Bridging From Multi-dimensionality of Idioms to Their Embodiment

Mahsa Morid

Supervisor: Dr. Laura Sabourin

A thesis submitted in partial fulfillment of the requirements for the

Doctorate in Philosophy degree in Linguistics

Department of Linguistics

Faculty of Arts

University of Ottawa

Abstract

In this thesis, I investigate idiom processing from two angles through three different studies. First, I approached idiom processing from a constraint-based perspective. According to this view, not all idioms are alike: they can differ regarding lexical, and linguistic characteristics, such as their level of familiarity. In this first study, I investigated the underlying processes during the comprehension of idioms with different characteristics. I used the Event Related Potential (ERP) technique, which has high temporal resolution, to investigate this issue. I provided evidence that idioms' characteristics impact their processing. More specifically, idioms which are more familiar to language users (i.e., the ones that are encountered more frequently) showed processing facilitation compared to less familiar idioms. Also, idioms with plausible literal interpretation showed processing advantages over idioms which are less likely to be interpreted literally. The second aim of the current thesis was to investigate idiom processing from an embodied account of language processing. According to this view, various sources of information (including linguistic, affective, and sensory-motor) are available and used during the comprehension of language. While, this view has become popular in many language processing studies, studies of idiom processing are still at the beginning of this journey. To be able to investigate idiom processing while considering the role of affective and sensory-motor factors, we require access to norming data. In the second (descriptive) study, I conducted a large-scale survey and collected measures of valence, arousal, concreteness, and imageability for a set of English idioms, by both native speakers of English and proficient second language speakers. In the last study, I explored how the emotional status of idioms and their concreteness contributes to their processing, and whether this contribution is modulated by idiom familiarity. We found that the impact of non-linguistic sources of information (affective and sensory-motor) is determined

by idiom familiarity, such that low familiar and high familiar idioms show different behaviour where these factors are concerned. For highly familiar idioms, behaviour aligns with the findings on word processing: for example, idioms with more positive valence showed facilitative processing. Unlike highly familiar idioms, valence had an inhibitory impact on idioms with low familiarity level, such that greater valence increased the reading time.

Keywords: Idiom processing, event-related potentials, self-paced reading, familiarity, literal plausibility, valence, arousal, concreteness, imageability

Co-Authorship Statement

The chapters of this dissertation are manuscripts that have been prepared for submission to scientific journals. The presented data are based on a series of collaborative research projects; however all manuscripts have been primarily written by Mahsa Morid. Chapter 2 has been published by *Canadian Journal of Experimental Psychology*, authored by Mahsa Morid, Nadia Bachar, & Laura Sabourin. Chapter 3 has been written in preparation to be submitted to *Applied Psycholinguistics* Journal, authored by Mahsa Morid & Laura Sabourin.

Chapter 4 has been published *Journal of Psycholinguistic Research* and is currently under revision, authored by Mahsa Morid & Laura Sabourin.

Acknowledgments

This has been a long journey. I have dreamed about writing this section of my dissertation and each time I started by thanking my great supervisor. Prof. Laura Sabourin, I sincerely thank you for your constant support and understanding during this journey. I would have not been able to accomplish this goal without your wise guidance and kind and supportive manner.

I would like to thank all my collogues, the ERPLing lab members, past and present, who never turned me down when I had a question and when I needed support. Santa Vinerte, Michèle Burkholder, Joy Yu Ying Li, Gabrielle Manning, Leah Gosselin, Ariane Senécal, I thank you for your academic support, but above all for all the chats and laughs, and great memories.

thanks to the amazing friends who have been with me the whole way, both in and out of grad school.

I would like to thank my family, my husband, my mother and my sisters, who have been a constant source of love and support. And finally, my dear daughter, Hana, who joined me in the middle of this journey. You slowed me down in the most incredible way. You were my sunshine during grey days.

Table of Contents

List of Tables

List of Figures

1 Introduction

> *"Edward do you know how many dogs the Russians sent into space before a man* ***walked on the moon****?... we are going to find out what went wrong [with our experiment] and we'll try again, and we'll fail again, because that's what progress looks like. Progress looks like a bunch of failures and you can have feeling about that, because it's sad, but you can't fall apart. And then one day we will succeed, and we will* ***walk on the moon****! Figuratively, anyway."*

This excerpt from the TV series, Grey's Anatomy, exemplifies one of many instances of figurative language we encounter on a daily basis. Differentiating between two instances of "walk on the moon" and reaching the proper meaning seems effortless for native speakers of the respective language (here English). Just like other seemingly-effortless aspects of language processing, what happens behind the scene is not as easy to explain. Unlike the literal interpretation of "walk on the moon", the idiomatic interpretation does not follow the simple rule of 2 + 2 = 4. In other words, the combination of the meaning of each word does not get us to the final figurative meaning of "walk on the moon".

For over half a century, different models have been proposed to explain the processing of idioms. I will start by giving the reader an overview of these models and situate the current dissertation within the field of idiom processing. This will be followed by a more detailed discussion of these models. Figure 1.1 depicts different generations of idiom processing models. The studies reported in the current dissertation are conducted within the framework of two different accounts of language processing as specified by the blue arrows. First, we considered a well-stablished

model of idiom processing (multidetermined model (Libben & Titone, 2008)) and we investigated the neurophysiological underpinnings associated with idiom comprehension by zooming in on their multi-determined nature (Chapter 2). Chapter 3 and Chapter 4 connects the processing of idioms to the embodied account of language processing. In other words, we aimed to go beyond the linguistic factors that the multi-determined model suggests affect idiom processing, and consider non-linguistic factors (emotional and sensory-motor) in idiom processing. In Chapter 5, I will present a general discussion and a proposal for future studies.

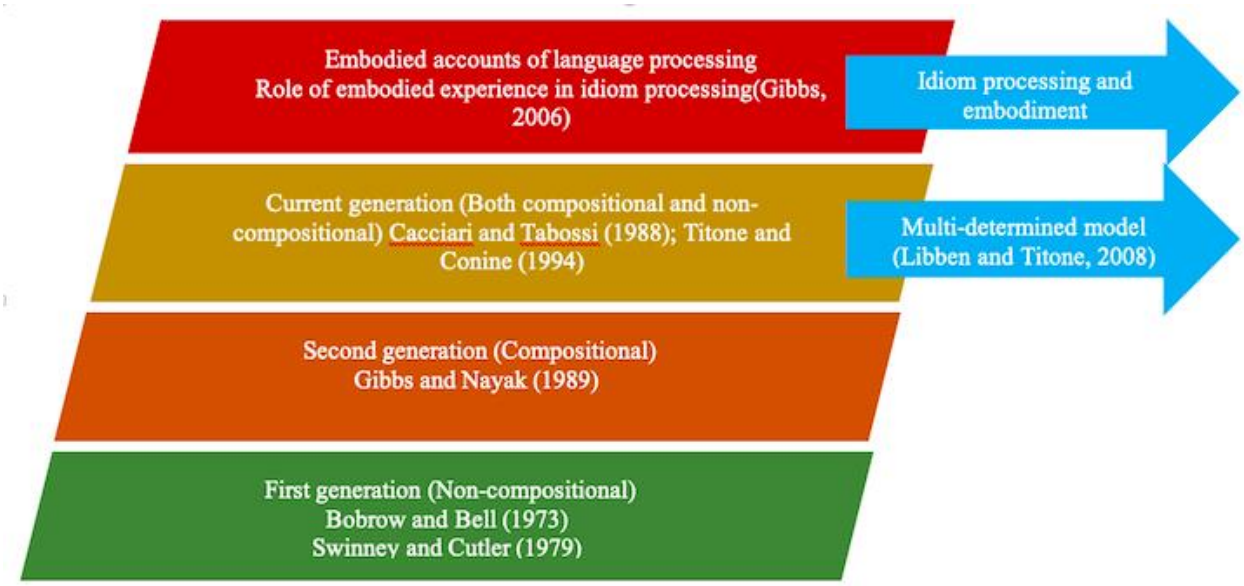


Figure 1.1 Situating the current study within the field of idiom processing. Blue arrows indicate the areas within which the studies presented in the current dissertation are conducted.

1.1 Definition of Idioms

When speakers use figurative language, their intended meaning cannot be understood by combining the meaning of individual words (Glucksberg, 1991). Along with other stylistic

figures (metaphors, proverbs, sarcasm, etc.), idioms (e.g., "*be a babe in the woods*", "*be a piece of cake*") are a type of figurative language. The exact definition of idioms is still debated and there is no consensus on how to differentiate idioms from other types of figurative expressions (Gibbs & Colston, 2012; Cacciari & Tabossi, 2014). Despite this lack of agreement, most researchers consider two main characteristics as defining features of idioms: 1) they are multi-word expressions whose meaning does not necessarily derive from the meaning of their constituents; 2) their meaning is conventionalized and cannot change through context.

1.2 Models of Idiom Processing

Models of idiom representation and processing[1] are generally divided based on their position regarding compositionality of idioms (Panou, 2017). First generation models, consider idioms as fixed expressions whose processing is non-compositional in nature. The common characteristics of these models is that they consider idiomatic and literal processing to be mutually exclusive (Cacciari & Tabossi, 2014). These models vary on the priority they give to the literal or idiomatic meaning activation (idiom-list hypothesis: literal first (Bobrow & Bell, 1973); lexical representation hypothesis: parallel activation of literal and idiomatic meaning (Swinney & Cutler, 1979); direct-access hypothesis: figurative first (e.g., Gibbs, 1980)). Despite these differences, these models all consider idioms as long words whose meaning are accessed directly from the lexicon. The middle generation of idiom processing models, are characterized by the

[1] In this section, the intention is not to give a comprehensive overview of all idiom processing models, but rather to give the reader sufficient and relevant information to be able to situate this study within models of idiom processing, and language processing more generally.

emphasis on pure compositionality of idiom processing[2] (Gibbs & Nayak, 1989; Hamblin & Gibbs, 1999). Moving forward, the latest generation of idiom processing models (known as hybrid models) state that idiom comprehension may happen through both compositional and non-compositional processes (Cacciari & Tabossi (1988); Titone & Conine (1994)). One common characteristics of this class of models is that they do not consider idioms as homogenous expressions, rather they all agree that idioms possess features that affect their processing. For example, a familiar idiom (i.e., an idiom which is frequently used or heard by the speakers) may be easier to process because the idiomatic meaning can, more easily, be directly retrieved.

Multi-determined model. Libben and Titone (2008) conducted a series of 4 studies to examine the effect of various characteristics that might impact idiom processing. In their first study, they collected norming measures on 210 English idioms. Participants read idioms and rated them for

1. Familiarity (the frequency with which they encounter idioms)
2. Meaningfulness (how well they know the meaning of idioms)
3. Literal Plausibility (whether the idiom has a plausible literal interpretation)

[2] Notice that some models (e.g., configuration hypothesis) have been classified under both compositional and non-compositional views by different researchers, because they bear characteristics that pertain to both.

4. Decomposability[3] (how the meaning of idioms is related to the meaning of its constituents).

among other factors. Using both off-line and on-line measures, Libben and Titone (2008) examined the impact of those factors on the processing of idioms. Based on the results of these studies, they proposed the multidetermined model of idiom processing: multiple sources of information are used at various stages of processing to compute the figurative meaning. These sources of information include both direct retrieval and compositional analysis. For example, compared to low familiar idioms, idioms which are highly familiar, are more likely to be configured in the memory and thus the activation and retrieval of the idiomatic meaning occurs sooner and easier. On the other hand, when idioms are low familiar, other factors such as possible literal interpretation may impact their processing. For low familiar idioms, with lower possibility of direct retrieval, the processing is more likely to be compositional. If such idioms have a possible literal interpretation, then the processor needs to select between two possible meanings, imposing processing load. However, in the absence of such competition, low literally plausible idioms will be easier to comprehended.

1.3 Using ERPs to Study Language Processing

In the 1980s, Event Related brain Potentials (ERPs) began to be used in studies of language processing (reviewed in Kaan, 2007). Due to its high temporal resolution this technique serves as a suitable tool to study fast occurring language processing. Besides temporal resolution, ERPs have another advantage over behavioural methods; the multi-dimensional nature of ERP

---

[3] In their study, Libben and Titone (2008) provide global decomposability, noun decomposability and verb decomposability. Whenever we mention decomposability, we mean global decomposability.

responses (Kaan, 2007) allows researchers to make a clearer inference about the underlying processes, since various ERP components have been related to a specific type of process. In contrast, ERPs are not as strong in terms of spatial resolution. This is partly due to the nature of this technique Luck (2014), and partly due to the fact that language-related ERPs occur relatively later than other types of processing and the associated components usually have a long-latency. This makes it difficult to locate what neural generators contribute to the generation of a specific ERP component (Luck & Kappenman, 2011). Nevertheless, over the years, studies focusing on the same ERP components have made it possible to relate a specific component, with a specific latency and topographic distribution, to a particular process or set of processes (Kutas & Hillyard, 1980).

1.3.1 ERP Components Reported in Idiom Processing Studies

The P600

The P600 component was initially reported in response to syntactic manipulations (Osterhout & Holcomb, 1992). The P600 is a positive shift, which occurs around 500 ms post-stimulus onset, and typically reaches its peak around 600 ms lasting for several hundred milliseconds. This component is typically observed over posterior electrode sites, although some studies report a more anterior distribution (e.g., Kaan & Swaab, 2003). Although the sensitivity of this component to syntactic manipulations appears to be a perfect complement to the semantic nature of the N400 (i.e., N400 for semantic aspect and P600 for syntactic aspect of language processing), more recently it has been shown that this component can be elicited in response to semantic violations (e.g., Kim & Osterhout, 2005) or pragmatic violations (Kuperberg et al., 2003). In summary, the P600 component does not reflect a unitary phenomenon (Canal et al., 2017) and

based on its functional role, shows different scalp distributions (e.g., repair mechanisms show a parietal, and reanalysis mechanisms have a fronto-central distribution) (Kaan & Swaab, 2003).

Given that the functional interpretation of the P600 remains open to interpretation (Luck & Kappenman, 2011), studies on idiom processing with focus on the P600 are typically less clear about their exact predictions or interpretations. Some studies point to this component's sensitivity to domain-general factors (e.g., Laurent at al., 2006) or its potential sensitivity to the structure of highly constraining context of idioms (e.g., Liu at al., 2010).

The P300

In contrast to the language related N400 and P600, the P300 (including the P3a an d P3b subcomponents) is reported in studies using oddball paradigms to investigate how stimulus information changes the brain electrical patterns (Luck & Kappenman, 2011). According to the context-updating theory, the P300 reflects brain activities underlying revision mechanisms. Alternatively, the P300 has been considered as an index of the sensitivity to the amount of resource allocation, (i.e., the amount of attentional resources that are engaged during task performance). Subsequently, the latency of this component is proportional to stimulus and task requirements. Its peak latency changes over the scalp with shorter frontal latency compared to parietal sites (Luck & Kappenman, 2011).

Some studies of idioms have considered the P300 component (e.g., Vespignani et al., 2010). These studies suggest that during idiom processing, language users process the words compositionally up to the point that they identify an idiom, after which the actual upcoming input (i.e., the individual constituents of idioms) is compared with the idiom-configuration stored in semantic memory resulting in elicitation of a P300.

The N400

Kutas and Hillyard (1980), in a landmark study, asked participants to read sentences containing a semantic anomaly (e.g., He spread the warm bread with socks) and compared them with semantically well-formed sentences. The sentences containing a semantic anomaly elicited a negativity that reached its peak at around 400 ms after the onset of the target word and was maximal over centro-parietal electrode sites. This component, labeled N400, has since been observed and reported by over 1000 published papers (Kutas & Federmeier, 2011). Two broad accounts have been proposed in regard to the N400s nature of processing. Kutas and colleagues (2006) proposed that the amplitude of the N400 is modulated by the "ease or difficulty of retrieving stored conceptual knowledge associated with a word (or other meaningful stimuli), which is dependent on both the stored representation itself, and the retrieval cues provided by the preceding context" (p. 669). Therefore, when a participant reads a sentence such as "He spread the warm bread with…", the word "butter" elicits a small N400, because the sentence context pre-activates the semantic features of that word, making its retrieval easier and less effortful. Whereas the word "cream" would elicit a larger N400, and a word such as "socks" would elicit the highest N400. The second account regarding the processing nature of the N400, proposed by Hagoort (2005) posits that the N400 does not merely reflect ease of retrieval, but it reflects the "semantic integration and unification process". In other words, according to this account, the word "butter" in the above example would elicit a smaller N400 because it can more easily be integrated into the sentence context than "cream" and "socks". Given the processing nature of the N400, it is not surprising that most ERP studies on idiom processing have focused on this component (Canal et al., 2017).

Post-N400 Positivity (PNP)

The term Post-N400 positivity (PNP), is used by some researchers as a theoretically neutral term to refer to any enhanced positivity, that is evident either after the N400 (600-900 ms), or with "substantial temporal overlap" with the N400 (Thornhill & Van Petten, 2012). It is hypothesized that this component reflects the cost of incorrect predictions by comprehenders. As opposed to the P600 component that has been attributed to re-analysis mechanisms, much less is known about the PNP (Van Petten & Luka, 2012). Nevertheless, the results of some empirical studies suggest that this positivity, which appears over frontal and left hemisphere electrode sites, is larger in response to unpredictable words, and appears to be larger with congruent but unexpected words compared to predictable words (Thornhill & Van Petten, 2012). Importantly, this frontal positivity appears to be distinct from posterior positivity which is observed in response to ungrammatical or anomalous stimuli (Brothers, Swaab, & Traxler, 2015). In one idiom processing study, Canal et al., (2017), reported a positive shift in frontal electrode sites in response to the first and the last word of a multi-word idiomatic expression.

1.4 Objectives of the Current Dissertation

1.4.1 First Aim of the Current Thesis (Chapter 2)

Processing of idioms with respect to their multi-determined model has been studied widely using different measures such as self-paced reading, priming, and eye-tracking (Libben &Titone, 2008, Titone & Libben, 2014; Titone et al., 2019). The aim of this study was to investigate to what extent the neurophysiology of idiom comprehension changes as a function of the idiom's characteristics. We particularly focused on the N400 component since, as stated above, its amplitude has been interpreted as an index of ease or difficulty of integration of the meaning of a

word into the general context (Kutas & Hillyard, 1980). The result of this study provided neurophysiological evidence in support of the multi-determined model of idiom processing.

1.4.1.1 Embodied Accounts of Language Processing

The models of idiom processing that we discussed in the previous section have their roots within linguistics theories (Titone et al., 2015). For example, Bobrow and Bell (1973) has built their model motivated by early linguistic theories (Chomsky, 1981; Clark & Lucy, 1975) and Swinney and Cutler's model (1979) was influenced by other linguistics theories (Nunberg, 1978). Psycholinguistic approaches to idiom processing have been influenced by the ideas proposed by Lakoff and Johnson (1980) in their book "Metaphors we live by". The idea which has been theorized as a "conceptual metaphor" argues that the nature of our conceptual system is metaphorical (Lakoff & Johnson, 2020). For example, in a culture, *argument* might be understood as a battle, when the people involved attack each other. So, the argument is conceptualized metaphorically as *argument is war*. In an imaginary culture, however, argument could be realized as *dance* during which the people involved try to create a performance rather than attacking each other. Two major conclusions of this view is that: the metaphorical concepts (e.g., *argument is war* versus *argument is dance*) help us to *understand one kind of experience in terms of another kind of experience* and essentially a less concrete experience (i.e., *argument*) is understood through more concrete ones (i.e., *war* or *dance*). Based on this view, cognitive embodiment argues that all cognitive processes, including language processing, have links with affective and motor areas of the brain (Vigliocco et al., 2009; Williamson, 2022). More specifically, embodied theories of semantic representation and processing argue that during the processing of meaning, the same neural structures that relate to perception and action gets activated automatically. Recently, and based on this view, researchers suggest that the emotional

content of idioms (valence: the extent to which an emotion is pleasant/positive or unpleasant/negative, and arousal: whether the evoked emotion is perceived as exciting or calming) and concreteness (the degree to which the meaning of a word or expression is understood through perception and action) may impact their processing (e.g., Citron et al., 2016; Findlay & Carrol, 2018).

1.4.1.2 Idiom Processing and Embodiment, Why it Matters!

Idioms usually convey meanings that are different from the meaning of its constituents. Importantly, it implies that the individual words that constitute idioms do not necessarily transfer their characteristics to the idiom. For example, the level of concreteness of that idiom might be different from the level of concreteness of the individual words of the idiomatic expression. Figure 1.2 schematically shows how the level of concreteness of the individual constituents of the expression "racked her brain" is different from the concreteness level of the idiom[4].

The left panel of Figure 2 shows the concreteness of the individual components of the

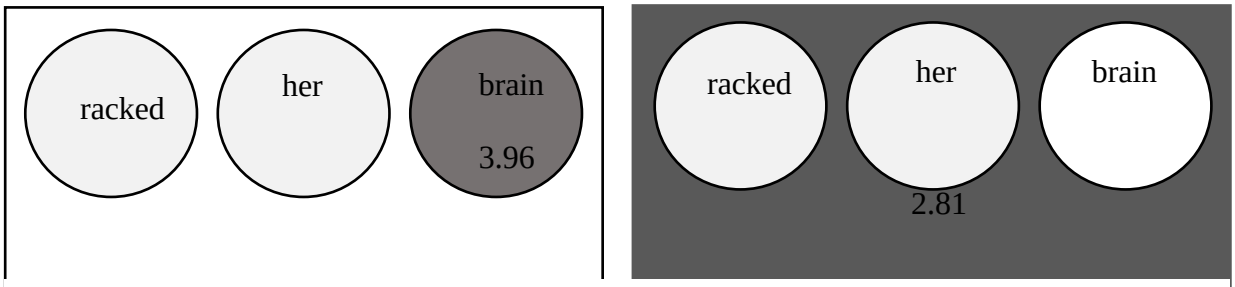


Figure 1.2 Comparing the concreteness of the individual component of "racked her brain" with the concreteness of the expression "racked her brain".

[4] The concreteness rating of the examples in this section are taken from the survey study that will be discussed in chapter 3. The concreteness of the individual words are taken from (Brysbaert & Warriner, 2014) .

The right panel shows the concreteness of the idiomatic expression “racked her brain”.

When the participants read/hear the idiom “racked her brain”, if the expression is unfamiliar to them and if there is no supporting context before the idiom, the word “brain” might be analyzed as an individual unit. However, if the language user is familiar with this expression, and more importantly, if the prior context makes the expression more predictable, then by reaching the end of the expression, the meaning of “racked her brain” might get activated. These two scenarios show that the language user is processing two words/concepts with varying level of concreteness (3.96 and 2.81 for “brain” or “racked her brain”, respectively).

1.4.2 Second Aim of the Current Thesis (Chapter 3)

In order to investigate the role of affective and sensory-motor variables in idiom processing, we developed affective (valence and arousal) and sensory-motor (concreteness and imageability) norms for 210 English idioms rated by native English speakers (L1) and English second language speakers (L2). Chapter 3 discusses the obtained ratings along with the relation between the collected norms (i.e., affective and sensory-motor variables) and the relation between these norms with previously collected psycholinguistic norms. We will also discuss the relation between L1 and L2 ratings. The collected ratings allowed us to conduct the third study (Chapter 4).

1.4.3 Third Aim of the Current Thesis (Chapter 4)

In the study presented in chapter 4, we investigated idiom processing from the perspective of the embodied theories of cognition—more specifically, the models that suggest multiple sources of information (linguistics, sensory-motor, and emotional) account for semantic representation and processing. The main aim of this study was to investigate how different sources of information—

affective, sensory-motor, as well as linguistic information— contribute to the processing of idioms. We found that different sources of information (sensory-motor and affective) contribute to the processing of idioms. Importantly, the impact of these non-linguistic sources of information is mediated by the level of familiarity of idiomatic expressions.

The details regarding each study, along with the review of the related literature is presented in the subsequent chapters.

References

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

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition, 136*, 135-149.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, *46*(3), 904-911.

Cacciari, C. & Glucksberg, S. (1991). Understanding idiomatic expressions: the contribution of word meanings. In G.B. Simpson (Ed.), *Understanding Word and Sentence* (pp. 217-240). Amsterdam: Elsevier.

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

Cacciari, C., & Tabossi, P. (2014). *Idioms: Processing, structure, and interpretation*. Psychology Press.

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.

Chomsky, N. (1981). Principles and parameters in syntactic theory. *Explanation in linguistics*, 32-75.

Citron, F. 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.

Clark, H. H. & Lucy, P. (1975). Understanding what is meant from what is said: a study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behaviour, 14*, 56–72.

Findlay, H., & Carrol, G. (2019). Contributions of semantic richness to the processing of idioms. *The Mental Lexicon, 13*(3), 311-332.

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

Gibbs Jr, R. W., & Colston, H. L. (2012). *Interpreting figurative meaning*. Cambridge University Press.

Gibbs Jr, R. W., & Nayak, N. P. (1989). Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive psychology*, *21*(1), 100-138.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in cognitive sciences*, *9*(9), 416-423.

Hamblin, J. L., & Gibbs, R. W. (1999). Why you can't kick the bucket as you slowly die: Verbs in idiom comprehension. *Journal of Psycholinguistic research*, *28*(1), 25-39.

Kaan, E. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, *1*(6), 571-591.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of cognitive neuroscience*, *15*(1), 98-110.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of memory and language*, *52*(2), 205-225.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive brain research*, *17*(1), 117-129.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In *Handbook of psycholinguistics* (pp. 659-724). Academic Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors we Live by*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (2020). Conceptual metaphor in everyday language. In *Shaping Entrepreneurship Research* (pp. 475-504). Routledge.

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

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

Liu, Y., Li, P., Shu, H., Zhang, Q., & Chen, L. (2010). Structure and meaning in Chinese: An ERP study of idioms. *Journal of Neurolinguistics*, *23*(6), 615-630.

Luck, S. J. (2014). An Introduction to the Event-Related Potential Technique. MIT Press.

Luck, S. J., & Kappenman, E. S. (Eds.). (2011). *The Oxford handbook of event-related potential components*. Oxford university press.

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

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, *31*(6), 785-806.

Panou, D. (2017). A Critical Assessment of L1 Idiom Comprehension and Processing Models. *Journal of Applied Linguistics and Language Research, 4*(3), 1-12.

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of verbal learning and verbal behavior, 18*(5), 523-534.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83*(3), 382-392.

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting Bilingual and Monolingual Idiom Processing. In R. R. Heredia & A. B. Cieslicka (eds.), *Bilingual Figurative Language Processing*, pp. 171–207. Cambridge: Cambridge University Press.

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

Titone, D., & Libben, M. (2014). Time-dependent effects of decomposability, familiarity and literal plausibility on idiom priming: A cross-modal priming investigation. *The Mental Lexicon, 9*(3), 473-496.

Titone, D., Lovseth, K., Kasparian, K., & Tiv, M. (2019). Are figurative interpretations of idioms directly retrieved, compositionally built, or both? Evidence from eye movement measures of reading. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 73*(4), 216.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190.

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

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition, 1*(2), 219-247.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods, 45*(4), 1191-1207.

Williamson, T. R. (2022). Embodiment Effects in Idiom Processing. *Cambridge Occasional Papers in Linguistics, 14*(1), 180-199.

2 Capturing the Multi-determined Nature of Idiom Processing Using ERPs[5]

Abstract

The multi-determined model (Titone & Libben, 2014) suggests that processing of idioms depends on multiple linguistic factors (e.g., familiarity, literal plausibility, decomposability). According to this model, these sources of information modulate the comprehension of idioms at different time courses. In the current study, we investigated whether these linguistics factors modulate the neurophysiological underpinnings associated with processing of different types of idioms. Adult native speakers of English read sentences that contained idioms with high and low level of familiarity and literal plausibility while their EEG was recorded. Event-related potentials data showed that idioms with low level of familiarity elicited larger negativity starting from 300 ms post stimulus onset and lasted for about 200 ms. A similar negativity, but which started later (at around 400 ms post stimulus onset) was also observed for idioms with a low level of literal plausibility. These results are consistent with the multi-determined model of idiom processing indicating the role of these linguistics factors over different time courses. Finally, the observed negativity for low familiar and low literally plausible idioms was greater over the right hemisphere. Accordingly, the possible role of the right hemisphere in processing idioms will be discussed.

Keywords: Formulaic language, Idioms, Comprehension, Multi-determined model, Event-related brain Potentials

---

[5] Morid, M., Bachar, N., & Sabourin, L. (2021). Capturing the multi-determined nature of idiom processing using ERPs. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *75*(2), 155.

2.1 Introduction

Figurative language plays an important role in every-day verbal communication (Kuiper, 2009). We purposefully play with words, combining them in such a way that they can reach a figurative interpretation, far from their actual meaning (Cacciari & Tabossi, 2014). Our brain is capable of deriving this figurative meaning and even enjoys this little game we can play with words. Idioms (e.g., "hit the sack") are a special type of figurative language, whose meaning has been conventionalized (Cacciari & Tabossi, 1998) and whose figurative meaning does not (necessarily) derive from the meaning of its constituents (Jackendoff, 1995). The general question of interest for psycholinguists is: how are idioms represented in the mind and how are they processed? (Vespignani, Canal, Molinaro, Fonda, & Cacciari, 2010; Cacciari, Padovani, & Corradini, 2007; Tabossi, Fanari, & Wolf, 2009). Most models of idiom processing attempt to describe this processing in light of what is known about literal language processing and are mainly concerned with a compositional *versus* non-compositional analysis during idiom comprehension. Generally, they favour one of three broad views: 1) idioms are represented as a single word and thus their processing is non-compositional (Bobrow & Bell, 1973; Swinney & Cutler, 1979), 2) idioms are processed compositionally, similar to regular language processing (Gibbs, Nayak, & Cutting, 1989; Hamblin & Gibbs, 1999), and 3) idiom processing is a combination of both compositional and non-compositional (direct retrieval) analyses (Cacciari & Tabossi, 1988; Titone & Connine, 1994). According to the third view, known as the hybrid view, idioms are processed compositionally until they become recognizable to language users after which direct retrieval becomes available. The introduction of the hybrid view shifted attention toward idiom-specific factors that facilitates their direct retrieval. However, the emphasis on these factors varies from one study to another. For example, some studies claim that the

predictability of idioms (i.e., the likelihood of predicting the last word of the expression) plays a key role in the availability of the direct retrieval (Cacciari & Tabossi, 1988). On the other hand, other studies suggest that multiple factors (such as language users' familiarity with idioms and the existence of a literal interpretation for the idiomatic expression) modulate the listener's comprehension. Moreover, these factors seem to assert their role at various time courses during processing (Libben & Titone, 2008; Titone & Libben, 2014; Titone, Lovseth, Kasparian, & Tiv, 2019). Various measures and techniques (i.e., behavioural measures, eye-tracking, Event-Related brain Potentials (ERP)) have previously been used to test these proposed models. Before outlining the current study, the following section will provide a brief overview of ERPs, the primary method used in the current study.

2.1.1 The Electrophysiology of Idiom Comprehension

ERPs are found during real-time recording of the brain's neural activity (electroencephalography, or EEG). The EEG signals are time-locked to a special event across multiple trials and are then averaged to form ERPs. Due to its high temporal resolution, this technique is ideal to study rapidly-occurring language processing. Besides temporal resolution, ERPs have another advantage over behavioral methods; the multi-dimensional nature of ERP responses (Kaan, 2007) allows the researchers to make clearer inference in regard to underlying processes, as various ERP components (e.g., N400, P600, etc.) have been related to specific processes. To date, few studies on idiom processing have utilized this technique. A subset of the literature focuses directly on one specific model of idiom processing (e.g., assessing the *salience hypothesis*; Laurent et al., 2006), while others use idioms as a vehicle to address more general issues in language processing (e.g., exploring *predictive forward-looking mechanisms* in language comprehension; Vespignani et al., 2010). Since these studies focus on differing

mechanisms that give rise to distinct ERP components, the time course and topography of ERPs being investigated varies (e.g., P300, N400, P600, and Post N400 Positivity known as PNP; for an overview of PNP see Van Petten & Luka, 2012). The remainder of this section reviews some of these studies.

Rommers, Dijkstra, and Bastiaansen (2013) conducted an ERP study to explore *semantic unification*, (i.e., integration of the meaning of the upcoming words into the larger units (Hagoort, 2005)), during idiom comprehension. They compared Dutch idiomatic (familiar idioms with low transparency) sentences with their matched literal counterparts, each with three conditions: Correct, Related, and Unrelated (For example, Correct: *lamp*; Related: *candle*; Unrelated: *fish*). To create the related condition, they replaced the target word with a semantically related word; in the unrelated condition the target word was replaced by a semantically unrelated word. The analysis focused on three time-windows, namely the Early N400 window (300-400 ms), Late N400 window (400-500 ms), as well as an additional post N400 window (500-800 ms). They found a negativity in 300-500 ms at centroparietal electrode sites for the three literal condition. The negativity was graded such that it was more negative for the Related compared to the Correct conditions, and more negative for the Unrelated compared to the Related condition. On the other hand, in the Idiomatic condition, a significant increase in negativity was observed for the Related and Unrelated condition compared to the Correct condition, however, the difference between these conditions was not statistically significant. According to Rommers and colleagues (2013), the absence of a difference between these conditions in the idiomatic context indicates that the literal meaning of the words is not fully activated or, alternatively, the semantic unification process for those words is less engaged compared to the literal condition.

In order to directly compare the processing of literal and idiomatic language, Canal and colleagues (2017) embedded ambiguous Italian idioms (i.e., idioms with literally plausible meaning) into two contexts, idiomatic and literal. Additionally, a control condition was included which comprised of the last word of the expressions in a literal context. ERPs were time-locked to the first word of the expressions. Consistent with the prediction of the configuration hypothesis (see Cacciari & Tabossi, 1988), they argued that in the idiomatic context, the literal meaning of the words will be processed compositionally up until the point of recognition, when the reader identifies the idiom. After this point, the idiomatic meaning will be retrieved from semantic memory and integrated into the ongoing context. In the N400 (300-500 ms) time window for word 3 (the last word of the idiomatic expression; W3), greater positivity for the idiomatic condition than the literal condition at anterior electrode sites was observed. The researchers claim that the frontal distribution of this positivity does not align with the canonical N400 which led them to conclude that there was no N400 effect. Instead they interpreted the observed findings as a PNP. No effect of condition was found for the late time window. Unexpectedly, they found the same frontal early PNP for W1. Therefore, participants most likely started the reanalysis of the meaning of the idiomatic expression (the process that is claimed to be reflected by a PNP) immediately after the recognition of the idiom at W1. Overall, the researchers argued that the PNP effects found in this study can be accounted for in terms of revision mechanisms in the idiomatic context, indicating that idiom processing is more demanding than literal language processing, especially for idioms whose literal interpretation is available.

2.2.2 The Present Study

While the studies reviewed above give us important insights on the neurophysiological underpinnings of idiom comprehension, they typically focus on specific types of idioms, such as familiar idioms whose literal meaning is always available. These studies generally compare the processing of that subcategory of idioms with matched literal sentences.

In the present study, we investigated the neurophysiological underpinnings associated with idiom comprehension by zooming in on their multi-determined nature. As such, instead of comparing the processing of idioms with literal language, we asked how and to what extent the neurophysiology of idiom comprehension changes as a function of idiom's characteristics. In particular, two factors (familiarity and literal plausibility) that have been reliably shown to impact the processing of idioms in prior behavioral and eye-tracking studies (e.g., Titone & Libben, 2014) were considered. We selected idioms from a database (Libben & Titone, 2008), such that they varied with respect to familiarity and literal plausibility. The supporting context of each idiom was also manipulated, in order to examine the impact of environment on idiom processing. Among ERP components, we particularly focused on the N400 component since its amplitude has been interpreted as an index of ease or difficulty of integration of the meaning of a word into the general context (Kutas & Hillyard, 1980). Thus, if these particular aspects of idioms actually modulate their processing, as previous behavioral studies suggest, we would expect that these processing differences should be captured by quantitative and/or qualitative differences indexed by the respective N400 amplitude and its topographical distribution.

2.2 Methods

2.2.1 Participants

Thirty-two students (mean age = 19.8 years, range = 18-21, 28 women) were recruited through the Integrated System of Participant Research (ISPR) , and received partial course credit for their participation. Prior to recruiting the participants, this study was reviewed and approved by the ethics board. All participants were right-handed, native speakers of English with normal or corrected-to-normal vision and no history of language or neurological disorders. The administered Language Background Questionnaire (Sabourin, Leclerc, Lapierre, Burkholder & Brien, 2016) revealed that most participants knew a language other than English, but did not start learning this second language before the age of six. Data from nine participants were lost or excluded from analysis due to technical difficulties at the testing time (n=2), excessive blinking (n=2), or noisy data which led to the exclusion of over 75% of data in one or more experimental conditions during artifact rejection (n=5).

2.2.2 Materials

Idiomatic expressions were taken from a database (Libben & Titone, 2008) that provides the normative characteristics of 219 idioms. These idiomatic expressions are structurally similar: they are all constructed in the form of "She/He/It $verb_{past\ tense}$ x noun" (e.g., "It slipped his mind"). Among these, 40 idioms which varied in their level of familiarity and literal plausibility were selected: High Familiarity (HF, e.g., "hit the sack"), Low Familiarity (LF, e.g., "chew the cud"), High Literally Plausible (HL, e.g., "kick the bucket"), and Low Literally Plausible (LL, e.g., "rule the roost"), resulting in 10 per condition. To examine the impact of environment, the supporting context of each idiom was manipulated. Idioms were embedded in a sentence context, and the supporting context was present before (Context-before) or after (Context-after) the

idiomatic expression. In all cases, the critical word was the final noun of the idiom (see Table 2.1 for examples).

Table 2. 1 *Examples of stimuli. ERPs were time-locked to the underlined words. HF: High Familiar; LF: Low Familiar; HL: High Literally plausible; LL: Low Literally Plausible*

Condition	Sentence
HF-Context-before	Mary was so tired, so as soon as she got home, she *hit the* ***sack*** and slept until the next day.
HF-Context-after	Mary *hit the* ***sack*** as soon as she got home, because she was so tired.
LF-Context-before	John wasn't sure about the job offer, so he *chewed the* ***cud*** before deciding.
LF-Context-after	john *chewed the* ***cud*** before deciding to accept the job offer because he wasn't sure about it.
HL-Context-before	After being sick for months, John *kicked the* ***bucket*** yesterday.
HL-Context-after	John *kicked the* ***bucket*** yesterday, after being sick for months.
LL-Context-before	As the oldest of her siblings, Mary *ruled the* ***roost*** when her parents weren't home.
LL-Context-after	Mary *ruled the* ***roost*** as the oldest of her siblings and bossed them when their parents weren't home.

Additionally, 40 control sentences were also created. For idioms with high level of Literal Plausibility (LP), the same expression was used in a sentence context that promoted its literal meaning (e.g., idiom: "kick the bucket", control: "John was frustrated with mopping and kicked the *bucket* across the hallway".). For Low LP idioms whose literal meaning is not available, the

last word of the idiomatic expression was used in a sentence context (e.g., idiom: “rule the roost”, control: “The farmer has a problem with wolves, so the chickens return to the *roost* at night.”).

Twenty filler sentences, matched for length, were included. Each participant saw 100 sentences (40 Idiomatic, 40 Control, 20 Filler), but each sentence only appeared as Context-before or Context-after. Cloze probability values were calculated via an online norming task (using Google online forms). The task consisted of two lists of the same sentences from the experimental lists which were presented up to the critical word. Twenty native English speakers were asked to read the sentences and fill-in-the-blank with the first word that came to mind. Cloze probability values for list 1 were .33, SD = .36 and .35, SD = .40, for idiomatic and literal sentences, respectively. Cloze probability values for list 2 were .37, SD = .29 and .31, SD = .35, respectively.

2.2.3 Procedure

Prior to the experiment, each participant completed a Language Background Questionnaire (Sabourin et al., 2016), followed by a handedness questionnaire. After being equipped with the ERP cap, they were seated in a soundproof, electrically shielded room approximately one meter from a computer screen. They were verbally instructed to minimize blinks and facial movements. Participants read sentences via the Rapid Serial Visual Presentation (RSVP) paradigm, during which each word was presented at the center of the screen in white letters on a black background. Each trial began with a fixation cross (+) which remained on the screen until the participant was ready to start reading the sentence by pressing any key on the keyboard. Each word remained on the screen for 300 ms with a 300ms black screen following each word. One-third of the sentences were followed by a yes/no comprehension question, to ensure participants full

attention. Prior to the experimental stimuli, participants completed 4 practice trials. The experimental portion of the study lasted approximately 30 minutes.

2.2.3.1 EEG Recording

The online EEG was recorded from 62 electrodes using a 64-channel Quick-Cap by Compumedics. To capture blinks and horizontal eye movements, four electrodes were placed around the eyes on the participants' face. Two additional electrodes were placed on the left and right mastoids (M1 and M2), used for off-line re-referencing of data. All impedances were kept bellow 10 KΩ. EEG recordings were amplified through SynAmps with a low pass filter of 200 Hz (no high pass filter), digitized on-line with a sampling frequency of 1000 Hz. The data was collected and stored by Neuroscan for off-line analysis.

2.2.3.2 ERP Analysis

ERP data were referenced off-line to the average of the left and right mastoids. A notch filter of 60 Hz and a low pass filter of 0.1 Hz at zero phase was applied to the continuous data. The continuous EEG was segmented into epochs of 1100 ms (100 ms pre-critical-word onset until 1000 ms post-critical-word onset. An average baseline of 100 ms pre stimulus onset was subtracted. Trials were examined for artifacts, including blinks and excessive muscle activity. If the overall rejection of trials exceeded a researcher-defined threshold of 25%, data from the participant was excluded from the analysis (n=5). An average of 1.31% of data across remaining participants was rejected. These artifacts were comparable across conditions (Context-before: 1.09%, Context-after: 1.74%, HF: 1.53%, LF: 1.31%, HL: 1.09%, LL: 1.09%). EEG signals, across all conditions, were time-locked to the noun in each Idiomatic expression or the equivalent noun used in Control conditions (the underlined words in Table 2.1). Following

Rommers and colleagues (2013), the data was extracted within three time-windows (Early N400: 300-400 ms; Late N400: 400-500 ms; Post N400: 500-800 ms).

Twenty-eight electrodes were included in the analysis (see Figure 2.1). For each time window, mean amplitude was averaged over quadrants (Left-Anterior (LeftAnt): FC1, FC3, FC5, F1, F3, F5, AF3; Right-Anterior (RightAnt): FC2, FC4, FC6, F2, F4, F6, AF4; Left-Posterior (LeftPost): CP1, CP3, CP5, P1, P3, P5, PO3 ; Right-Posterior (RightPost): CP2, CP4, CP6, P2, P4, P6, PO4).

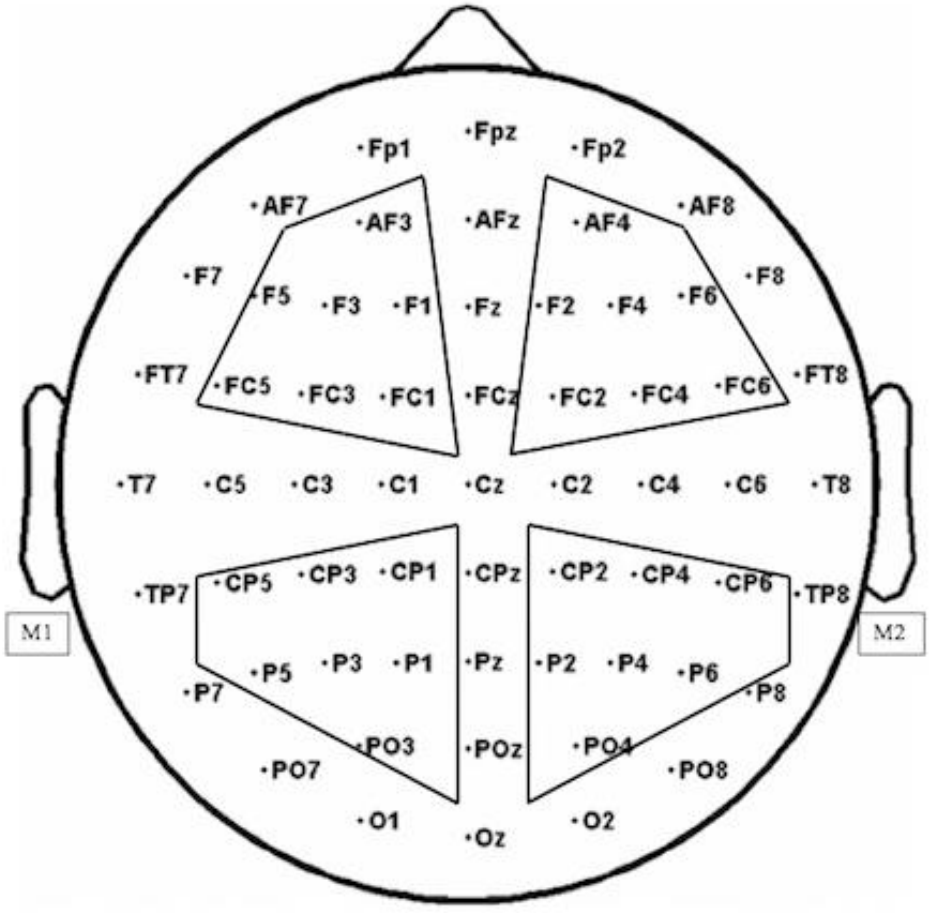


Figure 2.1 Electrode layout. Lines indicate quadrants used for mean amplitude analyses on the ERPs.

Separate within-subject repeated measure ANOVAs were conducted for each time window, consisting of three comparisons. First, the effect of Context was examined for Idiomatic condition. Since this is the only condition for which we could compare the processing of idioms with literal (Control) sentences, the effect of Context was also examined for literal sentences. The data, for each condition, were subjected to 2×2×2 repeated measure ANOVAs with factors Hemisphere (Left, Right), Anteriority (Anterior, Posterior), and Context (Before, After). To examine the effect of Familiarity and Literal Plausibility on the processing of idioms, 2×2×2 repeated measure ANOVAs were conducted on Idiomatic conditions with factors Hemisphere (Left, Right), Anteriority (Anterior, Posterior), and Familiarity (High Familiar: HF, Low Familiar: LF) for the effect of Familiarity; and Hemisphere (Left, Right), Anteriority (Anterior, Posterior), and Literal Plausibility (High Literal: HL, Low Literal: LL) for the effect of Literal Plausibility.

2.3 Results

Prior to ERP analysis, participant's responses to comprehension questions were investigated to ensure that they achieved an accuracy of at least 80%. The average accuracy was 90.31% with no participant falling outside 2 SDV (.03) from the mean.

2.3.1 Effect of Context

The main effect of Context was not statistically significant across all time windows for the Idiomatic condition (Early N400: $F(1, 22) = .01$, $p = .89$; Late N400: $F(1, 22) = 1.10$, $p = .30$; Post N400: $F(1, 22) = 3.12$, $p = 0.09$). Similarly, no significant interaction was observed between Context and other factors (Hemisphere and Anteriority). For Control sentences, the main effect of Context was not statistically significant in any time window (Early N400: F(1, 22)

= .32, P = .57; Late N400: $F(1, 22) = 1.39$, $P = .25$; Post N400: $F(1, 22) = .11$, $P = .74$), nor was any interaction between Context and other factors.

2.3.2 Effect of Familiarity

Low Familiar idioms elicited larger negativities relative to High Familiar idioms, lasting from 300 to 500 ms (Figure 2.2). This main effect of Familiarity was significant in the Early N400 time window ($F(1, 22) = 15.46$, $p < .001$). No interactions between Familiarity and additional factors were observed. While the same main effect of Familiarity continued to be significant in the Late N400 time window ($F(1, 22) = 13.64$, $p < .001$), there was also a significant interaction between Familiarity and Hemisphere ($F(1,22) = 5.00$, $p = .03$). While the amplitude of the signal for High and Low Familiar idioms is similar over the left hemisphere (High Familiar < Left: -1.90 μV>; Low Familiar <Left: -1.94 μV>), the effect over the right hemisphere show a greater negativity for Low Familiar idioms compared to High Familiar idioms (High Familiar < Right: -.42 μV>; Low Familiar < Right: -2.67>).

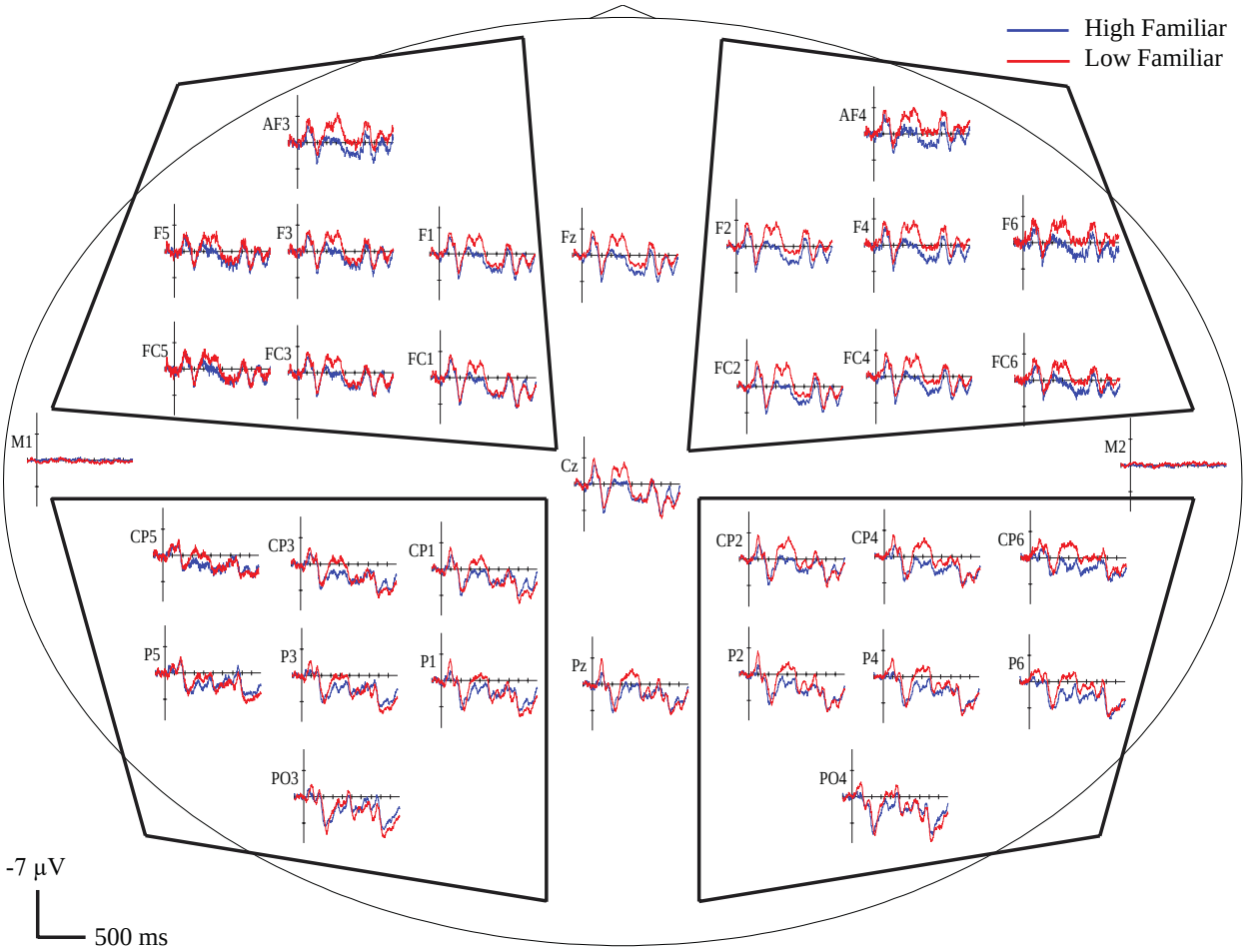


Figure 2.2 Grand-averaged (n = 23) ERPs for High Familiar (blue) and Low Familiar (red) idioms.

In the PNP time window, the main effect of Familiarity was not significant ($F(1, 22) = .06$, $p = .79$) nor was any interaction between Familiarity and additional factors.

2.3.3 Effect of Literal Plausibility

Similar to Familiarity, idioms with Low Literal Plausibility (LP) elicited a greater negativity relative to High Literally Plausible idioms. However, as can be seen in Figure 2.3, this negativity began later. The main effect of LP was not significant in the Early N400 time-window ($F(1, 22) = .31$, $p = .57$) and it started later at the Late N400 time-window ($F(1, 22) = 10.31$, $p = .004$).

Within this same time window, a marginally significant interaction between LP and Hemisphere was found ($F(1, 22) = 4.26$, $p = .05$).

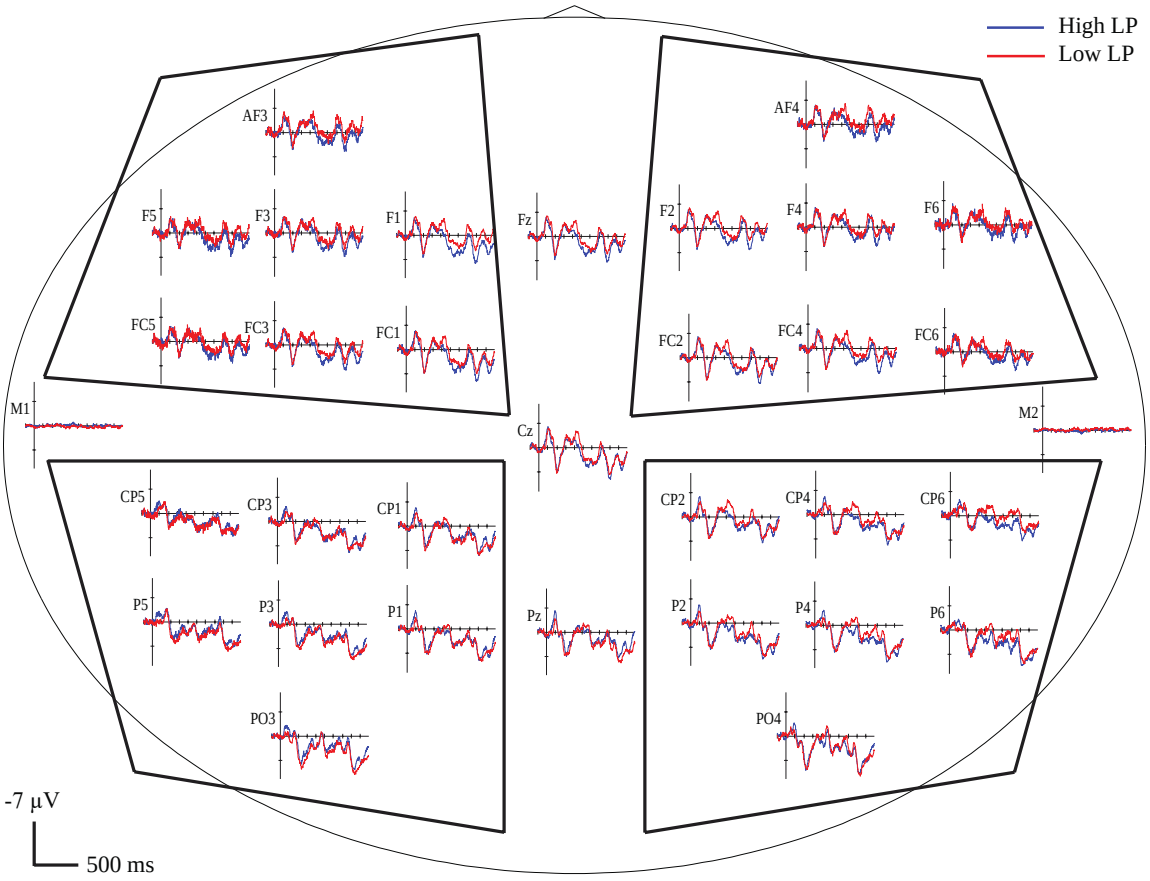


Figure 2.3 Grand-averaged (n = 23) ERPs for High LP (blue) and Low LP (red) idioms.

For the Post N400 time window, the main effect of LP was not significant ($F(1, 22) = .77$, $p = .38$) nor was any interaction between LP and differing factors.

2.4 Discussion and Conclusion

According to a multi-determined model, idiom processing is affected by different linguistic attributes (e.g., familiarity and literal plausibility), which can assert their roles at different stages of idiom processing (Titone & Libben, 2014). In this study, we investigated whether the multi-determined nature of idiom processing can be captured by the electrophysiological manifestations of the semantic processing of their constituents. This was done by considering

idioms with various levels of familiarity and literal plausibility (i.e., the variables that define a subclass of idioms), and comparing the electrophysiological underpinnings associated to the processing of those idioms. We selected structurally similar idioms and compared the obtained ERPs on the target word.

The result of the current study showed that the level of familiarity with idiomatic expressions results in a processing advantage indexed by a significant reduction in signal amplitude for highly familiar idioms compared to low familiar idioms, comparable with previous behavioural and eye-tracking studies (Libben & Titone, 2008; Titone et al., 2019). This effect started as early as 300 ms post target word onset. Regarding the effect of literal plausibility, literature supports two conclusions. While some research reports that the existence of a literally plausible meaning results in a processing disadvantage for these idioms (e.g., Titone & Libben, 2014), other research suggests a facilitative effect of literal plausibility (e.g., Mueller & Gibbs, 1987). The findings of the current study align with prior research suggesting a processing advantage for idioms whose literal meaning is available. This effect was indexed by a lower signal amplitude for idioms with a high level of literal plausibility. Interestingly, the impact of literal plausibility began at a later point, around 400 ms after post word onset, supporting the idea that different attributes asserts their roles at varying stages of processing.

2.4.1 Role of the Right Hemisphere in Figurative Meaning Comprehension

The assumption that the Right Hemisphere (RH) has a special role in the comprehension of non-literal language, (e.g., jokes, sarcasm, and figurative language), began with the observations of RH brain damaged patients who showed difficulty understanding non-literal language (Coulson, 2012). Despite these observations, the special role of the RH in understanding particular aspects of figurative language, (e.g., metaphor comprehension), in healthy populations is mixed (Kacinik

& Chiarello, 2007). For example, some argue that metaphoric meaning is initially activated in both left and right hemispheres, but this activation maintains only in the RH while it decays in the left hemisphere (Anaki, Faust, & Kravets, 1998). Others argue against the privileged role of the RH in metaphor comprehension (e.g., Coulson & Williams, 2005). Importantly, even studies providing evidence against the distinctive role of the RH in figurative language processing suggest that different types of metaphors, based on their novelty and complexity, may recruit different brain areas.

The result of the current study supports the greater involvement of the RH. However, this involvement depends on the type of idiomatic language. It seems that, when an idiom is unfamiliar, the resulting semantic complexity triggers the recruitment of the areas of the brain, (i.e., right hemisphere), that are responsible for comprehending novel and non-salient aspect of language (Coulson & Van Petten, 2007). Despite evidence from fMRI studies which shows greater involvement of the RH in the processing of ambiguous idioms, (i.e., idioms with literally plausible meaning; e.g., Zempleni et al., 2007), we found higher RH activation for idioms with a lower level of literal plausibility. As it has previously proposed, one possibility is that increased literal plausibility may promote the activation of a literal meaning for idioms (Titone & Libben, 2014). An alternative assumption is that literally implausible idioms, just like low familiar idioms, require brain resources that are responsible for novel meaning creation. This consequently results in higher activation of the RH for such idioms.

While this study, to our knowledge, is the first to investigate the neurophysiology of idiomatic expressions with respect to their complex nature, the factors that have not been considered in the current study require further investigation. First, as the multi-determined model of idiom processing proposes, there are other factors, such as decomposability (the extent to

which the meaning of idiom's constituents contribute to the final figurative meaning) that assert their role at later stages (when the intended meaning of an idiom is known) of figurative comprehension. As such, it is important to consider the effect of all possible linguistic factors in the future studies. Additionally, considering these linguistic dimensions as continuous rather than categorical variables could reveal interesting results. Moreover, in future studies, the neurophysiology of idiomatic expressions can be tracked for all of the idiom's constituents. This will consequently reveal more detailed information about the exact timing when each factor comes into play. Finally, while we did not find any significant effect of context, we anticipate that the effect of this factor might appear once idioms are embedded in a more informative sentence context.

To conclude, in the current study we showed that consistent with a multi-determined model, during idiom processing, multiple linguistic variables come into play, which assert their roles at different stages of processing. Any proposals which attempt to explain the nature of idiom processing without taking into account the role of these factors may result false generalizations about the processing of this complex form of language.

References

Anaki, D., Faust, M., & Kravetz, S. (1998). Cerebral hemispheric asymmetries in processing lexical metaphors. *Neuropsychologia*, *36*(4), 353-362.

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

Cacciari, C., Padovani, R., & Corradini, P. (2007). Exploring the relationship between individuals' speed of processing and their comprehension of spoken idioms. *The European Journal of Cognitive Psychology*, *19*, 417–445.

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

Cacciari, C., & Tabossi, P. (2014). *Idioms: Processing, structure, and interpretation*. Hoboken: Taylor and Francis.

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.

Coulson, S. (2012). Cognitive Neuroscience of Figurative Language. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics* (Cambridge Handbooks in Psychology, pp. 523-538). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139029377.027

Coulson, S., & Van Petten, C. (2007). A special role for the right hemisphere in metaphor comprehension?: ERP evidence from hemifield presentation. *Brain research*, *1146*, 128-145.

Coulson, S., & Williams, R. F. (2005). Hemispheric asymmetries and joke comprehension. *Neuropsychologia*, *43*(1), 128-141.

Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of memory and language*, *28*(5), 576.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in cognitive sciences*, *9*(9), 416-423.

Hamblin, J. L., & Gibbs, R. W. (1999). Why you can't kick the bucket as you slowly die: Verbs in idiom comprehension. *Journal of Psycholinguistic research*, *28*(1), 25-39.

Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E-J. vander Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp. 133–165). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kaan, E. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, *1*(6), 571-591.

Kacinik, N. A., & Chiarello, C. (2007). Understanding metaphors: Is the right hemisphere uniquely involved? *Brain and Language*, *100*(2), 188-207.

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

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.

Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research*, *1068*(1), 151-160.

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

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

Nasrallah, M., Carmel, D., & Lavie, N. (2009). Murder, she wrote: enhanced sensitivity to negative word valence. *Emotion, 9*(5), 609.

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

Sabourin, L., Leclerc, J. C., Lapierre, M., Burkholder, M., & Brien, C. (2016). The language background questionnaire in L2 research: Teasing apart the variables. In Proceedings of the *Annual Meeting of the Canadian Linguistics Association, Calgary, Canada. Available online.*

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of verbal learning and verbal behavior, 18*(5), 523-534.

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

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

Titone, D., & Libben, M. (2014). Time-dependent effects of decomposability, familiarity and literal plausibility on idiom priming: A cross-modal priming investigation. *The Mental Lexicon, 9*(3), 473-496.

Titone, D., Lovseth, K., Kasparian, K., & Tiv, M. (2019). Are figurative interpretations of idioms directly retrieved, compositionally built, or both? Evidence from eye movement measures of

reading. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 73*(4), 216.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190.

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

Zempleni, M. Z., Haverkort, M., Renken, R., & Stowe, L. A. (2007). Evidence for bilateral involvement in idiom comprehension: An fMRI study. *Neuroimage, 34*(3), 1280-1291.

3 Affective and Sensory-motor Norms for Idioms by L1 and L2 English Speakers[6]

Abstract

In the present study, we developed affective (valence and arousal) and sensory-motor (concreteness and imageability) norms for 210 English idioms rated by native English speakers (L1) and English second language speakers (L2). Based on internal consistency analyses, the ratings were found to be highly reliable. Furthermore, we explored various relations within the collected measures (valence, arousal, concreteness, and imageability) and between these measures and some available psycholinguistic norms (familiarity, literal plausibility, and decomposability) for the same set of idioms. The primary findings were: i) valence and arousal showed the typical U-shape relation, for both L1 and L2 data; ii) idioms with more negative valence were rated as more arousing; iii) the majority of idioms were rated as either positive or negative with only 4 being rated as neutral; iv) familiarity correlated positively with valence and arousal; v) concreteness and imageability showed a strong positive correlation; and vi) the ratings of L1- and L2-speakers significantly differed for arousal and concreteness, but not for valence and imageability. We discuss our interpretation of these observations with reference to the literature on figurative language processing (both single words and idioms).

Keywords: Idioms, valence, arousal, concreteness, imageability

---

[6] Morid, M., & Sabourin, L. (2023). Affective and Sensory-motor Norms for Idioms by L1 and L2 English speakers. *Applied Psycholinguistics* (submitted).

3.1 Introduction

In 1970, Fraser suggested that idioms are not homogenous when it comes to their transformability. Some idioms are more flexible and can easily undergo transformations (e.g., "*the towel was thrown in by him*" which is the passivized form of "*throw in the towel*" retains the idiomatic meaning "he gave up"), while others are inflexible (e.g., "*the bucket was kicked by him*" is less likely to be interpreted as "he died") (Gibbs & Gonzales, 1985). However, the diversity of idioms is not limited to their structural flexibility, indeed, idioms may vary along a large number of linguistic dimensions (e.g., Libben & Titone, 2008). For example, some idioms are more familiar to language users than others (e.g., "*is a piece of cake*" vs. "*go pear-shaped*"). Some idioms have literal interpretations but many lack this quality (e.g., "*cross one's fingers*" vs. "*be the apple of someone's eye*"). Furthermore, the meaning of some idioms can be extracted from the constituents that compose them, while others retain low decomposability (e.g., "*cover one's track*" vs. "*go cold turkey*"). Multiple studies in the field of psycholinguistics and neurolinguistics have indicated that these dimensions generally affect the processing of idioms (e.g., Cieślicka, & Heredia, 2011; Titone et al., 2019; Carrol & Conklin, 2020; Morid, Bachar, & Sabourin, 2021).

More recently, researchers have put forward the idea that idioms additionally vary along dimensions of emotional content. In general, it has been well-stablished that the emotional content of stimuli modulates their processing both at the word-level and the sentence-level (e.g., Citron, 2012; Arfé, Delatorre, & Mason, 2022). Emotional valence (the extent to which an emotion is pleasant/positive or unpleasant/negative) and arousal (whether the evoked emotion is perceived as exciting or calming) are considered two main dimensions that define emotions (Russell, 2003). Though researchers have access to databases which specify affective measures

(including emotional valence) for single words in various languages (e.g., Yao et al., 2017; Warriner et al., 2013; Stadthagen-Gonzalez et al., 2017; Imbir, 2016), idiom databases are remarkably scarce (Citron et al., 2016). The impact of emotional content on idiom processing is compelling given that these expressions are typically used in emotionally-charged conversations (Citron et al., 2016). As such, the primary aim of this study is to create the first known set of affective norms for English idioms.

Previous studies suggest that concreteness (the degree to which the meaning of a word or expression is understood through perception and action) is a crucial variable in the processing of emotional words (Barber et al., 2013). In particular, studies suggest that abstract words are more likely to refer to emotional states than concrete words (Altarriba, Bauer, & Benvenuto, 1999). Some researchers have proposed that, beside linguistic information, two major sources of experiential information (sensory-motor and affective) are involved in the process of word learning and representation (Vigliocco, Meteyard, Andrew & Kousta, 2009). These authors also argued that the representation of concrete words rely on sensory-motor information, while emotional information plays a crucial role in the representation and processing of abstract words (see Kousta et al., 2011 and Vigliocco et al., 2014 for behavioural and fMRI support, respectively). Importantly, it seems that the so-called 'concreteness effect' (i.e., the fact that participants process concrete words more rapidly than abstract words; Holcomb et al., 1999) can be replaced by an 'abstractness effect' (i.e., abstract words are processed more rapidly than concrete words), once the researchers control for a large number of lexico-semantic variables (including familiarity and imageability). Given the relation between affective factors, concreteness and imageability, a database containing norms for all these dimensions would be indispensable for future research. Thus, the second aim of the current study is to additionally

include concreteness and imageability ratings for the same set of English idioms, allowing researchers to control for these factors.

As mentioned above, valence and arousal are two main dimensions that define the structure of affect. The relation between valence and arousal is typically reported as quadratic (Citron, Weekes, & Ferstl, 2011; Citron et al., 2016) with both negative and positive words being more arousing than neutral words. Moreover, literature on single words suggest that words that are highly valenced and arousing tend to be more abstract (Vigliocco et al., 2014). Other studies report a positive correlation between arousal and imageability (Citron et al., 2014), or a negative quadratic correlation between arousal and concreteness (Montefinese, et al., 2014; Citron et al., 2016). As such, beyond simply providing a database of ratings for idioms, the third goal of this study is to explore the relation between these collected measures, as motivated by previous research on single words and non-literal expressions[7].

Lastly, given that idiomatic expressions are pervasive in everyday conversations (Citron et al., 2016), psycholinguistic researchers have increasinigly compared first language (L1) and second language (L2) idiom processing. Research shows that L2 learners usually encounter difficulty when learning and comprehending idioms (Abel, 2003; Titone et al., 2015). Though L2 idiom processing appears to bear similarities to L1 idiom processing (Heredia & Cieślicka, 2015), psycholinguistic and affective ratings by L1 and L2 speakers may vary. For example, compared to native speakers, English-learners rated English idioms as more decomposable

[7] The relation between linguistic variables with affective and sensory-motor variables was only explored for our L1 group, because the linguistic variables are only available for English native speakers.

(Abel, 2003). The final goal of this study is thus to provide both L1 and L2 affective and sensory-motor norms for the same English idioms. This will allow second language researchers to account for these factors in future idiom-processing studies.

3.2 Methods

3.2.1 Participants

A total of 555 students from [masked] (318 women, 216 men, 21 unspecified gender), between 18 and 23 years of age (*Mean* = 19.48) completed the online survey. Participants were recruited through the university's Integrated System of Participant Research (ISPR). They received partial course credit as compensation for their participation. Fifteen participants were excluded from the final analysis (see Data analysis section). The final sample thus contained 540 participants (314 women, 208, men, 18 unspecified gender; *Mean* age = 20.1, *range* = 18.4 – 23.7).

Participants completed an extensive Language Background Questionnaire (LBQ; Sabourin et al., 2016). Based on the information obtained from this LBQ, participants were divided into two groups (see Table 1): native (L1) English speakers (*n*=300), and second language (L2) English speakers (*n*=240). L1 speakers were participants who self-reported that English was their first language (over 90% exposure during infancy) and their current most dominant language. L2 speakers were participants who self-reported an alternative language as an L1 (less than 50% exposure to English during infancy) and indicated that English was currently their second-most dominant language. Note that both groups were highly proficient in English (see Table 3.1). Among L2 participants, 38% identified French as their L1; The rest indicated twenty one other languages as their L1 (for example, Arabic, Korean, Vietnamese, etc.). In cases where participants provided potentially inconsistent self-reports (for example they

indicated that English was their first language but also said they had low English proficiency), the participant's data was not included in the analysis.

Table 3.1 *Detailed information from Language Background Questionnaire*.

Information from LBQ	Inclusion criteria (L1)	Inclusion criteria (L2)
Native Language	English	Any language other than English
Second learned language (L2)	Any language	English
Self-rated English Proficiency	Mean = 4.98 [4.86-5]	Mean = 4.68 [4.53-4.71]
Age of Acquisition of English	From birth	Mean = 11.7 [9.4- 14.7]
Percentage of English usage during childhood	Mean = 98.3% [93.2- 100]	Mean = 20.2% [5- 30]

3.2.2 Materials

The experimental materials consisted of 210 idiomatic expressions from Libben and Titone (2008). The idioms from this database had previously been normed for the psycholinguistic dimensions of interest: familiarity, literal plausibility, and decomposability (defined in Table 3.2, below). As such, we could compare the pre-existing psycholinguistic dimensions with our novel affective and concreteness ratings. All idioms possessed the form of "She/He/It $verb_{past\ tense}$ x noun", where x was a preposition, an article, or a determiner (e.g., "*It slipped his mind*", "*She raised the devil*", "*He got a toehold*"). This uniformity ensured that length and phrasal complexity were well controlled (Libben & Titone, 2008). The full list of experimental materials can be found on the Open Science Framework repository at ASN Idioms.

Table 3.2 *Linguistic dimensions for idioms previously collected by Libben and Titone (2008) and considered in the present study*.

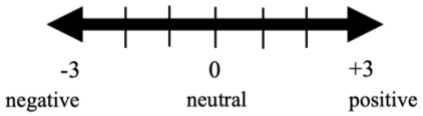
Measure	Definition
Familiarity	"Subjective frequency with which comprehenders encounter an idiom in its written or spoken form, regardless of their familiarity with the actual meaning of the phrase" (Libben & Titone, 2008, p. 1106).
Literal plausibility	"Refers to an idiom's potential for a literal interpretation. For example, some idioms, such as "*bite the bullet*", have a well-formed literal meaning, whereas other idioms, such as "*break her word*", only have a meaningful idiomatic interpretation" (Libben & Titone, 2008, p. 1106).
Decomposability	"Refers to how an idiom's words make independent semantic contributions to the overall figurative meaning" (Libben & Titone, 2008, p. 1106).

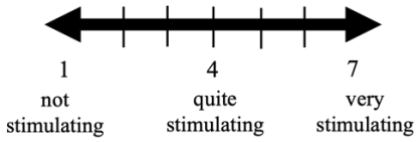
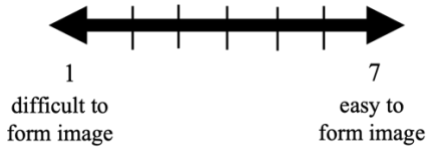
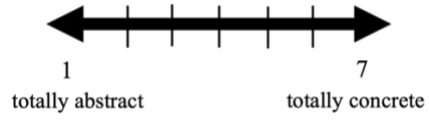
2.2.3 Procedure

We used Gorilla Experiment Builder (www.gorilla.sc) to create and host our online study (Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed, 2020). Participants first gave informed consent before being directed to the LBQ. To ensure that participants started the task with common knowledge about what an idiomatic expression is, they were instructed to read a short description about idioms, accompanied by some examples. Moreover, they were asked to rate the expressions based on the idiomatic meaning and not their literal meaning (e.g., "kick the bucket" should be rated based on the meaning "to die"). Participants were randomly assigned to one of four lists, corresponding to one of the four dimensions of ratings. That is, each participant was only asked to rate the full set of items (all 210 English idioms) according to a single measure (valence, arousal, imageability, or concreteness; see Table 5 for each list *n*).

The specific instruction for each measure was adapted from previous studies (Citron et al., 2016; Altarriba et al., 1999; Yao et al., 2017). The instructions were modified and accompanied by proper English examples to assure clarity. Note that ten native English speakers read and assessed the instructions for clarity before participants were recruited. Each instruction contained three main parts: a definition of the measure along with some examples, an explanation of the scale from which the participants were supposed to rate each idiom, and an explanation of the labels for each scale. These are briefly presented in Table 3.3. A screenshot of all instruction pages and one example of the questionnaire page may be found on the Open Science Framework repository at ASN Idioms.

Table 3.3 *Novel affective and sensory-motor measures for idioms collected in the present study*.

Measure	Definition	Instruction
Valence	Valence describes the extent to which an event is positive or negative. For example, the sentence "*The whole experience left a bad taste in my mouth*" describes a negative event, whereas the sentence "*She was feeling on top of the world after winning the tennis tournament*" describes something positive.	On a scale from -3 to + 3, rate how positive or negative the metaphorical meaning of the following sentences is. 
Arousal	Emotional Arousal describes to what extent an event is stimulating. For example, the sentence "*he screamed because of pain*" is perceived as very stimulating, while	Please rate the sentence for emotional arousal on a scale from 1 to 7 (consider metaphoric meaning).

	sentence "*she read a book*" might be perceived as not stimulating.	
Imageability	Imageability describes how difficult or easy it is to form an image of a word or an expression. For example, you might find the word "*flag*" a highly imageable word because it is easy to form an image of the word flag in your mind. The word "*charity*" on the other hand, might be rated low since it is difficult to form an image for this word.	Please rate the sentence for imageability on a scale from 1 to 7 (consider metaphoric meaning). 
Concreteness	A sentence conveys a concrete meaning when it denotes an experience that refers to one or more sensory modalities (e.g., touch, vision). For instance, the meaning of "*the blanket is so soft*" refers to the experience of touch. In contrast, the meaning of "*she had a sad thought*" is abstract since it cannot be associated with a sensory modality.	Please rate the extent to which the metaphorical meaning of the sentence refers to a state or event that can be experienced with the senses. 

3.3 Data Analysis

3.3.1 Data Pre-processing

Participants' data were removed from the analyses based on three exclusion criteria. First, if a participant responded "I do not know the meaning of this idiom" for over 50% of expressions, we did not include their data in the analysis (n= 7). Second, when a participant attributed the

same rating to over 85% of idioms or when the responses followed any noticeable patterns (e.g., the first 20 responses were the same and the second 25 responses were the same, etc.), we excluded them from the data analysis (*n*= 3). Finally, if the participant's responses on the LBQ did not allow us to confidently group their data into the L1 or the L2 group, we excluded their data from analysis (*n*= 5). Based on these criteria, 15 participants (2.7 %) were excluded from the data analysis.

Data analysis

The purpose of the current study was i) to explore the relation between affective variables; ii) to explore the relation among non-affective variables; and iii) to examine the relation between affective and sensory-motor variables with the psycholinguistic variables obtained in previous studies (Libben & Titone, 2008). We calculated Pearson partial correlations to explore these relations. Additionally, since previous studies consistently reported a quadratic relation between various affective measures (Citron, Weekes, & Ferstl, 2014; Ferré, Guasch, Moldovan, & Sánchez-Casas, 2012; Montefinese et al., 2014; Warriner, Kuperman, & Brysbaert, 2013), we conducted a quadratic regression predicting i) arousal from valence; and ii) familiarity from valence.

The last aim of the current study was to compare the ratings from L1 and L2 speakers. We conducted t-tests to statistically compare these groups. Note that, for both correlation and regression analyses, we compared pre-existing psycholinguistic dimensions with our novel affective and concreteness ratings for our L1 speakers only; this is because the previous literature has not collected the psycholinguistic variable ratings from second language informants.

3.4 Results and Discussion

3.4.1 Descriptive Statistics

The descriptive statistics for the ratings of L1 and L2 speakers are presented in Table 3.4. The final column ("Valid response %") indicates the proportion of obtained ratings for each variable; when participants indicate that an idiom is unknown and thus do not attribute it a rating, this response is not counted towards the "Valid response %". The full list of experimental materials can be found on the Open Science Framework repository at ASN Idioms.

Table 3.4 *Descriptive statistics for each variable and for L1 and L2 speakers.*

Group	Variable	Mean	SD	Median	Minimum	Maximum	Mean Valid Response %
L1	Arousal	2.80	0.55	2.77	1.40	4.28	89.14
	Valence	0.04	1.44	0.00	-2.59	2.72	84.74
	Concreteness	3.90	0.80	3.78	2.43	6.32	88.11
	Imageability	4.81	0.86	4.90	2.14	6.69	93.94
L2	Arousal	2.16	1.05	1.97	0.11	4.80	75.9
	Valence	0.01	1.42	0.22	-2.74	2.89	87.94
	Concreteness	4.68	0.86	4.67	2.50	7.00	86.69
	Imageability	4.76	1.19	4.82	2.00	7.00	84.75

In order to inspect whether the idiom's ratings are correlated with the ratings of their constituent (i.e., the noun and verb of each idiom; for example "*beat*" and "*breast*" in idiom "*he beat his breast*"), we calculated partial correlation between idiom's ratings and constituent's ratings. The constituent ratings (valence, arousal, concreteness, and imageability) were extracted

from available databases (Brysbaert & Warriner, 2014; Warriner, Kuperman, & Brysbaert, 2013). We did not find any significant correlation between idiom-related ratings with the respective ratings of the constituents.

Reliability measures

To assess the reliability of the ratings of the four variables that were included in the database, we calculated the intraclass correlation via Cronbach's alpha values (which is based on internal consistency) for each group (L1 and L2). The literature suggest that this analysis represents a more reliable measure than the split-half procedure (Citron et al., 2016). As such, individual participants were inputted as different variables, with idioms representing cases. The analysis showed high reliability for all variables in L1 group. Generally, the reliability values for L2 group were lower, however, they were all above acceptable level ($\alpha > 0.7$). The reliability measures for all variables and for L1 and L2 groups are reported in Table 3.5.

Table 3.5 *Measures of internal consistency (Cronbach's α) for variables whose rating was obtained in the current study by L1 and L2 group.*

Group	Variables	Cronbach's α	αs If Each Item Deleted	Participant No.
L1	Arousal	0.86	all > 0.85	72
	Valence	0.99	all > 0.97	77
	Concreteness	0.86	all > 0.85	77
	Imageability	0.89	all > 0.88	74
L2	Arousal	0.75	all > 0.73	57
	Valence	0.88	all > 0.85	63
	Concreteness	0.75	all > 0.72	60
	Imageability	0.81	all > 0.78	60

3.4.2 Relations Between Affective Variables

Of 210 idioms, 103 idioms were rated as positive, 103 as negative, and 4 idioms as neutral by the L1 group. The L2 group rated 122 idioms as positive and 88 idioms as negative. The numbers of positive and negative idioms are at odds with the previous studies in which most idioms were rated as negative (Citron et al., 2016). A t-test revealed that L1ers rated negative idioms as significantly more arousing than positive idioms, ($t(197) = -2.76$, $p < 0.01$). However, L2ers rated negative and positive idioms as similarly arousing (t (156) = 0.43, p = 0.43).

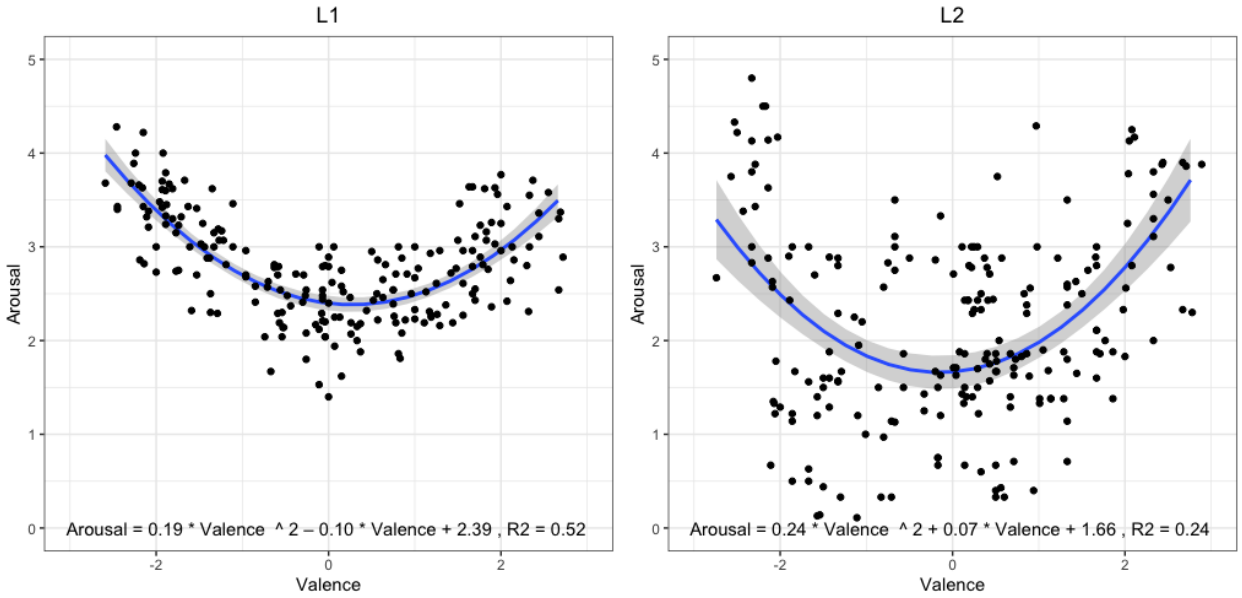


Figure 3.1 Valence ratings (-3 = very negative, 0 = neutral, +3 = very positive) plotted against arousal ratings (1 = not at all arousing, 7 = highly arousing) by L1 (left) and L2 (right) speakers.

Valence and arousal are plotted against each other in Figure 3.1. A visual inspection of these plots suggests that the relation between these variable (for both L1ers and L2ers) may be quadratic. To verify this impression, we conducted regression analyses wherein mean arousal was the dependent variable. Two regression models were compared. In model 1, all psycholinguistic and sensory-motor variables were included as predictors (L1 = imageability,

concreteness, familiarity, literal plausibility, decomposability, L2 = imageability, concreteness). In model 2, valence and valence square were also included as predictors.

For the L1 data, Model 1 accounted for 16% of the variance ($R^2 = 0.16$), ($F(5, 204) = 7.95$, $p < 0.001$). Model 2 accounted for 56% of the variance ($R^2 = 0.56$), ($F(7, 202) = 37.14$, $p < 0.001$), with both valence and valence square as significant predictors.

For the L2 data, Model 1 accounted for 3% of the variance ($R^2 = 0.03$), ($F(2, 207) = 3.81$, $p = 0.02$). Model 2 accounted for additional 23% of the variance ($R^2 = 0.26$), ($F(4, 205) = 18.17$, $p < 0.001$), with only valence square as significant predictor.

3.4.3 Relation Between Affective and Non-affective Variables

Table 3.6 represents the results of linear partial correlations between affective variables (arousal and valence) and non-affective (sensory-motor and psycholinguistic variables) for the L1 group, as well the relation between affective and sensory-motor variables for the L2 group.

Table 3.6 *Linear partial correlation between affective and psycholinguistic/sensory-motor variables rated by L1 speakers. LP: Literal plausibility.*

Group	Variable	Valence	Arousal
L1	Imageability	-0.01	0.14*
	Concreteness	-0.03	-0.13
	Familiarity	0.19**	0.35***
	LP	-0.02	0.07
	Decomposability	-0.05	-0.05
L2	Imageability	0.03	0.14*
	Concreteness	0.00	-0.14*

The numbers in the columns represent Pearson's *r* values; *p* values are expressed as specified. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

There was a significant positive correlation between arousal and imageability (for both L1 and L2 groups) indicating that participants found it easier to evoke a mental image for more arousing words. Generally, the relation between arousal and imageability for single words is mixed; both negative and positive effects are reported in the literature (Citron et al., 2016)and most studies call for further investigation of the relation between arousal and imageability.

For the L1 group, familiarity was significantly positively correlated to both arousal and valence. This indicates that idioms which are encountered more frequently were rated as being more arousing and more positive. Positive correlation between arousal and familiarity has indeed previously been reported for both single words and idioms (Montefinese et al., 2014; Citron et al., 2016).

Unlike one previous study which has explored the relation between decomposability and affective variables for German idioms[8] (Citron et al., 2016), we did not find any relation between decomposability or literal plausibility and either of the affective variables for the L1 group. These findings suggest that the emotional perception of an idiom is not impacted by the degree to which it possesses a literal interpretation. Similarly, the extent to which individual constituents contribute to the idiomatic meaning does not impact the perceived emotional or sensory-motor status of the idiomatic expressions.

[8] In the mentioned study, decomposability is termed as semantic transparency; Nevertheless, the functional definition of these measures is similar in both studies.

Inspecting Quadratic Relation Between Valence and Familiarity

Since valence is a bipolar dimension (i.e., it ranges from positive to negative), and given we found a significant correlation between valence and familiarity in our previous analysis, we explored, we examined whether the relation between these variables is best explained by a linear or a quadratic relationship. Two quadratic regression models were compared wherein familiarity was the dependent variable. In model 1, all psycholinguistic, sensory-motor variables and arousal were included as predictors. In model 2, valence and valence square were *also* included as a predictor. Model 1 predicted 32% of the variance ($R^2 = 0.32$), $F(5, 204) = 19.91$, $p < 0.001$. Model 2 accounted for an additional 3% of the variance ($R^2 = 0.35$), $F(7, 202) = 15.87$, $p < 0.001$ with valence only (not valence square) as an additional significant predictor. Therefore, a linear function best describes the relation between familiarity and valence. Estimated familiarity = 0.90 valence + 1.20 decomposability + 0.47 arousal + 0.21 imageability + 0.11. Studies on single words show the same linear relation between valence and familiarity (e.g., Citron et al., 2014). However, research also indicates that a quadratic function best explains the relation between valence and familiarity when it comes to idioms (Citron et al., 2016).

3.4.4 Relations Between Non-affective Variables

The linear correlations between non-affective variables are reported in Table 3.7. A significant positive correlation between concreteness and imageability was observed for both L1 and L2 groups. For the L1 group, imageability showed a significant positive correlation with familiarity. These findings suggest that it is easier to form a mental image of an idiom as it becomes more concrete and familiar. However, familiarity was not correlated to concreteness. This suggests that, unlike single words (Yao et al., 2017), idioms which are encountered more frequently are not necessarily more concrete.

Table 3.7 *Linear partial correlation between non-affective variables*

Group	Variable	Img.	Con.	Fam.	LP	Decomp.
L1	Img.		0.60***	0.26***	-0.07	-0.02
	Con.			0.007	0.07	-0.001
	Fam.				-0.005	0.38***
	LP					-0.19**
	Decomp.					
L2	Img.		0.23***			
	Con.					

Img: Imageability; Con.: Concreteness; Fam.: Familiarity; LP: Literal plausibility; Decomp.: Decomposability. The numbers in the columns represent Pearson's *r* values; *p* values are expressed as specified: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

3.4.5 Comparing L1 and L2 Ratings

The ratings produced by L1 and L2 speakers were compared using two-tailed t-tests. L1 speakers rated idioms as generally more arousing ($t(316.51) = -7.71$, $p < 0.001$) and less concrete ($t(415.63) = 9.54$, $p < 0.001$) compared to L2 speakers. There appeared to be more variance among L2 arousal ratings (see Table 3.4) than L1 ratings. The mean valence and imageability ratings did not differ between L1 and L2 group; nevertheless, a visual inspection of Figure 3.2 suggests that L1 imageability ratings has less variance than L2 ratings.

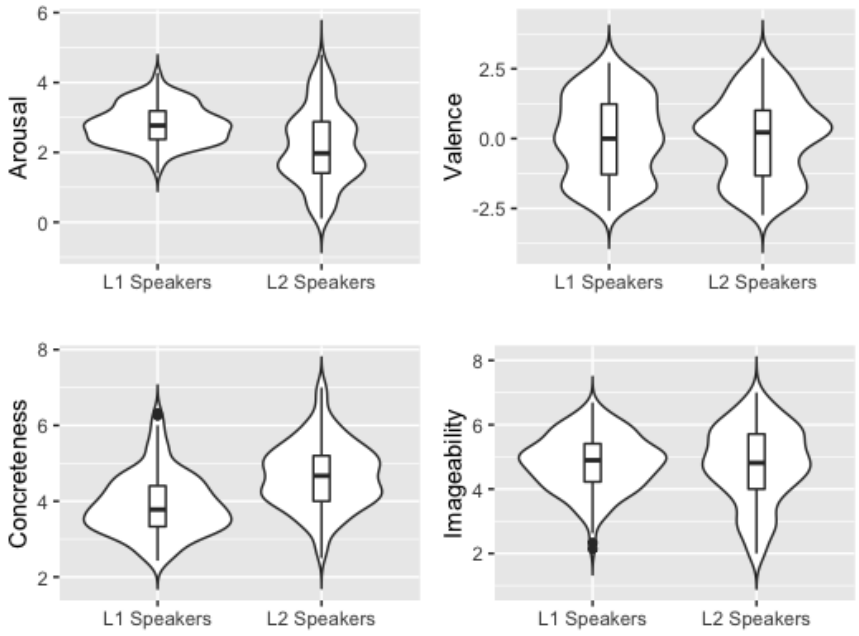


Figure 3.2 Violin and bar chart plots for ratings (arousal, valence, concreteness, and imageability) by L1 and L2 groups.

3.5 General Discussion

Considering the growing interest in understanding the relation between language and emotion, there is growing demand for reliable experimental material. The current study sought to establish affective and sensory-motor norms (valence, arousal, imageability, and concreteness) for 210 English idioms by both native speakers and highly proficient second language speakers. We aimed to describe the association between the collected norms (i.e., affective and sensory-motor variables) as well as to delineate the relationship between our novel data and previously collected psycholinguistic norms. Given the expanding literature on second language acquisition, processing and production, we collected data from L2 speakers and compared them to the idiom ratings of L1 participants. In the following paragraphs, we summarize the main findings of the current study.

Similar to the findings for single words and idioms (Kuppens et al., 2017), we observed a quadratic relation between valence and arousal. This quadratic relation shows that idioms that are highly valenced (either positive or negative) are perceived as highly arousing. This quadratic relation holds for both L1 and L2 groups (despite differences in variance), supporting the universality in the nature of this association.

For the L1 group, the relation between valence and arousal was asymmetrical (i.e., negative items had higher mean arousal than positive items). Similar results were observed by Citron and colleagues (2016), who postulated that this asymmetry was the result of the disproportionate number of negative idioms in their stimuli set. However, since the current study utilized a balanced set of items, it appears that negative idioms are indeed considered to be more arousing. Contrastively, for the L2 group, the mean arousal for negative and positive items were not significantly different. However, considering that the number of negative items (items with valence below zero) were much lower than positive items in the L2 group (negative: 88; positive: 122), the similar arousal for the negative and positive items found in the L2 data confirms that negative items were considered as being more arousing.

Why might L1 speakers be more likely to rate certain idioms as negative? While we do not have a definite explanation for this difference, we postulate that such differences between L1 and L2 speakers may be related to a number of factors, including the length of immersion in the L2 environment. For instance, Imbault and colleagues (2020) examined whether proficiency and length of immersion in English affected the valence and arousal ratings of English words by L2 speakers of English. These authors found that more proficient L2 speakers and those who lived in Canada (the L2 environment) for longer periods displayed more similar emotional responses

to L1 speakers. In our study, L2 participants had a wide variety of length of living in L2 environment. These factors merit further investigation in future research.

Note that only four items possessed neutral valence (valence = 0) in the present study. Similarly, Citron and colleagues (2016) reported that a mere three out of 619 German idioms carried neutral valence. These results support the idea that idioms are generally emotion-laden. This is particularly important for researchers who are interested in the affective processing of language. When it comes to idioms, researchers may have difficulty in compiling enough neutral items for their studies, and may thus need to consider the emotional content of their stimuli. Aside from the differences in the number of negative and positive items for the two groups, we found no significant difference between valence ratings for the L1 and L2 groups.

Unlike valence, the L1 and L2 groups attributed distinct arousal ratings to the idioms, with L1 speakers generally assigning higher arousal ratings. This result corroborates the findings on single words which shows greater cross-language variability for arousal and more generalized ratings for valence across languages (Eilola & Havelka, 2010; Redondo et al., 2007; Montefinese et al., 2014). Indeed, while the dimension of valence is perceived similarly across cultures (Russell, 1991), the variability in arousal ratings can be related to cross-cultural diversity (Montefinese et el., 2014). In the current study, L2 speakers originated from a wide range of language backgrounds (French, Persian, Arabic, Chinese, Vietnamese, Russian, etc.) and cultures, and tended to assign more variable arousal ratings (SD = 1.05) compared to the uniform L1 group (SD = 0.55). These findings suggest that second language researchers may use L1 valence ratings if they do not have access to L2 ratings, but that they should be cautious of using L1 arousal ratings for L2 speakers, as these groups are differently aroused by idioms.

An additional objective of the current study was to explore the relation between affective and non-affective idiom ratings. We observed positive correlations between arousal and imageability for both the L1 and L2 groups. Furthermore, both groups showed either a trend (L1 speakers) or a significant negative correlation (L2 speakers) between arousal and concreteness. Since prior studies show that the relation between arousal and sensory-motor variables may be non-linear, we examined whether a quadratic function best describes their association. Our result only marginally supported a quadratic relation between these variables. Given that the idioms in our study were generally attributed low to moderate arousal ratings (minimum = 1.40; maximum = 4.28), it is possible that such a compression in the dispersion of the data may have obscured the relationship between arousal and concreteness. Future studies should consider using a larger set of idioms (including idioms eliciting higher arousal responses) in order to capture the full picture.

Next, positive correlations between the affective factors and familiarity were observed: as participants rated idioms as more familiar, they tended to also consider them to be more arousing and more positive. This positive relation was previously found for German idioms (Citron et al., 2016). Though these authors found that a quadratic function best represented the relation between valence and familiarity, a linear function was a better fit for our own data. It is worth mentioning that in Citron and colleagues (2016), negative idioms made up more than two third of the items. When valence is more balanced, as in the current study, it appears that a linear relation is favored.

Finally, similar to previous studies (Paivio, 1991; Schwanenflugel et al., 1988), imageability and concreteness yielded a strong positive correlation. The robust association between these variables (as well as their relation to other variables, like arousal) suggests that

researchers should take special care while manipulating these factors; there is a danger of collinearity within certain types of analyses. Though imageability and concreteness are strongly related, they should still not be used interchangeably, since they represent two different constructs.

3.6 Conclusion, Limitations, and Future Directions

In sum, the present descriptive study compiles highly reliable affective and sensory-motor norms for a set of English idioms. It describes the relationship between collected measures, thereby enabling researchers to make systematic decisions about their stimuli in future research. In short, we have created the first known database consolidating affective and sensory norms for English idioms. It is worth mentioning that in the current study, most participants were young female university students. Considering that previous studies shows that both gender (Fischer, 2000) and age (Fairfield et al., 2017) might influence participants' ratings (especially affective ratings), future studies should seek to account for these potential factors.

References

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

Altarriba, Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. Behavior Research Methods, *Instruments, & Computers, 31*(4), 578–602.

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2018). Gorillas in our midst: Gorilla. sc. *Behavior Research Methods*.

Arfé, B., Delatorre, P., & Mason, L. (2022). Effects of negative emotional valence on readers' text processing and memory for text: an eye-tracking study. *Reading and Writing*, 1-26.

Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and language, 125*(1), 47-53.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods, 46*(3), 904-911.

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech, 63*(1), 95-122.

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.

Citron, F. 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.

Citron, F. M., Weekes, B. S., & Ferstl, E. C. (2014). How are affective word ratings related to lexicosemantic properties? Evidence from the Sussex Affective Word List. *Applied Psycholinguistics, 35*(2), 313-331.

Eilola, T. M., & Havelka, J. (2010). Affective norms for 210 British English and Finnish nouns. *Behavior Research Methods, 42*(1), 134-140.

Fairfield, B., Ambrosini, E., Mammarella, N., & Montefinese, M. (2017). Affective norms for Italian words in older adults: age differences in ratings of valence, arousal and dominance. PloS one, 12(1), e0169472.

Ferré, P., Guasch, M., Moldovan, C., & Sánchez-Casas, R. (2012). Affective norms for 380 Spanish words belonging to three different semantic categories. *Behavior Research Methods, 44*(2), 395-403.

Fischer, A. (Ed.). (2000). *Gender and emotion: Social psychological perspectives*. New York, NY: Cambridge University Press.

Gibbs Jr, R. W., & Gonzales, G. P. (1985). Syntactic frozenness in processing and remembering idioms. *Cognition, 20*(3), 243-259.

Heredia, R. R., & Cieślicka, A. B. (Eds.). (2015). *Bilingual figurative language processing*. Cambridge University Press.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 721.

Imbault, C., Titone, D., Warriner, A. B., & Kuperman, V. (2021). How are words felt in a second language: Norms for 2,628 English words for valence and arousal by L2 speakers. Bilingualism: *Language and Cognition, 24*(2), 281-292.

Imbir, K. K. (2016). Affective norms for 4900 Polish words reload (ANPW_R): Assessments for valence, arousal, dominance, origin, significance, concreteness, imageability and, age of acquisition. *Frontiers in psychology, 7*, 1081.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. Journal of Experimental Psychology: General, 140(1), 14.

Kuppens, P., Tuerlinckx, F., Yik, M., Koval, P., Coosemans, J., Zeng, K. J., & Russell, J. A. (2017). The relation between valence and arousal in subjective experience varies with personality and culture. *Journal of personality, 85*(4), 530-542.

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

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of the affective norms for English words (ANEW) for Italian. *Behavior research methods*, *46*(3), 887-903.

Morid, M., Bachar, N., & Sabourin, L. (2021). Capturing the multi-determined nature of idiom processing using ERPs. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *75*(2), 155.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(3), 255.

Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of ANEW (affective norms for English words). *Behavior research methods*, *39*(3), 600-605.

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological bulletin*, *110*(3), 426.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, *110*(1), 145.

Sabourin, L., Leclerc, J. C., Lapierre, M., Burkholder, M., & Brien, C. (2016). The language background questionnaire in L2 research: Teasing apart the variables. In Proceedings of the *Annual Meeting of the Canadian Linguistics Association, Calgary, Canada.*

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly*, 114-126.

Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M. A., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior research methods*, *49*(1), 111-123.

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting bilingual and monolingual idiom processing. In R. R. Heredia & A. B. Cies´licka (eds.), *Bilingual Figurative Language Processing*, pp. 171–207. Cambridge: Cambridge University Press.

Titone, D., Lovseth, K., Kasparian, K., and Tiv, M. (2019). Are figurative interpretations of idioms directly retrieved, compositionally built, or both? Evidence from eye movement measures of reading. *Canadian Journal of Experimental Psychology/Revue canadienne de Psychologie Experimentale , 73* (4), 216.

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: the role of emotion. *Cerebral Cortex*, *24*(7), 1767-1777.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition, 1*(2), 219-247.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods, 45*(4), 1191-1207.

Yao, Wu, J., Zhang, Y., & Wang, Z. (2017). Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behavior Research Methods*, *49*(4), 1374–1385.

4 Role of Affective Factors and Concreteness on the Processing of Idioms[9]

Abstract

In this study, we asked how the emotional status, i.e., valence and arousal, and concreteness of idioms contribute to their processing. Additionally, we asked whether the contribution of emotional factors and concreteness is modulated by other linguistic constraints, specifically idiom familiarity and decomposability, that has been shown to impact idiom processing. Participants read short idiomatic phrases (e.g., *he kicked the bucket*), word-by-word and for comprehension while their reaction time (RT) was recorded. The results showed that the emotional status of idioms contribute to their processing and this contribution is modulated by familiarity and decomposability levels of idioms in different ways. In particular, the impact of valence (i.e., the degree an idiom is pleasant/unpleasant) was modulated by familiarity, and the impact of arousal was modulated by decomposability. We did not find strong evidence for the contribution of concreteness for idiom processing. Our findings are aligned with theories of semantic representation, which suggest that besides linguistic information, sensory-motor and affective information are fundamental in representing meaning.

Keywords: Idiom processing, valence, arousal, concreteness, familiarity, self-paced reading task

[9] Morid, M., & Sabourin, L. (2023). Role of Affective Factors and Concreteness on the Processing of Idioms. *Journal of Psycholinguistic Research*.

4.1 Introduction

Idiomatic expressions like “she threw a fit” and “she cracked a joke” typically refer to *abstract* events and are more *emotionally* engaging than literal language (Citron et al., 2016).

The impact of emotional factors and concreteness on language processing has received considerable attention at both word-level and sentence-level processing (e.g., West & Holcomb, 2000; Barber, Otten, Kousta, & Vigliocco, 2013, for concreteness; Hinojosa, Moreno, & Ferre, 2020, for language and emotion). Surprisingly, the impact of these factors on the processing of idioms, as expressions that are inherently a preferred means of language for describing emotional state (Clore & Ortony, 1987) and abstract events, has largely been ignored in all but a few studies (e.g., Findlay & Carrol, 2019).

The Multi-Determined Model (Libben & Titone, 2008) of idiom processing suggests that idiomatic expressions differ in many respects, such as level of familiarity, decomposability, and literal plausibility, and this variability impacts their representation and processing (Titone & Libben, 2014). It has been hypothesized that semantic representation consists of a combination of linguistic information, as well as experiential information. The particular sources of experiential information are comprised of sensory-motor (the extent to which a concept can be learned through sensory-motor interaction with the world; that is, the extent to which it is concrete) and affective information (e.g., whether the concept is negative or positive, i.e., its valence) (Vigliocco et al., 2009). Based on this hypothesis, the current study is an attempt to consolidate previously known linguistic constraints on idiom processing with affective factors and concreteness as other possible sources of information that modulate language representation and processing. We wished to assess whether idiom processing is affected by varying degrees of concreteness and affect, the same way it is affected by linguistic constraints.

4.1.1 Affective Factors and Language Processing

Two main components have been identified as fundamental to the nature of affect: valence (the level of pleasantness of the stimulus; the extent to which the stimulus is positive or negative) and arousal (the intensity of emotion provoked by a stimulus; the extent to which it is calming or exciting) (Russell, 2003; Warriner, Kuperman, & Brysbaert, 2013). A growing body of literature investigates the effect of valence (whether positive and negative stimuli are processed differently) and arousal (whether the emotional intensity of the stimuli influences language processing) on the lexical and semantic aspects of language during single-word and sentence comprehension (see Hinojosa et al., 2020 for a review). Some studies focusing on interaction between language and emotion at the single-word level have found that negative stimuli (e.g., a word such as 'suicide') elicit slower reaction times (RTs) compared to positive stimuli (e.g., 'sunshine'): this was found in various tasks, for example with color naming in an emotional Stroop task, as well as in a lexical decision task (e.g., Williams, Mathews, & MacLeod, 1996). Other studies show enhanced sensitivity to negative stimuli (e.g., Nasrallah et al., 2009), or an advantage for both positive and negative stimuli over neutral stimuli (e.g., 'fabric') (reported in Hinojosa et al., 2020). Findings on the relation between valence and arousal are also mixed: some studies have reported an interaction between valence and arousal, such that both negative and positive words are highly arousing while neutral words are the least arousing words (e.g., Larsen, Mercer, Balota, & Strube, 2008), whereas others found no such relationship (e.g., Citron, Weekes, & Ferstl, 2014). Kuppens et al., (2013) reported six possible relations between valence and arousal (negative linear, positive linear, V shape, etc.): by means of statistical modeling, the authors conclude that there was no static relation between the two factors. The discrepancy among these studies is likely due to the fact that stimuli selection varies between studies and may

be more or less stringently controlled. For example, it has been shown that frequency, which is among the most important factors for word recognition, tends to modulate emotional effect (Kuperman et al., 2014), such that emotional factors exert larger effects in words with lower frequency than in words with higher frequency. There is also evidence showing that if valence is held constant, high arousing words capture more attention than words with low arousing level (Delaney-Busch et al., 2016).

On a smaller scale, studies have also considered the role of emotional factors at the sentence level during syntactic and semantic unification processes. For example, Fraga et al., (2017) conducted an ERP study to investigate the effect of emotional adjectives on sentence processing during a gender agreement judgment task. They found that the presence of emotional adjectives facilitated the detection of gender mismatches, confirming that emotionally-laden words seem to behave differently from neutral words, not only in isolation, but also during sentence processing. To examine the effect of emotion during semantic unification, some studies have incorporated emotional words within a sentence context, which was either emotionally neutral (Martin-Loeches et al., 2012) or non-neutral (positive or negative) (Delaney-Busch & Kuperberg, 2013). These studies showed that the integration of emotion words was more cognitively demanding when the sentence context was neutral. All in all, the literature provides compelling evidence in support of the role of affective factors on language processing at word and sentence level. However, how affective factors contribute to the processing of idioms is still understudied with just a few exceptions (e.g., e.g., Findlay & Carrol, 2019).

4.1.2 Concreteness and Language Processing

A word is considered concrete when it has available sensory referents (Schwanenflugel & Stowe, 1989). For example, the word ‘bread’ recalls an item that can be directly tasted, smelled, and

touched. Compared to concrete words, abstract words, e.g., 'justice', lack a direct sensory referent. The concreteness effect refers to the observations showing that concrete words are processed faster and more accurately than abstract words (Holcomb et al., 1999). Moreover, it has been shown that the time required to comprehend a sentence is shorter when the sentence is concrete (West & Holcomb, 2000), while sentences constructed of abstract words generally take longer to read (Schwanenflugel & Stowe, 1989). Two popular accounts of representational and processing differences between abstract and concrete words are the dual-coding theory and the context-availability hypothesis. According to the dual-coding theory (Paivio, 1991), the type of available information for concrete words is different from that of abstract words. Concrete words have access to information in both the "verbal linguistic" system and "nonverbal imagistic" system, while abstract words only have access to information stored in the verbal linguistic system. Thus, the processing advantage of concrete words arises from their available connections to multiple sources of information. The context-availability hypothesis (Kieras, 1978; Bransford & McCarrell, 1977), on the other hand, argues that the *quality* of available information, not the type of information, is different for concrete and abstract words. In other words, the advantage of concrete words over abstract words stems from the greater contextual association in semantic memory for concrete words compared to abstract words. Kousta, et al. (2011) showed that when controlling for imageability and context-availability, two important measures used to operationalize dual-coding and context-availability hypotheses, an abstractness effect replaces a concreteness effect (i.e., abstract words show a processing advantage compared to concrete ones). The authors argued that because abstract words are typically more emotionally valenced (Kousta, Vinson, & Vigliocco, 2009), they have greater access to the emotional experiential information, which results in a processing advantage for abstract words.

Alongside the contribution of affective factors and concreteness to the processing of words, some researchers have also considered the combination of these factors. Snefjella and Kuperman (2016) evaluated the influence of valence, arousal, and concreteness on word recognition. Using large corpora and mega-studies, they considered the average valence, arousal, and concreteness of the contexts in which a word appears, compared to the valence, arousal, and concreteness of the word itself. Their findings suggested that the properties of the context were a stronger predictor of word processing than the properties of the individual words. Findings also suggest that once valence is considered as a predictor for the processing of abstract items, it may override the impact of other factors such as concreteness (Vigliocco et al., 2009). Furthermore, Sheikh and Titone (2013) suggest that when words have strong association with both affective and sensory-motor information (concreteness), processing will benefit from only one of these sources of information. Finally, on the interaction between emotion and concreteness, Kim, Sidhu, & Pexman (2020) showed that valence interacts with concreteness in a recognition memory task, such that only abstract words were influences by the effect of valence in a memory task[10].

4.1.3 Idiom Processing

According to models of sentence processing, two main operations are involved during literal sentence comprehension: firstly, word meanings are accessed, and secondly, those meanings are compositionally integrated to reach the final message (Hagoort, 2005). However, this view is greatly challenged by idioms whose final message is not composed of the individual meanings of

[10] The paper discusses how this effect is different in adults and 7-8 year old children.

its constituents. One view is that idioms are represented as a single word, and thus their processing is non-compositional (Bobrow & Bell, 1973; Swinney & Cutler, 1979). The opposite view asserts that idioms are processed compositionally, like literal language processing (Gibbs, Nayak, & Cutting, 1989; Hamblin & Gibbs, 1999). Hybrid models of idiom processing proposed a middle ground, by considering the possibility of both compositional and non-compositional (direct retrieval) analysis (Cacciari & Tabossi, 1988; Titone & Connine, 1994). According to these views, idioms are processed compositionally until they become recognizable to language users, after which the direct retrieval becomes available.

The introduction of hybrid views shifted attention toward the idiom-specific factors that facilitate their direct retrieval. However, the emphasis on these factors varies from one study to another. For instance, it is shown that the predictability of idioms plays a key role in the availability of direct retrieval of idioms. Predictability is defined as the likelihood of recognizing an idiom before the phrase offset (i.e., the last word of the idiomatic phrase, for example, *bag* in *let the cat out of the [bag]*). In other words, direct retrieval is a privilege for idioms that can be predicted and recognized before their offset (Cacciari & Tabossi, 1988). Alternatively, other studies suggest that the predictability of idioms is strongly related to the subjective familiarity associated with idioms (Titone & Connine, 1994). Moreover, it is further suggested that the effect of different idiomatic dimensions is task-specific. For example, the decomposability of idioms (i.e., the extent to which the meaning of the idiom is related to the meaning of its constituent) only exerts an effect in tasks that direct participants' attention to the meaning of idioms (when participants are required to provide overt judgments on idiom meanings), but not in online experimental tasks (Libben & Titone, 2008). More recently, using eye-tracking technique, Titone and colleagues (2019) investigated the processing of idioms. They asked participants to read expressions with both literal

and idiomatic interpretations (e.g., *kick the bucket*). These phrases were followed by a context which biased participants toward either literal or idiomatic interpretations. By analyzing eye movement at different regions, they found that for comprehending idioms both compositional and non-compositional analysis are involved at different time course. Importantly, they concluded that as increased familiarity may increase the likelihood of direct retrieval of idiomatic phrases, increased decomposability may inhibit direct retrieval as it may encourage the literal interpretation. The impact of these linguistic constraints on the processing of idioms has been widely tested and has been established using various methods and techniques (e.g., Titone et al., 2019 using eye-tracking; Morid, Bachar, & Sabourin, 2021 using ERPs). Other factors relevant to idiom processing are now under investigation. For one, Findlay and Carrol (2019) asked whether "semantic richness" contributes to the processing of idiomatic expressions. While the definition of semantic richness varies in different studies, these authors focused on imageability and sensory experience ratings (SER). In a cross-modal priming lexical decision task, participants saw an idiomatic expression (e.g., "he blew a fuse") or a filler sentence of the same structure (e.g., "he ate the cake"). The idiomatic expressions were followed by a target word (500 ms after the idiom's offset), which was either related or unrelated to their figurative meaning. Results showed that imageability[12] had an inhibitory effect on the activation of the figurative meaning. They further considered the role of other factors including emotional valence, and physical interaction. Most relevantly to our study, they found a facilitative effect of valence for idioms with a higher familiarity level. In other words, for highly familiar phrases, figuratively related

[12]In a separate study, they asked participants how strongly each word evoked a taste, sight, touch, smell, or sound. These ratings were used in the cross-modal study.

words required shorter RT than non-figuratively related words; this effect was negligible for low familiarity idioms.

4.1.4 The Present Study

The present study investigated idiom processing from the perspective of the embodied theories of cognition—more specifically, the models that suggest multiple sources of information (linguistics, sensory-motor, and emotional) account for semantic representation and processing (Vigliocco, 2009). Models such as the multi-determined model (Libben & Titone, 2008) of idiom processing provide insight on how multiple linguistic constraints (e.g., familiarity, decomposability, etc.) modulate idiom processing, and how idioms which differ along those measures are processed differently. If sensory-motor and affective sources of information impact the processing of idioms like they impact other aspects of language, we hypothesized that, similarly to linguistic constraints, differing degrees of concreteness and emotionality would result in processing differences for idioms. Thus, the main aim of this study was to investigate how different sources of information— affective, sensory-motor, as well as linguistic information— contribute to the processing of idioms.

We initially explored the impact of constituent affect and concreteness (see methods section for a detailed description on the variables and how we calculated them) on the processing of literal and idiomatic sentences. The aim of this question was to compare idiomatic sentences with structurally similar literal sentences to see whether the influence of valence, arousal, and concreteness that is normally found for words holds for sentence processing and whether this impact relates to the final interpretation of sentences being literal or figurative.

We then explored the impact of phrasal level affective and sensory motor information on the processing of idioms. If idioms (at least to some degree) benefit from direct retrieval, then

this might be the phrasal level factors, i.e., the emotional status and concreteness of the phrase, that impact their processing. Additionally, we explored whether the impact of affective factors and concreteness is modulated by familiarity and decomposability of idiomatic phrases. Given that less familiar idioms are more likely to be analyzed compositionally, and highly familiar idiom's meaning (at least to some degree) is subject to direct retrieval, we hypothesized that familiarity might modulate the impact of non-linguistic sources of information. However, we did not develop any hypothesis regard to the nature or direction of this influence. We also considered the modulating impact of decomposability on the effect of non-linguistic (affective and sensory-motor) sources of information on idiom processing. Considering that the meaning of more decomposable idioms has a stronger connection to the meaning of individual words compared to non-decomposable phrases which are likely to be represented in semantic memory to some degree, we hypothesized that idioms with low and high level of decomposability may be impacted by affective information in different ways.

4.2 Methods

4.2.1 Participants

Forty-six students (mean age = 20.2 years, 35 female) were recruited through the Integrated System of Participant Research (ISPR) at [masked]. They received partial course credit as compensation for their participation. Participants completed a Language Background Questionnaire (Sabourin et al., 2016), which revealed that most participants knew a language other than English but did not start learning that second language before the age of three. Data from one participant was excluded from the analysis, due to low accuracy score (scored below 75%) under the criteria explained in the results section.

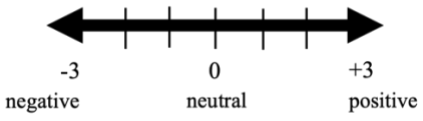
4.2.2 Materials

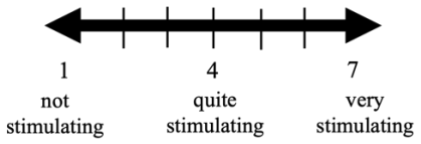
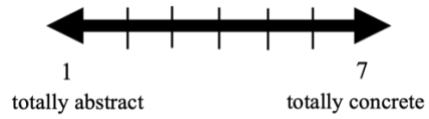
We initially chose 80 idiomatic expressions from a database that provides norming measures for 210 idioms (Titone & Liben, 2008). Since we were interested in the modulating impact of familiarity, we selected 40 of the most familiar (familiarity > 4 on a 5-point scale) and 40 of the least familiar idioms (familiarity < 2.5) from this dataset to ensure that idioms from both ends of this continuum were included. All idiomatic expressions were in the form of "She/He/It $verb_{past\ tense}$ x noun" (e.g., *It slipped his mind*). To create literal sentences, we combined the verbs and nouns such that the resulting sentences did not have an idiomatic meaning (e.g., "he *swallowed* his pride", and "he sugared the *pill*", resulted in "he *swallowed* his *pill*"). The final stimuli for each condition did not have the same number for each verb/noun used. For example, the verb "*got*" is used four time for the idiomatic condition and two times for the literal condition and the verb "had" is used two times for the idiomatic condition and 6 times for the literal condition. We asked three native English speakers to read these sentences and rate them for the level of figurativeness on a scale from 1 to 5. Any sentence with a greater rating than 1.5 was replaced with a new sentence which was evaluated by a different native speaker. This procedure assured us that the literal sentences did not convey any figurative interpretation. From the initial list of 80 idioms, 54 idioms were used to create literal sentences. These idioms maintained the balance between low and high familiarity (27 each). The final experimental stimuli consisted of 54 idiomatic expressions, and 54 literal sentences. We included 40 non-idiomatic filler sentences with the same structure (e.g., *he met his supervisor*) to decrease the proportion of idioms to non-idiomatic sentences. Following a within-subject design, each participant saw the entire list of material (54 idiom, 54 literal, 40 filler; 148 total) in a randomized order.

To obtain constituent valence, arousal, and concreteness, we obtained relevant measures for the component words of the idioms (nouns and verbs) from various databases (concreteness from Brysbaert & Warriner, 2014; valence and arousal from Warriner, Kuperman, & Brysbaert, 2013).

To obtain idiom-related measures, we asked 300 participants in a separate study ([masked]) to read idioms and rate them for several characteristics including valence, arousal, and concreteness. For valence, we asked them to rate how positive or negative the metaphorical meaning of each idiom; for arousal, we asked them to rate how stimulating the metaphorical meaning of each idiom; and for concreteness, we asked them to rate the extent to which the metaphorical meaning of the sentence refers to an experience, state or event that can be experienced with the senses. Table 1 shows the definition of each measure along with description of scales on which each measure is rated.

Table 4. 1 *Idiom related measures along with the definition and scale for each measure*

Measure	Definition	Instruction
Valence	Valence describes the extent to which an event is positive or negative. For example, the sentence "*The whole experience left a bad taste in my mouth*" describes a negative event, whereas the sentence "*She was feeling on top of the world after winning the tennis tournament*" describes something positive.	On a scale from -3 to + 3, rate how positive or negative the metaphorical meaning of the following sentences is. 

Arousal	Emotional Arousal describes to what extent an event is stimulating. For example, the sentence "*he screamed because of pain*" is perceived as very stimulating, while sentence "*she read a book*" might be perceived as not stimulating.	Please rate the sentence for emotional arousal on a scale from 1 to 7 (consider metaphoric meaning). 
Concreteness	A sentence conveys a concrete meaning when it denotes an experience that refers to one or more sensory modalities (e.g., touch, vision). For instance, the meaning of "*the blanket is so soft*" refers to the experience of touch. In contrast, the meaning of "*she had a sad thought*" is abstract since it cannot be associated with a sensory modality.	Please rate the extent to which the metaphorical meaning of the sentence refers to a state or event that can be experienced with the senses. 

To ensure participants' full understanding, the instructions were accompanied with examples. We calculated the reliability of these measures based on the internal consistency (Cronbach's alpha); for all measures, we obtained an alpha value over 0.8, indicating that the obtained measures are highly reliable. Finally, measures of familiarity and literal plausibility were obtained from Libben & Titone (2008). Table 2 depicts examples of stimuli, constituent-related

measures[13], and idiom related measures. Since the measures are obtained through different scaling, the relevant rating with which the measurements have been elicited is also stated under each

measure. A complete list of sentence stimuli can be found on the Open Science Framework repository at Idiom_Affect&Concreteness [14].

Table 4.1 *Stimuli and measurements*

Type	Example Sentences	Measurement	Min	Max	Mean	SD
Idiom	He spilled the beans. She packed her bags.	Constituent Valence Rating: (1,9)	3.95	7.39	5.43	0.71
		Constituent Arousal Rating: (1,9)	3.23	5.48	4.02	0.57
		Constituent Con. Rating: (1,5)	2.47	4.74	3.96	0.56
		Idiom Valence	-2.55	2.78	0.41	1.32

[13] Please note that constituent-related measure refers to average of verb and noun ratings. They are provided in this so the reader is able to compare idiom-related ratings with the ratings averaged over constituents.

[14] The link to the open science framework is "https://osf.io/86y2r/?view_only=005c0669f9054435abac112fed99b715 ".

		Rating: (-3,3)				
		Idiom Arousal Rating: (1,7)	2.27	4.73	3.49	0.55
		Idiom Con. Rating: (1,7)	2.73	5.56	4.08	0.64
		Familiarity Rating: (1,5)	1.57	4.97	3.23	1.26
		Literal Plausibility Rating: (1,5)	0.81	4.9	2.86	1.16
Literal	She spilled the <u>mustard</u>. He lost his <u>bag</u>.	Constituent Valence Rating: (1,9)	3.33	7.12	5.35	0.78
		Constituent Arousal Rating: (1,9)	3.64	5.35	3.99	0.64
		Constituent Con Rating: (1,5)	2.66	5.88	4.07	0.84

Example sentences for the literal and idiomatic conditions. Second column shows each measurement along with the rating with which the measurements have been elicited. Min & Max: minimum and maximum values for each measurement. Mean: the average of each measurement over the total number of sentences for each condition (N = 54). SD: standard deviation for the total number of sentences for each condition.

Table 3 represents the correlation among all variables for the subsample of idioms used in the current study.

Table 4.2 *Correlations among variables for the subset of idioms used in this study (N = 54)*

	ConsVal	ConsAr	ConsCon	IdiomVal	IdiomAr	IdiomCon	Fam.	LP
ConsVal		-0.29*	-0.21*	0.07	0.12	-0.02	-0.04	0.004
ConsAr			-0.25	0.19	0.14	0.01	0.24	-0.38*
ConsCon				-0.12	-0.05	0.19	-0.33*	0.44**
IdiomVal					0.17	0.10	0.26	0.03
IdiomAr						0.14	0.22	0.05
IdiomCon							0.11	0.19
Fam.								-0.16
LP								

ConsVal : Constituent Valence; ConsAr: Constituent Arousal; ConsCon: Constituent Concreteness; IdiomVal: Idiom Valence; IdiomAr: Idiom Arousal; IdiomCon: Idiom Concreteness; Fam: Familiarity; LP: Literal Plausibility; The numbers in the columns represent Pearson's *r* values, and *p* values are expressed as specified below the table. * $p < .05$; ** $p < .01$

4.2.3 Procedure

All participants were tested individually, beginning with the consent form followed by the Language Background Questionnaire (Sabourin et al., 2016). We opted for a self-paced reading task rather than a lexical decision task, which is a task that is commonly used in the study of emotional aspect on language processing, as it has been suggested that lexical decision task may direct participants to focus on single-word features, instead of connecting them to the larger body of language (Teng, Wallot, & Kelty-Stephen, 2016). Taking these considerations into account, participants completed a moving window self-paced reading task programmed in E-Prime version 3 (Psychology Software Tools, Pittsburgh, PA) where they were simply asked to read sentences at a natural pace. During this task, participants first saw four dashes representing the words of the sentence. They initiated the sentence by pressing the spacebar on the keyboard; each time they pressed this key, the next word appeared and the previous word was replaced by dashes. We instructed participants to read each sentence with the aim of comprehending what was written. We explicitly asked that they not stop once they had arrived at the last word but rather to hit the spacebar again once they had understood the sentence to move to the next screen. Participants started with four practice trials in the presence of the researcher and were provided with further instructions if needed. To ensure participants' full attention, one-third of the sentences were followed by a yes/no comprehension question.

4.2.4 Analysis

All analyses were performed using the RT of the last word of the sentence (always a noun)[15]. Trials with RTs below 200 ms and above 1500 ms were excluded. Table 4 shows the RT to all words and for both conditions.

Table 4.3 *Mean Participants' response time (RT) to each word of the phrase for both conditions.*

	RT			
	Word 1	Word 2	Word 3	Word 4
Literal	350.3	373.1	365.4	471.7
Idiom (all)	351.4	372.5	367.2	474.8
Idiom (High familiar)	350.9	370.5	364.7	464
Idiom (Low familiar)	350.8	373.4	369.7	485.6

We set this criterion to assure that only reaction times that reflect the comprehension of the sentences are included and the ones that are too short or too long are excluded[16]. This resulted in the exclusion of 4% of all data points. The accuracy of responses was then calculated to all

15 We collected Reaction Time to all words of the expressions and initially performed analysis on all words. However, since we did not find any statistically significant effect, we do not report them here and only the analysis on the last word of the expression will be presented.

16 Literature suggest that valid RTs do not typically start before 200 ms (Berger et al., 2021), because this is the minimum time that is required to encode stimuli. Thus we set our lower cutoff to 200 ms. For the upper cutoff, we excluded RTs that fell outside "*Mean*+2**SD*" range.

comprehension questions for each participant; an accuracy rate threshold of 75% was set, resulting in the exclusion of one participant from the analysis. RT data were analyzed using linear mixed effect models implemented in lme 4 R package (Bates, Mächler, Bolker, & Walker, 2015) in R (version 4.1.2), and effects plots were produced using the Effects package (version 4.0-3; Fox & Weisberg, 2018). Since there was correlation between some of variables, we scaled all continuous variables to reduce collinearity (Titone et al., 2019). Subject and item were included as random effects, with trial order as a control variable[17]. For each model, the estimated coefficient (*b*), standard error (*SE*), *t*- and *p*-values (using Satterthwaite approximations for degrees of freedom) are reported. We provide complete model outputs on the Open Access repository at Idiom_Affect&Concreteness .

4.3 Results

We first considered whether the affective level (valence and arousal) and concreteness of the individual component of the sentences (the noun and the verb) exert influence on the RT and whether this influence is different for literal and idiomatic sentences. We fitted a linear mixed-effect model with Type (Idiom versus Literal), Verb Valence, Verb Arousal, Verb Concreteness, and the interaction between Type and each of the other variables as fixed effects. We found a main effect of Verb Valence ($b = 15.29$, $SE = 6.99$, $t = 2.18$, $p < 0.05$). The interaction between Type and Verb Valence was also significant ($b = -20.40$, $SE = 9.41$, $t = -2.16$, $p < 0.05$). As

[17] Trial number was not significant in any model, indicating that the order of the representation of trials did not impact the results.

Figure 1 shows, by increasing Verb Valence[19], RT increased for idioms and decreased marginally for literal sentences.

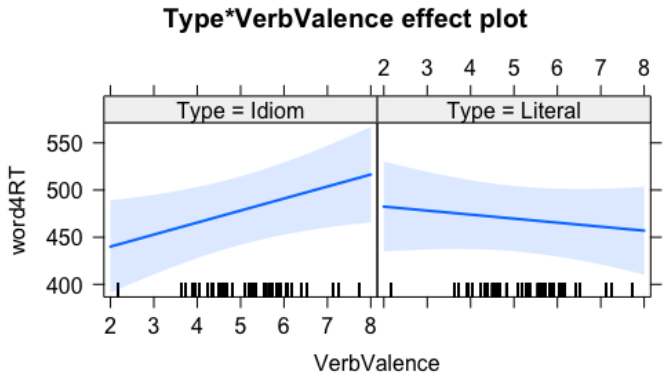


Figure 4.1 Partial effects plot of RT as a function of Verb Valence for literal and idiomatic conditions.

We then fitted another model with Type (Idiom versus Literal), Noun Valence, Noun Arousal, Noun Concreteness, and the interaction between Type and each of the other variables as fixed effects. We found a main effect of Noun Arousal (b = 11.65, SE = 5.22, t = 2.23, $p < 0.05$). For both sentence Type (idiom and literal), by increasing Noun Arousal, RT increased (see Figure 2). The interaction between Type and Noun Arousal was not significant.

[19] Please note that whenever we mention increasing valence we mean when valence becomes more positive or more pleasant, compared to more negative and unpleasant valence.

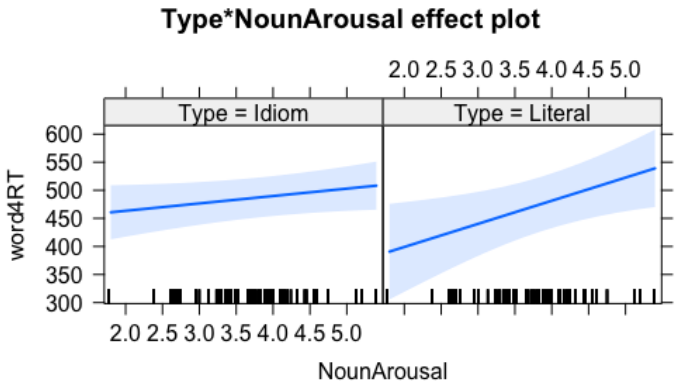


Figure 4.2 Partial effects plot of RT as a function of Noun Arousal for literal and idiomatic conditions.

We then considered the impact of phrasal level factors[20] (Idiom Arousal, Idiom Valence, and Idiom Concreteness) on the speed of processing of idiomatic sentences. We thus fitted a linear mixed-effect model with these factors as fixed effects. Since we were also interested in exploring the modulating effect of Familiarity and Decomposability, we additionally entered these two factors along with their interaction with idiom-related factors as fixed effect. Literal plausibility was entered as a control variable. We found a main effect of Familiarity (b = -22.22, SE = 7.07, t = -2.72, $p < 0.05$); increasing Familiarity decreased RT. There was a main effect of Idiom Valence (b = -13.31, SE = 6.12, t = -2.05, $p < 0.05$), and an interaction between Familiarity and Idiom Valence (b = -18.40, SE = 7.14, t = -2.57, $p < 0.05$); for less familiar

[20] Please note that when we considered phrase-level factors, we created models on idiom-only data since phrasal level factors are only applicable for idioms. The purpose here is not to compare idiomatic sentences with literal sentences but to explore how idiom-related factors impact their processing and how they interact.

idioms increasing Valence (larger positive valence) increased RT, and for highly familiar idioms, increasing Valence decreased RT (see Figure 3).

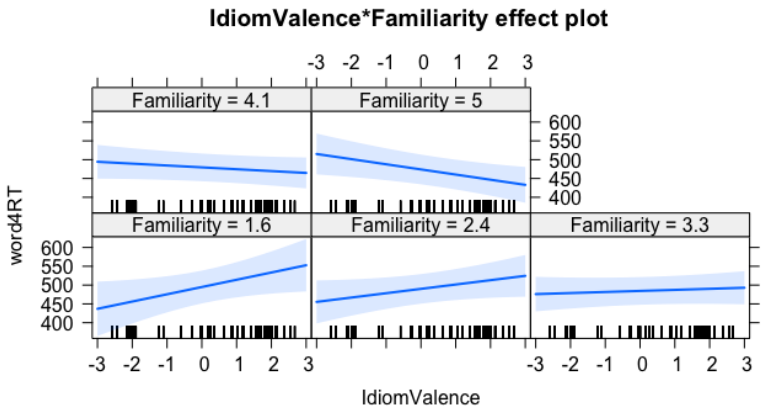


Figure 4.3 Partial effects plot of RT as a function of Idiom Valence and Familiarity.

Motivated by the possibility raised in the introduction regarding the processing differences between high and less familiar idioms and given the result of the previous model (interaction between Familiarity and Idiom Valence), we explored the impact of idiom-related factors for low and highly familiar idioms, separately. We fitted two other models on less familiar (idioms with Familiarity lower than 2.5) and highly familiar(idioms with Familiarity higher than 4) data. Fixed factors were the same as in the previous model, except that we did not include Familiarity or its interaction, as its impact was already investigated. As in the previous model, Literal Plausibility was included as control variable. For highly familiar idioms, there was a significant main effect of Idiom Arousal (b = 79.37, SE = 6.75, t = 2.15, $p < 0.05$); increasing Arousal increased RT, and an interaction between Idiom Arousal and Decomposability (b = -112.65, SE = 5.85, t = -2.09, $p < 0.05$); for idioms with low Decomposability level, increasing Arousal increased RT and for idioms with high Decomposability, increasing Arousal decreased

RT (see Figure 4). For less familiar idioms, we only found main effect of Idiom Arousal (b = -63.68, SE = 8.16, t = -2.26, p < 0.05); increasing Arousal decreased RT (see Figure 5).

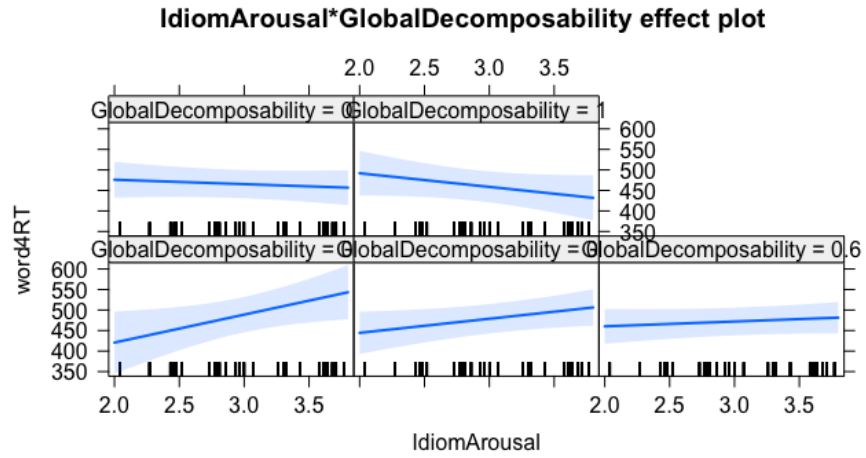


Figure 4. 4 Partial effects plot of RT as a function of Idiom Arousal and Decomposability for high Familiar Idioms.

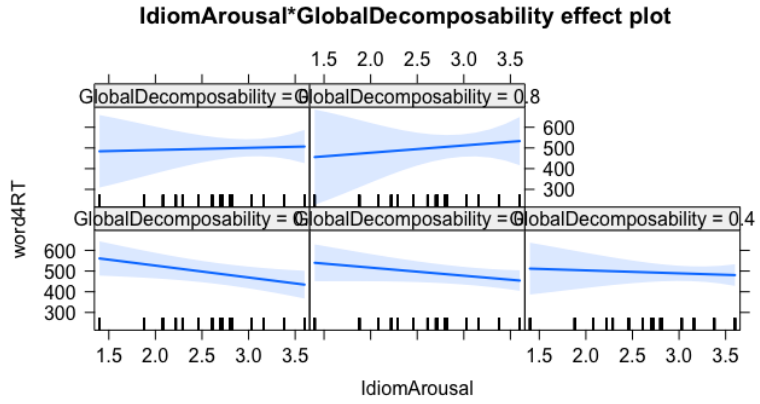


Figure 4.5 Partial effects plot of RT as a function of Idiom Arousal and Decomposability for low Familiar Idioms.

4.4 Discussion and Conclusion

The Multi-Determined model of idiom processing suggests that idioms vary along linguistic characteristics, such as familiarity and decomposability, and these differences impact

their processing (Titone & Libben, 2014; Titone et al., 2019). On the other hand, general models of semantic representation suggest that besides linguistic information, other classes of information (sensory-motor and affective) also contribute to the representation of knowledge. Our results suggest that, generally, when both affective and concreteness factors are considered, affective factors, and not concreteness, exert influence on sentence processing for both literal and idiomatic sentences. Moreover, these results support both models by suggesting that the processing of idioms is impacted by affective factors (i.e., valence and arousal). Importantly, this influence is modulated by idioms' familiarity and decomposability.

We first discuss our findings relating to the comparison between idiomatic and literal conditions. We did this comparison to explore whether, similar to single words, the emotional status and concreteness of the words modulate the processing of longer units of language (sentences) and if this impact is influenced by the final interpretation being literal or figurative. We found that when the sentence contained a verb with positive valence (pleasant) this had a facilitative effect on the processing of literal sentences and an inhibitory effect on the processing of idiomatic phrases. The effect of arousal was similar for both sentence types; more arousing nouns had an inhibitory effect on comprehension. What this suggests is that regardless of sentence type, the emotional status of the constituents impacted sentence processing whereas concreteness did not influence their processing. This is in line with Vigliocco (2009) and Sheikh and Titone (2013)'s findings suggesting that when words have strong association with both affective and sensory-motor information (concreteness), processing will benefit from only one of these sources of information, not both. This is an important finding which shows that overriding the effect of concreteness by the impact of emotional factors, holds for both literal and idiomatic sentence processing. Additionally, when we looked at the RT of the last word of each sentence,

we found the effect of verb valence (the second word of the sentence) and noun arousal (the last word of the sentence) on the processing of sentences. Prior studies on single word processing show that information on emotional properties of the stimuli (arousal) is processed at earlier stages, (~ 200 ms after the presentation of the target), while valence appears to be analyzed later (~ 500 ms after the presentation of the target) (Molinaro, 2020). It seems that, given our task, we captured the effect of verb valence on the RT to words (which appeared more than 500 ms after the verb), but the effect of verb arousal which appears earlier was not captured here. On the other hand, we captured the effect of noun arousal. If the sentence context was larger, we may have been able to see the effect of noun valence on the next words within context. To make any strong claim in this regard, future studies need to use a time sensitive method such as ERPs.

Further, the effect of arousal was the same for both sentence types. In general, there is more consensus on the impact of arousal on word processing: more arousing words capture and hold more attention than low arousing words (e.g., Delaney-Busch et al., 2016). Similarly, at sentence level, and regardless of sentence type, higher level of arousal resulted in slower processing.

Finally, the impact of verb valence had opposite directions for literal and idiomatic conditions. In literal conditions, in line with studies on single words which report facilitative effect of pleasant stimuli on processing, more positive verb valence had a facilitative effect on the processing of literal condition. The literature attributes the delayed processing of stimuli with lower valence (unpleasant stimuli) to "automatic vigilance" (Kuperman, 2014): prolonged attention to unpleasant stimuli, which helps people to avoid engaging in harmful situations or unpleasant social interactions, results in slower RTs (Kuperman, 2014). The reverse effect of verb valence can be attributed to speakers' general expectation about idioms. Studies in the field

of discourse analysis show that speakers tend to use idioms to convey negative events such as complaints (Drew & Holt, 1998, cited in Citron, 2016). Additionally, there are fewer idioms that convey positive meaning (Citron, 2016). This evidence suggests that speakers might develop global expectations for idioms as expressions that are normally used to convey negative concepts. This assumption accords with the theories which assume that knowledge of words and concepts encompass different sources of experiential information, including the emotional state of words (Vigliocco et al. 2009, cited in Sheikh and Titone, 2013). The conflict between the subconscious expectation of speakers about idioms as negative statements, and the positiveness of verb valence, resulted in a processing cost for idioms with positive verb valence. This assumption is consistent with Nicolle and Goel's (2013) findings about valence which argues that valence is "cognitively penetrable" (i.e., it is influenced by the consistency of speakers' overall comprehension of the sentence with their general beliefs). All in all, the current results suggests that constituents' affective characteristics impacts the processing of sentences with both literal and figurative interpretations.

We now, turn to the effect of phrasal level affect and concreteness (i.e., idiom valence, idiom arousal, and idiom concreteness) on the processing of idioms. When we considered data from all idioms, we found that idiom valence impacted their processing and this was modulated by the familiarly level of the idiom. In particular, when valence of the phrase became more positive, it had an inhibitory effect on the processing of less familiar idioms. As mentioned above the inhibitory effect of positive valence for idioms can be related to speakers' general expectations about idioms as negative stimuli. But why do idioms with higher level of familiarity show a reversed effect? As idioms become more familiar, their processing (at least to some degree) becomes non-compositional (e.g., Titone et al., 2019). This allows for faster processing

and integration of the meaning. Thus, for higher familiar idioms there is not enough time for the higher-order knowledge (expectations for idioms being negative) to clash with idiom's positivity. As such, highly familiar idioms benefit from positive valence similar to single words. If our assumption is true, this shows that the processing of high and less familiar idioms are different beyond the reasons stated in previous studies. In other words, the processing differences between them influence how emotional source of information come to play.

Finally, given the observed interaction between familiarity and idiom valence, we separately looked at data with high and low familiarity level. For both high and less familiar idioms, idiom level arousal impacted their processing, however, in opposite directions. When looking at data including only the highly familiar idioms, arousal had an inhibitory effect. Whereas when we looked at data containing only the less familiar idioms, arousal was facilitative. Again, it seems that for highly familiar idioms, whose processing more likely follows a non-compositional route (Titone et al., 2019), the effect of arousal is similar to words: arousal level has an inhibitory effect. Additionally, for highly familiar idioms, the impact of arousal is modulated by decomposability. When idioms are less decomposable, their processing more likely follows a non-compositional route (Titone et al., 2019) and becomes more similar to single words. While we cannot provide robust explanation on why idioms with lower level of familiarity and higher level of decomposability show opposite effects of the impact of valence and arousal, we can confidently suggest that the processing of idioms are influenced by their emotional status. Importantly, this influence depends on how linguistic characteristics of idioms foster either compositional or non-compositional analysis of idioms.

To conclude, we investigated idiom comprehension while considering the impact of word-level and phrase-level affective and sensory-motor factors (valence, arousal, and

concreteness) on their processing. The result suggested that that whenever idioms have characteristics which foster non-compositional analysis (higher familiarity and lower decomposability) the impact of emotion becomes similar to what is usually reported for single words: inhibitory effect of high arousal and facilitative effect of positive valence. However, when idioms lack the characteristics which facilitates direct retrieval of their meaning (i.e., they are very low in familiarity level or they are highly compositional) the impact of emotional status of their processing is different. These findings are consistent with hybrid models of idiom comprehension (e.g., Libben & Titone, 2008; Titone & Libben, 2014; Titone et al., 2019), which assert that idiom comprehension follows two routes: direct retrieval of highly familiar, non-decomposable idioms and compositional analysis of less familiar decomposable idioms. The results also confirm models of semantic representation which suggest that beside linguistic information, affective sources of information influence meaning representation and language processing (e.g., Vigliocco et al., 2009). All in all, while previous models such as the multi-determined model (Libben and Titon, 2008) point to the importance of linguistic information during idiom comprehension, here we showed that affective factors appear to be impact their processing while they are essentially modulated by linguistic information.

We acknowledge that there are some caveats in our methodology. Most importantly, in this study, we have used very short sentence contexts while we looked at the last word of each sentence as the target word. The short sentence context inhibited us from including any spill-over region and looking at words after the target words for possible effect of the studied factors. Also when the last word of a sentence is used as the target word, the result might be influenced by the wrap-up effect. Although these caveats might restrict generalizability of our findings, we believe it does not invalidate the main findings of the current study.

One line of future directions is to consider a larger sentence context where the target words appear in non-final position and thus is immune from any potential wrap-up effect (though see Stowe et al, 2018 for another view). Also, when idioms appear in a supporting context, it might be possible to find the impact of current studied factors at earlier stages of processing, for example on the verb. Moreover, in future studies, high temporal techniques such as ERPs or a technique that would allow for looking at different stages of processing such as eye-tracking should be used to be able to track the processing of idioms while considering the role of affective and sensory-motor information.

References

Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and language, 125*(1), 47-53.

Bates, D., Machler, M., Bolker, B. & Walker, S. (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1-48.

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

Bower, G. H. (1992). How might emotions affect learning. *The handbook of emotion and memory: Research and theory*, 3, 31.

Bransford, J. D., & McCarrell, N. S. (1977). A sketch of a cognitive approach to comprehension: Some thoughts about understanding what it means to comprehend. In Johnson-Laird, P. N., and Wason, P. C. (eds.), *Thinking: Readings in Cognitive Science*, Cambridge University Press, Cambridge, UK, pp. 377-399.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, *46*(3), 904-911.

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

Citron, F. 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.

Citron, F. M., Weekes, B. S., & Ferstl, E. C. (2014). Arousal and emotional valence interact in written word recognition. Language, *Cognition and Neuroscience*, *29*(10), 1257-1267.

Clore, G. L., Ortony, A., & Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of personality and social psychology*, *53*(4), 751.

Delaney-Busch, N., & Kuperberg, G. (2013). Friendly drug-dealers and terrifying puppies: Affective primacy can attenuate the N400 effect in emotional discourse contexts. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(3), 473-490.

Drew, P., & Holt, E. (1998). Figures of speech: Figurative expressions and the management of topic transition in conversation. *Language in society*, *27*(4), 495-522.

Findlay, H., & Carrol, G. (2019). Contributions of semantic richness to the processing of idioms. *The Mental Lexicon*, *13*(3), 311-332.

Fox, J., & Weisberg, S. (2018). Visualizing fit and lack of fit in complex regression models with predictor effect plots and partial residuals. *Journal of Statistical Software*, *87*, 1-27.

Fraga, I., Padrón, I., Acuña-Fariña, C., & Díaz-Lago, M. (2017). Processing gender agreement and word emotionality: New electrophysiological and behavioural evidence. *Journal of Neurolinguistics*, *44*, 203–222.

Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of memory and language*, *28*(5), 576.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in cognitive sciences*, *9*(9), 416-423.

Hinojosa, J. A., Moreno, E. M., & Ferré, P. (2020). Affective neurolinguistics: towards a framework for reconciling language and emotion. *Language, Cognition and Neuroscience*, *35*(7), 813-839.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological

investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 721.

Holt, D. J., Lynn, S. K., & Kuperberg, G. R. (2009). Neurophysiological correlates of comprehending emotional meaning in context. *Journal of cognitive neuroscience*, *21*(11), 2245-2262.

Kieras, D. (1978). Beyond pictures and words: Alternative information-processing models for imagery effect in verbal memory. *Psychological Bulletin*, *85*(3), 532.

Kim, J. M., Sidhu, D. M., & Pexman, P. M. (2020). Effects of emotional valence and concreteness on children's recognition memory. Frontiers in Psychology, 11, 615041.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14.

Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473-481.

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: valence and arousal affect word recognition. *Journal of experimental psychology. General, 143*(3), 1065–1081. https://doi.org/10.1037/a0035669

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological bulletin*, *139*(4), 917.

Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: Importance of word arousal. *Emotion*, *8*(4), 445–452. https://doi.org/10.1037/1528-3542.8.4.445

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

Martín-Loeches, M., Fernández, A., Schacht, A., Sommer, W., Casado, P., Jiménez-Ortega, L., & Fondevila, S. (2012). The influence of emotional words on sentence processing: Electrophysiological and behavioral evidence. *Neuropsychologia, 50*(14), 3262-3272.

Molinaro, N. (2020). "Words and emotions in sentence context": a commentary on Hinojosa, Moreno and Ferré (2019). *Language, Cognition and Neuroscience, 35*(7), 862-864.

Morid, M., Bachar, N., & Sabourin, L. (2021). Capturing the multi-determined nature of idiom processing using ERPs. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 75*(2), 155.

Nasrallah, M., Carmel, D., & Lavie, N. (2009). Murder, she wrote: enhanced sensitivity to negative word valence. *Emotion, 9*(5), 609.

Newcombe, P. I., Campbell, C., Siakaluk, P. D., & Pexman, P. M. (2012). Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns. *Frontiers in human neuroscience, 6*, 275.

Nicolle, A., & Goel, V. (2013). What is the role of ventromedial prefrontal cortex in emotional influences on reason. *Emotion and reasoning*, 154.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(3), 255.

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://support.pstnet.com/.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, *110*(1), 145.

Sabourin, L., Leclerc, J. C., Lapierre, M., Burkholder, M., & Brien, C. (2016). The language background questionnaire in L2 research: Teasing apart the variables. In Proceedings of the *Annual Meeting of the Canadian Linguistics Association, Calgary, Canada.*

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly*, 114-126.

Sheikh, N. A., & Titone, D. A. (2013). Sensorimotor and linguistic information attenuate emotional word processing benefits: an eye-movement study. *Emotion, 13*(6), 1107.

Snefjella, B., & Kuperman, V. (2016). It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition, 156*, 135-146.

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of verbal learning and verbal behavior, 18*(5), 523-534.

Teng, D. W., Wallot, S., & Kelty-Stephen, D. G. (2016). Single-word recognition need not depend on single-word features: Narrative coherence counteracts effects of single-word features that lexical decision emphasizes. *Journal of Psycholinguistic Research, 45*(6), 1451-1472.

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

Titone, D., & Libben, M. (2014). Time-dependent effects of decomposability, familiarity and literal plausibility on idiom priming: A cross-modal priming investigation. *The Mental Lexicon, 9*(3), 473-496.

Titone, D., Lovseth, K., Kasparian, K., and Tiv, M. (2019). Are figurative interpretations of idioms directly retrieved, compositionally built, or both? Evidence from eye movement

measures of reading. *Canadian Journal of Experimental Psychology/Revue canadienne de Psychologie Experimentale , 73* (4), 216.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition, 1*(2), 219-247.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods, 45*(4), 1191-1207.

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *Journal of Cognitive Neuroscience, 12*(6), 1024-1037.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological bulletin, 120*(1), 3.

5 General Discussion and Future Directions

In this dissertation, I addressed the processing of idioms from two angles. First, I explored the neurophysiological underpinnings during idiom comprehension. In specific, I investigated whether the factors that have been shown to impact the processing of idioms at behavioural level, modulate the neurophysiological underpinnings associated with processing of different types of idioms (Chapter 2). Considering that the field of language processing, in general, is increasingly considering the relation between linguistic information and other sources of experiential information (sensory-motor and affective), I dedicated the rest of this dissertation to provide the required sensory-motor and affective norms for a set of English idioms (Chapter 3) and to explore the role of those factors during idiom processing in a self-paced reading task (Chapter 4). In this chapter, I will give a brief summary of the main findings. Next, I will present a proposal on how this line of research can be continued in the future.

5.1 Summary of the Findings

In the first study (Chapter 2), we looked at idioms with various levels of familiarity and literal plausibility and compared the electrophysiological underpinnings associated to the processing of those idioms. Consistent with the predictions of constraint-based approaches (e.g. the multi-determined model of idiom processing; Libben & Titone (2008)), we found that the level of familiarity with idiomatic expressions results in a processing advantage. We also found a processing advantage for idioms with a plausible literal interpretation. Importantly, the impact of these factors did not start at the same time (familiarity was shown to impact earlier stages of processing). While the influence of these factors has received support in many behavioural studies, this study was the first to provide neurophysiological support for the multi-determined model of idiom processing.

The field of psycholinguists has recently attended to the relation between language and emotion (van Berkum, de Zubicaray, & Schiller, 2019). According to the Affective Language Comprehension (ALC) model, "emotion pervades every step of the language comprehension process" (van Berkum, de Zubicaray, & Schiller, 2019, p. 756) and thus, as these same authors argue, the field of psycholinguistics "cannot afford to ignore" (p. 738) the relation between language and emotion. Given the importance of considering the role of emotion (and sensory-motor information) during language comprehension, in the second study (Chapter 3), we developed a large scale survey and asked both native speakers of English and proficient L2 speakers of English (N = 550) to rate 210 English idioms for valance, arousal, concreteness, and imageability. Besides providing the norms, we explored various types of relations between the collected measures (e.g., between affective and non-affective measures) as well as the relation between collected norms by L1 and L2 speakers. For example, we showed that the same type of U-shape relation between valence and arousal, that is usually reported for single words, holds for idioms and this relation holds for both L1 and L2 data. Moreover, we found that L1 and L2 ratings were similar for some ratings (valence and imageability) and different for some others (arousal and concreteness). This database provides a useful tool for researchers to explore the relation between linguistic and other sources of information during idiom processing.

In the last study (Chapter 4), we explored how the emotional status of idioms and their concreteness contribute to their processing, and whether this contribution is modulated by idiom familiarity. We found that the impact of non-linguistic sources of information (affective and sensory-motor) is determined by idiom familiarity, such that low familiar and high familiar idioms show different behaviour. For high familiar idioms, the behaviour aligns with the findings on word processing. For example, high familiar idioms with more positive valence showed

facilitative processing. On the other hand, valence had an inhibitory impact on idioms with low familiarity level such that greater valence increased the reading time.

5.2 Contribution of This Dissertation to the Cycle of Research

In order to advance science, every field needs both theory building (inductive) and theory testing (deductive) research (Bhattacherjee, 2012), to complete the cycle of research, as shown in the following figure.

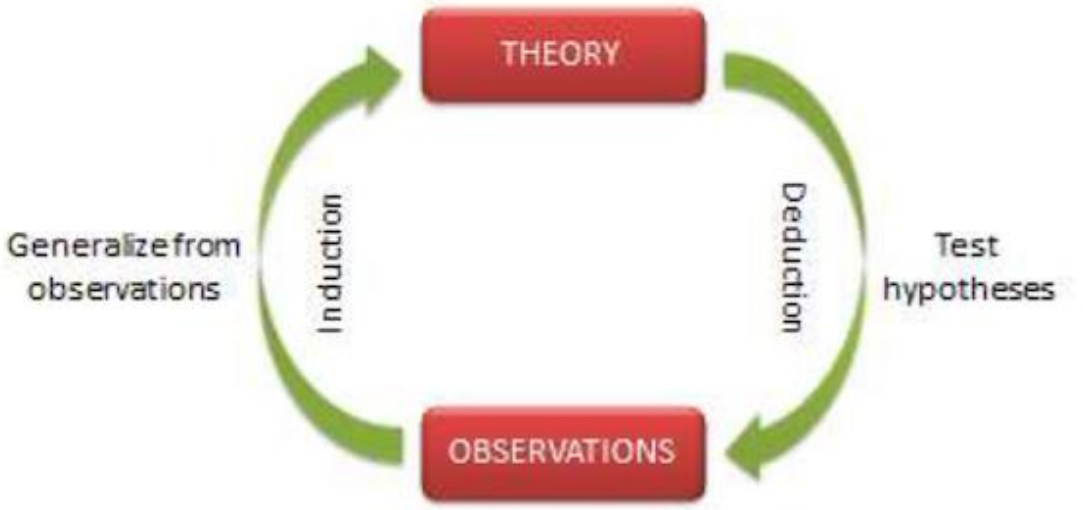


Figure 5.1 The cycle of research, adapted from (Bhattacherjee, 2012)

The first part of this research was deductive, when we provided neurophysiological evidence in support of previously established models of idiom processing. The second part of the current dissertation was conducted with the hope to provide both tools and evidence for developing a new generation of idiom processing models; a model which considers the role of non-linguistic (specifically affective) information in the processing of idioms. While such models/hypotheses has already been proposed for language processing (e.g., Vigliocco et al., 2009), to the best of our knowledge, such a model for idiom processing has not yet been fully developed. We

acknowledge that the current work is still not at the stage of model building, but we have provided useful tools and evidence to pave this way.

5.3 Implications and Future directions

We hope that the findings of this dissertation inform the researchers about the importance of incorporating both linguistic and non-linguistics factors when they study idiom processing as only by doing so, we will be able to unravel the extent to which understanding language relies on the interaction between mind, body, and emotion. Using multiword expressions that have non-literal interpretations is a particularly compelling and novel way to investigate embodied accounts of language processing. If a word used in a non-literal way recruits the same sensory-motor region of the brain as their literal counterpart, and if the emotional status of multiword expressions impacts their comprehension above and beyond the emotional status of individual words this provide strong evidence in support of association between linguistic and experiential information in understanding language.

One way to expand the current studies is to use ERPs to investigate the neurological underpinning during idiom processing while manipulating affective (valence and arousal) and sensory-motor (concreteness) level of idioms. Using ERPs give us important insight on when sensory and motor information come to play during figurative comprehension. For example, if researchers observe that the activation of sensory-motor areas happens *early* (around 250 ms), then this activation is automatic and is an integral part of semantic processing. If the activation happens during the N400 or P600 time window, then the contribution of sensory-motor information is related to semantic retrieval and/or semantic integration.

To extend the finding of this dissertation, the impact of the same factors (in specific emotional and sensory-motor factors) as understudies factors in the area of L2 idiom processing.

Moreover, by considering L2 speakers with various language profile (e.g., early and late L2 speakers; short versus long immersion in L2 environment; low and high proficiency) we will be able to uncover at what level two languages may share representation and organization and how these are impacted by non-linguistic factors.

Finally, future studies need to consider the role of non-linguistic factors while embedding idiomatic expressions in a larger context. As mentioned earlier, in our last study, we did not find impact of noun arousal, but it could have been related to our short sentences. By providing readers with a larger context, we will be able first investigate how a supporting context may interact with the influence of non-linguistic factors, and second, whether the impact of some of these factors might be observable on the words that come after the idiomatic expressions.

References

Barsalou, L. W. (2008). Grounded cognition. *Annual review of psychology*, *59*(1), 617-645.

Bhattacherjee, A. (2012). *Social science research*. USF Tampa Bay Open Access Textbooks, Tampa, FL.

Dudschig, C., De la Vega, I., & Kaup, B. (2014). Embodiment and second-language: Automatic activation of motor responses during processing spatially associated L2 words and emotion L2 words in a vertical Stroop paradigm. *Brain and language, 132*, 14-21.

Ghio, M., & Tettamanti, M. (2016). Grounding sentence processing in the sensory-motor system. In *Neurobiology of language* (pp. 647-657). Academic Press.

Gianelli, C., & Fischer, M. H. (2015). Motor measures of embodiment: The case of number concepts. In *Conceptual and interactive embodiment* (pp. 107-119). Routledge.

Gibbs Jr, R. W. (2013). The real complexities of psycholinguistic research on metaphor. *Language Sciences, 40*, 45-52.

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

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788-804.

Schmidt-Snoek, G. L., Drew, A. R., Barile, E. C., & Agauas, S. J. (2015). Auditory and motion metaphors have different scalp distributions: an ERP study. *Frontiers in human neuroscience*, *9*, 126.

van Berkum, J. J. A. (2019). Language comprehension and emotion: Where are the interfaces, and who cares? In G. I. de Zubicaray, & N. O. Schiller (Eds.), *The Oxford Handbook of neurolinguistics* (pp. 735–766). Oxford University Press.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition, 1*(2), 219-247.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General, 135*(1), 1.

Appendices

Appendix A: Experimental stimuli for ERP study (Chapter 2)

Sentence	Critical word
After receiving a promotion, John was walking on sunshine all day.	sunshine
As the oldest of her siblings, Mary ruled the roost when her parents weren't home.	roost
John was unsure at his new job, but he learned the ropes quickly.	ropes
John wasn't sure where the building was, but he hedged his bets and turned right.	bets
Mary's friend forgot to bring her wallet and Mary had to foot the bill for dinner.	bill
After being scratched by a cat, the dog licked its wounds under the tree.	wounds
After the storm, John and Mary mended their fences with their neighbours.	fences
John's dog doesn't like his medication, so John sugared the pill to help him.	pill
Mary just got her braces off and showed her teeth to her friends.	teeth
Peter wouldn't start the car so John forced his hand and they drove away.	hand
After being sick for months, John kicked the bucket last night.	bucket
After keeping her project a secret, Mary accidentally showed her cards to John.	cards
After showing up late to work for months, John got the sack yesterday.	sack
John was fired after the bank manager discovered he had been lining his pockets for month	pockets
John wasn't sure about the job offer, so he chewed the cud before deciding.	cud
Everyone came to the party because Mary's friends had spread the word about her birthday	word
John found out he had failed the math test and threw a fit after class.	fit
Mary was very upset, so John cracked a joke to cheer her up .	joke
The dog wouldn't stop barking while John slept. It drove him nuts every night.	nuts
When Mary didn't return his phone call, John took a hint and stopped calling her eventually.	hint
John and Mary were bored, but Mary had a ball that they could play with.	ball
John wasn't able to tie his shoes, so Mary taught him to tie the knot properly.	knot
John's front steps were slippery after the snowstorm, so he broke the ice with a shovel.	ice
John's hands were full with groceries, so he used his head to hold the door open.	head
Mary was excited for her trip to France. She packed her bags a week in advance.	bags
Jane said she didn't want to go to the party, and Mary got the message finally.	message
John baked a cake. Everyone said it hit the spot at the potluck.	spot
John didn't need Mary to explain why she ignored his calls. He got the picture eventually.	picture
John has a test tomorrow, so he hit the books for hours.	books
Mary was tired. When she got home, she hit the sack until the next day.	sack

Appendix B: A snapshot of norms of valence, arousal, concreteness, and imageability for 210 English idioms by L1 and L2 speakers (Chapter 3).

L1 ratings

	Arousal			Valence			Concreteness			Imageability		
Idiom	Mean	SD	ValidresponseF	Mean	SD	Validrespons	Mean	SD	Validrespons	Mean	SD	Validrespons
He beat his breast	2.12	1.83	60.71	-0.11	1.01	51.85	4.21	2.15	57.58	5.00	1.92	83.33
She raised the devil	3.60	2.16	89.29	-1.89	0.95	88.89	3.03	1.92	90.91	4.17	1.59	95.83
He showed the flag	1.62	1.69	75.00	0.15	1.98	59.26	3.05	1.35	69.70	4.39	2.15	75.00
He lost the thread	1.67	1.61	85.71	-0.67	1.81	66.67	3.32	1.75	78.79	4.09	1.88	95.83
He saved his bacon	2.25	2.19	85.71	1.19	1.05	77.78	2.71	1.63	87.88	4.52	1.99	87.50
He battled the storm	2.22	1.97	96.43	0.89	2.07	92.59	3.78	2.14	96.97	5.08	1.56	100.00
He curled the lip	2.70	1.61	82.14	-0.59	1.83	59.26	4.84	1.62	78.79	5.24	2.05	87.50
He pulled his punches	2.80	1.61	89.29	-0.04	1.67	74.07	3.27	1.70	78.79	3.86	1.85	87.50
He got his goat	1.94	1.86	64.29	0.07	2.25	51.85	3.00	1.62	57.58	4.30	1.92	83.33
He got a toehold	1.53	1.68	53.57	-0.11	1.58	51.85	3.65	1.58	54.55	3.22	1.96	75.00
He cut the mustard	1.40	1.76	53.57	0.00	1.36	51.85	2.43	1.65	45.45	3.38	1.67	66.67
They mended their fences	2.64	1.78	89.29	2.11	0.84	92.59	3.76	1.74	84.85	4.48	2.09	95.83
He wagged his tongue	2.50	2.11	78.57	0.56	1.50	66.67	4.56	2.16	75.76	4.95	1.86	87.50
She gnashed her teeth	2.88	1.86	89.29	-1.41	1.03	81.48	5.80	0.68	78.79	5.44	1.72	75.00
He overplayed his hand	3.43	1.57	100.00	-1.63	0.96	92.59	3.30	1.68	93.94	3.73	1.70	91.67
He greased the wheels	2.08	1.72	92.86	0.85	1.07	70.37	4.64	1.85	84.85	4.81	2.06	87.50
He walked on sunshine	3.36	2.13	100.00	2.44	1.30	96.30	3.47	2.42	96.97	4.91	1.63	91.67
He bought the farm	2.18	2.04	78.57	0.37	1.66	62.96	3.56	1.58	63.64	4.75	1.97	83.33
He ruled the roost	2.19	1.38	57.14	0.26	1.51	44.44	2.44	1.46	63.64	3.30	1.89	83.33
He split a gut	2.08	2.17	85.71	-0.26	2.11	74.07	3.47	1.61	66.67	3.81	1.81	87.50
She swept the board	2.61	2.10	82.14	1.56	1.10	77.78	4.50	1.83	90.91	5.00	1.66	91.67
He dropped a brick	2.41	1.59	78.57	-0.85	1.18	77.78	4.44	1.91	81.82	5.19	2.18	87.50
He learned the ropes	2.36	1.55	89.29	1.89	1.04	96.30	3.18	1.68	96.97	4.65	1.67	95.83

L2 ratings

	Arousal			Valence			Concreteness			Imageability		
Idiom	Mean	SD	Validrespons	Mean	SD	Validrespons	Mean	SD	Validrespons	Mean	SD	ValidresponsePercent
He beat his breast	4.17	1.38	69.02	-2.03	2.21	86.63	4	1.39	50.00	4.5	1.07	73.62
She raised the devil	0.13	2.02	84.47	-1.57	2.20	50.35	5.6	1.75	83.33	4.75	0.90	53.25
He showed the flag	1.67	1.42	11.50	-1.29	1.61	64.36	3.4	1.66	83.33	6.75	0.90	90.49
He lost the thread	1.4	2.04	89.03	-1.57	2.29	29.72	4.4	1.84	83.33	4.75	1.22	47.47
He saved his bacon	0.33	1.55	18.44	0.6	2.27	40.14	3.67	1.31	50.00	5	1.43	8.18
He battled the storm	1.38	1.60	92.23	1.14	1.54	18.46	5.5	1.52	100.00	3	0.94	75.90
He curled the lip	1.88	1.78	33.22	-1.43	1.62	22.92	4.2	1.24	83.33	4.75	1.59	90.80
He pulled his punches	0.71	1.87	75.29	0.71	1.67	82.13	7	1.00	50.00	4.33	1.04	92.73
He got his goat	1	1.51	0.51	-1.01	1.92	14.04	4.75	0.95	66.67	4.33	1.64	51.12
He got a toehold	1.2	1.22	42.11	-0.14	2.18	77.82	2.67	1.88	50.00	3	1.77	76.14
He cut the mustard	2.33	1.14	60.96	0.25	1.41	80.61	3.5	0.93	66.67	3.67	1.03	52.01
They mended their fences	2.5	1.91	25.60	0.83	1.68	2.72	4.6	1.61	83.33	4	1.50	42.33
He wagged his tongue	2.29	1.63	37.29	-1.33	1.94	28.16	5.67	1.73	50.00	4.67	0.97	26.60
She gnashed her teeth	1.29	2.05	17.03	-2	2.00	66.13	4	1.64	100.00	4.6	1.47	66.81
He overplayed his hand	3.43	1.71	4.05	-2.29	1.31	6.21	4	1.49	50.00	2.67	1.22	28.40
He greased the wheels	3.33	1.94	61.65	-0.14	2.03	22.35	5	1.53	50.00	6	1.87	11.59
He walked on sunshine	3.86	1.46	8.40	2.71	1.98	44.04	3.83	1.40	100.00	5	1.04	36.51
He bought the farm	0.97	1.41	52.18	-0.8	1.89	57.41	4.25	1.88	66.67	5.33	1.60	41.31
He ruled the roost	1.67	1.95	65.01	0.5	1.45	51.91	3.75	1.80	66.67	2.67	1.19	74.01
He split a gut	1.67	1.46	94.49	-0.2	1.94	44.05	4	1.68	33.33	3.25	1.17	93.88
She swept the board	1.71	1.56	97.82	0.01	2.26	96.15	3.4	1.25	83.33	4.25	1.78	32.17
He dropped a brick	1.95	1.24	40.00	-1.09	2.06	17.46	4.8	1.36	83.33	5.8	1.23	51.88
He learned the ropes	2.43	1.68	68.36	0.29	1.56	22.32	4	1.49	100.00	4.6	1.04	64.83

Appendix C: Experimental stimuli for ERP study (Chapter 4)

IDIOMS					LITERAL			
He	beat	his	breast.		He	beat	his	leg.
He	broke	the	ice.		He	broke	his	head
She	broke	her	word.		She	broke	her	leg.
She	called	the	shots.		She	called	for	questions.
He	carried	the	torch.		They	carried	the	flag.
She	changed	her	mind.		He	changed	his	pants.
He	clipped	her	wings.		They	changed	the	show.
She	cooked	his	goose.		He	clipped	the	thorns.
She	covered	her	tracks.		She	cooked	the	wings.
He	cracked	a	joke.		He	covered	the	toehold.
He	curled	the	lip.		He	curled	the	board.
He	dropped	a	line.		He	dropped	the	pants.
She	dusted	his	pants.		He	dusted	his	bag.
She	gave	her	word.		She	gave	the	eggs.
She	got	his	goat.		She	gave	a	hint.
She	got	the	message.		He	got	the	ball.
He	got	the	picture.		She	got	the	spot.
He	got	a	toehold.		She	greased	her	lips.
He	greased	the	wheels.		He	had	a	brain.
She	had	a	ball.		She	had	a	habit.
She	had	a	lark.		He	had	that	habit.
He	hit	the	books.		He	had	a	question.
She	hit	the	sack.		It	had	a	roost
It	hit	the	spot.		They	had	a	torch.
She	kicked	some	butt.		He	hit	her	breast.
He	kicked	the	habit.		He	hit	his	butt.
She	killed	the	time.		He	hit	the	fence.
She	laid	an	egg.		She	hit	his	head.
He	lifted	her	spirits.		He	kicked	the	goose.
He	lost	his	cool.		He	killed	a	lark.
He	lost	his	nerve.		It	laid	an	egg.
He	lost	the	thread.		He	lifted	the	sack.
They	mended	their	fences.		He	lost	his	bag.
She	packed	her	bags.		She	lost	her	books.
They	sat	on	thorns.		He	mended	his	sail.
He	shook	a	leg.		She	packed	the	beans.
He	showed	the	flag.		It	sat	on	nest.
It	slipped	his	mind.		She	shook	her	butt.
She	spilled	the	beans.		They	showed	the	track.
She	spilled	her	guts.		He	slipped	on	ice.
He	split	a	gut.		She	spilled	the	eggs.
She	spoke	her	mind.		He	spilled	the	mustard.
They	spread	the	word.		She	split	the	line.
She	stole	the	show.		He	spoke	some	words.
He	sugared	the	pill.		She	spread	the	cream.
He	swallowed	his	pride.		He	stole	the	gauntlet.
She	swept	the	board.		She	swallowed	her	pill.
He	took	a	beating.		She	swept	the	track.
He	took	a	hint.		She	took	some	bacon.
He	took	the	plunge.		She	took	a	picture.
She	took	the	veil.		She	took	some	beans.
She	trimmed	her	sails.		She	trimmed	her	brow.
She	used	her	head.		She	used	that	word.
He	won	the	day.		He	won	the	bet.