1.5
25b	literal	unfamiliar	your grandmother's doily	a grandmother's small decorative cloth	5.25	1.5
25c	figurative	familiar	your grandmother's blanket	a useless activity that no one wants to do	5.25	1.5
25d	literal	familiar	your grandmother's blanket	a grandmother's warm decorative cover	5.25	1.5
26a	figurative	unfamiliar	repair the wooden banister	to do something unexpectedly nice	4.9	2.45
26b	literal	unfamiliar	repair the wooden banister	to restore the handrail of the staircase made of wood	4.9	2.45
26c	figurative	familiar	repair the wooden table	to do something unexpectedly nice	4.9	2.45
26d	literal	familiar	repair the wooden table	to restore the rectangular table made of wood	4.9	2.45
27a	figurative	unfamiliar	the putty on the wall	a truth one can no longer ignore	4.5	2.75
27b	literal	unfamiliar	the putty on the wall	the soft paste located on the wall	4.5	2.75
27c	figurative	familiar	the glue on the wall	a truth one can no longer ignore	4.5	2.75
27d	literal	familiar	the glue on the wall	the sticky paste located on the wall	4.5	2.75
28a	figurative	unfamiliar	the damaged bassoon	one's personal and secret wish	5	1.4
28b	literal	unfamiliar	the damaged bassoon	a broken woodwind instrument	5	1.4
28c	figurative	familiar	the damaged drum	one's personal and secret wish	5	1.4
28d	literal	familiar	the damaged drum	a broken percussion instrument	5	1.4
29a	figurative	unfamiliar	how the gavel hits the counter	an unfortunate and uncomfortable situation	5.2	2.4
29b	literal	unfamiliar	how the gavel hits the counter	how the wooden hammer strikes the work table	5.2	2.4
29c	figurative	familiar	how the knife hits the counter	an unfortunate and uncomfortable situation	5.2	2.4
29d	literal	familiar	how the knife hits the counter	how the metal blade strikes the work table	5.2	2.4
30a	figurative	unfamiliar	a sash to match the look	the standards of behavior followed by most people	4.8	2.6
30b	literal	unfamiliar	a sash to match the look	a strip of satin cloth that goes with an outfit	4.8	2.6
30c	figurative	familiar	a glove to match the look	the standards of behavior followed by most people	4.8	2.6
30d	literal	familiar	a glove to match the look	a hand covering that goes with an outfit	4.8	2.6

10.6 APPENDIX F

Appendix F refers to the experimental items in Chapter 7. Sentences were split into beginning and ending for presentation.

Item	condition	Beginning	Ending
1	idiom	John needed to	deliver the goods.
1	literal	John needed to	deliver the pizza.
2	literal	I told him to	get a drink.
2	idiom	I told him to	get a grip.
3	literal	John appears to	have the code.
3	idiom	John appears to	have the blues.
4	idiom	He wanted to	have a go.
4	literal	He wanted to	have a party.
5	idiom	John was ready to	hit the road.
5	literal	John was ready to	hit the button.
6	literal	Jessica's uncle	kicked the ball.
6	idiom	Jessica's uncle	kicked the bucket.
7	literal	It's clear that he	knew the names.
7	idiom	It's clear that he	knew the score.
8	idiom	John worried he might	lose his touch.
8	literal	John worried he might	lose his keys.
9	idiom	The young man tried to	make a pass.
9	literal	The young man tried to	make a poster.
10	literal	I heard he wants to	pop the balloon.
10	idiom	I heard he wants to	pop the question.
11	literal	John always likes to	pull my hair.
11	idiom	John always likes to	pull my leg.
12	idiom	Peter finally	saw the light.
12	literal	Peter finally	saw the film.
13	idiom	He decided not to	sit the fence.
13	literal	He decided not to	sit on the bench.
14	literal	I think that John	spilled the milk.
14	idiom	I think that John	spilled the beans.
15	literal	We will need to	split the chocolate.
15	idiom	We will need to	split the difference.
16	idiom	She will have to	take the stand.
16	literal	She will have to	take the bags.
17	idiom	John just saw them	tie the knot.
17	literal	John just saw them	tie the bow.
18	literal	Sam is going to	turn the burgers.
18	idiom	Sam is going to	turn the tables.
19	literal	She told him not to	waste his paint.
19	idiom	She told him not to	waste his breath.
20	idiom	Peter couldn't just	play the game.
20	literal	Peter couldn't just	play the piano.
21	idiom	I think he would	give the world.
21	literal	I think he would	give the answer.
22	literal	She just about	went to London.
22	idiom	She just about	went to pieces.
23	literal	Last night he really	took the car.
23	idiom	Last night he really	took the cake.
24	idiom	At work last week John	turned a corner.
24	literal	At work last week John	turned a screw.
25	idiom	Sandy tried to	break the ice.
25	literal	Sandy tried to	break the glass.
26	literal	We will need to	cover the food.
26	idiom	We will need to	cover the territory.

Item	condition	Beginning	Ending
27	idiom	Today John will	do the honors.
27	literal	Today John will	do the laundry.
28	idiom	Monday she really	dropped a bombshell.
28	literal	Monday she really	dropped a plate.
29	literal	I heard he might	have cold hands.
29	idiom	I heard he might	have cold feet.
30	idiom	He don't think he	has the heart.
30	literal	He don't think he	have the tickets.
31	literal	Writing notes helped Sally not to	lose her glasses.
31	idiom	Writing notes helped Sally not to	lose her thread.
32	literal	Her cousin really	missed the basket.
32	idiom	Her cousin really	missed the mark.
33	idiom	I am sure she	played a part.
33	literal	I am sure she	played a violin.
34	idiom	He might have to	pull the plug.
34	literal	He might have to	pull the rope.
35	literal	She told me to	save my money.
35	idiom	She told me to	save my skin.
36	idiom	I don't think he	wears the pants.
36	literal	I don't think he	wears the watch.
37	idiom	That's his chance to	show his flag.
37	literal	That's his chance to	show his art.
38	literal	He asked if it	ring a phone.
38	idiom	He asked if it	rings a bell.
39	literal	I hope he can	lend an egg.
39	idiom	I hope he can	lend an ear.
40	idiom	Sam found it hard to	make a living.
40	literal	Sam found it hard to	make a list.