Is a Job (Like) a Jail? Differences in Metaphor Versus Simile Processing and Comprehension in L1 and L2 English Speakers

Iola Patalas

A Thesis
In the Department of Psychology

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Psychology) at
Concordia University
Montreal, Quebec, Canada

April 2024

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: *Iola Patalas*

Entitled: *Is a Job (Like) a Jail? Differences in Metaphor Versus Simile Processing and Comprehension in L1 and L2 English Speakers*

and submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY *(Psychology)*

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Xavier Ottenwaelder

_____ External Examiner
Dr. Albert Katz

_____ Examiner
Dr. Andrew Chapman

_____ Examiner
Dr. Norman Segalowitz

_____ Examiner
Dr. Natalie Phillips

_____ Co-Supervisor
Dr. Andreas Arvanitogiannis

_____ Thesis Supervisor
Dr. Debra Titone

Approved by _____
Dr. Andrew Chapman, Graduate Program Director

06/13/2024 _____
Dr. Pascale Sicotte
Dean of the Faculty of Arts & Science

Abstract

Is a Job (Like) a Jail? Differences in Metaphor Versus Simile Processing and Comprehension in L1 and L2 English Speakers

Iola Patalas, Ph.D.
Concordia University, 2024

When listeners hear a metaphor such as *jobs are jails*, they typically understand the intended meaning even though such sentences are not literally true. Within psycholinguistics, there has been a lingering debate over how such phrases are understood, and how they differ from other related forms, such as similes (e.g., *jobs are like jails*). Accordingly, pragmatic theorists suggest that literal meaning of metaphors must first be rejected in order to attain understanding of intended metaphorical meaning from context and world knowledge. In contrast, direct-access theorists argue that both metaphor and simile are understood automatically, without prior parsing of literal semantic meaning. Relevant here, relatively little research has ascertained what meanings are attained during the time-course of metaphor or simile processing, and further, how this differs between first and second language speakers (henceforth, L1 and L2, respectively). To pursue this issue, this thesis presents three studies that investigate both the moment-by-moment online processing of metaphors and similes and their ultimate comprehensibility, in both L1 and L2 English speakers. STUDY 1 describes two cross-modal lexical decision experiments spanning four time points during the course of processing (vehicle word onset, vehicle recognition point, 500ms and 1000ms post-recognition point) in two samples of L1 English speakers, comparing the priming of literal and figurative meanings of metaphors and similes in high- and low- aptness and familiarity conditions. It showed that aptness and familiarity modulated which meanings were activated during metaphor and simile processing, and that literal meanings were activated faster and lingered later than figurative meanings. STUDY 2 repeated the same experimental design in a sample of L2 English speakers, with four time points collapsed into two (early and late) to determine whether L2 speakers process metaphors the same way L1 speakers do. We demonstrated that L2 speakers did not appreciably prime figurative or literal meanings during online processing of metaphors, and only primed literal meanings while processing similes. STUDY 3 probed whether L1 and L2 speakers found metaphors and similes globally comprehensible when aptness and familiarity were manipulated and when given ample time to make offline judgments about these sentences. It found that L1 speakers judged highly familiar metaphor more comprehensible than similes that had the same constituents. However, L2 speakers preferred simile when sentence familiarity was high or aptness was low. Together, STUDIES 1 to 3 highlight that online processing and offline comprehension of metaphor and simile differed according to language background and sentence attributes. Specifically, familiarity was important for both online processing of metaphor and simile for L1 speakers. In contrast, L2 speakers relied more heavily on semantic decomposability to make sense of figurative expressions.

Acknowledgements

I extend my profound gratitude to Dr. Debra Titone, who provided endless guidance about the construction of my thesis, supported me through tough times, and helped me understand how to make my research questions meaningful and speak about them confidently. I thank her for encouraging me when I needed it most and I thank her for her patience and faith in me. I also deeply thank Dr. Marco Silvio Guiseppe Senaldi, who taught me almost everything I know about statistics, assisted me in developing good experiments, and provided endless moral support.

This thesis also would not have been possible without the many years of guidance from Dr. Roberto de Almeida, who helped me construct thorough and novel experimental protocols, underscored the importance of intellectual rigour, and taught me not to be afraid of going against the tide. I thank him for his support in presenting my early findings at conferences and teaching me important lessons about cognition and experimental design.

I could not have completed this work without the guidance of Dr. Norman Segalowitz and Dr. Natalie Phillips, whose expertise helped me distill my research questions and whose kindness allowed me to find a path forward. Thank you also to Dr. Andreas Arvanitogiannis, who took it upon himself to make sure I could continue my research in times of uncertainty.

My thanks also go out to all the volunteers who contributed to data collection and consolidation, particularly Kristen Charles, Caitlyn Antal, Kyan Salehi, Alexa Falcone, Isabelle Oppong, and Nathan Boulianne.

Finally, I thank my family for their patience and support, and Michael Johancsik for listening to me patiently, lending me his computer, and helping me collect my thoughts whenever I needed someone to bounce ideas off of.

Dr. de Almeida contributed to the experimental design of STUDIES 1 and 2, and data collection for STUDIES 1 and 2 was performed in his lab at Concordia University. The experimental design of STUDY 3 was conceived together with Drs. Titone and Senaldi and data collection was performed at Dr. Titone's lab at McGill University. Drs. Titone and Senaldi provided guidance for the thesis structure and data analysis of STUDIES 1 to 3, and also provided comments on early drafts of this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Human communication is often perceived as effortless, but this subjective impression hides a potentially complex set of cognitive capacities and operations that make successful comprehension possible (see Katz, 2024; Perlovsky, 2009). For example, when we hear a sentence such as *my lawyer is a shark*, we understand almost immediately what the speaker means, even though lawyers, by definition, cannot be sharks. Sharks do not earn law degrees, and becoming a lawyer is a uniquely human endeavour. So how is it possible that a hearer understands what a speaker means by this sentence – and what cognitive processes lead to this understanding? Crucially, given that people communicate with each other using first languages learned from birth, and second languages learned subsequently, do these cognitive processes differ for first- and second-language speakers (henceforth, L1 and L2, respectively)? Does it matter at what age an L2 is learned, or do people with native-like fluency in an L2 learned slightly later in life still differ from L1 speakers in their cognitive processing of a given language?

Such questions are central to understanding human language and cognitive function generally. Moreover, they can manifest across many different levels of language and across a wide variety of linguistic phenomena, though crucially, in this dissertation I focus on one particular aspect of language – nonliteral language. Nonliteral language encompasses many different kinds of linguistic phenomena, including metaphors such as *my lawyer is a shark* as well as irony, sarcasm, metonymy, and idiom (Katz, 2024). It is an important branch of the language sciences because it is common in natural language (Colston & Gibbs, 2021; Lakoff & Johnson, 1980), and highlights the ways in which human language users can easily go beyond component word meanings to build an interpretation that is much broader than the purely literal,

propositional meaning, or even one that is entirely different. While nonliteral language is often thought of as the domain of literature, rhetoric, or poetry, it is ubiquitous in human speech and can help convey meaning that is not as easily, or accurately, conveyed by literal speech alone (Colston & Gibbs, 2021). Metaphors are known to appear as often as every 25 words in both written and spoken language (Fuks, 2021), indicating that they are an integral, and likely necessary, component of human communication. As Lakoff and Johnson (1980) famously wrote, "metaphor is pervasive in everyday life, not just in language but in thought and action." Indeed, perhaps metaphor is so pervasive in language because it reflects how we think and understand ourselves and the world around us. Metaphor can be used to communicate abstractions, social concepts, and emotions when literal language fails (see Caesanto & Bottini, 2014; Maass et al., 2014; Meier et al., 2014), and this may reflect fundamental cognitive processes.

Beyond simply reflecting the inherent realities of human cognition, metaphor use in wider society can also be used to reframe how we think and direct human behaviour and relationships. Maass et al. (2014) described how metaphors commonly used in the media and in everyday conversation can deepen biases and social inequities by entrenching associations between social outgroups and negative stereotypes – the word *beast* used to describe certain types of criminals, for instance, can dehumanize them and thus subtly affect attitudes about how criminals should be dealt with in society, changing the types of punishments considered acceptable for certain crimes. Giving imagistic qualities to pre-existing stereotypes can further entrench those stereotypes and make them feel more real or concrete – most consistently where vertical metaphors for power structures are concerned, such as concepts of *upper class* and *lower class* (Maass et al., 2014).

Metaphors about brightness and darkness or temperature are often used to convey information about mental states and emotional valence, which can alter social perception and communication (Meier et al., 2014). Similarly, metaphorical expressions about spatial relationships may ease memory encoding and retrieval of abstract educational concepts by making them appear more concrete, like associating rightward flow of time and space (Caesanto & Bottini, 2014). Caesanto and Bottini (2014) also argued that metaphor use can also change the way people think about abstract concepts over time.

Such concrete uses make metaphor a powerful tool for expressing and understanding a number of abstractions in both educational and social contexts. This can have significant consequences for individuals and society which are not often consciously considered – or even noticed – by the people who employ them and are affected by them. For instance, scientists are increasingly studying the effects of certain types of metaphor use on medical practice (e.g., Declercq & van Poppel, 2024; Fuks, 2021; Munday et al., 2022; Wackers et al., 2021), not just in an individual patient-practitioner context, but also in the context of the entire focus of medical research and practice itself (Fuks, 2021). Fuks (2021) explains that military metaphor (e.g., the *war on cancer*, or on a personal level *battling cancer*) has pervaded both patient-provider interactions and the medical literature, and that this status quo has deleterious effects on both individual patient care – by setting the patient against their own body and convincing doctors that they are operating on or treating an enemy in the form of the physical being of the patient – and the larger research environment – by divorcing holistic patient care and experience of illness from the larger *battle* against the nonspecific environment of warfare against a general concept of disease. This framing can both have deleterious effects on patient health and recovery – because it dehumanizes patients in their own minds – and on medical and pharmaceutical

research and practice generally, by shifting focus from the individual lifestyle and health needs of patients toward catchall disease elimination solutions like drugs and surgeries (Fuks, 2021). However, when certain metaphors are used by patients and providers, they can positively impact patient experiences of, e.g., chronic pain (Munday et al., 2022). Deliberately changing the framing of disease models in research by changing the metaphors used to describe illness and health could, therefore, have dramatic effects on the way medicine is understood and communicated and the way that patients understand their relationships with their own bodies and the medical system.

Maass et al. (2014) explained that by far the most common form of metaphor used in natural language is conventional metaphor, i.e., metaphor in the form X is Y – although this claim is disputed (see Al-Azary et al., 2021). In the sentence *my job is a jail*, the vehicle word (*jail*) imparts some additional meaning to the topic word (*job*), leading the listener to understand, for example, that the speaker feels trapped or limited by their job, that it is unpleasant or oppressive. What is not yet well-understood, however, is how this meaning is attained on the cognitive level – does a listener first understand the sentence as literal and then reject the literal sentence meaning, or is there another cognitive mechanism which allows listeners to understand the figurative meaning of a metaphor directly? In order to understand how this meaning is effectively understood, we must first understand what cognitive resources comprehenders recruit while listening to the sentence and which meanings – literal, figurative, both, or neither - are activated in the first moments after hearing it.

Metaphor Versus Simile

Philosophers dating back to Aristotle (trans. 1926) posited that conventional metaphors in the form *X is Y* (e.g., *my job is a jail*) were a modified or shorthand form of simile in the form *X*

is like Y (*my job is like a jail*), and this is a common view to this day (see Pambuccian & Raney, 2021). In the case *X is Y*, it is generally understood that the sentence is not literally true (except, perhaps, in the unlikely situation where the speaker works in a prison), but rather conveys a semantically related, figuratively distinct interpretation. In the case *X is like Y*, it is entirely possible to reach a literal interpretation – something can always be *like* something else in some sense. Both are interpreted by listeners or readers as conveying something potentially true – and semantically similar – about the world, but do comprehenders arrive at those interpretations using the same cognitive process? Pambuccian and Raney (2021) argued that when they are familiar, both metaphors and similes are interpreted automatically, while unfamiliar similes and metaphors are both understood as comparison statements and processed as such.

Linguistically, it is not readily apparent that *my job is a jail* and *my job is like a jail* should be understood the same way. While adding the word *like* invites a comparison process between jobs and jails, the metaphorical statement suggests that the vehicle predicates something about the topic, as in the sentence *my job is an occupation*, or suggests that a jail represents a superordinate category to which my job belongs. Although neither interpretation is on its face semantically merited, listeners reach an interpretation all the same, which is typically highly consistent across listeners (see Roncero & de Almeida, 2015). This has led to a number of theories about the process that underlies metaphor comprehension.

Theories of Metaphor Comprehension: Overview

If Lakoff and Johnson (1980) were correct in asserting that metaphor expresses how we think, then the question of how we think about metaphor becomes critical to our understanding of human cognition itself (see Katz, 2024). Take the question of mental representation of concepts – beyond language and communication, how do we remember and recognize discrete

types of things in the world and categorize them? How does a child conceptualize what a cat is, store that concept in memory, and recognize a new cat when he or she sees one? This basic question underlies much of our research on human cognition itself, and language research can help us answer these questions, albeit indirectly, by probing how concepts are communicated and tokened in the memory.

If cognitive concepts are atomic (see de Almeida, 2001; de Almeida & Antal, 2021; Fodor and Pylyshyn, 1988), then it is likely that the concept of a JAIL would map onto the word *jail* as a discrete concept with a discrete meaning. If, however, a concept such as JAIL is a collection of loosely connected features and attributes (e.g. Glucksberg, 2000; Kumar et al., 2021) then it should be possible to understand a sentence such as *my job is a jail* automatically, as a sort of comparison of the attributes associated with JOBS and JAILS. Theories about metaphor comprehension map closely onto these contrasting theories of mental concepts and inform further research about cognition beyond the linguistic sphere.

The classical view of metaphor comprehension developed in the 1970s by philosophers such as Davidson (1978), Grice (1975) and Searle (1979), termed the pragmatic model, holds that comprehenders first interpret a statement literally, and then upon realizing it cannot be literally true, search for alternate meanings from context and broader world-knowledge, in what is known as a pragmatic process. This comprehension process is necessarily slower than literal processing as it requires that a mismatch in stated and intended meaning be detected, and then subsequently resolved using slower pragmatic processes. According to pragmatic theory, linguistic structures are "shallow" representations (de Almeida & Lepore, 2018), dependent on linguistic computations performed on symbols (in this case, words), which provide no automatic access to pragmatic meanings.

Searle (1979) explained that a literal sentence such as *Sally is tall* can be easily understood as long as the speaker and listener have a shared referent for *tallness* – in this case, tallness relative to women or human beings, and not relative to giraffes or buildings. As long as there is no mismatch in shared context, no additional cognitive resources need be recruited to understand the sentence literally and for that understanding to be shared between speaker and comprehender. In contrast, understanding a sentence such as *my job is a jail* requires an active search on the part of the listener for a referent in the real world which might render the sentence comprehensible, as well as salient features of the meanings of *job* and *jail* which might be commonly shared. Searle distinguishes between *sentence meaning* (the denotative meaning where *jail* is understood literally) and *utterance meaning* (what the speaker intended to convey) – and suggests that most listeners will ultimately search for utterance meaning in an attempt to attain a shared understanding with the speaker.

In line with Lakoff and Johnson (1980), Davidson (1978) proposed that rather than having a denotative meaning, metaphors are understood *imagistically*, i.e., the listener imagines a job as a jail, in a process that is more sensory or conceptual than propositional. Davidson (1978) explained: "Absurdity or contradiction in a metaphorical sentence guarantees we won't believe it and invites us, under proper circumstances, to take the sentence metaphorically." While much has been made of this distinction, both theories agree that the fundamental nature of a metaphor is, first, not to be believed.

An important caveat to the pragmatic model is that it allows for an exception to the process of literal understanding, rejection, and subsequent search for alternate meanings. As Searle (1979) explained, certain nonliteral phrases which are used enough in common parlance become *lexicalized* – that is, they are used frequently enough to denote a single atomic meaning

like a single word would and at this point can be processed automatically. Therefore, a highly lexicalized phrase like *time is money* may not be subject to the same cognitive processes as *my job is a jail*, because it is stored in memory as a discrete phrase with its own atomic meaning.

In contrast to the pragmatic model of metaphor comprehension, a more recent set of theories has arisen which posits that metaphors are not initially understood by processing, and subsequently rejecting, literal meanings. Termed the direct-access model, this set of theories suggests that utterance meaning is understood directly, in a process that is just as fast as literal sentence processing and no more cognitively demanding (e.g. Colston & Gibbs, 2021; Gibbs, 1994; Glucksberg & Keysar, 1990; Wolff & Gentner, 2000; see Gibbs & Colston, 2012, for a review). Two main processes have been proposed to account for this direct processing: the first, termed *categorization*, is a process in which the topic is taken as a subordinate member of a superordinate category denoted by the vehicle (Glucksberg & Keysar, 1990; Glucksberg, 2003), while the second, termed *comparison*, is a process in which properties or attributes of the topic and vehicle words are mapped until overlapping properties are found, in a non-hierarchical structure (Pambuccian & Raney, 2021; Wolff & Gentner, 2000).

The main difference between categorization and comparison processes is that comparison is non-directional, meaning that alignment of common "salient properties" (Wolff & Gentner, 2000) of the words *job* and *jail* should inform comprehension of sentence meaning – a process that should closely align with that used to comprehend a simile such as *my job is like a jail*. A common critique of this model is that it should render metaphors fully reversible – *my job is a jail* should mean the same thing as *my jail is a job* and yet this appears not to be the case (Chiappe et al., 2009). In contrast, direct-access theories promoting categorization processes posit that metaphor comprehension, unlike simile comprehension, is directional, and that *jail* is

interpreted as a shorthand for a larger category of things to which *jobs* belong – e.g., unpleasant and oppressive environments. This aligns with predication theory as proposed by Kintsch (2001), which asserts that metaphorical predication is essentially indistinguishable from literal predication (e.g., *Sally is tall*), and a conventional metaphor is simply a different kind of class-inclusion statement.

A third set of theories, broader in scope but here termed *exhaustive access* theories of metaphor comprehension, propose that both literal and figurative meanings are accessed to varying extents during metaphor processing, without necessarily deliberating on the overall speed of comprehension processes. One such model, termed the *dual-access* model by Carston (2010), expands on Davidson's (1978) contention that metaphor comprehension is imagistic in nature. Carston (2010), building on the idea that there are fundamental differences between extended literary metaphor or mixed metaphor and conventional metaphor, argued that two competing processes occur during metaphor comprehension. The first is a process of ad-hoc interpretation where the metaphor is treated as propositional in nature and the listener relies on pragmatic processes to attain a figurative meaning as quickly as possible, and the second is a long-lasting imagistic interpretation which facilitates ongoing understanding of extended metaphor by holding the literal meaning in mind for an extended period (Carston, 2010).

To demonstrate why this distinction is important, Carston (2010) gives an example of mixed metaphor from Tirrell (1989): *if you find a student with a spark of imagination, water it*. A reader might quickly create an ad-hoc construction for the phrase *spark of imagination* with a meaning similar to *seed* – i.e., a *spark of imagination* is something small that will grow and develop over time; similarly, *water it* means, metaphorically, to nurture it in order for it to grow. Thus, if figurative meanings were accessed automatically and also persisted over time, this

sentence should not be jarring to a reader as it carries a fairly obvious meaning: if your student shows the beginnings of creativity, nurture that creativity in your student. However, Carston (2010) argues that this sentence is quite jarring to read, because the literal/imagistic meaning of *spark* persists even after the ad-hoc meaning of the phrase containing it is constructed, and *water a spark* carries the opposite meaning to *nurture a spark*. According to Carston (2010), the intuition that such mixed metaphors are nonsensical indicates that literal word meaning activation should be more robust and long-lasting than ad-hoc figurative interpretations of individual metaphor phrases. That is, the fact that *watering a spark* is imagined at all suggests that the real meaning of *spark*, or an image of a spark, persists in the mind even after *utterance meaning* for the first metaphor has been reached. This stands in contrast to the direct-access view that words are tokened as collections of features or attributes rather than as discrete, atomic concepts.

In a similar vein, Sperber and Wilson (2008) have suggested another account of metaphor comprehension frequently termed the *relevance theory* of metaphor. According to Sperber and Wilson (2008), all language conveys meaning through *implicatures* and can imply meanings either weakly or strongly – but they contend that metaphor is not a unique case of language in this sense. Instead, they suggest that most metaphor phrases simply entail a collection of weaker implicatures rather than a smaller set of strong implicatures which are inferred when listening to a purely literal sentence. Like Carston (2010), Sperber and Wilson (2008) suggested that most metaphorical sentences lead to ad-hoc interpretations which can be strong or weak depending on the specific metaphor in question and its overall sentential and environmental context.

In contrast, an alternative exhaustive-access theory put forth by Giora (2003), the *graded salience theory*, suggests that it is not the whole sentence or phrase that is key to comprehension

but the meaning of the vehicle word itself. Giora (2003) proposed that there are multiple competing meanings of many words, and that these need not be literal – if a word is frequently used to denote a metaphorical meaning, then that metaphorical meaning may be stored in the mental lexicon as a core meaning of the word itself. Thus, the fact that *my job is a jail* makes no literal sense need not enter into comprehension processes at all – if *jail* has a frequently used figurative meaning, then that figurative meaning may be instantly selected by a listener depending on which of the possible meanings is more salient in a given linguistic community. A group of law enforcement officers may interpret the sentence literally, but most others would understand it figuratively. Thus Giora (2003) predicts that both meanings should be initially activated, but the meaning that is ultimately selected is context-insensitive and dependent on the conventionality and frequency of each possible word meaning. This leads us to a key question – are all metaphors equally likely to be interpreted figuratively?

Are All Metaphors the Same?

Several features contribute to the processing of metaphor which can affect how they are processed, including familiarity, aptness, and conventionality. Familiarity refers to how frequently a metaphor is used in speech and text, which can have profound consequences on processing speed. According to pragmatic theorists such as Searle (1979), highly familiar metaphors, such as *time is money*, may be fully lexicalized and thus interpreted automatically as though they were a single lexical item denoting a single meaning independent of the component word meanings. A low familiarity metaphor, such as *eyelids are curtains*, may require a completely different comprehension process. This means that the figurative meaning of highly familiar metaphors may be directly retrieved from lexical memory rather than compositionally built word-by-word, leading to significantly faster activation of figurative meanings.

In a key study of online metaphor processing using cross-modal priming, Blasko and Connine (1993) found that the figurative meaning of highly familiar metaphors was comprehended faster overall than for novel metaphors. This effect has been replicated across multiple studies (e.g., Blasko & Briihl, 1997; Damerall & Kellogg; 2016), and a recent resource-depletion study using a Stroop task found that familiar, but not novel metaphors were associated with automatic rather than effortful processing (Pambuccian & Raney, 2021). Multiple theories for this processing time difference have been proposed, but of note, and in keeping with Carston's (2010) theory that metaphors can be processed both imagistically and linguistically, Al-Azary and Katz (2021) found that high-familiarity metaphors were more likely to be understood in the abstract, while novel metaphors were more likely to be understood in terms of bodily-action associates, or "embodied." This difference in cognitive representation could offer an alternative to the standard pragmatic view that unfamiliar metaphors are processed more slowly than familiar metaphors because familiar metaphors are "lexicalized" and novel metaphors require a process of rejection of the literal meaning (Searle, 1979).

Aptness, typically defined as the extent to which a metaphor vehicle captures properties of the topic, increases the extent to which a plausible interpretation of a figurative sentence can be reached (Roncero et al., 2016). A high aptness metaphor such as *genes are blueprints* should increase the extent to which a figurative meaning could be compositionally built relative to a low aptness metaphor like *love is a child*.

Numerous studies have underscored the importance of aptness for metaphor comprehension, especially for novel metaphor. In Blasko and Connine (1993), figurative meanings of novel metaphor were activated preferentially when the metaphors were highly apt but not when they were less apt. Although aptness tends to be correlated with familiarity

(Thibodeau & Durgin, 2011), studies separating aptness from familiarity and conventionality have found that aptness, rather than familiarity, best predicted metaphor comprehensibility in Alzheimer's patients (Roncero & de Almeida, 2014), eased reading times relative to conventionality (Jones & Estes, 2006), and predicted a preference for using metaphor over simile both in experimental and naturalistic settings (Chiappe & Kennedy, 1999; Roncero et al., 2016). Kintsch (2001) questioned the view that aptness and familiarity are critically important for metaphor comprehension, suggesting instead that familiar metaphors in studies such as that conducted by Blasko and Connine (1993) are not "real" metaphors because they compare two concepts that are closely linked in the semantic space, and suggests that such examples represent an intermediary form between a metaphor (which compares two unrelated concepts) and a literal statement.

Conventionality, often used interchangeably with familiarity, is used by some researchers to denote the extent to which certain vehicle words are used to convey metaphorical meanings, and thus the conventionality measure captures some properties of both aptness and familiarity (Bowdle & Gentner, 2005). This is believed by some researchers (e.g. Bowdle & Gentner, 2005) to be a better measure of metaphoricity of a sentence than measures of topic and vehicle relatedness like aptness or familiarity, but the conventionality measure omits considerations of the full phrase meaning and how full sentences are composed during the comprehension process.

It is worth noting that there is an important distinction between the real-time activation of meaning and ultimate comprehension, demonstrated clearly in the idiom literature. For example, Libben and Titone (2008) found that increased familiarity and increased decomposability (an analog of aptness which refers to how much the individual words in a phrase semantically contribute to overall phrase meaning) of an idiom phrase jointly determined whether the idiom

was rated as meaningful – however, familiarity and decomposability do not always correlate, and the mechanisms through which they facilitate comprehension may be different. In a cross-modal priming idiom study, Titone and Libben (2014) found that increased familiarity led to more figurative priming early on, while increased decomposability led to less figurative priming. This suggests that both decomposability (or aptness) and familiarity ease ultimate comprehension, but through different mechanisms – familiarity appears to increase immediate comprehension speed of figurative meanings, while aptness facilitates ultimate comprehension. In another study of metaphorical and literal verb usage, Columbus et al. (2015) found that low-familiarity metaphorical verbs were read more slowly than verbs used literally, but this difference was smaller for high-familiarity verbs, and the effect was modulated by individual differences in executive function. Thus, the cognitive processes ascribed to idioms are likely to be relevant to other figurative forms, such as metaphor and simile (and vice-versa).

Mapping Metaphor and Simile onto Metaphor Comprehension Theories

Unlike metaphors, similes can be directly understood as literal comparison statements. Adding the word *like* to *my job is a jail* immediately renders the sentence both literally plausible and easily comprehensible as a statement comparing some features of the job to a jail, which could thus have major impacts on sentence processing. The word *like* serves as a contextual cue that encourages comprehenders to conceptually fuse the conceptual-semantics of the topic and vehicle in the moment. Despite this difference in syntactic structure, the sentence *my job is like a jail* is typically understood as conveying a meaning which is semantically highly similar to that conveyed by the sentence *my job is a jail* (Pambuccian & Raney, 2021). This small difference in surface meaning could have major implications depending on which theory of figurative language processing is employed, and thus comparing metaphor and simile directly can help

distinguish which theory of metaphor processing reflects the reality of the underlying cognitive processes.

Pragmatic models predict that processing of similes should be faster than processing of metaphors as there is no anomaly detection event to trigger the rejection, and inhibition, of literal meanings in similes prior to a search for utterance meaning (Roncero et al., 2021; Searle, 1979). Therefore, according to pragmatic models, sentence comprehension following simile processing need not involve suppression of literal meanings in later comprehension stages and may not even require activation of figurative meanings at all.

Direct-access models, on the other hand, make variable predictions about simile processing relative to metaphor processing. Comparison models of metaphor processing (e.g., Wolff & Gentner, 2000) consider metaphors to be comparison statements much like similes, and this should result in near-identical processing between simile and metaphor. Categorization or predication models (e.g., Glucksberg & Keysar, 2000; Glucksberg, 2003; Kintsch, 2001; Utsumi & Sakamoto, 2011), on the other hand, posit that metaphor is fundamentally different from simile in structure and meaning as they hold that metaphor, but not simile, involves hierarchical relationships between topic and vehicle. This need not incur additional processing costs for metaphor relative to simile (or vice-versa), but suggests that comprehension of simile and metaphor involves different underlying processes and different relationships between topic and vehicle.

One direct-access theory which attempts to merge these two disparate processes into one overarching model is the Career of Metaphor theory proposed by Bowdle and Gentner (2005). According to Career of Metaphor theory, novel metaphors function as comparison statements like similes, while more conventional or lexicalized metaphors begin to function as

categorization statements (Bowdle & Gentner, 2005). It is important to note that according to Bowdle and Gentner (2005), it is conventionality of the vehicle, and not the familiarity of the entire metaphor, which drives this difference between novel and conventional metaphors, in line with the graded salience theory proposed by Giora (2003). According to these theories, the vehicle word itself holds multiple potentially salient meanings which are activated by external context, and it is the salience of these single-word meanings which determines how a metaphorical sentence is ultimately understood.

Bowdle and Gentner (2005) relate the findings of a reading-times study which found that novel similes were read faster than novel metaphors, suggesting that the word *like* facilitates comprehension of comparison statements, while conventional metaphors were read faster than conventional similes. They suggest this finding is attributable to the fact that conventional vehicles invite categorical processing, and the word *like* in similes impedes this categorical processing in conventional figurative language, while facilitating it in novel figurative sentences. They go on to state that similes are indicative of a "weaker relationship between the target and base concepts" (Bowdle & Gentner, 2005). While pragmatic theorists would suggest that slower processing of novel metaphors is due to an inherent grammatical mismatch between sentence structure and meaning, Bowdle and Gentner reject this idea, stating that the slower processing of novel metaphor (relative to simile) is instead caused by the search for a nonexistent categorical relationship between vehicle and target (Bowdle & Gentner, 2005).

In contrast, Jones and Estes (2006) found that aptness, rather than conventionality, better predicted reading times for novel metaphor targets, suggesting that the relatedness of topic and vehicle is more important for metaphor comprehension than the relative salience of figurative and literal vehicle meanings. Another explanation for the difference between categorical and

comparison processes of metaphor has been proposed by Steen (2008), who suggests that intentional or deliberate metaphors (such as *Juliet is the sun*) are processed using comparison processes, while nondeliberate metaphors (such as *we have come a long way*) are mostly processed using categorization processes.

Recent empirical findings suggest that factors such as familiarity, aptness and context modulate not only the processing speed but also the preference for metaphors or similes in different contexts (Chiappe & Kennedy 1999; Pambuccian & Raney, 2021; Roncero et al., 2016). In an eye-tracking study, Ashby et al. (2018) found that metaphors were read slower than similes, with longer first-pass times, go-past times, and regressions from the vehicle – and that this effect was more pronounced for unfamiliar topic/vehicle pairs in one of two experiments. A self-paced reading study by Roncero et al. (2021) employing negated metaphors with explanations following the metaphor/simile found that negated metaphors (e.g., *jobs are not jails*) were read faster than negated similes (e.g., *jobs are not like jails*), suggesting that metaphors are viewed as less inherently true than similes with the same constituents. In addition, they found that explanations negating a figurative property were read faster following metaphor, while explanations negating a literal property were read faster following a simile (Roncero et al., 2021). These findings are consistent with the theory that similes activate more literal interpretations than do metaphors. Consistent with these results, an eye-tracking study by Olkoniemi et al. (2022) found that there were more gaze regressions when reading metaphor than simile, and that this effect was modulated by familiarity.

An eye-tracking study by Durgin and Gelpi (2017) found that while comprehensibility ratings did not substantially differ between similes and metaphors with the same constituents, comprehensibility ratings predicted both first-pass and second-pass durations for metaphor

vehicles and only second-pass durations for simile vehicles, possibly indicating that similes are first interpreted as literal comparisons while metaphors are first interpreted figuratively. Taken together, these empirical results are somewhat consistent with the predictions of Career of Metaphor theory but suggest that it is the tendency to interpret similes literally, and not a categorization/comparison mismatch, that makes novel similes easier to comprehend than novel metaphors.

Despite the literature reviewed above and over half a century of research on figurative language (see Katz, 2024, for a review), many questions remain to be investigated. For example, it is unclear whether the word *like* in similes functions as a contextual cue to facilitate the activation of figurative meanings by inviting comparisons. Thus, while similes are on their face more literally comprehensible than metaphors, it is unclear whether the literal meaning is always preferentially activated in simile comprehension or whether it is activated alongside figurative meaning. It is also possible that individual differences in language experience – bilingualism, for example – modulate the comprehension processes to differentially impact metaphor and simile processing.

Metaphor and Simile in Second Language (L2) Speakers

To date, most of the empirical literature on metaphor and simile processing focused on comprehension processes in L1 speakers of a given language. As multiculturalism becomes the norm across the globe, and as people increasingly have to navigate jobs, healthcare, and social interactions in a second or third language, it is more important than ever to consider how language background affects metaphor comprehension and production. Understanding these processes is critical for L2 learning and teaching as well as for understanding how differences in metaphor comprehension may affect medical care, decision making, and the communication of

social and political concepts across different language groups.

For L2 learners, this can mean that comprehension of metaphor is key to successful communication in an L2 and to navigating structures like the medical system. It has been argued, for example, that effective health care relies on a shared narrative understanding of disease and health by patient and doctor, and that metaphor is used both by patients to convey symptoms to the doctor and by doctors to communicate the reality of a medical condition to the patient (Coulehan, 2003; Fuks, 2021). For example, telling a patient *you have a time bomb in your chest* can dramatically alter the way in which a patient understands their risk profile and can even alter clinical outcomes (Coulehan, 2003). Patients with chronic pain conditions had a different experience of their chronic pain conditions depending which metaphors they and their providers used to frame their experience of illness (Munday et al., 2022).

Understanding how to frame symptoms and health goals in a personal way, with the patient in the position of a driver in their own health journey, could alter how medical providers approach patient care (Fuks, 2021), but this may depend on an L2 speaker's ability to use figurative language intentionally in patient-provider interactions, and a provider's ability to communicate effectively with a patient. In an increasingly multicultural society, it is important to understand how L2 speakers comprehend metaphor as this can affect their health outcomes and the choices medical practitioners make about how to communicate about disease.

Outside of medicine, a burgeoning field now exists which investigates how metaphors aid in comprehension of social relationships and ingroup/outgroup dynamics (Maass et al. 2014), political concepts (Ottatti et al., 2014), emotion (Sherman et al., 2014), and time/space relationships (Casasanto, 2014). Some of these supra-linguistic metaphorical instincts appear to be culturally bound, however – for example, one study found that in a trust experiment, subjects

were likely to give less money to a stranger to distribute when playing a simple trust game in a corner of a room that has been sprayed with fish scent versus a corner scented by fart spray or a neutral smelling corner (Lee & Schwarz, 2012). However, fishiness is not associated with suspicion in every language and culture (Lee & Schwarz, 2012) so this physical association between scent and trust should, theoretically speaking, only hold for certain cultures and not others.

On the other hand, warmth appears to be associated with trust in every culture, as evidenced by the fact that holding a warm cup of tea makes people behave pro-socially in experiments, regardless of culture (Schnall 2014). Caesanto and Bottini (2014) report that left-right associations in politics appear to have originated in language and subsequently become universal, but people think of time as moving right to left or left to right depending in which direction they write. This suggests that some, but not all metaphors should be easily comprehensible across languages; however, relatively little empirical work has investigated cross-linguistic metaphor comprehension processes directly.

There are good reasons to infer the presence of processing differences for figurative language between L1 and L2 speakers. As many researchers have demonstrated, the familiarity or conventionality of a metaphor or simile may affect the ease of interpretation of its meaning, and perhaps entirely change the interpretation of the meaning itself – yet the familiar meanings of certain phrases or topic/vehicle relationships may not translate across language and culture. A word which may take on multiple meanings in one language (*fishy* meaning both smelling of fish and suspicious, for example) may not carry those meanings for someone with a different linguistic background. An L1 speaker may be faster at determining whether a sentence makes literal sense while an L2 speaker may assume that there is a gap in their literal understanding. It

is also unclear how such effects might change with L2 fluency and experience. Unfortunately, very little work has thus far determined whether bilingualism changes metaphor processing in a comprehender's L1 relative to monolingualism.

Some empirical work has suggested that there are additional costs of figurative language processing, relative to literal language processing, in L2 speakers. In ERP studies, both novel and familiar processing evoked late positivity in L2, but only novel metaphor evoked late positivity in L1 (Jankowiak et al., 2017), suggesting that most metaphors are processed as unfamiliar in an L2. In the early N400 window, there were processing deficiencies for metaphor in L2 and not L1 (Jankowiak et al., 2021), indicating that L2 speakers experience increased difficulty retrieving figurative meanings in early processing.

For example, an eye-tracking study by Senaldi et al. (2022) demonstrated that replacing a noun in an idiom with the equivalent noun in a participant's L2 (e.g., *break the glace* instead of the idiomatic *break the ice* in French-English bilinguals) incurred a larger processing cost in early processing stages than did replacing a noun in a comparable literal phrase. Another study comparing high- and low-proficiency L2 speakers found that familiarity modulated metaphor processing in high-proficiency L2 speakers, but not low-proficiency L2 speakers (who struggled with metaphor comprehension generally relative to L1 speakers); for simile, familiarity modulated speed of processing for low-proficiency (and not high-proficiency) L2 speakers, with L2 proficiency predicting overall ease of simile comprehension (Olkoniemi et al., 2022). Based on these results it appears that simile functions more like literal language for proficient L2 speakers compared to metaphor and idiom.

These findings underscore the importance of considering not only general language background (L1 vs. L2) but also level of L2 proficiency when studying differences between L1

and L2 comprehension of figurative expressions. There is relatively little research comparing figurative language processing across different levels of language proficiency. Thus, in the following studies, we investigate whether there are processing differences for metaphor and simile between L1 speakers of English and L2 English speakers who approach L1-like levels of English proficiency.

Rationale for the Studies Presented in this Dissertation

Despite decades of theoretical and empirical work investigating metaphor and simile processing, contemporary science finds itself at an impasse regarding the underlying mechanisms which drive metaphor and simile processing (see Colston & Gibbs, 2021; Roncero et al., 2021; Katz, 2024, for a review). This is in part because the bulk of experimental work on metaphor processing to date has relied upon *offline* measures such as comprehensibility ratings and corpus analyses, or crude measures such as comprehension speed which do not provide direct insight into underlying cognitive processes nor differentiate between automatic meaning activation and effortful comprehension processes. Such global processing measures cannot help to distinguish between competing theories of the moment-by-moment cognitive processes which underlie ultimate comprehension, only indirectly hint at the level of processing cost involved. As Katz (2024) points out, many online processing studies assume that greater cognitive resources are used for figurative language compared to literal language because it is more complex or harder to interpret, but to test this assumption it is necessary to determine what meanings precisely are tokened during the course of metaphor processing, and many studies have failed to directly test the concepts tokened at each moment during figurative language processing.

While some *online* reading studies have shown that sufficient biasing context can close the gap in reading time between metaphorical and literal sentences (e.g., Inhoff et al., 1984), it is

impossible to determine from reading times alone whether this is due to direct access or facilitation of pragmatic processes – and indeed, reading-time studies which do not employ biasing contexts often find reading-time differences between metaphor and simile (e.g., Janus & Bever, 1985). Among the relatively few time-sensitive *online* studies conducted on metaphor processing, research employing event-related potentials (ERP) has demonstrated that figurative targets typically elicit larger N400 amplitudes than literal targets, indicative of higher processing costs (Bambini et al., 2016). However, ERP relies on timed serial word presentation, which introduces issues with ecological validity. Most research on metaphor comprehension, moreover, relies on reading tasks, while metaphor is also a major component of spoken language. Additionally, it is unclear to what extent language history impacts figurative language processing, and to what extent any such differences as may exist between L1 and L2 speakers depend on factors such as familiarity, aptness, and sentence structure.

Proponents of direct-access models of metaphor processing have claimed that the evidence for direct figurative access is overwhelming (e.g., Colston & Gibbs, 2021; Glucksberg, 2003), yet global processing time measures are a tenuous basis upon which to rest such an assertion. In order to understand what happens when a comprehender reads or hears a metaphor or simile, it is not sufficient to point to reading time differences (or the lack thereof) in figurative and literal language processing. Instead, we must look for ways to interrogate what meanings are activated as a figurative expression unfolds, leading up to the moment of ultimate utterance comprehension. This question has profound implications for our understanding of general cognitive processes as well as language comprehension.

The Present Thesis

The aim of the following studies was to distinguish which interpretations – figurative,

literal, or both – were accessed across time as listeners heard conventional metaphors and corresponding similes spoken in a natural cadence. In STUDIES 1 and 2, we compared similes and metaphors with the same constituents (differing only in the word *like*) at four different time points using a cross modal lexical decision paradigm using aurally presented metaphor/simile sentences as primes and either figuratively or literally related words as visual targets. In STUDY 3, we collected un-timed comprehension judgments for a set of English-language metaphors in both L1 and L2 English speakers, to determine whether global differences in interpretation of emotional valence, familiarity, aptness, and comprehensibility depended on the language background of readers.

Chapter 2: STUDY 1

Time-course of Metaphor and Simile Processing in L1 English Speakers

We employed a cross-modal lexical decision paradigm to compare metaphor and simile processing at the recognition point of the vehicle word (determined by a gating paradigm) and 500ms later. In a second group of participants, metaphor and simile processing was compared at the onset of the vehicle word and 1000ms after the recognition point. Using a similar experimental paradigm, Blasko and Connine (1993) demonstrated that aptness and familiarity modulated the ease of activation of figurative target words associated with metaphors, finding that metaphorical meanings as well as literal meanings related to the vehicle word were activated for highly familiar metaphors and low-familiarity, high-aptness metaphors but not for metaphors which were rated as both low-familiar and low-apt. However, they presented visual targets at the offset of the vehicle word for a relatively long time (250ms), making it impossible to distinguish the precise time-course of the activation of both figurative and literal word/utterance meanings.

In STUDY 1, we attempted to determine precisely which meanings (literal or figurative) were activated in the minds of listeners at the onset of the vehicle word, the moment a metaphor vehicle was first recognized, and then later, to ascertain whether figurative and literal meanings were accessed simultaneously or sequentially during the processing of figurative sentences with varying aptness and familiarity. To this end, we modified the procedure used by Blasko and Connine (1993) to make it as time-sensitive as possible, using forward- and backward-masked visual target presentations lasting only 80ms. Additionally, we used similes with the same constituents as nominal metaphor sentences (except for the word *like*) as controls for their metaphor counterparts to determine whether literal plausibility/decomposability of sentences altered participants' comprehension of their meaning over time.

The rationale for using a cross-modal lexical decision task is that it is both naturalistic and time-sensitive. Participants listened to an aurally presented metaphor or simile on headphones and pressed a button to respond "yes" or "no" to the question "Is this a word in English?" after seeing a concurrently presented visual target word on the screen (presented at one of four time points relative to the spoken phrase). The target words presented were related either to the figurative meaning of the metaphor/simile or to the literal meaning of the vehicle word in each sentence. This experimental paradigm has the advantage of discouraging conscious judgment about the meaning or comprehensibility of the metaphor/simile phrase – the lexical decision acts as a measure of activation of figurative or literal target meaning which should reflect automatic comprehension processes, and participants were not aware that they were making metaphoricity or comprehensibility judgments. Metaphor sentences were embedded among filler sentences of varying structure which did not contain metaphors, to further discourage listeners from identifying the purpose of the task.

Priming (operationalized as the time it takes to respond to the related target word, subtracted from the time it takes to respond to the unrelated control word) represents the ease of activation of the relevant figurative or literal target meaning in each sentential context. Thus, use of the cross-modal lexical decision task should provide insight into the precise time-course of activation of each relevant meaning of the metaphor or simile phrase.

Recognition points for vehicle words were determined by a norming task employing a gating paradigm (following Zwitserlood, 1989), and literally or figuratively related targets were taken from a set of published norms (Roncero & de Almeida, 2015) while unrelated control words for related targets were matched for written frequency, morphology, length, and number of syllables. The use of similes with identical constituent words to matched nominal metaphors

(except the word *like*) allows for a direct comparison between the metaphor and simile form of each sentence.

Following Blasko and Connine (1993), familiarity and aptness ratings from a set of published norms (Roncero & de Almeida, 2015) were split along the median value and binned as "high" and "low" familiarity/aptness, allowing us to determine whether specific features of each metaphor or simile affected which meanings were activated during comprehension. Frequency for target words was taken from Corpus of Contemporary American English (COCA; Davies, 2008-) while subtitle frequency was used for aurally presented topic and vehicle words (Brysbaert & New, 2009), allowing us to control for both verbal and written word frequency.

Experiment 1 of STUDY 1 compared 32 metaphors and similes at vehicle recognition point and 500ms later in a sample of L1 English speakers. Experiment 2 compared the same metaphors and similes at vehicle word onset and 1000ms post-recognition point, in a second sample of L1 English speakers. Priming was further analyzed as a function of aptness, familiarity, and constituent word frequency of metaphor/simile phrases.

Carston (2010) has suggested that ad-hoc meanings are rapidly constructed in the course of hearing or reading a metaphor, while literal or imagistic meaning activations persist as speech or text continues. To confirm these findings, two post-recognition time points were employed – 500ms post-recognition, to give insight on early processing, and a later time point (1000ms post-recognition) more in line with the presentation time employed by Blasko and Connine (1993), who presented visual targets for 250ms following vehicle word offset. Additionally, to ensure that it was the metaphor or simile as a whole rather than a common feature shared by topic word and vehicle word, that primed figurative meanings, we tested whether figurative meanings were primed directly following vehicle word onset but prior to the vehicle word recognition point.

According to the pragmatic model (e.g., Searle, 1979), literal priming should be observed for both metaphor and simile sentences at the recognition point. At later points metaphors should prime only figurative meanings while literal meanings are inhibited, and similes should continue to prime literal meanings (since there is no mismatch between literal meaning and the simile form of each sentence) – except in the case of highly familiar metaphors, which should automatically activate figurative meanings. Conversely, according to direct-access models (e.g., Colston & Gibbs, 2021), we should expect to observe only figurative priming at the recognition point as well as later points for metaphor, and possibly also for simile (although some direct-access models do not rule out that similes can be interpreted as literal comparison statements). Exhaustive-access models would predict that listeners access both literal and figurative meanings at the recognition point and later in comprehension, but that this may differ depending on features of individual sentences such as aptness and familiarity. Salience theory (Giora, 2003), for instance, would predict that metaphor vehicles containing highly salient metaphorical meanings should prime figurative meanings earlier than less-familiar, or less-apt, metaphors.

In all cases, priming should not be expected at onset as there should be no associated meanings accessed prior to the recognition point of the vehicle word. If associated meanings are accessed at onset, this could reflect associations with the topic word in isolation, suggesting close semantic relationships between topic and vehicle or topic and global phrase meaning. Assuming the topic word itself is not sufficient to prime figurative meanings related to the whole metaphor/simile phrase or literal meanings of the vehicle word in each sentence, there should be no priming of either figurative or literal targets at the onset of the vehicle word.

According to the pragmatic model, there should be literal target priming across both metaphor and simile sentences at the recognition point, and little or no figurative target priming;

at the later time point, figurative targets should be primed in metaphor sentences and literal priming should be lower or nonexistent as utterance meaning is reached. On the other hand, direct-access models would predict significant figurative target priming at the recognition point for metaphor (and also simile, under comparison but not necessarily categorization models) which should continue at the later time point, and less literal priming. Neither the pragmatic model nor most direct-access models yield clear predictions for simile sentences, but based on past studies (e.g., Ashby et al., 2018; Roncero et al., 2021) we expected similes to behave more like literal sentences and facilitate more literal target priming overall across both time points, compared to metaphors.

Compared to past studies which allowed much more time for lexical decisions to be reached, our cross-modal lexical decision study was designed with sufficient time-sensitivity to investigate the moment-by-moment processing of metaphors and similes as comprehension unfolds over time. Prior studies like Blasko and Connine (1993) allowed listeners to respond after word offset, which likely could not capture the moment-by-moment processes of comprehension while listening to a figurative sentence. We expanded upon their experiment by adding similes as literal controls in order to compare literal and figurative processing directly. By employing a multimodal, naturalistic task with multiple consecutive time points, we hoped to determine which meanings are activated along the course of conventional metaphor and simile processing in L1 English speakers.

Method

Participants

Participants were 52 L1 English speakers between the ages of 18 and 59 (40F;12M, M_{age} = 24.81). Participants in experiment two were 31 L1 English speakers between the ages of 19

and 34 (26F;5M, M_{age} = 22.20). All participants reported normal or corrected-to-normal vision and hearing and met the following inclusion criteria: (1) They learned English before the age of 5 and identified English as their dominant language; (2) they learned English before any other language; (3) they reported no history of hearing or reading disability. Participants who were recruited via the Concordia University Psychology online participant pool were compensated with course credit, while all other participants were compensated with $10 for one hour of participation. Participants for two norming studies are described below.

Materials

Experimental materials consisted of 32 sentences containing metaphors and similes in the form *X is (like) Y* and 160 filler sentences. Metaphor/simile sentences were selected from Roncero and de Almeida (2015), which consists of a set of metaphor/simile topic/vehicle pairs with accompanying norms. The sentences chosen had a broad range of aptness and familiarity ratings, although the small number of sentences with very low aptness ratings (below 4.00 out of 10.00) were excluded on the basis that very low-apt metaphors may not be correctly interpreted as metaphorical and thus may not facilitate figurative priming.

Selection of Written Targets

All targets were selected from a published set of metaphor/simile norms (Roncero & de Almeida, 2015). The Roncero and de Almeida (2015) norming study asked participants to generate associates and explanatory words for both the simile and metaphor versions of each sentence and for the topic and vehicle words in isolation. For use as our figuratively related targets, we selected explanatory words generated by the highest number of participants for each metaphor, which did not appear as associates for the vehicle word in isolation. For example, the

top generated associate for both the sentence *lawyers are sharks* and the vehicle word *sharks* was *dangerous*, so we used the second most frequently generated associate, *mean*.

Similarly, for our literally related targets, we selected words from the Roncero and de Almeida (2015) norms which were generated as associates of the vehicle word in isolation, and which did not appear as explanatory words for the metaphor on the whole – e.g., the literally related target for *sharks* was *blood*, as the two most commonly listed words *dangerous* and *big* were also listed as associates for the full sentence *lawyers are sharks*. The unrelated control words selected to calculate priming effects were chosen among those matching the experimental targets according to the following criteria: (a) written frequency, based on the Corpus of Contemporary American English (COCA; Davies, 2008-), (b) number of letters, (c) number of syllables, and (d) morphological structure.

Exclusion of Frequently Co-occurring Target Words

To ensure that any potential priming effects between prime and target were not derived from an automatic association between the vehicle and target words (i.e., due to being frequently paired in speech, like *salt* and *pepper*), we conducted a norming experiment where each vehicle word was read aloud to 12 L1 speakers of English from the Montreal area, who were asked to say out loud the first word that came to mind. Their responses were collected and any word which was named more than twice was excluded from selection as a target for that vehicle word. For example, the word *bars* came up more than twice in response to the word *jail*, so it could not be used as a target word for the metaphor *jobs are jails*.

Selection of Prime Sentences

Metaphors and similes were embedded in longer sentences with explanatory contexts, with the word *because* following each vehicle word so that participant responses never occurred

during the explanatory region of the sentence; these sentences also began with generic *proposition-attitude* statements (e.g., *It is hardly a secret that lawyers are sharks, because with few exceptions, lawyers are bloodthirsty and ruthless*) (See Appendix A for a full list of experimental sentences and targets).

Filler sentences did not repeat the topic or vehicle words of any experimental sentences. Of these, 32 followed a similar sentence structure as experimental sentences, while 128 filler sentences did not syntactically resemble experimental sentences. Visual targets for filler sentences were 64 real English words and 96 "nonsense" strings of letters that did not resemble English words, of varied lengths to reflect the varied lengths of experimental targets. All sentences were read by a female L1 English speaker and recorded for aural presentation, with natural prosody and reading speed. Special attention was given to matching the prosody and timing of metaphor and simile pairs, to make them nearly identical in prosody and timing except for the word *like*.

Familiarity and aptness of sentence primes were ranked *low* or *high* based on ratings from the Roncero and de Almeida (2015) norms: we determined the median rating for all 32 metaphor sentences, with all sentences rated below the median value classified as *low* apt/familiar and all sentences rated above the median value classified as *highly* apt/familiar. Based on a 10-point scale with 10 meaning "highly apt/familiar" and 0 meaning "not at all apt/familiar," the high-familiarity stimuli had a mean rating of 6.68 (4.85-9.95), while low-familiarity stimuli had a mean rating of 3.50 (2.30-4.75); high-aptness sentences had a mean rating 8.06 (7.35-8.96) while low-aptness sentences had a mean rating of 6.26 (5.00-7.22).

Gating Study

We employed a gating paradigm to determine the recognition point of each vehicle word,

following the procedure developed by Grosjean (1980; see also Zwitserlood, 1989). Recordings of each vehicle word were cut into slices of increasing length, adding increments of 50ms each (i.e., first 50ms of the word followed by first 100ms of the word, etc.). These were played consecutively to 10 L1 speakers of English from the Montreal area over noise cancelling headphones. Participants were asked to write down what word they thought they were hearing after each consecutive slice was presented. Their responses were collected and recognition times for each word were defined as the moment when 80 percent of participants correctly identified the prime word (with or without pluralization). This paradigm has been shown to be sensitive to the recognition point of spoken words regardless of phonemic or syllabic boundaries, while being sententially context-insensitive (see Warren & Marslen-Wilson, 1987; Zwitserlood, 1989). It is important to note that only vehicle words in isolation were presented, without any preceding context.

Selection of Target Presentation Times

During the lexical decision task, the word onset was defined as the beginning of the vehicle word while the recognition point was defined as 40ms prior to recognition time, to account for screen refresh rate as well as to allow for the possibility that in the gating study, the vehicle word could have become recognizable anytime within the 50ms slice presented to participants. The later time points were defined as 500ms following recognition point and 1000ms following recognition point. The 500ms interval was chosen on the basis that Blasko and Connine (1993) presented targets at vehicle word offset, and our aim was to add to their findings by investigating what meanings are activated during the listening process before listeners have had time to think consciously about the meaning of the sentence. The 1000ms interval was chosen as it was typically closer to or shortly following most word offsets, in line with Blasko

and Connine (1993), as the presentation time of 500ms post-recognition point typically fell before vehicle offset but substantially after the vehicle recognition point.

Design

Each topic/vehicle pair was presented in either a metaphor- or simile-containing sentence, along with a figuratively related target, literal target, or matched control target, at an early (*recognition* or *onset*) or late (*recognition + 500ms* or *recognition + 1000ms*) time point. In Experiment 1, participants saw targets either at *recognition point* or *recognition + 500ms*; in Experiment 2, participants saw targets at *onset* and *recognition + 1000ms* time points. Example sentences and corresponding targets are presented in Table 1.

Table 1

Examples of Stimuli for Cross-Modal Lexical Decision Task

AUDITORY PRIME		TARGET	
		FIGURATIVE	*LITERAL*
It is hardly a secret that *jobs are (like) jails* because generally speaking, jobs confine people for much of their lives.	*related*	BORING	DANGEROUS
	unrelated	CASUAL	COLLECTED
AUDITORY PRIME		*FIGURATIVE*	*LITERAL*
It is common knowledge that *memory is (like) a sponge* because in most cases, memory is more absorbent than you would expect.	*related*	SOAKING	DIRTY
	unrelated	PEERING	RAPID

A total of 16 counterbalanced lists were created following a 2 x 2 x 2 x 2 design for both experiments. Each block contained two experimental sentences in each condition along with all 160 filler sentences, 20 of which were followed by comprehension questions to ensure participants were attending to spoken stimuli. Each participant completed two blocks containing one list each – i.e., each participant heard both the simile and metaphor version of each sentence once, although they heard them within different target conditions. The sentences were randomized in order within each block of trials and participants were randomly assigned to each set of lists.

Procedure

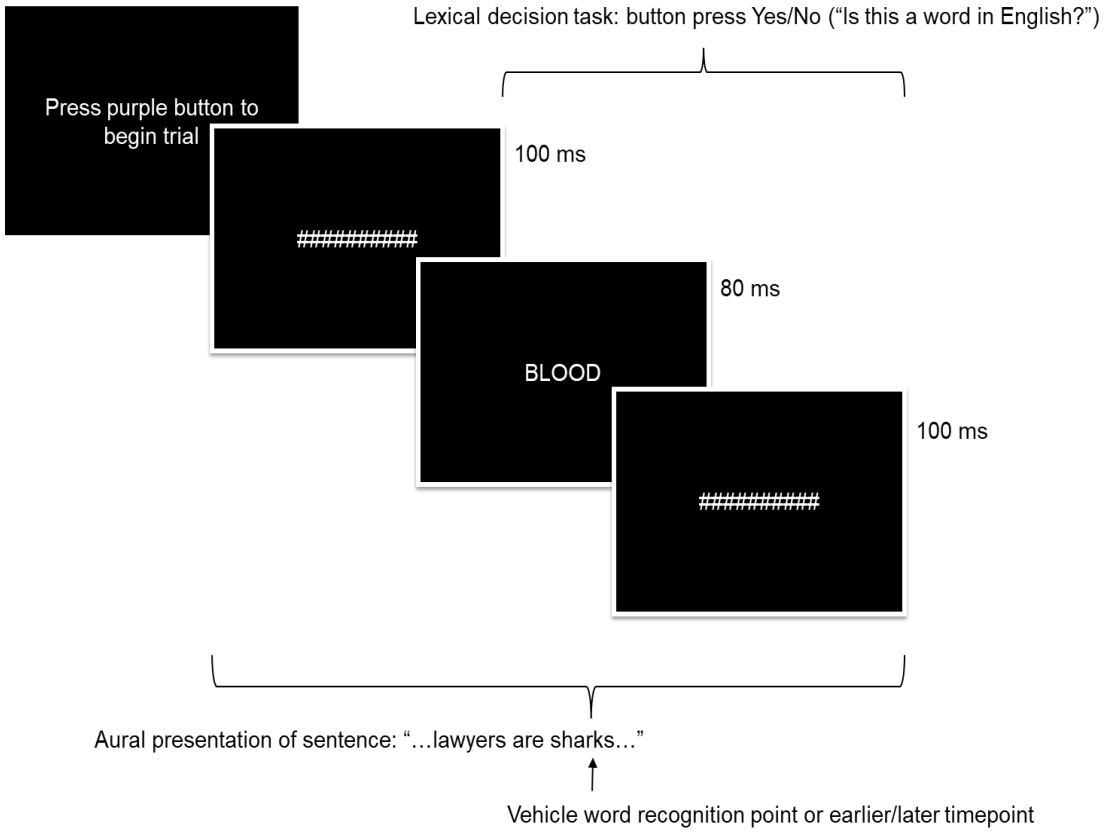
Participants were tested on an iMac computer using Psyscope X B57 (Cohen, MacWhinney, Flatt, & Provost, 1993) using an ioLab button box. After voluntary consent was obtained, each participant was seated in front of the screen in a dark room and instructed to attend to both the aurally presented sentences and visual stimuli on the screen. Sentences were presented over noise-cancelling headphones. Participants were instructed that their primary task was to identify whether the letters they saw on the screen constituted an English word and to press a button to indicate YES or NO as quickly and accurately as possible. They were also instructed that their secondary task was to answer comprehension questions about the sentences they heard over the headphones.

A schematic diagram of the procedure is provided in Figure 1. Each trial consisted of a prompt asking participants to press a button when they were ready for the next trial, followed by an aural presentation of each sentence. Target words/nonwords appeared in white 20-point Arial font text in capital letters on a black screen for 80ms each, preceded and followed by masks

(series of 10 hashes) which appeared for 100ms. This brief masked priming procedure was meant to reflect faster and more automatic processes of recognition rather than slower processes of judgment. Masked priming (see Forster, 1999) reflects early processes of lexical recognition which should be uncontaminated by other semantic factors. The backwards mask prevents visual aftereffects of the target word so that the recognition task is limited to only the 80ms during which the target word appears on the screen. Each participant was presented with five practice trials, after which the experimenter answered any remaining questions about the procedure.

Figure 1

Schematic for Stimulus Presentation in Studies 1-2



Data Analyses

Data Exclusions for STUDY 1 and STUDY 2

Analysis of reaction times (RTs) was restricted to correct trials (i.e., those where participants correctly identified the target as an English word) while incorrect trials were omitted (13.81% of all data points). As is common in cross modal lexical decision paradigms (e.g., Friedmann et al., 2008), all reaction times above 2 seconds were discarded prior to data analysis (1.82% of all data points). For each participant, mean reaction time was computed and trials with reaction times over 2.5 standard deviations above or below that participant's mean were removed (2.6% of all trials). Based on a priori decisions, we discarded blocks of trials where participants answered fewer than 70% of comprehension questions correctly or answered fewer than 60% of trials correctly (42 out of 244 total blocks of trials)[1].

Statistical Analyses

Analyses for trials containing figurative targets and literal targets were performed separately because experimental targets for the figurative and literal targets were not matched to each other (only to their unrelated controls) – i.e., the figurative and literal targets for each sentence could have a different number of letters/syllables or different morphology. Additionally, this allowed us to simplify the models to ease interpretation of results. For both literal and figurative targets, two separate analyses were conducted: first we analyzed reaction times overall for each condition, controlling for frequency of targets and, subsequently, a secondary analysis was performed with low- and high-aptness and low- and high-familiarity conditions added to each model. For each model, the estimated coefficient (β), standard error (SE), t, and p values are reported. Significance was evaluated using Satterthwaite approximations implemented in the lmerTest package (Kuznetsova et al., 2017).

[1] Note that data cleaning was performed on aggregated data for STUDY 1 and STUDY 2 prior to statistical analyses.

Results

Table 2

Summary of STUDY 1 Experiment 1 findings

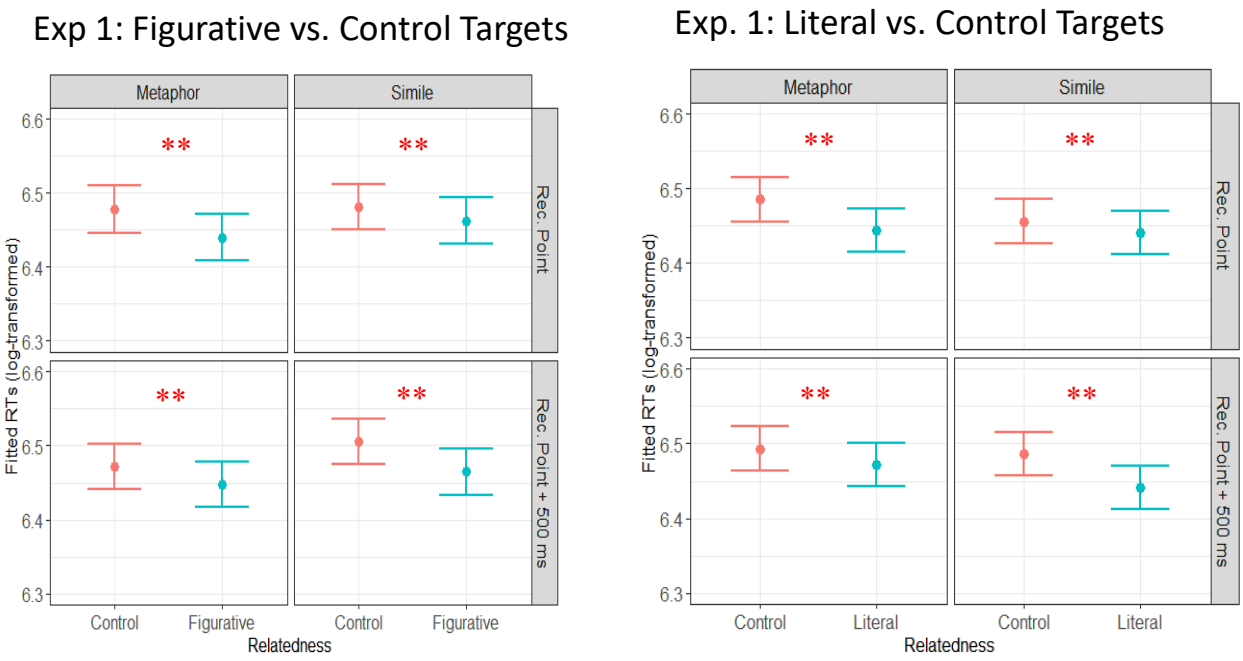
		Recognition point	*Recognition point + 500ms*
Target type	**Sentence type**		
Figurative	*Metaphor*	Priming for high familiarity targets	Priming for low familiarity targets
	Simile	Priming for high familiarity targets	Priming for low familiarity targets
Literal	*Metaphor*	Priming for high aptness targets	Priming for high aptness targets
	Simile	Priming for low aptness targets, negative priming for high aptness targets	Priming for both high and low aptness targets

Results for Experiment 1 are summarized in Table 2 while results for Experiment 2 are summarized in Table 3. We fitted linear mixed-effects models for figurative targets and literal targets separately using the lme4 (Bates et al., 2015) package in R (R Core Team, 2012), with subjects and items (vehicles) entered as random effects with maximally specified random intercepts and slopes and consecutively removed random slopes if we encountered convergence issues (see Barr et al., 2013). For each analysis, log-transformed reaction times were regressed against a continuous frequency variable, trial order variable, and categorical fixed effects of priming (control/experimental targets), sentence type (metaphor/simile) and time point (recognition point/+500ms post-recognition or onset/+1000ms post-recognition), as well as all

first-order and second-order interactions between priming, sentence condition, and time point. Categorical values were sum-coded (for sentence type, met: - 0.5, sim: + 0.5; for priming, control: - 0.5, related: + 0.5; for time point, early: - 0.5; late: + 0.5). For the purpose of our analysis, we were only interested in main effects of the control/experimental target condition or

Figure 2

Experiment 1: Time-course of Figurative and Literal Target Activation in L1 participants



Target priming for figurative and literal targets at vehicle word recognition point and at recognition point +500ms, in metaphor and simile sentences for L1 English speakers. Error bars represent +/-1 SEM. Asterisks represent significant main effects with * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

interactions which included the control/experimental target condition as our goal was to investigate priming effects and not raw reaction time differences between conditions.

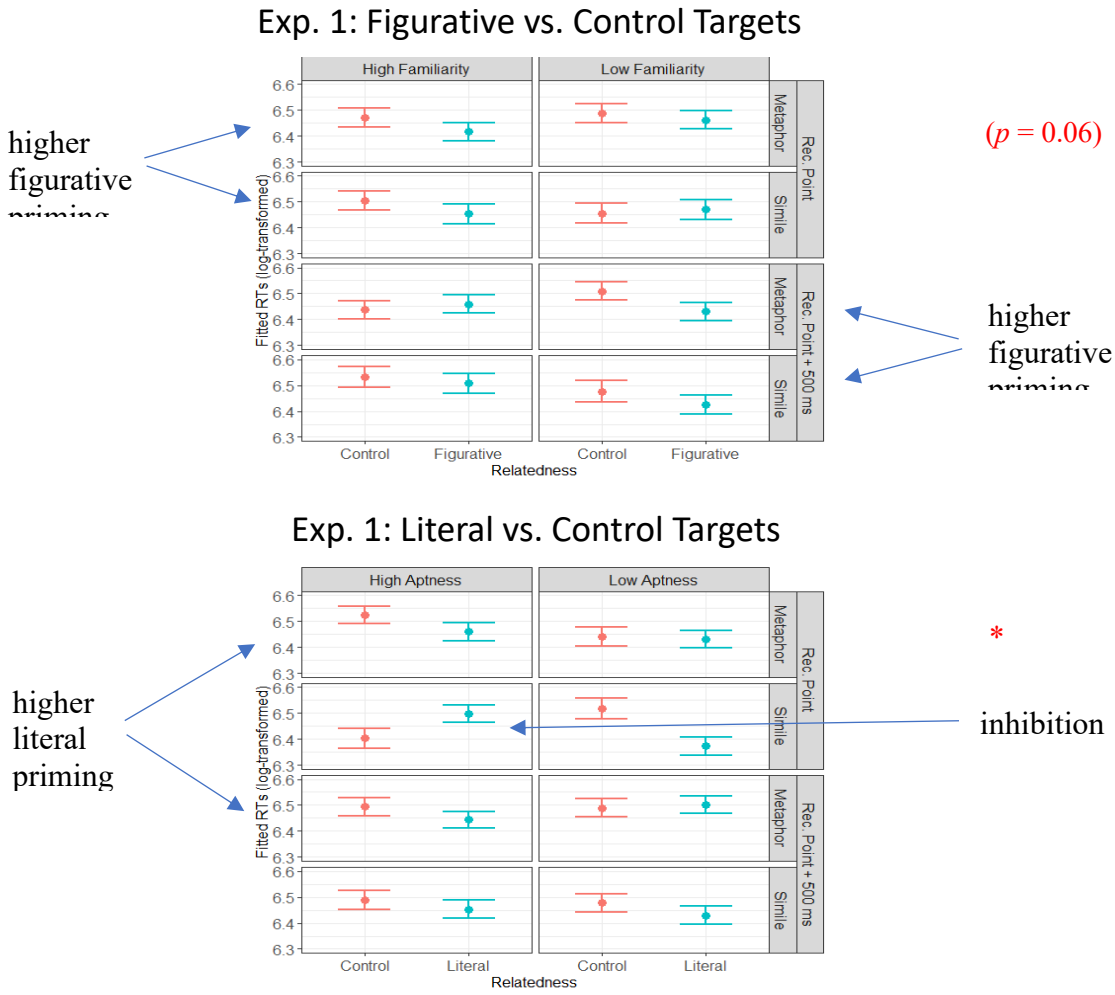
Experiment 1 (recognition point/+500ms)

As can be seen in Figure 2 (left panel), there was a significant main effect of priming for figurative targets across conditions ($\beta = -0.033$, $t = -2.59$, $SEM = 0.012$, $p = 0.0099$), indicating

that participants responded to figurative targets faster than they did to control targets. For literal targets (Fig. 2, right panel), there was likewise a main effect of priming across conditions (β = -0.031, t = - 2.80, SE = 0.011, p = 0.0053).

Figure 3

Experiment 1: Time-course of Figurative and Literal Target Activation vs. Familiarity and Aptness in L1 participants



Target priming for figurative and literal targets at vehicle word recognition point and at recognition point + 500ms, in metaphor and simile sentences as a function of sentence aptness and familiarity in L1 English speakers. Error bars represent +/- 1 SEM. Asterisks represent significant interaction effects with * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

Subsequently, we fitted additional models for both figurative and literal targets with subjects and items entered as random intercepts and log-transformed reaction times regressed against a continuous frequency variable as well as categorical fixed effects of priming, time point, sentence type, aptness (high/low), and familiarity (high/low). Aptness and familiarity were sum-coded (for familiarity, high familiar: - 0.5; low familiar: + 0.5; for aptness, high apt: - 0.5; low apt: +0.5). In addition to main effects, we investigated all first-order, second-order, and third-order interaction terms between priming, time point, sentence type, and (separately) familiarity/aptness. Results are summarized in Table 2.

For figurative targets (Fig. 3, upper panel), we found a main effect of priming (β = - 0.031, t = - 2.62, SE = 0.012, p = 0.0089) as well as a marginally significant interaction between priming, time point, and familiarity (β = - 0.011, t = - 1.89, SE = 0.058, p = 0.060). There was more priming for highly familiar metaphor and simile sentences at the recognition point relative to low-familiar sentences; 500ms later, this effect was inverted and there was more priming for low-familiar sentences than high-familiar sentences.

For literal targets (Fig. 3, lower panel), a four-way interaction effect was found between priming, time point, sentence type, and aptness (β = 0.22, t = 1.99, SE = 0.011, p = 0.046). For both metaphor and simile, there was more overall priming observed at the earlier time point than the later time point. For metaphors, there was significant priming at both time points in the high-aptness, but not the low-aptness condition, whereas for similes, we observed priming across both aptness conditions at the later time point. At the recognition point, there was priming for low-aptness similes but reverse priming/inhibition for high-aptness similes.

Experiment 2 (vehicle onset/+1000ms)

We found a significant interaction effect of priming and sentence type for figurative targets across conditions (β = - 0.065, t = - 3.26, SE = 0.028, p = 0.018), indicating that there was priming of figurative targets in the simile condition but not in the metaphor condition (see Fig. 4, left panel). For literal targets (Fig. 4, right panel), however, there was only a main effect of priming across conditions (β = - 0.037, t = - 2.87, SE = 0.013, p = 0.0042). Results are summarized in Table 3.

Table 3

Summary of STUDY 1 Experiment 2 findings

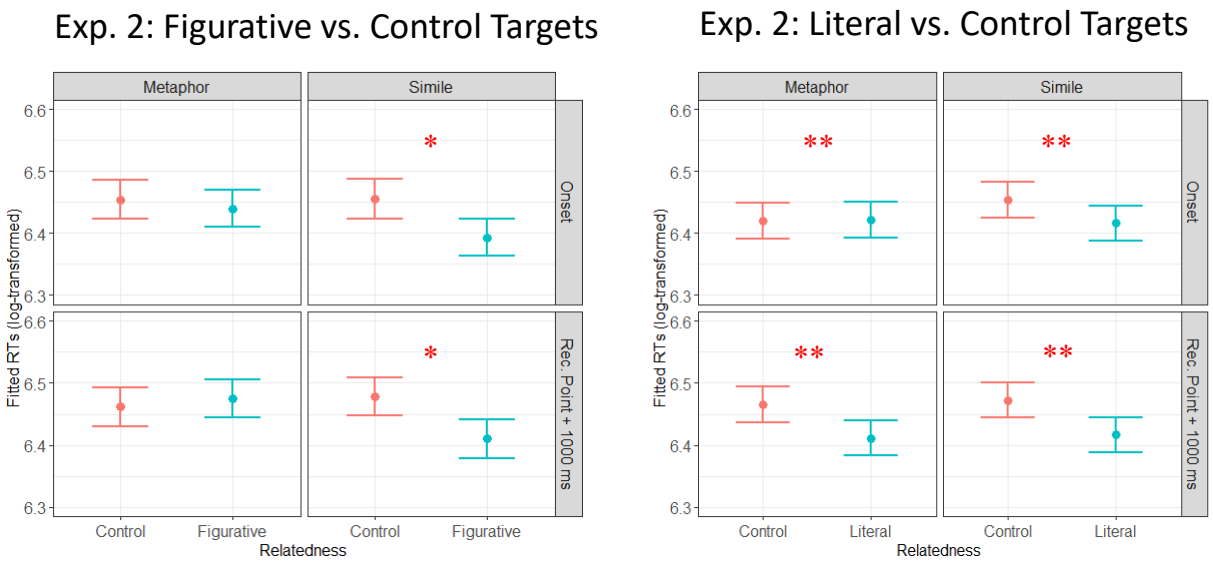
		Vehicle word onset	*Recognition point + 1000ms*
Target type	**Sentence type**		
Figurative	*Metaphor*	No priming	No priming
	Simile	Priming across aptness and familiarity conditions	Priming across aptness and familiarity conditions
Literal	*Metaphor*	Overall priming across target types	Overall priming across target types
	Simile	Priming for low aptness targets, negative priming for high aptness targets	Priming for both high and low aptness targets

We additionally ran linear mixed-effects regression analyses for both figurative and literal targets with subjects and items entered as random intercepts regressed against log-transformed reaction times, with a continuous frequency variable as well as categorical fixed effects of priming, time point, sentence type, aptness (high/low), and familiarity (high/low), as well as all

first-order, second-order and third-order interaction terms between priming, time point, sentence type, and familiarity/aptness. In this subsequent analysis we found no further modulating effects of familiarity or aptness.

Figure 4

Experiment 2: Time-course of Figurative Vs. Literal Target Activation in L1 participants



Target priming for figurative and literal targets at vehicle word onset and at recognition point + 1000ms, in metaphor and simile sentences in L1 English speakers. Error bars represent +/- 1 SEM. Asterisks represent significant main effects or interaction effects with * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Discussion

In Experiments 1 and 2 of STUDY 1, we examined online processing of figurative and literal targets related to the meanings of metaphor and simile sentences (which were identical but for the word *like*) using a cross-modal lexical priming paradigm where reaction times to visually presented written targets following aural presentation of sentence primes were measured. The results generally supported the idea that priming happens earlier in sentence processing for targets literally related to vehicle words, but for figuratively related targets the ease of processing depends on metaphor/simile familiarity. Additionally, ease of literal meaning processing for

similes varied with simile aptness, with high-aptness similes showing less literal target priming than low-aptness similes.

More specifically, when aptness and familiarity were not included in the models, both literal and figurative targets showed priming across time points (recognition point and 500ms post-recognition) and sentence conditions in Experiment 1, indicating that even using an 80ms presentation interval with forward and backward masking, semantic priming was possible for English L1 speakers for both figurative and literal meanings of metaphor and simile sentences. However, including sentence familiarity and aptness in the models demonstrated that these priming effects were indeed modulated by factors like familiarity and aptness, consistent with dual-processing models of metaphor processing. Figurative priming was observed at the recognition point for highly familiar metaphors and similes but not for low-familiar sentences, while for low-familiarity sentences figurative priming was observed later, suggesting that highly familiar metaphors/similes may be lexicalized and their intended meanings processed much faster than the meanings of novel metaphors and similes.

Interestingly, this pattern of results differs somewhat from the results of Blasko and Connine (1993), which found significant figurative priming for highly familiar metaphors at a time point more comparable to our later time point, and no overall figurative priming for low familiarity sentences. This may be explained by their use of a relatively longer target presentation time which may have allowed participants to recruit more conscious comprehension strategies. Another key difference may be that the metaphorical vehicle word was sentence-terminal in the Blasko and Connine (1993) study. This may have caused participants to focus all their cognitive resources on interpretation of the metaphor itself. In STUDY 1, sentences continued after the end of the metaphor/simile phrase, which may have prevented lingering

attention on the metaphor phrase once utterance comprehension was initially reached. Diverting participants' attention away from the vehicle word of the metaphor using a continuing sentential context, as we did in our study, may have encouraged participants to settle on their initial interpretation of each figurative phrase instead of consciously focusing on its interpretation.

Another potential explanation for the different pattern of results is that Blasko and Connine (1993) presented target words at a time point significantly later than either our early or intermediate-late time point (500ms post-recognition), and our results indicate that figurative meanings are primed and then fade relatively rapidly during the course of comprehension – so their study may not have been time-sensitive enough to catch figurative priming in the temporal "sweet spot" when figurative meanings are primed most strongly. Assuming this interpretation is correct, their findings could indicate that literal meanings are rejected more strongly in lexicalized metaphors (thus preventing figurative interpretations from fading quickly), while activation of figurative meaning in novel metaphor is relatively weaker or briefer. However, our study did not find an interaction between familiarity and figurative priming at the latest time point (1000ms post-recognition), which casts some doubt on this possibility.

While Blasko and Connine (1993) previously found that low-familiarity metaphors activated figurative meanings only when they were highly apt, we did not find an interaction between aptness and familiarity for figurative targets – figurative meanings were reliably activated for all low-familiarity metaphors, albeit slower than for highly familiar metaphors. It is possible that this is because all the metaphor primes in our study were at least moderately apt (rated between 5.00 and 7.22 on a 10-point scale), and extremely low-apt metaphors may not prime figurative meanings at all. This was unavoidable since the set of norms from which our stimuli were selected was based on an internet corpus search and thus all our stimuli were

sourced from natural language use by English speakers (Roncero & de Almeida, 2015); to investigate the effect of very low aptness it may be necessary to construct new metaphor phrases which would rarely be used in natural speech. Another possibility is that the later time point employed in the Blasko and Connine (1993) study may have reflected a point during comprehension where figurative meanings for low-apt sentences had already been activated, and subsequently rejected as nonsensical. In order to distinguish between these explanations, we employed a later time point in Experiment 2 to investigate whether figurative meanings continued to be primed later in comprehension for low-apt sentences.

For literal targets, there was more overall activation at the early time point, which diminished over time, except in the case of highly apt similes which showed negative priming at the recognition point. This effect is consistent with the predictions of the pragmatic model of metaphor processing, suggesting that literal meanings are activated immediately upon hearing the vehicle word, but that the literal meaning of the vehicle word is gradually replaced with figurative meanings as comprehension of the utterance meaning is reached by listeners. However, for metaphors, literal meaning activation was stronger for highly apt metaphors at both time points, suggesting that low-apt metaphors may be harder to process generally – there may be an inhibition effect on literal meanings due to the seemingly anomalous nature of novel metaphor constructions. If this is the case, then it can be assumed that phrase meaning, rather than single-word meaning, is given priority in metaphor processing – contrary to the theories of Giora (2003) who suggests that figurative meanings derive from the relative frequency of multiple interpretations of single words which are used figuratively.

For similes, this general inhibitory effect of low aptness was not observed, suggesting that similes are more likely to lead to literal interpretations generally, regardless of aptness. We

suggest that the inhibition effect at the recognition point for highly apt similes may be due to an anticipatory effect of the word *like* in similes – i.e., hearing the word *like* may prime the listener to expect a figurative phrase and this may lead to anticipation of the figurative phrase meaning in highly lexicalized figurative phrases such as *time is (like) money* (see Pambuccian & Raney, 2021, for a discussion of how the word *like* may invite figurative comparisons). In order to test this theory, we employed an earlier time point in Experiment 2 to test whether priming could be observed prior to the recognition point in anticipation of a figurative vehicle word in similes.

Overall, these results indicate that literal meanings related to the vehicle word are accessed at the first moments that the vehicle word in similes/metaphors is recognized and this activation fades over time, while figurative meanings related to the whole phrase are accessed later on in novel metaphors and similes and almost immediately in highly familiar or lexicalized metaphors/similes. This suggests there is some degree of simultaneous co-activation of literal and figurative meanings during the processing of metaphor and simile phrases, inconsistent with the claims of *Career of Metaphor theory* (Bowdle & Gentner, 2005), which posits that metaphors are interpreted either as categorization or comparison statements.

It appears, rather, that there is some degree of exhaustive access to multiple word/phrase meanings over the course of figurative language processing, with literal meanings activated earlier and for longer than figurative meanings – indicative of the primacy of literal meaning during sentence processing. Activation of the relevant figurative meanings appears to be brief, as it was already lower 500ms after the recognition point of highly familiar metaphor/simile vehicles. This is most consistent with the *dual-access theory* proposed by Carston (2010), which posits that ad-hoc meaning activation for metaphor phrases occurs rapidly and briefly while literal (and perhaps imagistic) meaning comprehension persists, aiding in the comprehension of

complex poetic and literary metaphor. While numerous studies (e.g., Inhoff et al., 1984; Janus & Bever, 1985; Wolff & Gentner, 2000) have investigated reading-time and comprehension differences between metaphorical and literal sentences, or tested activation after phrase offset (e.g., Blasko & Connine, 1993), their findings may be more indicative of deliberate comprehension processes than of *online* processing, so it is unclear how long figurative and literal meaning activation persists when a metaphor or simile is embedded within a longer passage of speech.

To ascertain whether these patterns apply to even more extreme early and late time points, we tested literal and figurative priming at word onset and 1000ms post-recognition point. Contrary to our findings in Experiment 1, Experiment 2 found that priming at the vehicle onset and 1000ms post-recognition point were insensitive to aptness and familiarity of the metaphor sentence. We found overall literal priming at both time points but figurative priming only for similes, at both time points. These results are consistent with our observations in Experiment 1, as they suggest there is a temporal "sweet spot" where figurative meanings of metaphors are activated at the point of comprehension, and the activation of figurative meanings decays rapidly thereafter, lasting less than 1000ms post-recognition of the vehicle word.

Outside of this narrow window, there was more activation of literal meanings of the vehicle word, consistent with the pragmatic model of metaphor comprehension which suggests that the literal meaning of words has primacy over figurative meanings. This result is also consistent with the *dual-access model* of metaphor comprehension proposed by Carston (2010), which suggests that literal meanings persist longer than ad-hoc meaning constructs for individual metaphor phrases. It is also possible that the word *because* later in the sentence interfered with activation of the figurative meaning of the metaphor when listening continued, as listeners may

have taken this as a cue to stop searching for meanings and listen for the explanation given by the speaker.

Our findings stand in contrast with those of Blasko and Connine (1993), who found that figurative meanings for familiar metaphors and highly apt novel metaphors were still activated after the offset of the vehicle word in a cross-modal lexical decision experiment. However, this may be readily explained by the differences in stimulus design between the two experiments – unlike Blasko and Connine (1993), we embedded metaphor phrases in a longer sentential context which discouraged conscious deliberation upon the meaning of the metaphor phrase. It is possible that sentence-terminal figurative language invites further (conscious) speculation upon its meaning, while a sentence which continues after the metaphor phrase may limit the attentional resources expended to interpret figurative meanings. There is some evidence that executive control mediates the speed at which metaphors are interpreted in various contexts (Columbus et al., 2015), suggesting that the effect of attention may contribute to the comprehension speed of certain metaphors.

For similes, on the other hand, figurative meanings and literal meanings were both activated at the onset and 1000ms post-recognition. The early activation of figurative meanings in the simile condition is consistent with our hypothesis that the word *like* invites the listener to anticipate the vehicle word, perhaps activating more meanings which are potentially related thematically to the topic word. The vehicle word onset may potentially represent a phase of sentence processing when literal meanings for the anticipated vehicle word are activated but not yet inhibited, as they are at the recognition point of highly familiar similes, suggestive of a brief point during simile processing when utterance meaning is optimally cognitively represented. It is also likely that literal meanings are less inhibited in later stages of simile comprehension as

similes are literal sentences which invite the listener to think literally about the meaning of the vehicle word even after figurative meanings have been activated.

However, a question remains: do the patterns observed in STUDY 1 apply equally to L2 learners? Studies on metaphor (Jankowiak et al., 2017; 2021; Tang et al., 2022) and idiom (Senaldi & Titone, 2022; Senaldi et al., 2022; Zhu & Minda, 2021) have found processing differences for figurative sentences in an L2, but we know of no other studies to date investigating the time-course of metaphor comprehension in L2 speakers. In STUDY 2, we addressed this question by repeating the same experiments with L2 English speakers.

Chapter 3: STUDY 2

Time-course of Metaphor and Simile Processing in L2 English Speakers

In STUDY 2, we investigated whether the patterns of meaning activation for metaphors and similes observed in STUDY 1 also applied to highly proficient English L2 speakers. As has been shown in past studies investigating metaphor and idiom, figurative language processing differs between L1 and L2 speakers, but to date, most research on online processing of figurative language in L2 focuses on idiom and relatively few studies investigating online metaphor processing in an L2 have been conducted. Werkmann Horvat et al. (2021) found that L2 participants did not show priming for highly lexicalized metaphorical targets in a cross-modal priming study, but did show priming for literal targets when hearing the same sentences. Another study found that metaphor dyads were processed more slowly by L2 speakers than by L1 speakers (Jankowiak et al., 2021). However, there is some evidence that L2 proficiency modulates figurative language comprehension (Tiv et al., 2021); thus it is unclear whether this applies to all L2 speakers equally.

Multiple electrophysiological studies have demonstrated processing differences between L1 and L2 metaphors. An ERP study by Jankowiak et al. (2017) found that early processing of metaphor differed for L1 but not L2 speakers – with both novel and conventional metaphors evoking late positivity in an L2 while only novel metaphor evoked late positivity in the L1 – suggesting that familiarity may play a larger role in metaphor processing in L1 speakers compared to L2 speakers. Another ERP study in Chinese-English bilinguals similarly found that L2, but not L1 metaphors evoked more negative N400s in the parietal region and less late positivity when participants were presented with complex scientific metaphors (Tang et al., 2022). These findings suggest that L2 speakers may have greater difficulty compositionally

building figurative meanings of metaphors, and, where possible, directly retrieving the figurative meanings of lexicalized or high-familiar metaphors (also see Senaldi & Titone, 2022). Although there is less of a literature on simile processing in L2, we hypothesized that online simile comprehension may be easier for L2 listeners because of the cueing provided by the word *like*.

We employed the same cross-modal lexical decision paradigm used in STUDY 1 to compare metaphor and simile processing in English L2 speakers with very high English proficiency. Due to the low number of participants, we combined the two earlier time points from STUDY 1 (onset and recognition point) into one "early" time point, and the two later time points (+500ms post-recognition and +1000ms post-recognition) into one "late" time point for our analysis. We hypothesized that processing of figurative meanings for metaphors would be slower overall for L2 English speakers relative to L1 speakers, and that familiarity would affect figurative meaning activation less relative to L1 speakers (however, there is some evidence that conventionality does ease idiom processing in L2 speakers; see Milburn et al., 2021). For similes, we expected that the word *like* would facilitate figurative meaning activation relative to metaphor sentences.

Method

Participants

Participants were 25 high-proficiency L2 English speakers between the ages of 19 and 35 (21F;4M, M_{age} = 22.80). Participants all reported normal or corrected-to-normal vision and normal hearing and met the following inclusion criteria: (1) They learned English before the age of 5 and identified English as their dominant language; (2) they learned one or more languages before English; (3) they reported no history of hearing or reading disability. Participants were

recruited via Concordia University's online participant pool or via online job listings and were compensated with course credit or $10.

Materials, Design and Procedure

Experimental materials, design and procedure were identical to those in STUDY 1, except that the early visual target presentation point was collapsed over word onset and recognition point, while the late presentation point was collapsed over 500ms post-recognition point and 1000ms after the recognition point, as determined by the gating study presented in Experiment 1.

Data Analysis

For a summary of data exclusions see Chapter 2. As in STUDY 1, separate analyses were performed for literal and figurative targets, and analyses were first performed on all trials without aptness/familiarity conditions, then subsequently performed with aptness/familiarity conditions included in the models. For each model, the estimated coefficient (β), standard error (SE), t, and p values are reported. Significance was evaluated using Satterthwaite approximations implemented in the lmerTest package (Kuznetsova et al., 2017).

Results

We fitted linear mixed-effects regression models for figurative targets and literal targets separately, using the lme4 (Bates et al., 2015) package in R (R Core Team, 2012), with subjects and items (vehicles) entered as random effects using a maximal random effects structure for intercepts and slopes (starting from a maximally specified random structure and consecutively removing slopes in case of convergence issues). In all analyses, log-transformed reaction time values were regressed against a continuous frequency variable, continuous trial order variable and categorical fixed effects variables of priming (control/experimental targets), sentence

condition (metaphor/simile) and time point (early/late), as well as first-order interactions between priming, sentence condition, and time point. Categorical values were sum-coded (for sentence condition, met: - 0.5, sim: + 0.5; for priming, control: - 0.5, related: + 0.5; for time point, early: - 0.5; late: + 0.5). As in STUDY 1, we were only interested in priming, not raw reaction time values, and thus report only main effects of the control/experimental target condition or interactions which included the control/experimental target condition. Results are summarized in Table 4.

Table 4

Summary of STUDY 2 Findings

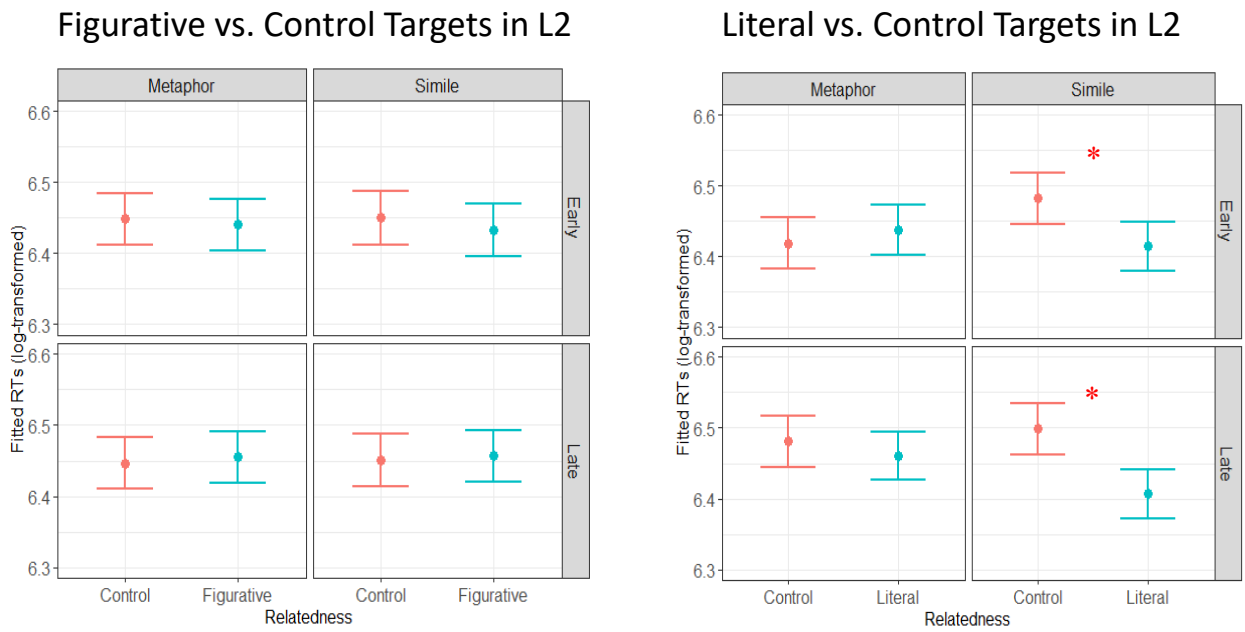
		Early	*Late*
Target type	**Sentence type**		
Figurative	*Metaphor*	No priming	No priming
	Simile	No priming	No priming
Literal	*Metaphor*	No priming	No priming
	Simile	Priming across aptness and familiarity conditions	Priming across aptness and familiarity conditions

For figurative targets, we found no significant main effect of priming or first-order interactions (Fig. 5, left panel). In the literal target condition (Fig. 5, right panel), we found a significant interaction between priming and sentence type, (β = - 0.079, t = - 2.30, SE = 0.035, p = 0.022), with significant priming only for simile sentences and not for metaphors.

We then conducted follow-up analyses for figurative and literal targets with subjects and items entered as random intercepts and log-transformed reaction times regressed against a continuous frequency variable and continuous trial order variable; categorical priming, time point, aptness (high/low), and familiarity (high/low) were entered as fixed effects, as well as all first-order, second-order and third-order interaction terms between priming, time point, sentence type, and familiarity/aptness. As above, aptness and familiarity were also sum-coded (for familiarity, high familiar: - 0.5; low familiar: + 0.5; for aptness, high apt: - 0.5; low apt: +0.5).

Figure 5

Time-course of Activation of Figurative and Literal Targets in L2 participants



Target priming for figurative and literal targets at early (onset/recognition point) and late (recognition point +500/+1000ms) time points, in metaphor and simile sentences in L2 English speakers. Error bars represent +/- 1 SEM. Asterisks represent significant interaction effects with * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

This analysis yielded no additional effect of familiarity or aptness on priming.

Discussion

In STUDY 2, we investigated moment-by-moment processing of metaphors and similes in in a sample of L2 English speakers, using a cross-modal lexical decision task to investigate priming of literally and figuratively related visual targets while participants listened to metaphor and simile sentences of varying levels of familiarity and aptness. The purpose of employing the cross-modal lexical decision task was to shed light on what figurative and literal meanings were activated during metaphor/simile processing in English L2 participants prior to conscious efforts to comprehend sentence meaning.

In highly proficient L2 English speakers, we found no overall priming of figurative meanings in either metaphor or simile sentence conditions, while priming of literal targets occurred only in the simile, but not the metaphor condition. It is possible that the lack of figurative priming in English L2 speakers could be due to the fact that the four time points from Experiments 1 and 2 (in STUDY 1) were collapsed into "early" and "late" time points and there is less overall priming of figurative targets at onset and 1000ms post-recognition point; however, this appears unlikely as 17 of the 25 participants in STUDY 2 were run on the time points from Experiment 1 (recognition point/+500ms post-recognition), which in L1 participants resulted in significant figurative priming for both metaphor and simile sentences. It appears more likely, therefore, that despite very high proficiency in English, L2 English speakers have more difficulty compositionally building meanings of metaphors and similes in their L2 and are also less likely to directly comprehend figurative meanings of lexicalized metaphors/similes (as demonstrated by the lack of familiarity effects).

These results are in line with the findings of Jankowiak et al. (2017), who demonstrated that both novel and figurative metaphors evoked a late positivity response in the brains of L2

speakers, and Jankowiak et al. (2021), which found that metaphor dyads were processed more slowly by L2 speakers than L1 speakers. Taken together with the outcome of STUDY 2, this suggests that L2 speakers, even high-proficiency bilinguals, have deficits in immediate online processing of figurative metaphor meanings.

This result also calls into question the underlying assumptions of direct-access metaphor processing theories (e.g. Colston & Gibbs, 2021) and some exhaustive-access models, at least in L2 speakers. The results of STUDY 2 demonstrate that even highly proficient L2 English speakers do not appear to easily access alternative figurative meanings for metaphor vehicles, even in highly familiar sentences such as *time is (like) money*. If certain vehicle words have highly salient figurative meanings which out-compete their literal meanings (as suggested by, e.g., Giora, 2003), then these meanings should also be apparent to fluent English speakers who learned another language before English, but our results suggest that this is not the case. In a study of idiom processing in an L2, Senaldi and Titone, (2022) also found that bilinguals had more difficulty directly retrieving lexicalized meanings of common idioms in their L2, so there may be global differences in processing of figurative language in an L2 regardless of language proficiency.

L2 English speakers also demonstrated no literal target activation in the metaphor condition, although they activated literal meanings in the simile condition. This may be due to the higher composability of simile phrases relative to metaphors, which may lead to literal meaning activation for simile vehicles. The literal non-plausibility of metaphors may interfere with the activation of literal meanings in metaphor in English L2 speakers, suggesting that literal plausibility of sentences may be more important for meaning comprehension in even highly proficient L2 speakers.

One limitation of STUDY 2 is that it investigated real-time figurative language processing only in highly proficient L2 speakers of English who report English as their dominant language that was learned before age 5. Despite this limitation, there appear to be clear differences in online processing of metaphors and similes by L1 and L2 speakers of English even at this very high level of proficiency.

It is unclear whether the differences between L1 and L2 speakers in metaphor/simile comprehension apply solely to online processing or whether there are also effects on general comprehension and offline processing of figurative meaning. This distinction has important consequences for real-world scenarios such as metaphor use in medical settings, news stories or casual speech, as a delay in online processing of figurative meanings could have minimal impact on L2 speakers' ability to understand figurative language in real-world settings, but a general deficit in (effortful) comprehension of figurative language would have wider-reaching effects on L2 speakers' ability to understand the many forms of figurative language employed in daily communication. For this reason, it is important to understand whether these differences in L1/L2 metaphor comprehension are limited to moment-by-moment processing of speech or whether they also affect general comprehension after the 1000ms threshold we tested in STUDIES 1 and 2.

Chapter 4: STUDY 3

Comparing Comprehension Ratings of Metaphor and Simile in L1 and L2 English Speakers

In STUDY 3, we asked the question: how does global comprehension of metaphors and similes vary between L1 and L2 speakers of English as a function of aptness and familiarity, as well as English proficiency? To investigate differences in overall comprehension of metaphors and similes (with the same constituents except the word *like*), and the relationship between comprehension, aptness and familiarity of these figurative phrases in readers of varying language backgrounds, STUDY 3 employed a survey paradigm including the metaphors and similes from STUDIES 1 and 2. Specifically, this judgment task determined how L1 and L2 English speaking participants' comprehensibility scores for 84 metaphors/similes interacted with their own aptness and familiarity ratings for those sentences. As in STUDIES 1 and 2, metaphor and simile sentences were taken from a set of norms published by Roncero and de Almeida (2015).

STUDIES 1 and 2 demonstrated online processing differences for metaphor and simile between L1 English speakers and high-proficiency English L2 speakers – L1 speakers were far more sensitive to aptness and familiarity than L2 speakers and demonstrated a priming effect for both figurative and literal targets in a cross-modal lexical decision task, while L2 speakers primed only literal meanings of simile sentences and no priming of any figurative targets was observed for L2 speakers. Other online processing studies of metaphor in an L2 (e.g. Jankowiak et al. 2017; 2021; Tang et al., 2022) have also demonstrated deficits in processing of metaphor by L2 speakers of a language using ERP. A recent cross-modal lexical priming study comparing literal language like *invest time* to highly lexicalized ("dead") metaphors such as *invest effort* in L2 also found that unlike L1 speakers, who processed both lexicalized metaphors and literal

primes faster than unrelated primes, L2 speakers only showed priming in the literal condition, and not for highly familiar lexicalized metaphors (Werkmann Horvat et al., 2021).

However, deficits in online processing of figurative language in L2 speakers do not necessarily mean that L2 speakers cannot ultimately comprehend metaphorical language in their L2 – it may simply mean that comprehending utterance meaning requires slower, more effortful processing. This is a key distinction for fields such as healthcare and second-language teaching because an inability to comprehend metaphorical language can have far-reaching consequences for L2 speakers (see Fuks, 2021; Maass, 2014). As Lakoff and Johnson (1980) described, figurative language is used in all aspects of life and is an integral component of natural language, so deficits in figurative language processing could fundamentally alter the lives of L2 speakers who must navigate healthcare, education, and socialization in their L2.

Studies on figurative language comprehension in an L2 usually focus on idiom rather than metaphor, although both idiomatic and non-idiomatic metaphors and similes are common in speech and writing (Lakoff & Johnson, 1980). The deficits commonly found in L2 idiom processing research are often hypothesized to be a result of compositional difficulty of idioms (Carrol & Conklin, 2014; Senaldi & Titone, 2022), and could be a result of most L2 speakers interpreting metaphorical language via a process of direct translation from their L1, as suggested by Taki (2013), who found Iranian-English bilinguals commonly reported translating idioms in order to understand them. If this is indeed the main comprehension strategy employed by L2 learners, aptness should be more important for metaphor comprehension than familiarity in the L2, as aptness facilitates the composition of phrases which make sense to readers/listeners.

Moreover, similes should be easier to compose than metaphors, since similes are clear comparison statements unlike metaphors which can rely on polysemy or literally untrue

constructions like *lawyers are sharks*. Conversely, if compositionality is not key to L2 understanding of figurative language, highly familiar or lexicalized metaphors should be easier for L2 speakers to comprehend due the frequency with which lexicalized metaphors are encountered in natural language.

Some studies have shown that students have difficulty comprehending even highly lexicalized metaphors in their L2, based on students' own self-report of difficulty of comprehension (Littlemore, 2001; Littlemore et al., 2011). It is unclear whether this effect applies to all bilinguals or only bilinguals with relatively low L2 proficiency, and to our knowledge no studies have compared bilingual comprehension of metaphors and similes of varying levels of familiarity and aptness.

To further determine whether the online processing deficits in L2 participants demonstrated in STUDIES 1 and 2 also apply to final sentence comprehension, STUDY 3 employed a survey paradigm which asked L1 and L2 English speakers to rate the comprehensibility of 84 metaphor and simile sentences, and also collected familiarity, aptness, and valence ratings for each sentence. These ratings were then compared to literal and anomalous sentences to determine how comprehensible both metaphors and similes were for L1 and L2 speakers compared to literal and anomalous sentences. Responses were collected from L1 speakers, as well as L2 speakers with a range of self-reported English proficiency and who had acquired English as an L2 at different times to further investigate whether age of acquisition and English exposure level in daily life affected comprehension of figurative English language.

Method

Participants

Participants were 36 bilingual L1 English speakers (24F;11M;1O, M_{age} = 24.69) with an average age of acquisition of their L2 of 7.11 years, and 24 L2 English speakers (16F;5M;3O, M_{age} = 25.79) with an average age of acquisition of English of 6.42 years. All participants reported written and verbal fluency in English, normal or corrected-to-normal vision and no history of learning disabilities or attention disorders. Participants were recruited from a variety of online research and job listings groups and received $15 for their participation.

Materials, Design and Procedure

Materials included 84 metaphor sentences and 84 simile sentences (identical to the metaphors except for the word *like*) taken from a set of published norms (Roncero & de Almeida, 2015) in two counterbalanced lists. We also generated 42 literal sentences and 42 anomalous sentences to act as controls for experimental materials; these control sentences were syntactically similar to the metaphor/simile sentences (i.e., taking the form *x is/are y*) (See Appendix B for a full list of stimuli).

We used Qualtrics (Qualtrics, Provo, UT) to construct two versions of an online survey containing one of two counterbalanced lists, each containing 42 metaphor and 42 simile versions of each figurative sentence as well as all 42 literal and all 42 anomalous control sentences. The survey first asked participants to answer a number of demographic questions and then a battery of questions about their language background including age of acquisition of all languages spoken and percent exposure to each language in various settings. After demographic and language background information was collected, participants were asked to answer questions about 168 sentences in two separate blocks (See Appendix C for a full list of survey questions).

The first block asked participants to rate each sentence on comprehensibility (along a 7-point scale from 1, not at all comprehensible, to 7, maximally comprehensible). Participants were

given the example *Jane is an airplane* as a not at all comprehensible sentence, and *Jane is a doctor* as a maximally comprehensible sentence. Second, participants were asked to rate the same sentences on familiarity (on a 7-point scale from 1, not at all familiar, to 7, maximally familiar).

The second block asked participants to rate the same sentences (from 1-7) on whether they were interpretatively apt and emotionally charged. Participants were told that *politics is a jungle* is a highly apt sentence because the meaning of the second content word (*jungle*) captures salient properties of the first content word (*politics*); in contrast, *politics is a desk* is less straightforwardly apt because the meaning of the second content word (*desk*) has no obvious bearing on any salient property of the first content word (*politics*). For emotional valence we defined 1 as "negatively emotionally charged" and 7 as "positively emotionally charged" where the sentence *puppies are joyful* is a positively emotionally charged sentence and *war is death* is a negatively emotionally charged sentence. For all ratings, participants were encouraged to use intermediate ratings for sentences that were somewhere in between maximal and minimal values.

Data Analysis

Analysis was limited to participants who spoke at least two languages in order to control for effects of monolingualism/bilingualism. Seven participants were removed from analysis for being pure English monolinguals. In order to exclude participants who were not attending to the task, we excluded participants who rated anomalous sentences as on average 3.5/7 or higher on familiarity (14 participants). One participant was removed from the data for having self-contradictory answers on the demographic questions. Answers from the remaining 60 participants were pooled for analysis of language differences and then analysed separately for effects of language background on comprehension. For each model, the estimated coefficient (β),

standard error (SE), t, and p values are reported. Significance was evaluated using Satterthwaite approximations implemented in the lmerTest package (Kuznetsova et al., 2017).

Results

To test language differences in comprehension, a linear mixed effects regression analysis was conducted on comprehensibility ratings for metaphor/simile sentences using the lme4 (Bates et al., 2015) package in R (R Core Team, 2012). Items (sentences) and subjects were entered as random effects with random intercepts, with aptness, familiarity and valence entered as fixed effects. Sentence type and language group were treated as categorical variables and sum-coded (simile: + 0.5 vs metaphor: - 0.5; L1: - 0.5 vs L2: + 0.5). Aptness, familiarity, and valence were entered as scaled effects using the average score from L1 English speakers for each item. Comprehensibility scores were regressed against categorical terms of sentence type (metaphor/simile) and language background (L1/L2) as well as one-way interactions between sentence type and language background. Additionally, we explored two- and three-way interactions between sentence type, language background, and aptness as well as interactions between sentence type, language background and familiarity variables to avoid interactions between aptness/familiarity scores themselves. We did additionally explore two- and three-way interactions between sentence type, language background, and emotional valence to control for potential interactions with valence scores although valence was not a focus of our analysis.

Our analysis produced an interaction effect of sentence type, language group and familiarity on sentence comprehensibility ($\beta = 0.74$, $t = 3.86$, $SE = 0.19$, $p = 0.0001$) (see Fig. 6, upper panels). To further investigate this interaction, we ran additional models on each language group subset in isolation and found that, in L1 speakers, there was a significant interaction between sentence type and familiarity ($\beta = - 0.41$, $t = - 3.54$, $SE = 0.12$, $p = 0.0041$) (see

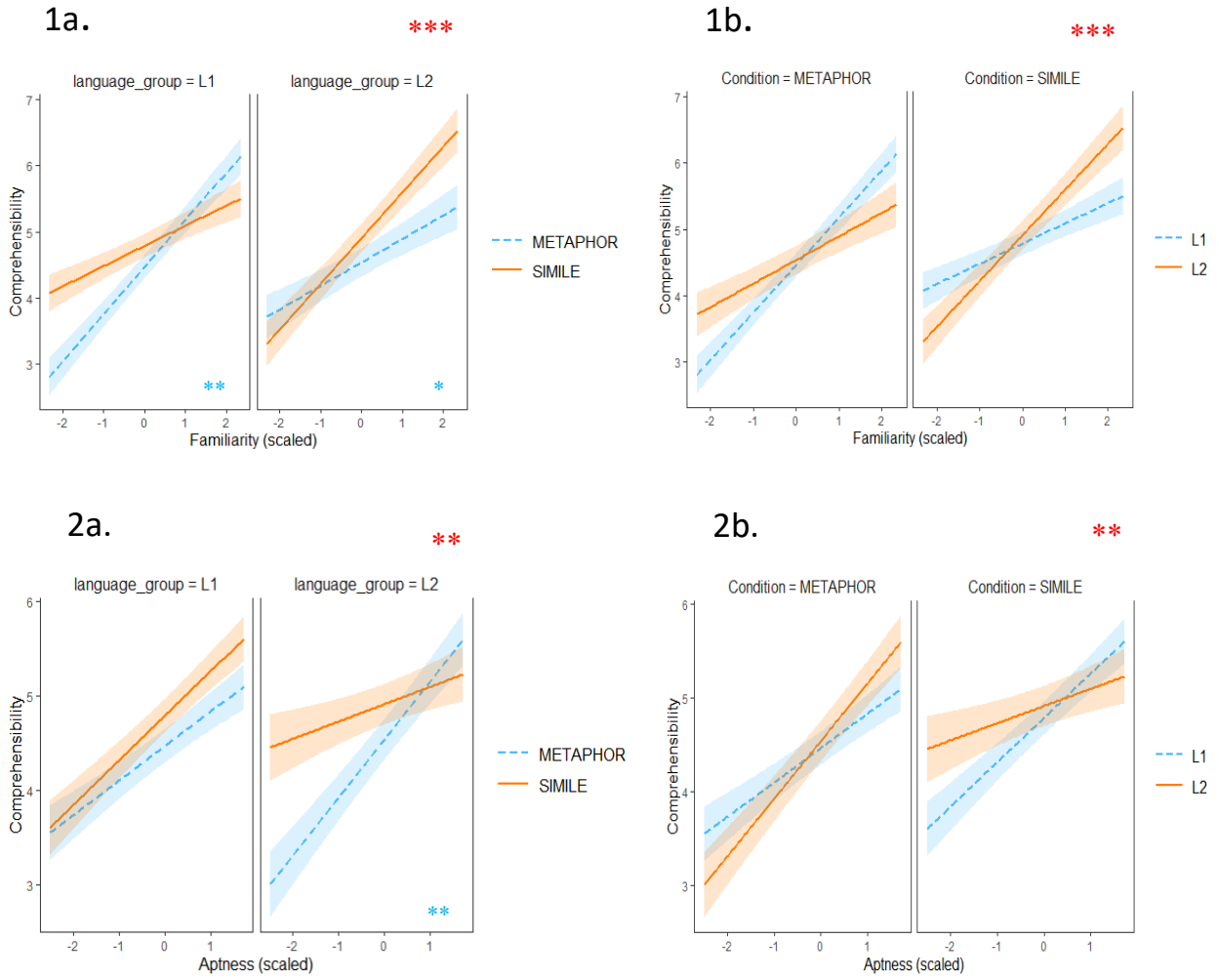
Appendix D) with L1 speakers rating high-familiarity metaphors more comprehensible than high-familiarity similes, and low-familiarity metaphors less comprehensible than low-familiarity similes. L2 English speakers rated low-familiarity metaphors and similes similarly comprehensible, but rated high-familiarity similes more comprehensible than high-familiarity metaphors ($\beta = 0.32$, $t = 2.05$, $SE = 0.16$, $p = 0.041$) (see Appendix D).

A three-way interaction effect of sentence type, language group and aptness on comprehensibility was also found ($\beta = -0.065$, $t = -2.87$, $SE = 0.19$, $p = 0.0048$) (see Fig. 6, lower panels). When additional analyses were performed on each language group subset, we found no interaction between aptness and sentence type in L1 English Speakers. In the sample of L2 English speakers we did find an interaction between sentence type and aptness, with L2 speakers rating highly apt metaphors and similes similarly comprehensible but rating low-apt similes more comprehensible than low-apt metaphors ($\beta = -0.42$, $t = -2.71$, $SE = 0.15$, $p = 0.0067$) (see Appendix D).

To test the effect of language background on L1 and L2 English speakers' comprehensibility ratings, we ran two separate linear mixed effects regression models on each language group, including age of acquisition of English (for L2 speakers) or of their L2 (for L1 English speakers) and trial order as scaled continuous fixed effects and their one-way and two-way interactions with each other and valence, aptness and familiarity separately; item (sentence) and subject were entered as random effects into the model. For the L1 English speakers we again used the average aptness/valence/familiarity ratings from L1 participants for each item but for the L2 English speakers we used the average L2 ratings for each item.

Figure 6

Comprehensibility Ratings for Metaphor and Simile in L1 and L2 English Speakers



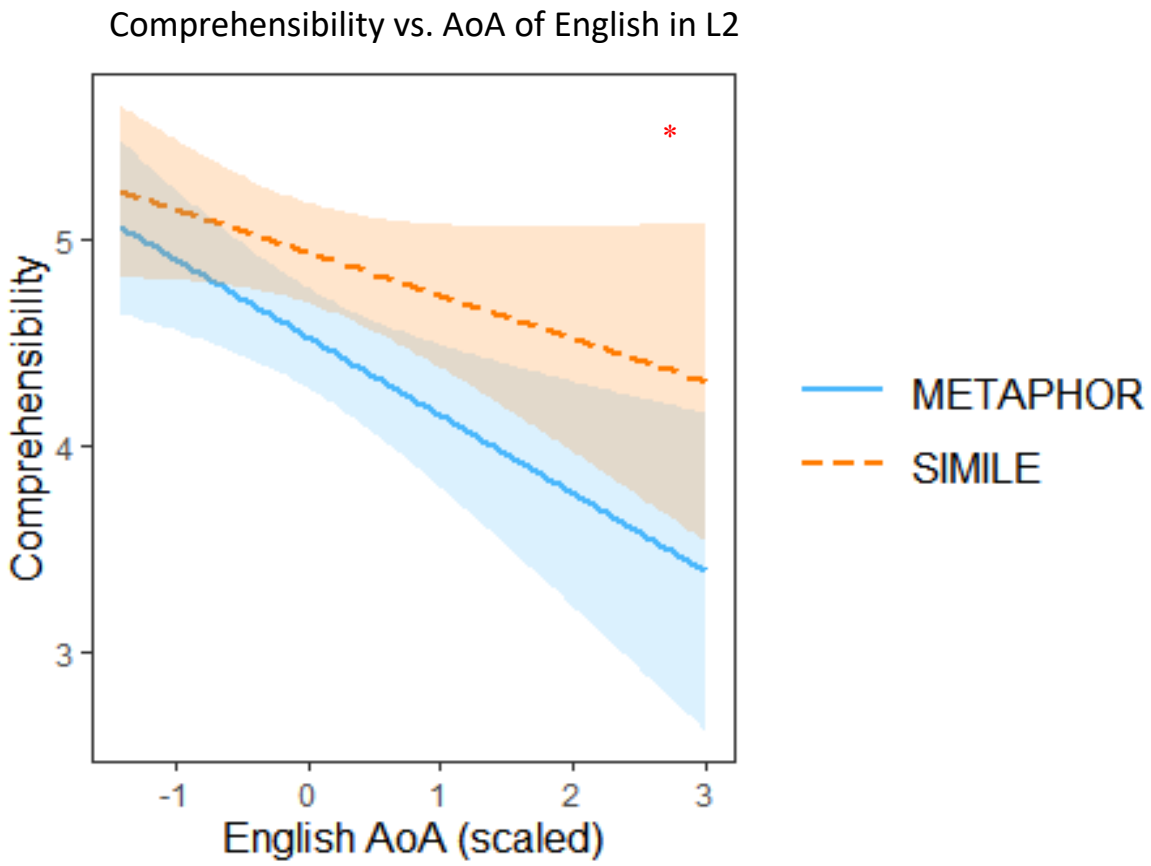
Comprehensibility ratings for sentences as a function of sentence type (metaphor/simile), language group (L1/L2), familiarity (1a, b), and aptness (2a, b). +/- 1 SEM represented by shaded areas. Red asterisks represent significant interaction effects, with * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. (Blue asterisks indicate p-values derived from additional models run on language groups in isolation).

The analysis of L2 English speakers' comprehensibility scores produced a significant interaction between sentence type and age of acquisition ($\beta = 0.17$, $t = 2.15$, $SE = 0.079$, $p = 0.032$) (see Fig. 7); the later L2 participants learned English, the lower their overall comprehensibility scores, with late English learners rating similes more comprehensible than

metaphors. In L1 English speakers, the age of acquisition of their L2 produced no significant main effects or interactions with comprehensibility ratings of English metaphors/similes.

Figure 7

Comprehensibility Scores for Metaphor and Simile Sentences as a Function of Age of Acquisition of English (in L2 English Speakers)



Comprehensibility ratings for sentences as a function of sentence type (metaphor/simile) and age of acquisition of English in L2 English speakers. +/- 1 SEM represented by shaded areas. Asterisk represents significant interaction effect, with * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Discussion

We performed linear mixed-effects model analyses to test differences in global metaphor/simile comprehension between L1 and L2 English speakers when given unlimited time to rate the comprehensibility of figurative sentences in a survey task. Participants were asked to

rate familiarity, aptness, emotional valence, and overall comprehensibility of sentences and both L1 and L2 English speakers' comprehensibility ratings were regressed against L1 English speakers' aptness and familiarity ratings to elucidate relationships between the aptness and familiarity of a figurative sentence and its overall comprehensibility to both L1 and L2 speakers.

Overall, our findings indicate that greater familiarity and aptness of a figurative sentence both increased comprehensibility scores for metaphor/simile, for L1 as well as L2 English speakers. Our results indicated that L1 English speakers rated high-familiarity metaphors significantly more comprehensible than high-familiarity similes, and low-familiarity similes more comprehensible than low-familiarity metaphor. This result is broadly compatible with the idea of lexicalized or "dead" metaphor famously put forward by Searle (1979), which posits that certain metaphorical phrases are repeated so frequently in spoken/written language as to become lexicalized in and of themselves – once a metaphorical meaning is lexicalized, the metaphorical meaning of the phrase should presumably become more salient to a fluent speaker of the language than the literal meaning which could be derived from that phrase.

If we imagine that an L1 English speaker hears *time is money* and instantly understands the intended utterance meaning (which is the figurative meaning), then that phrase should be more immediately comprehensible than a phrase which is not lexicalized, such as *time is like money*. The addition of *like* to the sentence introduces additional novelty to a frequently encountered sentence, potentially making the sentence appear less comprehensible to a fluent L1 speaker than the metaphorical expression. Thus, our findings are compatible with the idea that certain metaphorical phrases are fully lexicalized – to the extent that the corresponding simile, being less frequently employed in natural language, seems less comprehensible. It stands to reason that L2 speakers, who have less lifetime exposure to these English phrases, should not

necessarily find these lexicalized phrases more comprehensible in their metaphor iteration than in their simile iteration.

Moreover, our results indicated that in the case of novel metaphor – e.g., *love is a child* – L1 English speakers preferred the simile version and found it more comprehensible than the metaphor version of the same utterance. This is in accordance with the conclusion that metaphor is favoured over simile if it is highly lexicalized; novel metaphor, which is not lexicalized, should naturally be less comprehensible than novel simile, since the word *like* in simile renders it more logically comprehensible. Assuming the pragmatic theory of metaphor comprehension is correct and phrases are first parsed literally before being subject to pragmatic processes (Davidson, 1978; Grice, 1975; Searle, 1979), L1 English speakers should prioritize literal meaning when first hearing a sentence, and thus give higher comprehensibility ratings to highly composable sentences which make sense literally, like any simile employing the word *like* (which renders the sentence literally comprehensible in all cases).

Even when a reader is given time to ponder the meaning of the sentence, a sentence which is literally true should still be more comprehensible assuming its figurative meaning is novel and unfamiliar. Indeed, this is what we found when probing the global comprehensibility ratings of L1 English speakers – only highly familiar (lexicalized) metaphors were rated more comprehensible than similes with the same constituents. This suggests that sentence decomposability is subordinate to its common usage in a reader's L1, but when a sentence is not commonly used, decomposability of the sentence contributes to comprehensibility. Although aptness did correlate with higher comprehensibility overall in L1 speakers in our study, there was no interaction with sentence type, suggesting that higher aptness does not specifically affect which sentence type (metaphor or simile) is preferred by L1 speakers.

L2 English speakers rated low-familiarity metaphors and similes similarly comprehensible, but rated high-familiarity similes more comprehensible than high-familiarity metaphors. There was also an interaction between aptness and comprehensibility in L2 speakers, who rated highly apt metaphors and similes similarly comprehensible; however, low-apt similes were rated more comprehensible than low-apt metaphors. Taken together, these results suggest that L2 speakers found simile more comprehensible than metaphor, and did not prefer the lexicalized versions of common metaphors like *time is money*. This could indicate that even fluent L2 speakers have difficulty incorporating alternative, figurative meanings of words and phrases when they have learned the literal meanings of those words/phrases first. It is possible that L2 speakers, even at a high level of fluency, conceptually represent word meanings in their first language and mentally translate word meanings from their L2 to their L1 when communicating in an L2 (as Taki, 2013, suggests), leading to increased processing difficulty when a phrase does not make literal sense and is not commonly used in their L1.

As aptness decreased, L2 speakers likely relied even more on the compositional nature of a sentence to make sense of it; thus, literally comprehensible/decomposable similes were more easily comprehended by L2 speakers, compared to literally nonsensical metaphors. Interestingly, this finding is also consistent with findings from the idiom literature: Senaldi et al. (2022) found in an eye-tracking study that idioms with the terminal noun represented in the L2 of English-French bilinguals (e.g., *break the glace* rather than *break the ice*) incurred additional processing costs compared to literal sentences. This suggests that direct retrieval is impaired when a figurative phrase is presented in a non-standard form, and code-switching forces readers to rely on compositional parsing of the sentence.

Senaldi and Titone (2022) posited that L1 readers initially directly retrieved figurative meanings of idioms, and then later decomposed them into constituent words to further comprehend them, but L2 readers relied less on direct retrieval, and the present study extends this finding to nominal metaphor. Although our study used sentences fully in the L2, our results suggest that L2 speakers still preferentially parsed L2 phrases compositionally even when they were lexicalized, but this effect seemed to be especially strong when the metaphor or simile was less apt and therefore more difficult to comprehend. In the case of similes, compositional parsing is made easier by the word *like* and thus the sentence should be more easily comprehended in an L2.

Finally, comprehensibility scores for each language group were regressed against L2 age of acquisition, and average familiarity/aptness scores for that group, to investigate the relationship between language background and comprehension. Using L2 participants' own familiarity and aptness ratings in the model allowed us to test whether this effect was due to L2 participants being more familiar with different figurative expressions than L1 participants, but even when L2 familiarity and aptness ratings were used there was no interaction between familiarity or aptness, age of acquisition and metaphor/simile comprehensibility.

L2 participants who learned English later in life had lower comprehensibility scores overall than earlier English learners, and comprehensibility scores for metaphor and simile diverged, with later English learners rating similes more comprehensible than metaphors compared to earlier English learners. This finding suggests that the later an L2 speaker learns their L2, the more reliant they are on compositional parsing of figurative language and the less likely they are to fully comprehend nominal metaphor. Those in our sample who learned English

earlier did not find similes significantly more comprehensible than metaphors overall, and may be less reliant on sentence decomposability for sense-making.

This does not necessarily indicate that early L2 learners rely as heavily on direct retrieval as L1 speakers do, as participants were given unlimited time to make a judgment – only that they ultimately judged metaphors and similes as equally comprehensible. On the other hand, there was no effect of L1 English speakers' L2 age of acquisition on their comprehensibility ratings, suggesting that simply learning an L2 earlier or later in life is unlikely to dramatically affect figurative language processing in their L1.

Late L2 learners' preference for simile over metaphor could have applications in, e.g., healthcare settings or educational materials aimed at L2 speakers. Specifically, in situations where figurative language conveys important concepts accurately, using the simile form of a nominal metaphor to ease comprehension could be an easy way for healthcare providers or educators to modify their normal way of speaking without losing the meaning conveyed by a figurative phrase. Considering that figurative language is frequently used in speech by L1 speakers, using the simile rather than the metaphor form of certain figurative expressions could form a simple compromise and maximize effective communication of concepts which are often best or most simply conveyed metaphorically.

Overall, our findings are consistent with work from the idiom literature (e.g. Senaldi & Titone, 2022; Senaldi et al., 2022) suggesting that figurative language in an L2 is less immediately comprehensible than in an L1, likely because L2 parsing of figurative language is more compositional than in an L1 and less likely to depend on direct retrieval of a phrase. We also found that this effect was stronger the later participants learned their L2, even when L2 participants' own familiarity and aptness ratings were regressed against their sentence

comprehensibility scores. Taken together, these findings suggest that direct retrieval of lexicalized metaphors is not a significant contributor to figurative language comprehension in L2 participants, since they rated even highly familiar simile forms more comprehensible than corresponding metaphors with identical topic and vehicle words. In many of these cases, the metaphor form is more common than the simile and thus if direct retrieval was in play, metaphors should be easier to comprehend than similes. L1 English speakers, on the other hand, rated highly familiar metaphors more comprehensible than their simile form, suggesting a lesser reliance on compositional parsing when encountering familiar figurative phrases.

Because STUDY 3 only investigated comprehensibility ratings of nominal metaphors in the form *x is y* and corresponding similes in the form *x is like y*, it is unclear whether our results would apply to other forms of metaphor like dyads and metonymy or extended poetic metaphor. However, our results suggest that the more decomposable a figurative phrase is, the more easily it should be understood by an L2 speaker. Future studies should investigate the differences between L1 and L2 comprehension of other forms of metaphor and how metaphor comprehension is affected by context in L1 versus L2 speakers.

Chapter 5: General Discussion

In this dissertation, I investigated differences in both online and offline metaphor and simile processing and comprehension measures in L1 and L2 speakers of English. Taken as a whole, my dissertation addressed one overarching question: how are metaphors and similes processed in real-time, and ultimately comprehended, by L1 and L2 speakers of English, and how is this modulated by the way these expressions can differ vis-à-vis aptness and familiarity of each sentence? STUDY 1 employed a time-sensitive cross-modal lexical decision task to investigate the time-course of literal and figurative meaning activation in L1 English speakers as participants heard copular metaphors and similes using the same constituents except for the word *like* at four different time points. STUDY 1 also investigated whether the activation of literal and figurative meanings was affected by familiarity and aptness of the figurative expression. STUDY 2 investigated the time-course of literal and figurative meaning activation in fluent English L2 speakers as they heard the same copular metaphors and similes, collapsed over two early and late time points. STUDY 3 investigated global comprehensibility of copular metaphors and similes with the same topic and vehicle words, in both L1 and L2 speakers of English, and how familiarity and aptness affected the comprehensibility of figurative sentences.

In the discussion to follow, I present evidence from STUDIES 1 and 2 demonstrating that activation of literal and figurative meanings of metaphor is time-sensitive and dependent on aptness and familiarity; I also discuss how online metaphor and simile processing differs between L1 and L2 speakers. I next discuss how ultimate metaphor/simile comprehension differs by sentence type between L1 and L2 speakers, and how this relates to online processing. I then integrate these findings with various competing models of metaphor processing. In doing so, I argue that these findings collectively support a mixed or dual-access model of metaphor

processing in which both literal compositionality and ad-hoc rapid figurative meaning construction play a role, and explain how this differs for L1 and L2 speakers. I conclude with a discussion of the implications of these findings for communication aimed at L2 speakers.

Time-Course of Activation of Literal and Figurative Meaning

In STUDIES 1 and 2, participants were presented with a metaphor or simile prime sentence embedded in context aurally over headphones while figuratively-related, literally-related, or unrelated target words were presented on a screen (following Blasko & Connine, 1993). Participants were asked to press a button indicating whether the target word was an English word or not, and priming (the difference between time to respond to an unrelated word versus a related word) was taken as an indication of whether that figurative or literal meaning was activated at a specific point during sentence listening. Although the study followed protocols from the seminal work of Blasko and Connine (1993), our updated version of their experiments employed a far more time-sensitive task to investigate the exact timing of figurative and literal meaning activation as listeners heard metaphors in real-time, and used similes as direct literal controls for metaphor sentences in order to compare the processing of literally plausible versus purely figurative sentences. L1 speakers of English (STUDY 1) and English L2 speakers (STUDY 2) followed identical protocols, but due to a lower number of participants the four time points for English L1 speakers (word onset, recognition point, 500ms post-recognition and 1000ms post-recognition) were collapsed into two time points (early and late).

L1 speakers of English displayed significant priming across all conditions when aptness and familiarity were not controlled. When aptness and familiarity were included in the models, L1 English speakers had significant priming of figurative meanings at the vehicle word recognition point for high-familiar metaphor and simile, and later activation of figurative

meanings for low-familiar metaphor/simile. Outside of the recognition point and +500ms post-recognition time window, L1 participants did not demonstrate priming of figurative meanings, suggesting there is a temporal "sweet spot" for figurative meaning activation during sentence listening in L1 speakers, and that the moment of activation of figurative meaning depends on metaphor/simile familiarity.

Literal meanings, however, were activated at recognition point for metaphor sentences, while highly apt similes demonstrated inhibition of literal meaning at the recognition point. Literal meanings were also activated at vehicle word onset and at +1000ms post-recognition point in L1 speakers, but not at +500ms post-recognition point, suggesting that literal meanings corresponding to a figurative sentence are activated early on, briefly suppressed when figurative meanings are activated, and then persist past the point where figurative meanings are no longer activated. Our findings provide evidence for the lexicalization, i.e., direct retrieval of figurative meaning, of highly familiar metaphor as well as the corresponding simile, and suggest that slower pragmatic processes are engaged for metaphors which are less familiar. However, literal meaning was activated early on in metaphor/simile processing and continued to be activated after figurative meaning activation faded, suggesting primacy of literal meaning outside of the "sweet spot" for figurative activation.

The findings for L2 English speakers in STUDY 2 differed substantially, suggesting that L2 speakers do not process figurative language in the same way L1 speakers do. L2 speakers demonstrated no overall figurative priming at any time point, while literal meanings were primed only for simile sentences. This suggests that in the first moments after hearing a metaphor, L2 speakers struggled to comprehend the meaning of the metaphor outright, perhaps suggesting that the non-composability of copular metaphor phrases interferes with L2 speakers' comprehension

of the sentence as a whole, and maybe even its constituent words. L2 speakers did demonstrate priming of literal meanings in the simile condition, suggesting that the literal comprehensibility of a simile was easier for them to initially process, although they did not immediately grasp its (intended) figurative meaning. The higher composability of simile phrases may facilitate literal meaning activation for L2 speakers, making some interpretation of simile meaning immediately accessible.

Moreover, there were no interactions between familiarity, aptness, and priming in our L2 cohort, suggesting that L2 speakers did not benefit from the same lexicalization process as L1 participants, and even highly familiar metaphors did not activate figurative interpretations in the early stages of processing. Thus, frequent exposure to a metaphor alone does not seem to induce primacy of figurative meaning in L2 speakers as it does in L1 speakers, suggesting that composability, not direct access to a lexicalized meaning, is more critical to L2 comprehension of both simile and metaphor.

Taken together, these results suggest that L2 speakers experience significant processing difficulties compared to L1 speakers when initially encountering figurative sentences, no matter how familiar or apt. This effect is stronger in the case of metaphor, which is not literally comprehensible or composable, and may interfere with overall linguistic processing in L2 speakers. The sample of L2 participants in our study were very early English learners, implying that this effect holds even for highly fluent L2 English speakers with very early exposure to English.

Global Comprehension of Metaphor/Simile Meaning

To test whether initial processing differences between L1 and L2 English speakers affect ultimate comprehension of figurative sentences, STUDY 3 employed an offline rating task

testing the comprehension of 84 metaphor and simile sentences by both L1 and L2 speakers. Specifically, in STUDY 3, we asked participants to rate each sentence on a Likert scale from 1 to 7 on measures of comprehensibility, familiarity, aptness, and emotional valence. Unlike in STUDIES 1 and 2, there was no hard time limit given to participants and they were asked to actively assess whether a sentence was comprehensible, rather than immediate meaning activation being probed. This allowed us to assess whether impaired or slowed online processing of figurative meaning carried over to ultimate comprehension of sentential meaning and whether initial processing deficits in L2 participants affected ultimate comprehension of figurative sentences.

When the L1 and L2 participants were given ample time to consider the meaning of each sentence, there were still significant differences in their comprehensibility ratings for each sentence type, with aptness and familiarity exerting differential effects on L1 and L2 comprehension of metaphors and similes. Overall, aptness and familiarity (based on the average ratings of L1 speakers) increased comprehensibility scores for both L1 and L2 English speakers. However, L1 speakers showed an advantage for highly familiar metaphor over highly familiar simile, suggesting that lexicalization has a positive effect on ultimate metaphor comprehension for L1 English participants. L2 participants, conversely, rated similes more comprehensible than metaphors with increasing familiarity, suggesting there is no lexicalization effect for L2 speakers when comprehending metaphor. This result is consistent with our findings from STUDIES 1 and 2, demonstrating that there is a significant lexicalization effect (i.e., direct or rapid access to the figurative meaning) for L1 English speakers when a metaphor is sufficiently familiar, and that this carries over to ultimate comprehension, while the same is not true for L2 speakers.

Unlike L1 speakers, L2 speakers also rated low-apt similes as more comprehensible than low-apt metaphors, suggesting an L2 advantage for phrases that are more decomposable or (potentially) literally true. This suggests that when the topic and vehicle are less obviously related, L2 English speakers rely more heavily on composability to make sense of the sentence and have a difficult time parsing sentences which cannot be literally true, i.e., metaphors. These findings were consistent with the results from STUDY 2, suggesting that composability or literal plausibility of a sentence is much more important for L2 speakers than it is for L1 speakers when processing the same figurative sentences.

Indeed, the only priming evident in STUDY 2 in our L2 cohort was literal priming during online processing of similes, and it appears that this preference for compositional parsing carries over to ultimate comprehension, with L2 speakers continuing to find similes easier to comprehend even when given time to consider full sentence meaning. The results from STUDY 3 suggest that L2 speakers were ultimately able to comprehend both metaphors and similes when given enough time, assuming the metaphor/simile in question was sufficiently apt, but when metaphors were less apt L2 participants struggled, ultimately, to comprehend them, even given time to think about sentence meaning.

What is made clear by STUDIES 1-3 in the present dissertation is that parsing of figurative meaning is not universal across L1 speakers of a language and L2 speakers, either at the initial processing stage or at the stage of ultimate sentence comprehension. Although metaphor permeates natural language (Colston & Gibbs, 2021; Lakoff & Johnson, 1980), it cannot be assumed that it is understood equally well or quickly by L1 and L2 speakers, and the processes by which L2 speakers arrive at utterance meaning for metaphor and simile appear to be quite different than those employed by L1 speakers. Most notably, L1 speakers appear to easily

and near-instantaneously understand the intended meaning of lexicalized, familiar utterances like *time is money* while L2 speakers may never attain native-like processing of these familiar phrases.

Regardless, the results of the above experiments can help us understand what happens in the brain during metaphor and simile processing and distinguish which common models of metaphor processing most accurately describe the cognitive processes involved in metaphor processing in the brain. Notably, outside of lexicalized metaphors, literal meanings seem to be activated earlier and for longer in both L1 and L2 speakers when first hearing a metaphor or simile, and appear to be active longer in L1 speakers even when they rapidly activate figurative meanings.

In this sense, there are significant commonalities between L1 and L2 speakers where comprehension of metaphor is concerned: there appears to be an overall primacy of literal meaning in the early processing stages, and both metaphor and simile comprehensibility is largely dependent on the aptness and familiarity of these figurative expressions in both L1 and L2 speakers. Outside of lexicalized metaphor in L1 speakers, our results indicate that there is a delay in the activation of figurative meanings even in L1 speakers and that figurative meaning is ultimately difficult for L2 speakers to attain, calling into question the idea that figurative meanings are accessed directly during metaphor processing, in either L1 or L2 speakers of a language.

Evidence for Models of Metaphor Comprehension

There is ongoing debate between psycholinguists about how metaphors are processed, predominantly between proponents of pragmatic models (e.g., Davidson, 1978; Grice, 1975; Roncero et al., 2021; Searle, 1979) and direct-access models (e.g., Colston & Gibbs, 2021; Gibbs

& Colston; 2012; Glucksberg, 2003; Glucksberg & Keysar, 1990; Wolff & Gentner, 2000). Pragmatic theorists suggest a three-stage process of comprehension in which metaphors are first processed as fully literal sentences, then rejected as literally untrue, and subsequently listeners use context and pragmatic cues to reach comprehension of the intended utterance meaning (see Searle, 1979). Proponents of direct-access theories argue that figurative language is processed directly without any need for rejection of the literal, propositional meaning (Colston & Gibbs, 2021; see Gibbs & Colston, 2012, for a review), although there is some disagreement in the field about how precisely this direct access to the figurative meaning is attained (see Katz, 2024).

For instance, some argue that metaphors are understood as comparisons between topic and vehicle (e.g., Wolff & Gentner, 2000) where common features of topic and vehicle words are non-hierarchically compared and the most frequent overlapping meanings or attributes give rise to the figurative interpretation of the phrase. In contrast, categorization or predication models (e.g., Glucksberg & Keysar, 2000; Glucksberg, 2003; Kintsch, 2001; Utsumi & Sakamoto, 2011) suggest that the vehicle term in a metaphor represents a hierarchically superordinate category to which the topic belongs – in the sentence *jobs are jails*, for example, *jails* represent a category of oppressive or limiting environments to which *jobs* metaphorically belong. In the former instance, similes are considered mostly indistinguishable in meaning from metaphors as they both compare topic and vehicle non-hierarchically, whereas in the latter instance, similes and metaphors differ in meaning.

The Career of Metaphor theory proposed by Bowdle and Gentner (2005) marries comparison and categorization models, suggesting that novel metaphors (like simile) function as comparison statements, and familiar or conventional metaphors function as categorization statements (and thus the metaphorical form of a figurative statement should be preferred if it is

familiar, while simile should be preferred if a phrase is unfamiliar). However, pragmatic theories of metaphor comprehension, like Career of Metaphor theory, allow for faster processing of highly familiar metaphors through a process of lexicalization, where common metaphorical phrases become "dead" (Searle, 1979) and are processed directly according to their commonly understood figurative meaning. The key difference between direct-access models and pragmatic models is that pragmatic models assume the literal meaning must be accessed and subsequently rejected before utterance meaning can be reached, while direct-access models do not assume any rejection of literal meaning is necessary for comprehension (Colston & Gibbs, 2021; see Gibbs & Colston, 2012, for a review).

The findings from STUDY 1 suggest that literal meanings are indeed activated earlier than figurative meanings in moment-by-moment metaphor and simile processing, particularly when metaphors are novel: less-familiar metaphors and similes did not induce priming of figurative targets at the recognition point of the vehicle word, but highly familiar metaphors and similes both did. This is compatible with both Career of Metaphor theory (Bowdle & Gentner, 2005) and pragmatic theories positing some phrases are lexicalized (e.g., Searle, 1979). However, literal meanings were also activated alongside figurative meanings even in highly familiar metaphors, especially at earlier processing stages, with the only exception being high-aptness similes, which had inhibited literal priming at the recognition point.

Career of Metaphor theory (Bowdle & Gentner, 2005) predicts that similes should be better at triggering figurative meaning activation in the novel condition (due to being comparison statements) and metaphors should be better at triggering it in the familiar condition (as categorization statements), but our findings did not demonstrate a clear preference for metaphor or simile depending on familiarity. This also calls into question the notion that literal meanings

must be fully rejected before figurative meaning can be reached, as Searle (1979) suggests – and familiarity did not seem to inhibit literal meaning, as it should if these are "dead" phrases which only connote their figurative meaning.

It is important to note that not all theorists believe lexicalization means a phrase is "dead" or fully set in its meaning. Colston and Gibbs (2021) suggest that metaphorical idioms such as *John kicked the bucket* convey richer meanings than their literal counterparts (e.g., *John died*) can convey, and that therefore common figurative use of a phrase does not simply connote a specific literal meaning. Instead, they argue that metaphorical idioms continue to be used precisely because they convey more layers of meaning than a comparable literal phrase, whether that be through mental embodiment of a verb contained within the metaphor or through evoking a mental image (Colston & Gibbs, 2021) – thus, they assert that metaphors "precisely demonstrate what we mean," beyond what can be demonstrated by literal language. This theory is echoed by "embodiment" theory as presented by Al-Azary and Katz (2021), which suggests that some non-literal language is physically embodied in the mind of the listener or reader rather than being understood purely linguistically. The STUDY 1 findings appear to be compatible with this interpretation of lexicalized phrases.

However, the findings from STUDY 1 (Experiment 2) indicate that literal meanings were activated strongly both early (at vehicle word onset) and later (1000ms post-recognition point) in metaphor and simile comprehension – similes, but not metaphors, primed figurative targets at the late time point while both similes and metaphors primed literal targets both at vehicle word onset and 1000ms post-recognition. If figurative meaning is accessed directly, as direct-access theorists suggest, there should be no delay in reaching figurative meaning (compared to literal) for novel metaphors and literal meaning should not linger once figurative meanings have already decayed.

Our findings are therefore most compatible with exhaustive access theories such as the dual-access theory of metaphor comprehension proposed by Carston (2010), who proposes that fast ad-hoc pragmatic processes rapidly construct figurative meaning in the process of reading or listening while literal meaning remains consistently activated. Carston (2010) suggests that this process can account for extended literary and poetic metaphor, where multiple figurative constructs are layered in longer passages of speech or writing, allowing for both semantic understanding and imagistic or embodied properties triggered by literal meanings of constituent words.

Moreover, when the same experimental design was repeated with fluent L2 English speakers in STUDY 2, we found that L2 speakers did not display the same lexicalization effect for familiar metaphor, despite encountering familiar metaphor and simile often in their daily lives. Priming occurred only for literal targets in the simile condition in L2 speakers, suggesting that L2 speakers rely heavily on semantic compositionality of a phrase rather than familiarity in order to interpret both novel and familiar figurative sentences. Individual differences in language background appeared to modulate the moment-by-moment processing of figurative phrases, suggesting that there may be multiple pathways to attaining meaning used by L1 and L2 speakers. This finding also calls into question whether figurative meanings are typically directly accessed, as highly proficient L2 speakers seem to preferentially access literal meanings associated with similes.

In STUDY 3, we performed an offline task to gauge ultimate comprehension of simile and metaphor phrases in L1 and L2 speakers. In line with the predictions of Career of Metaphor Theory (Bowdle & Gentner, 2005), L1 speakers rated highly familiar metaphors more comprehensible than highly familiar similes, and preferred novel similes to novel metaphors.

However, our online task in STUDY 1 suggests that the reason for this is not, as Bowdle and Gentner (2005) suggest, that one is processed as a categorization statement and the other as a comparison statement, since literal and figurative meanings were co-activated at the first stages of processing. Instead, we suggest that in L1 speakers, decomposability of a sentence is subordinate to their experience of its common use in natural language, so (less decomposable) metaphorical forms are preferred for familiar phrases due to their familiarity. When a figurative construct is novel, L1 speakers rely more heavily on decomposability of the phrase to make sense of its meaning. Interestingly, increasing aptness did not seem to affect whether L1 speakers rated similes or metaphors more comprehensible.

L2 speakers, in contrast, rated high-familiarity similes more comprehensible than high-familiarity metaphors, while displaying no preference when sentences were unfamiliar. While they showed no preference for metaphor or simile in the high-apt condition, they rated similes more comprehensible than metaphors when the sentences were less apt. This suggests that L2 participants relied more heavily on compositionality of a sentence than L1 participants, especially when a figurative expression was not very apt, and did not appear to benefit from direct retrieval of commonly understood figurative meanings as L1 speakers did. L2 speakers who learned English later also rated similes more comprehensible than metaphors relative to those who learned English earlier, suggesting that the later an L2 is learned, the more important decomposability becomes to reach an interpretation of figurative language. STUDY 3 provides further evidence that compositionality plays a major role in L2 comprehension of figurative language, and suggests that overall, it is easier for L2 speakers to understand the figurative meaning of sentences which can be interpreted as literally true.

Implications for L2 Speakers

Our findings suggest that L2 speakers rely more on sentence decomposability to interpret the meaning of figurative language (also see Senaldi & Titone, 2022; Senaldi et al., 2022), perhaps because most L2 speakers mentally translate sentences word-by-word into their L1 (Taki, 2013). This has profound implications for the way they might interpret many common phrases when communicating in their L2, as metaphor is rife in natural language (Lakoff & Johnson, 1980). Indeed, while metaphors are commonly used in medical settings to convey important information about disease progression and treatment (Fuks, 2021), or in the news, educational or social settings to convey concepts not easily conveyed using literal language (Caesanto & Bottini, 2014; Maass et al., 2014; Meier et al., 2014), not all metaphors have direct cross-cultural or cross-linguistic equivalents.

To accurately convey such concepts to L2 speakers using figurative language, it may be beneficial to use simile over metaphor forms of common metaphor and idiom phrases when speaking to or writing for L2 speakers of a language. STUDY 3 of this thesis found that L2 speakers rated similes more comprehensible than metaphors, especially when they were highly familiar or low-apt. In instances where using appropriate metaphors to think about, e.g., a chronic illness may even affect disease outcomes (Munday et al., 2022), delivering the message in a form that is easily comprehensible may result in better health outcomes in L2 speakers and thus using highly decomposable literal phrases may make the difference in settings where a patient might otherwise be confused by the language used. In increasingly multicultural societies using simile over metaphor may be preferable in settings where many listeners may not process even highly lexicalized phrases as quickly or easily as literal language. Contrary to the assumptions of pragmatic theorists, it appears that lexicalized phrases are not always “dead” and metaphorical language still tokens multiple meanings even when it is commonly used in speech.

Future Directions

This thesis investigated the moment-by-moment processing of copular metaphors and similes, as well as their ultimate comprehension, in both L1 and L2 speakers of English. However, figurative language is not limited to copular metaphor and simile and encompasses everything from metonymy to irony to extended poetic metaphor. Future studies should investigate whether our findings about dual-access to literal and figurative meanings hold for other forms of figurative language, in both L1 and L2 speakers. Indeed, while there is a significant literature on idiom processing in bilinguals (e.g., Carrol & Conklin, 2014; Senaldi et al., 2022; Senaldi & Titone, 2022), less work has been done to investigate other forms of figurative language like simile, metaphor, and metonymy.

It would also be beneficial to perform more online experiments of metaphor processing embedded in longer biasing contexts, to ascertain whether figurative or literal meanings can be accessed more easily depending on surrounding linguistic context and environmental referents. Although some experiments have shown contextual effects on metaphor interpretation (e.g. Inhoff & Lima, 1984), such studies have usually involved offline processing. Increasingly time-sensitive methods are available to investigate the moment-by-moment processing of figurative language (e.g., ERP, eye-tracking, and multimodal priming methods), and such time-sensitive experimental designs can be used to resolve the debate between competing models of metaphor processing. In L2 speakers, studies using biasing context could be performed to determine whether L2 speakers can be triggered to understand cultural referents which may not exist in their L1 or culture of origin given sufficient linguistic context.

Finally, as suggested by Katz (2024), multimodal experimental paradigms could be used to probe the underlying cognitive concepts tokened by metaphor, as opposed to studies limiting

metaphor use to the realm of spoken or written language. Although STUDIES 1-2 of this thesis used a multimodal (auditory + visual) design to probe the interaction between spoken and written language, future studies should probe whether simile and metaphor are understood equally when metaphorical vehicles or topics are expressed pictorially, or through physical embodiment, as well as through spoken or written language. Multimodal studies probing figurative thought as well as language could help shed further light on how figurative language is represented conceptually, without relying on words in any given language to express the concepts that are cognitively represented by linguistic constructs. This could also help ascertain to what extent figurative concepts are understood linguistically by L1 and L2 speakers, and to what extent they cross language and cultural barriers.

Conclusion

Ultimately, this thesis concludes that literal meaning and syntactic composability has primacy over figurative meaning in most contexts, especially when a sentence is less familiar to comprehenders (whether that be due to how frequently it occurs in speech, or due to the language background of the comprehender). In L1 speakers, however, familiarity of a metaphorical sentence can lead to more rapid attainment of its figurative meaning during sentence processing, and to the impression that it is ultimately more comprehensible in its metaphorical form.

Rapid processing does not suggest, however, that literal meanings are discarded (as pragmatic theories suggest) or are bypassed entirely (as many direct-access theories suggest); rather, there is simultaneous dual access to figurative meanings which can be attained rapidly during a “sweet spot” of sentence processing by L1 speakers, and literal meanings which are activated earlier and continue to linger afterwards. The ability to easily attain these figurative meanings during processing, as well as preference for metaphorical forms, is attenuated in L2

learners, especially in those who have learned English later in life. What L1 and L2 speakers have in common is that the more novel a figurative sentence is, the more they rely on literal semantic decomposability of the sentence to ultimately comprehend its intended meaning.

Bibliography

Al-Azary, H., Gagné, C. L., & Spalding, T. L. (2021). Flute birds and creamy skies: The metaphor interference effect in modifier–noun phrases. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 75*(2), 175–181. https://doi.org/10.1037/cep0000251

Al-Azary, H., & Katz, A. N. (2021). Do metaphorical sharks bite? Simulation and abstraction in metaphor processing. *Memory & Cognition, 49,* 557-570. https://doi.org/10.3758/s13421-020-01109-2

Aristotle. (trans. 1926). *The art of rhetoric*. New York, NY: G. P. Putnam's Sons.

Ashby, J., Roncero, C., de Almeida, R. G., & Agauas, S. J. (2018). The early processing of metaphors and similes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology, 71*(1), 161-168. https://doi.org/10.1080/17470218.2016.1278456

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling metaphor from context: An ERP study. *Front. Psychol., 7*, 559. doi: 10.3389/fpsyg.2016.00559

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

Blasko, D. G., & Briihl, D. S. (1997). Reading and recall of metaphorical sentences: Effects of familiarity and context. *Metaphor and Symbol, 12*(4), 261-285. https://doi.org/10.1207/s15327868ms1204_4

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology, 19*(2), 295-308. https://doi.org/10.1037/0278-7393.19.2.295

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological review, 112*(1), 193-216. doi:10.1037/0033-295X.112.1.193

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977-990. https://doi.org/10.3758/BRM.41.4.977

Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition, 17*(4), 784–797. https://doi.org/10.1017/S1366728913000795

Carston, R. (2010). XIII—Metaphor: Ad Hoc Concepts, Literal Meaning and Mental Images. *Proceedings of the Aristotelian Society, 110*(3pt3), 295-321. doi:10.1111/j.1467-9264.2010.00288.x

Casasanto, D. (2014). Experiential origins of mental metaphors: Language, culture, and the body. In M. Landau, M. D. Robinson, & B. P. Meier (Eds.), *The power of metaphor: Examining its influence on social life* (pp. 249-268). American Psychological Association, viii, 301 pp.

Casasanto, D., & Bottini, R. (2013). Spatial language and abstract concepts. *WIREs Cognitive Science, 5*(2), 139-149. https://doi.org/10.1002/wcs.1271

Chiappe, D., & Kennedy, J. M. (1999). Aptness predicts preference for metaphors or similes, as well as recall bias. *Psychonomic Bulletin & Review, 6*(4), 668-676.

Chiappe, D., Kennedy, J. M., & Smykowski, T. (2009). Reversibility, aptness, and the conventionality of metaphors and similes. *Metaphor and Symbol, 18*(2), 85-105. doi:10.1207/S15327868MS1802_2

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments & Computers, 25*(2), 257–271. https://doi.org/10.3758/BF03204507

Colston, H. L., & Gibbs, R. W. (2021). Figurative language communicates directly because it precisely demonstrates what we mean. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 75*(2), 228–233. https://doi.org/10.1037/cep0000254

Columbus G., Sheikh, N. A., Côté-Lecaldaire, M., Häuser, K., Baum, S. R., & Titone, D. (2015). Individual differences in executive control relate to metaphor processing: An eye movement study of sentence reading. *Frontiers in Human Neuroscience, 8*, 1057. https://doi.org/10.3389/fnhum.2014.01057

Coulehan, J. (2003). Metaphor and medicine: Narrative in clinical practice. *Yale Journal of Biology and Medicine, 76*(2), 87-95.

Damerall, A. W., & Kellogg, R. T. (2016). Familiarity and aptness in metaphor comprehension. *The American Journal of Psychology, 129*(1), 49-64. https://doi.org/10.5406/amerjpsyc.129.1.0049

Davidson, D. (1978). What metaphors mean. *Critical Inquiry, 5*(1), 31-47.

Davies, M. (2008-) *Word frequency data*. [Data set]. The Corpus of Contemporary American English (COCA). https://www.wordfrequency.info.

de Almeida, R. G. (2001). Conceptual deficits without features: A view from atomism. *Behavioral and Brain Sciences, 24*(3), 482-483. doi:10.1017/S0140525X01284155

de Almeida, R. G., & Antal, C. (2021). How can semantics avoid the troubles with the analytic/synthetic distinction? In S. Löbner, T. Gamerschag, T. Kalenscher, M. Schrenk, & H. Zeevat (Eds.), *Concepts, frames and cascades in semantics, cognition and ontology* (pp. 103-120). Springer Open Access.

de Almeida, R. G., & Lepore, E. (2018). Semantics for a Module. In de Almeida, R. G. & Gleitman, L. R. (Eds.), *On Concepts, modules, and language*. New York: Oxford

Declercq, J., & van Poppel, L. (2024). Metaphors in interaction: Reusing, developing and resisting metaphors of illness, the body and medical treatment in chronic pain consultations. *Journal of Pragmatics, 221*, 168-182. https://doi.org/10.1016/j.pragma.2024.01.001.

Durgin, F. H., & Gelpi, R. A. (2017). When do vehicles of similes become figurative? Gaze patterns show that similes and metaphors are initially processed differently. *CogSci 2017: Proceedings of the 39th Annual Meeting of the Cognitive Science Society.* 1967-1972.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*(1-2), 3–71. https://doi.org/10.1016/0010-0277(88)90031-5

Forster, K. I. (1999). The microgenesis of priming effects in lexical access. *Brain and Language, 68*(1-2), 5-15. https://doi.org/10.1006/brln.1999.2078

Friedmann, N., Taranto, G., Shapiro, L. P., & Swinney, D. (2008). The leaf fell (the leaf): The online processing of unaccusatives. *Linguist Inq. 39*(3), 355-377. doi: 10.1162/ling.2008.39.3.355

Fuks, A. (2021). *The Language of Medicine.* Oxford: Oxford University Press.

Gibbs, R. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. New York: Cambridge University Press.

Gibbs, R. W., Jr., & Colston, H. L. (2012). *Interpreting Figurative Meaning*. New York: Cambridge University Press.

Giora, R. (2003). *On Our Mind: Salience, Context, and Figurative Language*. Oxford: Oxford University Press.

Glucksberg, S. & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review, 97*(1), 3-18. https://doi.org/10.1037/0033-295X.97.1.3

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends Cogn Sci, 7*(2), 92-96. doi:10.1016/S1364-6613(02)00040-2

Glucksberg, S., & Estes, Z. (2000). Feature accessibility in conceptual combination: Effects of context-induced relevance. *Psychonomic Bulletin & Review, 7*(3), 510-515. https://doi.org/10.3758/BF03214364

Grice, H. P. (1975). Logic and Conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and Semantics* (pp. 41-58). New York: Academic Press.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*(4), 267–283. https://doi.org/10.3758/BF03204386

Inhoff, A. W., Lima, S. D., & Carroll, P. J. (1984). Contextual effects on metaphor comprehension. *Memory & Cognition, 12*(6), 558-567. https://doi.org/10.3758/BF03213344

Jankowiak, K., Naranowicz, M., & Rataj, K. (2021). Metaphors are like lenses: Electrophysiological correlates of novel meaning processing in bilingualism. *International Journal of Bilingualism, 25*(3), 668-686. https://doi.org/10.1177/1367006921996820

Jankowiak, K., Rataj, K., & Naskręcki, R. (2017). To electrify bilingualism: Electrophysiological insights into bilingual metaphor comprehension. *PLoS ONE 12*(4): e017578. https://doi.org/10.1371/journal.pone.0175578

Janus, R., & Bever, T. (1985). Processing of metaphoric language: An investigation of the three-stage model of metaphor comprehension. *Journal of Psycholinguistic Research, 14*(5), 473-487.

Jones, L. L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language, 55*(1), 18-32. https://doi.org/10.1016/j.jml.2006.02.004

Katz, A. N. (2024). Selective view of the last 50 years of the experimental psycholinguistics of idiom, metaphor, and irony: A commentary. *Discourse Processes, 61*(1-2), 90-102. https://doi.org/10.1080/0163853X.2024.2319898

Kintsch, W. (2001). Predication. *Cognitive Science, 25*(2), 173-202. https://doi.org/10.1207/s15516709cog2502_1

Kumar, A., Steyvers, M., & Balota, D. A. (2021). A critical review of network-based and distributional approaches to semantic memory structure and processes. *Topics in Cognitive Science, 14*(12), 54-77. doi: 10.1111/tops.12548

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

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

Landau, M. J., Robinson, M. D., & Meier, B. P. (2014). *The power of metaphor: Examining its influence on social life*. American Psychological Association.

Lee, S. W. S., & Schwarz, N. (2012). Bidirectionality, mediation, and moderation of metaphorical effects: The embodiment of social suspicion and fishy smells. *Journal of Personality and Social Psychology, 103*(5), 737–749. https://doi.org/10.1037/a0029708

Libben, M., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition, 36*(6), 1103-1121. https://doi.org/10.3758/MC.36.6.1103

Littlemore, J. (2006). Metaphoric competence, second language learning, and communicative language ability. *Applied Linguistics, 27*(2), 268-294. doi:10.1093/applin/aml004

Littlemore, J., Chen, P. T., Koester, A., & Barnden, J. (2011). Difficulties in metaphor comprehension faced by international students whose first language is not English. *Applied Linguistics, 32*(4), 408-429. https://doi.org/10.1093/applin/amr009

Maass, A., Suitner, C., & Acuri, L. (2014). The role of metaphors in intergroup relations. In M. Landau, M. D. Robinson, & B. P. Meier (Eds.), *The power of metaphor: Examining its influence on social life* (pp. 153-177). American Psychological Association, viii, 301 pp. https://doi.org/10.1037/14278-008

Meier, B. P, Scholer, A. A., & Fincher-Kiefer, R. (2014). Conceptual metaphor theory and person perception. In M. Landau, M. D. Robinson, & B. P. Meier (Eds.), *The power of metaphor: Examining its influence on social life* (pp. 43-64). American Psychological Association. https://doi.org/10.1037/14278-003

Milburn, E., Vulchanova, M., & Vulchanov, V. (2021). Collocational frequency and context effects on idiom processing in advanced L2 speakers. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 75*(2), 169–174. https://doi.org/10.1037/cep0000214f3kRB

Munday, I., Kneebone, I., Rogers, K., & Newton-John, T. (2022). The language of pain: Is there a relationship between metaphor use and adjustment to chronic pain? *Pain Medicine, 23*(12), 2073-2084. https://doi.org/10.1093/pm/pnaa467

Olkoniemi, H., Bertram, R., & Kaakinen, J. K. (2022). Knowledge is a river and education is like a stairway: An eye movement study on how L2 speakers process metaphors and similes. *Bilingualism: Language and Cognition, 25*(2), 307-320. DOI: https://doi.org/10.1017/S1366728921000869

Ottatti, V., Renstrom, R., & Price, E. (2014). The metaphorical framing model: Political communication and public opinion. In M. Landau, M. D. Robinson, & B. P. Meier (Eds.), *The power of metaphor: Examining its influence on social life* (pp. 179-202). American Psychological Association, viii, 301 pp.

Pambuccian, F. S., & Raney, G. E. (2021). A simile is (like) a metaphor: Comparing metaphor and simile processing across the familiarity spectrum. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 75*(2), 182–188. https://doi.org/10.1037/cep0000242

Perlovsky, L. (2009). Language and cognition. *Neural Networks, 22*(3), 247-257. https://doi.org/10.1016/j.neunet.2009.03.007

Roncero, C., & de Almeida, R. G. (2014). The importance of being apt: Metaphor comprehension in Alzheimer's disease. *Frontiers in Human Neuroscience, 8,* Article 973.

Roncero, C., & de Almeida, R. G. (2015). Semantic properties, aptness, familiarity, conventionality, and interpretive diversity scores for 84 metaphors and similes. *Behavior Research Methods, 47*(3), 800-812. doi:10.3758/s13428-014-0502-y

Roncero, C., de Almeida, R. G., Martin, D. C., & de Caro, M. (2016). Aptness predicts metaphor preference in the lab and on the internet. *Metaphor and Symbol, 31*(1), 31-46. https://doi.org/10.1080/10926488.2016.1116908

Roncero, C., de Almeida, R. G., Pissani, L. & Patalas, I. (2021). A metaphor is not like a simile: Reading-time evidence for distinct interpretations for negated tropes. *Metaphor & Symbol, 36*(2), 85-98. doi:10.1080/10926488.2021.1882258

Schnall, S. (2014). Are there basic metaphors? In M. Landau, M. D. Robinson, & B. P. Meier (Eds.), *The power of metaphor: Examining its influence on social life* (pp. 225-247). American Psychological Association, viii, 301 pp.

Searle, J. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 83-111). Cambridge University Press.

Senaldi, M. S. G., Titone, D.A. (2022) Less direct, more analytical: Eye-movement measures of L2 idiom reading. *Languages, 7*(2), 91. https://doi.org/10.3390/languages7020091

Senaldi, M. S. G., Wei, J., Gullifer, J.W., & Titone, D. (2022). Scratching your *tête* over language-switched idioms: Evidence from eye-movement measures of reading. *Memory & Cognition, 50,* 1230-1256. https://doi.org/10.3758/s13421-022-01334-x

Sherman, G. D., & Clore, G. L. (2014). Do evaluative metaphors shape emotional thought? A call for new evidence. In M. Landau, M. D. Robinson, & B. P. Meier (Eds.), *The power of metaphor: Examining its influence on social life* (pp. 205-223). American Psychological Association.

Sperber, D., & Wilson, D. (2008). A deflationary account of metaphors. In Gibbs, R. W. (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 84-105). Cambridge University Press.

Steen, G. (2008). The paradox of metaphor: Why we need a three-dimensional model of metaphor. *Metaphor and Symbol, 23*(4), 213-241.

Taki, S. (2011). Cross-cultural communication and metaphorical competence. *International Journal of Language Studies, 5*(1), 47-62.

Tang, X., Shen, L., Yang, P., Huang, Y., Huang, S., Huang, M., & Ren, W. (2022). Bilingual processing mechanisms of scientific metaphors and conventional metaphors: Evidence via a Contrastive Event-Related Potentials study. *Sec. Psychology of Language, 13*. doi=10.3389/fpsyg.2022.894114

Thibodeau, P. H., & Durgin, F. H. (2011). Metaphor aptness and conventionality: A processing fluency account. *Metaphor and Symbol, 26*(3), 206-226. https://doi.org/10.1080/10926488.2011.583196

Tianshu, Z., & Minda, J. P. (2021). An investigation of idiom processing advantage using translated familiar idioms. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 75*(2), 162–168. https://doi.org/10.1037/cep0000245

Tirrell, L. (1989). Extending: The structure of metaphor. *Noûs, 23*(1), 17-34. doi:10.2307/2215827

Titone, D. A., & 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. doi:10.1075/ml.9.3.05tit

Tiv, M., Deodato, F., Rouillard, V., Wiebe, S., & Titone, D. (2021). Second language experience impacts first language irony comprehension among bilingual adults. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 75*(2), 126–132. https://doi.org/10.1037/cep0000230

Utsumi, A., & Sakamoto, M. (2011). Indirect categorization as a process of predicative metaphor comprehension. *Metaphor and Symbol, 26*(4), 299-313. https://doi.org/10.1080/10926488.2011.609120

Wackers, D. Y. M, Plug, H. J., & Steen, G. J. (2021). "For crying out loud, don't call me a warrior": Standpoints of resistance against violence metaphors for cancer. *Journal of Pragmatics, 174*, 68-77. https://doi.org/10.1016/j.pragma.2020.12.021

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics, 41*(3), 262-275. https://doi.org/10.3758/BF03208224

Werkmann Horvat, A., Bolognesi, M., & Kohl, K. (2021). The status of conventional metaphorical meaning in the L2 lexicon. *Intercultural Pragmatics, 18*(4), 447-467. https://doi.org/10.1515/ip-2021-4002

Wolff, P., & Gentner, D. (2000). Evidence for role-neutral initial processing of metaphors. *J Exp Psychol Learn Mem Cogn, 26*(2), 529-541. https://doi.org/10.1037/0278-7393.26.2.529

Zhu, T., & Minda, J. P. (2021). An investigation of idiom processing advantage using translated familiar idioms. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 75*(2), 162–168. https://doi.org/10.1037/cep0000245

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32*(1), 25-64. doi:10.1016/0010-0277(89)90013-9

Appendix A

Experimental materials employed in STUDY 1 (Chapter 2) and STUDY 2 (Chapter 3)

Aurally presented metaphors/similes (from Roncero & de Almeida, 2015 norms) are in boldface while surrounding context is presented in non-boldface text. Aurally presented prime sentences were accompanied by one of four capitalized visual target words (literally related, figuratively related, or matched unrelated controls for each literal/figurative word). Participant ratings for aptness and familiarity (from Roncero & de Almeida, 2015 norms) are presented in the table. In our analyses the aptness/familiarity ratings were split by median value and input as categorical values as high (H) and low (L).

Sentence	Target (fig)	Control (fig)	Target (lit)	Control (lit)	Aptness		Familiarity	
We can all agree that **education is (like) a stairway**, because in most cases, education helps people climb higher in life.	UPWARD	FIRMLY	EXERCISE	ARGUMENT	8.96	H	5.90	H
Most people believe that **families are (like) fortresses**, because generally speaking, families can shield us from difficulties.	STRONG	GROUND	BIG	FEW	7.78	H	3.25	L
It is common knowledge that **fingerprints are (like) portraits**, because everyone knows that fingerprints can identify individuals.	UNIQUE	ACTIVE	BEAUTIFUL	EFFECTIVE	6.13	L	5.05	H
It is well-known that **genes are (like) blueprints**, because it's been proven that genes are the building blocks of life.	MAPS	HATS	CONSTRUC-TION	HEADQUAR-TERS	8.39	H	7.85	H
It is a commonly held belief that **heaven is (like) a treasure**, because most people think heaven is full of beautiful things.	PLEASANT	ABSTRACT	EXPENSIVE	SENSITIVE	6.61	L	4.40	L
No one would deny that **insults are (like) daggers**, because most of the time, insults are used to hurt people.	HURTFUL	HAUGHTY	BLADE	MOUSE	7.91	H	6.40	H
It is hardly a secret that **lawyers are (like) sharks**, because with few exceptions, lawyers are bloodthirsty and ruthless.	MEAN	NEXT	BLOOD	CAUSE	8.61	H	7.70	H
Everyone agrees that **lawyers are (like) snakes**, because generally speaking, lawyers are slippery and sneaky.	SNEAKY	STUFFY	DEATH	SENSE	7.39	H	4.00	L

It is a general truth that **life is (like) a river**, because as we all know, life carries you along on its currents.	LONG	HIGH	BLUE	FINE	7.22	L	4.85	H
We all know that **love is (like) a drug**, because most of the time, love is extremely addictive.	HIGH	LEFT	DEATH	VOICE	8.26	H	8.60	H
It is common knowledge that **memory is (like) a sponge**, because in most cases, memory is more absorbent than you would expect.	LONG	HIGH	BLUE	FINE	6.70	L	7.50	H
The fact is, **minds are (like) computers**, because it's been proven that minds can process a lot of information.	INTELLI-GENT	RESIDEN-TIAL	EXPENSIVE	IMMEDIATE	7.48	H	7.65	H
We all know that **obligations are (like) shackles**, because generally speaking, obligations hold people back.	ANNOYING	ADJACENT	HEAVY	SHORT	6.09	L	2.80	L
It is a commonly held belief that **teachers are (like) sculptors**, because as everyone knows, teachers help shape children's minds.	BUILDERS	AIRPORTS	CLAY	FORT	7.91	H	5.35	H
We all know that **time is (like) money**, because generally speaking, time can either be spent well or wasted.	IMPORTANT	DIFFERENT	GREEN	CHIEF	8.43	H	9.95	H
Everyone agrees that **time is (like) a thief**, because as we all know, time can steal away our precious moments.	QUICK	SWEET	DANGEROUS	TECHNICAL	6.91	L	3.30	L
No one would deny that **trust is (like) glue**, because everyone knows that trust holds people together.	BINDS	ACHES	LIQUID	MOTION	6.30	L	2.80	L
It is obvious to everyone that **typewriters are (like) dinosaurs**, because as we all know, typewriters are objects from a bygone era.	ANTIQUES	CLUSTERS	DANGEROUS	COLLECTED	6.30	L	3.20	L
Most people believe that **wisdom is (like) an ocean**, because as everyone knows, wisdom is vast and mysterious.	VAST	PURE	BLUE	LATE	6.91	L	5.40	H
It is hardly a secret that clouds are (like) cotton, because generally speaking, clouds look soft and light.	WHITE	YOUNG	COMFY	RUDDY	5.22	L	5.95	H
It is well-known that **friendship is (like) a rainbow**, because in most cases, friendship brings us joy in dark moments.	RARE	SLOW	CURVY	FLAKY	5.74	L	3.65	L
It is hardly a secret that **jobs are (like) jails**, because generally speaking, jobs confine people for much of their lives.	BORING	CASUAL	DANGEROUS	COLLECTED	5.00	L	3.15	L

It is obvious to everyone that **salesmen are (like) bulldozers**, because as we all know, salesmen don't let anything stand in their path.	ANNOYING	ADJACENT	BIG	FEW	5.35	L	2.45	L
We all know that **anger is (like) fire**, because it has been proven that anger causes people to act destructively.	SCARY	EAGER	FLAMES	GRAINS	8.48	H	4.75	L
We can all agree that **beauty is (like) a passport**, because as everyone knows, beauty gives people freedom in life.	AD-VANTAGE	APART-MENT	BOOKLET	BRACKET	7.13	L	4.47	L
No one would deny that **cities are (like) jungles**, because as we all know, cities are big and hard to navigate.	CRAZY	MAGIC	ANIMALS	FACULTY	8.61	H	6.60	H
It is common knowledge that **exams are (like) hurdles**, because most people think exams are obstacles that students need to cross.	STRESSFUL	WORTH-LESS	HEIGHT	GLANCE	7.87	H	6.10	H
Most people believe that **eyelids are (like) curtains**, because most of the time, eyelids are useful for blocking out light.	OPEN	REAL	COLORFUL	ADEQUATE	7.35	H	3.30	L
It is a general truth that **faith is (like) a raft**, because generally speaking, faith keeps us afloat in times of uncertainty.	UNSTEADY	SADISTIC	FLAT	WARM	7.22	L	2.30	L
It is well-known that **knowledge is (like) light**, because as everyone knows, knowledge is illuminating.	ILLUMI-NATES	APPRECI-ATES	ELECTRICITY	ANNIVER-SARY	8.09	H	4.45	L
It is a commonly held belief that **music is (like) medicine**, because in most cases, music can help people heal.	HELPFUL	INSTANT	BITTER	ASLEEP	7.39	H	6.05	H
The fact is, **love is (like) a child**, because generally speaking, love is pure and innocent.	INNOCENT	MODER-ATE	ANNOYING	ADJACENT	5.30	L	3.80	L

Appendix B

Experimental materials employed in STUDY 3 (Chapter 4)

Metaphor, simile, anomalous, and literal sentences for STUDY 3 are presented below. Metaphor and simile sentences are taken from Roncero and de Almeida (2015), a set of norms based on online corpus analysis.

Metaphor	**Simile**
Alcohol is a crutch	Alcohol is like a crutch
Anger is fire	Anger is like fire
Anger is a heart	Anger is like a heart
Beauty is a passport	Beauty is like a passport
The Bible is a sword	The Bible is like a sword
Billboards are warts	Billboards are like warts
Christ is a door	Christ is like a door
Christians are salt	Christians are like salt
Cigarettes are time bombs	Cigarettes are like time bombs
Cities are jungles	Cities are like jungles
Clouds are cotton	Clouds are like cotton
Debt is a disease	Debt is like a disease
Deserts are ovens	Deserts are like ovens
Desks are junkyards	Desks are like junkyards
Dreams are water	Dreams are like water
Education is a stairway	Education is like a stairway
Exams are hurdles	Exams are like hurdles
Eyelids are curtains	Eyelids are like curtains
Faith is a raft	Faith is like a raft
Families are fortresses	Families are like fortresses
Fingerprints are portraits	Fingerprints are like portraits
Friendship is a rainbow	Friendship is like a rainbow
Genes are blueprints	Genes are like blueprints
Giraffes are skyscrapers	Giraffes are like skyscrapers
God is fire	God is like fire
God is a parent	God is like a parent
Greed is a buzzard	Greed is like a buzzard
Health is glass	Health is like glass

Hearts are closets
Heaven is a treasure
Highways are snakes
Insults are daggers
Jobs are jails
Knowledge is light
Knowledge is money
Knowledge is power
Knowledge is a river
Lawyers are sharks
Lawyers are snakes
Life is a beach
Life is a bottle
Life is a dream
Life is a joke
Life is a journey
Life is a river
Love is a child
Love is a drug
Love is a flower
Love is gold
Love is a melody
Love is a rainbow
Love is a rose
Memory is a river
Memory is a sponge
Men are fish
Minds are computers
Money is oxygen
Music is medicine
Obligations are shackles
Peace is a river
Pets are kids
Rage is a volcano
Runners are torpedoes
Salesmen are bulldozers
Schools are zoos
Science is politics
Sermons are sleeping pills
Skating is flying
Smog is a shroud

Hearts are like closets
Heaven is like a treasure
Highways are like snakes
Insults are like daggers
Jobs are like jails
Knowledge is like light
Knowledge is like money
Knowledge is like power
Knowledge is like a river
Lawyers are like sharks
Lawyers are like snakes
Life is like a beach
Life is like a bottle
Life is like a dream
Life is like a joke
Life is like a journey
Life is like a river
Love is like a child
Love is like a drug
Love is like a flower
Love is like gold
Love is like a melody
Love is like a rainbow
Love is like a rose
Memory is like a river
Memory is like a sponge
Men are like fish
Minds are like computers
Money is like oxygen
Music is like medicine
Obligations are like shackles
Peace is like a river
Pets are like kids
Rage is like a volcano
Runners are like torpedoes
Salesmen are like bulldozers
Schools are like zoos
Science is like politics
Sermons are like sleeping pills
Skating is like flying
Smog is like a shroud

Soldiers are pawns
Stores are zoos
Teachers are sculptors
Television is candy
Time is money
Time is a snail
Time is a thief
Tongues are fire
Tree trunks are straws
Trees are umbrellas
Trust is glue
Typewriters are dinosaurs
Winter is death
Wisdom is an ocean
Women are cats

Soldiers are like pawns
Stores are like zoos
Teachers are like sculptors
Television is like candy
Time is like money
Time is like a snail
Time is like a thief
Tongues are like fire
Tree trunks are like straws
Trees are like umbrellas
Trust is like glue
Typewriters are like dinosaurs
Winter is like death
Wisdom is like an ocean
Women are like cats

Filler (literal)

Houses are buildings
Chairs are furniture
Love is an emotion
Children are people
Owls are birds
Cycling is a sport
The heart is an organ
Printers are machines
Silver is an element
Pianos are instruments
Silk is a fabric
Accountants are workers
Painters are artists
Parenthood is a responsibility
Flowers are living things
Wood is a material
Coats are clothing
Doctors are professionals
Students are children
Murderers are criminals

Filler (anomalous)

Professors are coat hangers
Violins are heaters
Paper is a vehicle
Summer is a mammal
Whales are paperclips
Metal is food
Swords are dogs
Necklaces are shoes
Nurses are lampshades
Carpets are icicles
Lollipops are buckets
The wind is an onion
Love is a crouton
Violets are door frames
Bags are wild animals
The sky is a sandwich
Raccoons are ropes
Sprinters are toilet paper
Oranges are shipyards
Brotherhood is a shoe

Appendix C

Experimental procedure employed in STUDY 3 (Chapter 4)

The full text of the participant questionnaire in STUDY 3, including the full set of demographic questions and experimental questions, is presented below.

Welcome to our survey!
In the first part, you will be asked a few questions concerning yourself and your language background.
In the second part, you will be asked to make some decisions about a series of short sentences. You will see the same list of sentences twice, and each time you will be asked to make different decisions.
Please read the instructions carefully, so you are clear on what specific decisions you need to make.

PART 1: Demographic Questions

Q1. What is your age? Please indicate in years

Q2. Please indicate the first 3 characters of your postal code.

Q3. What is your gender identity?

Q4. Where were you born?

Q5. You indicated you were born in Canada. In which province or territory were you born?

Q6. In what year did you move to Canada?

Q7. What is your current status in Canada?

Q8. Where were your parents born?

Q9. What is your ethnic origin? Select all options that apply.
*Ethnic origin refers to a person's 'roots' and should not be confused with citizenship, nationality, language, or place of birth. For example, a person who has Canadian citizenship, speaks Punjabi (Panjabi) and was born in the United States may report Guyanese ethnic origin (Statistics Canada, 2016).

Q10. Please indicate your education level, or its equivalent (highest degree/qualification earned).

Q11. What is your current employment status?

Q12. Your mentioned that you are currently working. What is your occupation? For example: accountant, sales clerk, programmer, etc.

Q13. Do you have a religion or spiritual tradition?

Q14. What is your religion or spiritual tradition?

Q15. Do you regularly attend a place of worship?

Q16. Which of the following best indicates your annual household income before taxes?

Q17. Think of this ladder as representing where people stand in Canada. At the top of the ladder are the people who have the most money, most education, and most respected jobs. At the bottom are the people who have the least money, least education, and least respected jobs or no job. The higher up you are on this ladder, the closer you are to the people at the very top, and the lower you are, the closer you are to the people at the very bottom. Please select where you think you stand at this time in your life, relative to other people in Canada.

Q18. Think about the contexts and groups where you regularly spend time in your daily life. It could be a sports team, a social club, a leisure association, a professional organization, a group that offers activities to children (e.g., extracurricular activities), a community organization, a volunteer group, a yoga class, etc. Please list these contexts below. If more than two such contexts are present in your daily life, list those where you spend most time.

Q19. Are you in a committed romantic relationship? For example: married, common-law, cohabiting, serious relationship, etc.

Q20. What language(s) did your romantic partner learn at birth (first year of life)?

Q21. What language(s) did you learn at birth (first year of life)?

Q22. What language(s) other than English or French did you learn at birth (first year of life)?

Q23. Other than the language(s) you learned at birth and other than English or French, what language(s) do you know, or have known in the past? Do not enter English or French. If there are more than 2 languages, please enter the ones that are the most significant in your life today. Please note that you do not need to be proficient in these languages.

Q24. How old were you when you first started to learn... Please round to the nearest year with no comma or decimals.

Q25. What is your current speaking ability in all these languages?

Q26. In which single language do you feel most comfortable in terms of speaking?

Q27. What percentage of an average day do you speak each of your languages overall? All percentages must add up to 100%

Q28. What is your current reading/writing ability in all these languages?

Q29. In which single language do you feel most comfortable in terms of reading/writing?

Q30. What percentage of an average day do you read/write each of your languages overall? All percentages must add up to 100%

Q31. When you are at home, what percentage of the time do you use (speaking, hearing, writing, reading) each of your languages? Please note that by "at home" we mean with the people that live with you. All percentages must add up to 100%

Q32. When you interact with your romantic partner, what percentage of the time do you use (speaking, hearing, writing, reading) each of your languages? All percentages must add up to 100%

Q33. When you are with friends, what percentage of the time do you use (speaking, hearing, writing, reading) each of your languages? All percentages must add up to 100%

Q34. When you are with neighbors or acquaintances, what percentage of the time do you use (speaking, hearing, writing, reading) each of your languages? All percentages must add up to 100%

Q35. When you are at work, what percentage of the time do you use (speaking, hearing, writing, reading) each of your languages? All percentages must add up to 100%

Q36. When you are at school, what percentage of the time do you use (speaking, hearing, writing, reading) each of your languages? All percentages must add up to 100%

Q37. When you are in public settings (e.g., library, bank, stores), what percentage of the time do you use (speaking, hearing, writing, reading) each of your languages?

Q38. You mentioned regularly attending a place of worship. When you are in that environment, what percentage of the time do you use (speaking, hearing, writing, reading) each of your languages? All percentages must add up to 100%

Q39. You mentioned "" as a context where you regularly spend time. When you are in that environment, what percentage of the time do you use (speaking, hearing, writing, reading) each of your languages? All percentages must add up to 100%

Q40. Indicate the language(s) used by your teachers for majority instruction (i.e., not in foreign language classes) at each educational level. If one level of education does not apply to you, you may leave it blank.

PART 2: Language Questionnaire

In this block, you will make two separate decisions about each short sentence.

The FIRST DECISION is whether each short sentence is COMPREHENSIBLE in a global way, along a 7-point scale where 1 = “not at all comprehensible” and 7 = “maximally comprehensible”. Intermediate values of the scale, which we encourage you to use, would indicate in-between global comprehensibility. For example, “Jane is a doctor” is straightforwardly comprehensible. Thus, one might give that sentence a rating of 6 or 7.

In contrast, “Jane is an airplane”, is a less straightforwardly comprehensible sentence. Thus, one might give that sentence a rating of 1 or 2. Other sentences might be in-between, and if so, we encourage you to select the appropriate intermediate value of the scale.

The SECOND DECISION is whether each short sentence is FAMILIAR to you, along a 7-point scale, where 1 = “not at all familiar” and 7 = “maximally familiar”. Again, we encourage you to make this decision for each short sentence, and to make use of intermediate values of the scale.

Please rate how comprehensible and familiar the following sentences are.

Now you will make 2 additional decisions about the same set of sentences.

The FIRST DECISION is whether each short sentence is INTERPRETATIVELY APT in a global way, along a 7-point scale where 1 = “not at all apt” and 7 = “maximally apt”. Intermediate values of the scale, which we encourage you to use, would indicate in-between aptness. For example, “Politics is a jungle” is a highly apt sentence because the meaning of the second content word (jungle) captures salient properties of the first content word (politics). Thus, one might give that sentence an aptness rating of 6 or 7. In contrast, “Politics is a desk”, is less straightforwardly apt because the meaning of the second content word (desk) has no obvious bearing on any salient property of the first content word (politics). Thus, one might give that sentence a rating of 1 or 2. Other sentences might be in-between, and if so, we encourage you to select the appropriate intermediate version of the scale.

The SECOND DECISION is whether each short sentence is EMOTIONALLY CHARGED along a 7-point scale where 1 = “negatively emotionally charged”, 7 = “positively emotionally charged”, and 4 = “emotionally neutral”. For example, “Puppies are joyful” would be maximally emotionally charged in a positive way. Thus, one might give that sentence a rating of 6 or 7. In contrast, “War is death” would be maximally emotionally charged in a negative way. Thus, one might give that sentence an emotional rating of 1 or 2. Finally, “Tables have legs” would be neither positive nor negative in an emotional charged way. Thus, one might give that sentence the intermediate value of 4.

Please rate how apt and emotionally charged the following sentences are.

Thank you for your participation in this study. The goal of this experiment was to understand how people process metaphorical expressions. A metaphor, like that lawyer is a shark, is a figurative expression involving a more or less transparent mapping between a topic (lawyer) and a vehicle (shark) to denote a (typically abstract) feature of the topic (e.g. a lawyer's ruthlessness).

Metaphors can be more or less easy to understand, result more or less familiar to speakers, and convey different emotional values. The purpose of this survey was to understand how metaphors differ from each other based on these characteristics, and how the language background of each speaker impacts these results (e.g. how ratings change when metaphors are processed in someone's first vs second language). If you would like to read more about how metaphors are processed, we encourage you to read the following papers. If you have any questions or concerns about the study, you are welcome to contact the researchers
at the email address provided in the consent form.

Columbus, G., Sheikh, N. A., Côté-Lecaldare, M., Häuser, K., Baum, S. R., & Titone, D. (2015). Individual differences in executive control relate to metaphor processing: An eye movement study of sentence reading. Frontiers in human neuroscience, 8, 1057.

Glucksberg, S. (1998). Understanding metaphors. Current Directions in Psychological Science, 7(2), 39-43.

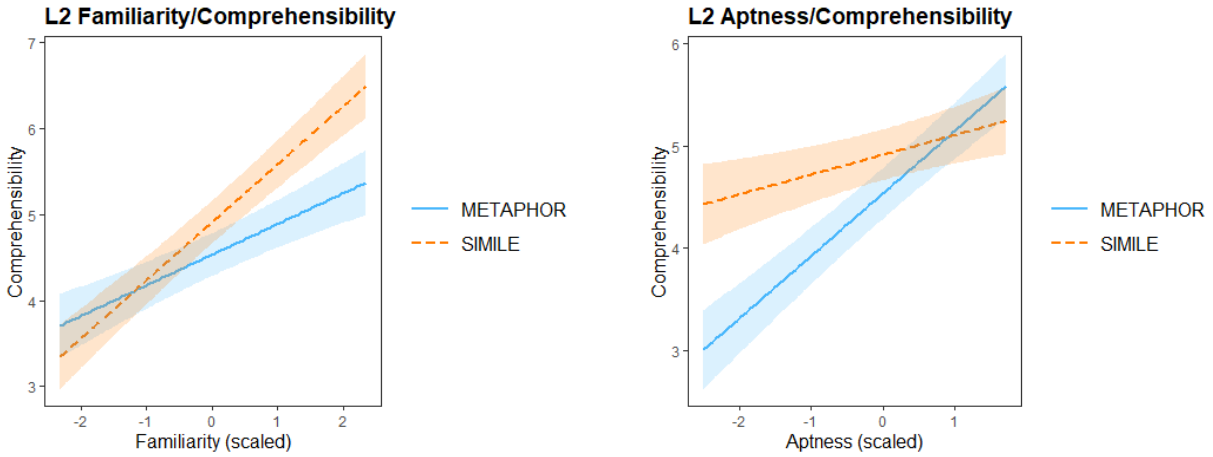
Now that you have been informed about the true nature of this study, you may now decide whether or not you would like the researcher to keep your data for use, or withdraw it and yourself from the study. If you withdraw from the study, all your data collected up until this point will be destroyed unless specified otherwise by you.

If you withdraw from the study, all your data collected up until this point will be destroyed unless specified otherwise.

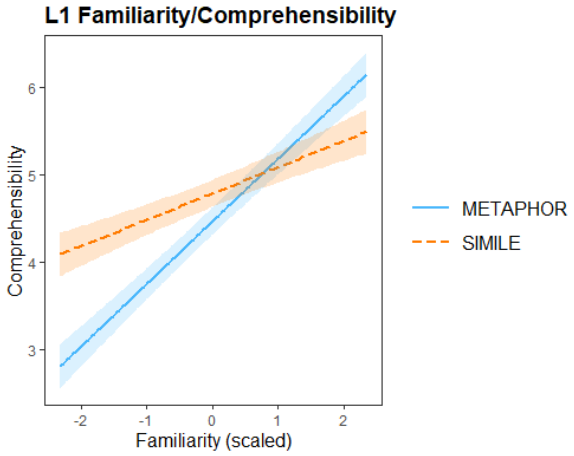
Appendix D

Additional figures for STUDY 3 (Chapter 4)

Figures corresponding to additional models for STUDY 3 are presented below. After initial linear mixed effect model analyses found three-way interaction effects of sentence type, language group and aptness/familiarity, survey responses were broken down by language background (L1 vs. L2) and further linear mixed effects models were performed testing the interactions between sentence type and aptness/familiarity in L1 and L2 language groups separately. Below are charts representing significant interactions between sentence type/aptness or sentence type/familiarity in the L1 and L2 language groups.



Comprehensibility ratings as a function of sentence type (metaphor/simile) and familiarity ratings (left) and aptness ratings (right) in L2 English speakers. +/- 1 SEM represented by shaded areas.



Comprehensibility ratings as a function of sentence type (metaphor/simile) and familiarity ratings in L1 English speakers. +/- 1 SEM represented by shaded areas.