

L1 & L2 extended metaphor comprehension: behavioural insights into the effects of
language nativeness and metaphor novelty

by

Stepan Kopytov

Bachelor of Arts, Moscow State Linguistic University, 2023

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF ARTS

in the School of Languages, Linguistics, and Cultures

©Stepan Kopytov, 2025

University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by
photocopy or other means, without the permission of the author.

We acknowledge and respect the Ləkʷəŋən (Songhees and Xʷsepsəm/Esquimalt)
Peoples on whose territory the university stands, and the Ləkʷəŋən and ƳSÁNEĆ
Peoples whose historical relationships with the land continue to this day.

L1 & L2 extended metaphor comprehension: behavioural insights into the effects of
language nativeness and metaphor novelty

By

Stepan Kopytov

Bachelor of Arts, Moscow State Linguistic University, 2023

Supervisory Committee

Dr. John Archibald, Supervisor
School of Languages, Linguistics, and Cultures

Dr. James Tanaka, Co-supervisor
Department of Psychology

Abstract

This thesis investigates how native (L1) and non-native (L2) English speakers comprehend metaphorical language, with a specific focus on nominal extended metaphors. While prior research has extensively examined single nominal metaphors, often of the “X is a Y” type, comparatively little is known about the processing of extended metaphoric expressions, particularly among bilinguals. This study addresses that gap by exploring if and to what extent metaphor conventionality and figurative context modulate comprehension in L1 and L2 speakers.

The primary experiment employed a semantic judgment task (SJT), in which participants were presented with short context paragraphs followed by target phrases. They were asked to judge whether the phrases were meaningful or meaningless, while both reaction times and accuracy rates were recorded. The experiment was conducted using PsychoPy v2024.2.4 and included both single and extended metaphor conditions, with target utterances falling into four categories: novel metaphors, conventional metaphors, literal sentences, and anomalous sentences. The results revealed several important findings. First, there was no evidence that extended metaphor contexts induced a shift in the processing mode for conventional metaphors. That is, the presence of context did not lead participants to engage in alignment-based (comparison-driven) processing. However, metaphorical context did have a clear facilitatory effect on the processing of both novel and conventional metaphors, though differences emerged between groups. Native speakers processed both types more quickly and accurately when metaphor-related contextual support was provided, while non-native speakers showed a similar trend, but the effect for conventional metaphors was not statistically significant. Notably, the study found that both L1 and highly proficient L2 speakers

benefited from metaphor conventionality in the single condition, contrary to earlier findings which suggested that L2 speakers are generally less sensitive to this semantic dimension. This suggests that at high levels of proficiency, L2 speakers may engage with familiar figurative expressions in ways that resemble native processing. At the same time, the data also indicated that L2 participants made less use of extended context, with smaller numerical gains and reduced sensitivity compared to their L1 counterparts.

These findings provide new behavioural evidence that additional vehicle-related elements in extended metaphors reduce processing costs. Moreover, they highlight the importance of language proficiency in shaping metaphor comprehension, suggesting that L2 users — at least those that are highly proficient — can achieve sensitivity to figurative conventions comparable to that of native speakers. This thesis contributes to and builds on our currently available knowledge of how metaphor structure, conventionality, and language status interact in cognitive processes associated with semantic ambiguity resolution in real-time language comprehension.

Table of contents

Supervisory Committee.....	.ii
Abstract.....	.iii
Table of Contents.....	v
List of Tables.....	viii
List of Figures.....	ix
Dedication.....	xi
1. Chapter one: introduction.....	1
2. Chapter two: top-down and bottom-up metaphor processing.....	7
2.1 Top-down approaches to metaphor processing.....	11
2.1.1 Analogical mapping approach.....	12
2.1.2 Categorisation approach.....	14
2.1.3 Alignment-first or abstraction-first processing: career of metaphor.....	16
2.2 Literal and nonliteral routes to metaphoric meaning.....	26
2.3 Extended metaphor comprehension.....	30
2.3.1 ‘Lingering of the literal’ and ‘lingering of the linguistic form’.....	31
2.3.2 Empirical evidence for ‘Carston’s account’.....	35
2.3.2.1 An eye-tracking study by Rubio-Ferdnández et al.....	35
2.3.2.2 A study by Ronderos and Falkum: a cross-modal lexical priming paradigm.....	40
2.4 Predictions about extended metaphor processing: bilingualism and anticipated effects of conventionality.....	43
2.4.1 Bilingual metaphor processing.....	43
2.4.2 Conventionality effects on extended metaphor processing.....	48
2.4.3 Conventionality effects within the framework of Relevance Theory.....	53
2.5 Research gap.....	55

2.6	Research objectives.....	57
2.7	Empirical predictions.....	57
3.	Chapter three: the current study.....	59
3.1	Experiment 1: pre-test of stimuli.....	59
3.1.1	Methods.....	59
3.1.2	Results and discussion.....	61
3.2	Experiment 2: validation study.....	69
3.2.1	Methods.....	69
3.2.2	Results and discussion.....	71
3.2.3	Limitations.....	80
3.3	Experiment 3.....	81
3.3.1	Methods.....	81
3.3.1.1	Design.....	81
3.3.1.2	Materials and procedure.....	83
3.3.1.3	Participants.....	85
3.3.2	Results and discussion.....	86
3.3.2.1	Reaction time results across all utterance types in the single condition.....	86
3.3.2.2	Accuracy rates across all utterance types in the single condition.....	90
3.3.2.3	Reaction time results: metaphor type x language group x condition.....	93
3.3.2.4	Accuracy rates: metaphor type x language group x condition.....	97
3.3.2.5	Effects of conventionality as a continuous variable.....	102
4.	Chapter four: general discussion.....	107
4.1	Contextual facilitation: extended metaphors reduce processing costs.....	108
4.2	Language background & conventionality effects: strategic processing differences.....	109

4.3	Limitations and future research.....	111
6.	References	114

List of Tables

Table 1: Pairwise t Values and p Values from Mixed Effects Regression.....	38
Table 2: Post-hoc Comparisons for Reaction Time (Bonferroni Corrected).....	73
Table 3: Post-hoc Comparisons for Accuracy (Bonferroni Corrected).....	74
Table 4: ANOVA: RT (Single Condition).....	87
Table 5: Post-hoc: RT (Within-Group).....	88
Table 6: Post-hoc: RT (Between-Group).....	89
Table 7: ANOVA: Accuracy (Single Condition).....	91
Table 8: Post-hoc: Accuracy within Groups.....	91
Table 9: Post-hoc: Accuracy between Groups.....	93
Table 10: ANOVA: RT (Metaphors Only, Both Conditions).....	95
Table 11: Post-hoc: RT (Metaphors Only, within Groups).....	96
Table 12: Post-hoc: RT (Metaphors Only, between Groups).....	97
Table 13: ANOVA: Accuracy (Metaphors Only).....	98
Table 14: Post-hoc: Accuracy (Metaphors Only, within Groups).....	99
Table 15: Post-hoc: Accuracy (Metaphors Only, between Groups).....	101
Table 16: Linear Regression: Reaction Time ~ Conventinality.....	103
Table 17: Logistic Regression: Accuracy ~ Conventinality.....	105

List of Figures

Figure 1: The Diagram of Annual Publications on Metaphor Processing.....	2
Figure 2: Metaphorical Paradigm.....	3
Figure 3: Novel and Conventional Metaphor Comprehension.....	4
Figure 4: Visualized Structure Mapping Theory.....	13
Figure 5: Categorization-based Processing Model (All men are wolves).....	15
Figure 6: Career of Metaphor.....	21
Figure 7: Comprehension Times for Experiment 2.....	21
Figure 8: Reaction times (ms) to novel metaphors, conventional metaphors, literal utterances, and anomalous utterances among intermediate and advanced L2 speakers.....	25
Figure 9: Reaction times (ms) to novel metaphoric, conventional metaphoric, literal, and anomalous word dyads in Polish (dark grey) and English (light grey).....	26
Figure 10: Factors Constituting Semantic Salience.....	28
Figure 11: Proposed Novel Extended Metaphor Processing Model.....	33
Figure 12: Examples of Experimental Stimuli.....	36
Figure 13: Example of a Critical Trial in All Conditions.....	41
Figure 14: Activation Pattern of Literal-related and Metaphor-related Features...42	
Figure 15: Processing Procedures for the Two Metaphor Types: Novel and Conventional.....	51
Figure 16: Anticipated Trends in Processing Times: Single and Extended Metaphors.....	51
Figure 17: Alternative Conventional Extended Metaphor Processing Model (the strong version).....	54
Figure 18: Comparison of Dimensions across Stimulus Types.....	62
Figure 19: Comprehensibility Distribution by Stimulus Type.....	63
Figure 20: Metaphoricity Distribution by Stimulus Type.....	64
Figure 21: Familiarity Distribution by Stimulus Type.....	65

Figure 22: Conventinality Distribution by Stimulus Type.....	66
Figure 23: Conventinality of Metaphors (Most to Least Conventional).....	67
Figure 24: Regression of Comprehensibility on Conventinality.....	68
Figure 25: Visual Timeline of an Experimental Trial (Experiment 2).....	70
Figure 26: Mean Reaction Times by Stimulus Type.....	72
Figure 27: Mean Accuracy by Stimulus Type.....	72
Figure 28: Reaction Times vs Mean Conventinality.....	75
Figure 29: Accuracy vs Mean Conventinality (Logistic Regression).....	76
Figure 30: Reaction Time vs Mean Familiarity.....	77
Figure 31: Accuracy vs Mean Familiarity (Logistic Regression).....	78
Figure 32: Examples of Stimuli for Novel Single and Novel Extended Metaphors.....	82
Figure 33: Visual Timeline of an Experimental Trial (Experiment 3).....	84
Figure 34: Mean RT with SE: Line Plot.....	86
Figure 35: Accuracy by Stimulus Type (Single Condition).....	90
Figure 36: Interaction: RT by Condition & Metaphor Type.....	94
Figure 37: Interaction: Accuracy by Condition & Metaphor Type.....	98
Figure 38: Reaction Time vs Conventinality: Linear Regression.....	102
Figure 39: Accuracy vs Conventinality: Logistic Regression.....	104
Figure 40: Slope of Accuracy ~ Conventinality.....	106

Dedication

To my family, friends, and mentors — thank you for your constant encouragement and support. And to all the multilingual minds that navigate more than one language and more than one world: this work is, in many ways, about you. May it stand as a small tribute to the power of metaphor to help us understand and express what is otherwise difficult to say...

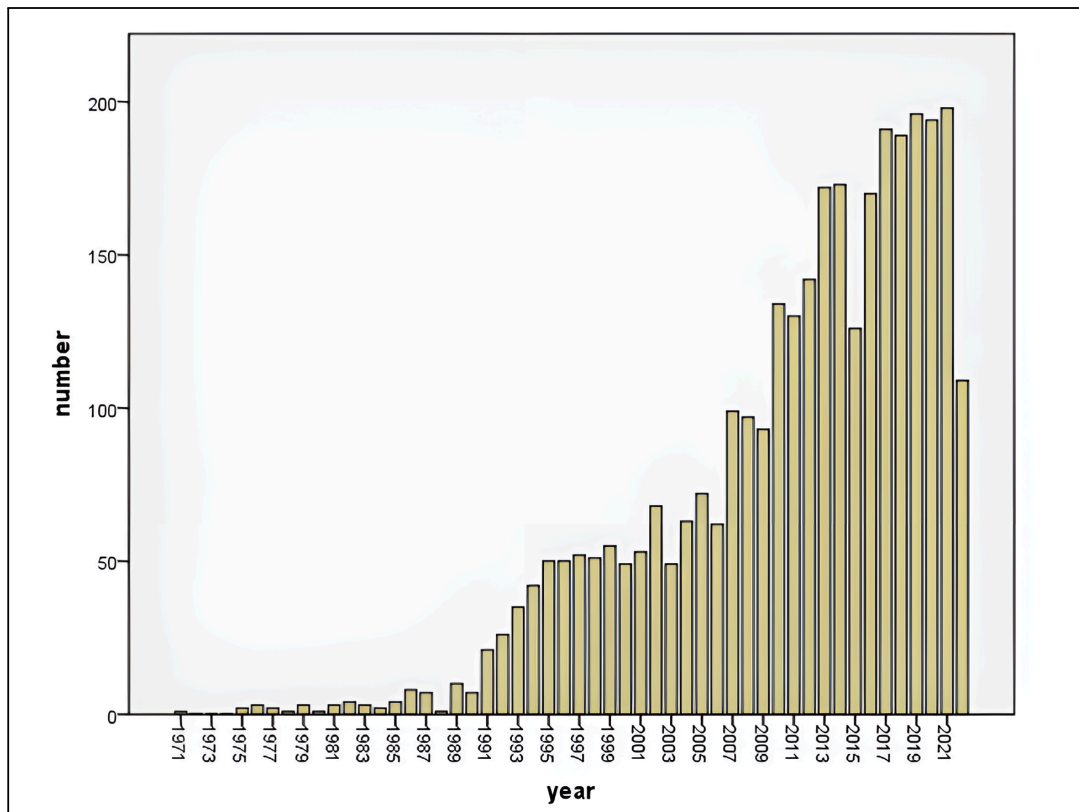
1. Chapter one: introduction

Figurative language processing undoubtedly is one of the most frequently investigated areas within the broader scope of semantic processing (Neves et al., 2021). Non-literal expressions – be those instances of metaphorical language, idioms, hyperbole or irony – are pervasive in everyday communication (Lakoff & Johnson, 1980), allowing speakers to convey complex ideas, emotions, and abstract concepts in a concise and vivid manner. Unlike literal language, which maintains a direct correspondence between words and their referents, figurative expressions require listeners to go beyond surface meanings and infer intended interpretations based on context and prior knowledge.

Over the past four decades, researchers from various fields — such as semantics (e.g., Recanati, 2004; Carston, 2002), cognitive science (e.g., Lakoff & Johnson, 1980; Fauconnier & Turner, 2002; Thibodeau & Boroditsky, 2011), and psycholinguistics (e.g., Swinney & Cutler, 1979; Glucksberg, 2001; Coulson & Van Petten, 2002; Bambini et al., 2012, etc.) — have explored the complex mechanisms underlying figurative language comprehension. These efforts have placed particular emphasis on metaphor (see Figure 1), which serves as a key area of investigation due to its pervasiveness in language and its reliance on abstract and context-dependent meaning-making processes.

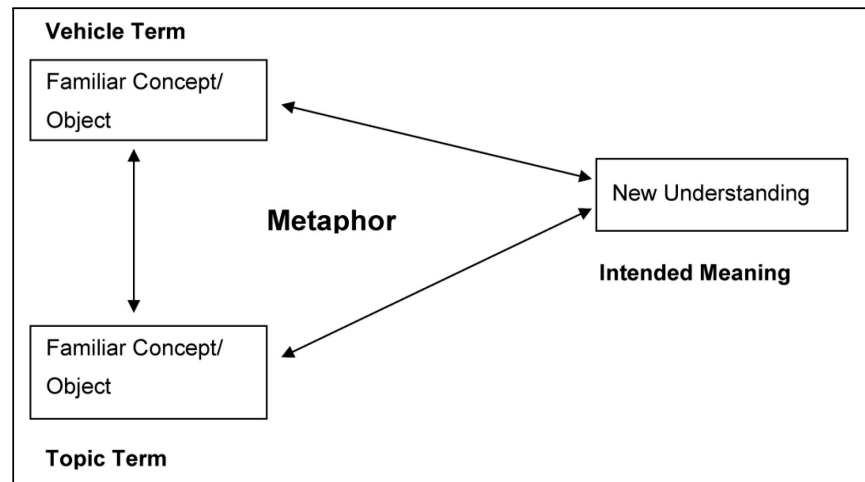
Figure 1

The Diagram of Annual Publications on Metaphor Processing (Peng & Khatin-Zadeh, 2023)



For instance, the metaphor *time is money* illustrates how people conceptualize time in terms of a tangible, valuable resource, shaping expressions like “*spending time*” or “*wasting time*”. Metaphors involve a topic¹ term and a vehicle term, where one concept (topic) is understood in terms of the other (vehicle). Such figurative mappings highlight the cognitive mechanisms that essentially allow speakers to mentally represent abstract concepts through more concrete domains (see Figure 2).

¹ In the literature on metaphor processing, the terms *topic* and *target* are used interchangeably, just as the terms *vehicle* and *base*.

Figure 2*Metaphorical Paradigm (Jensen, 2006)*

In my thesis, I focus on the intricate nature of metaphor comprehension (MC). Although the dominant amount of research done in the area of figurative language processing has been limited to the monolingual context (e.g., Bowdle & Gentner, 2005; Pynte et al., 1996; Gentner & Wolff, 1997), more and more papers have started to emerge recently, reporting the processing specificities of bilingual speakers as well (e.g., Heredia, R. R., & Cieślicka, 2015; Ardila et al., 2017; López et al., 2017).

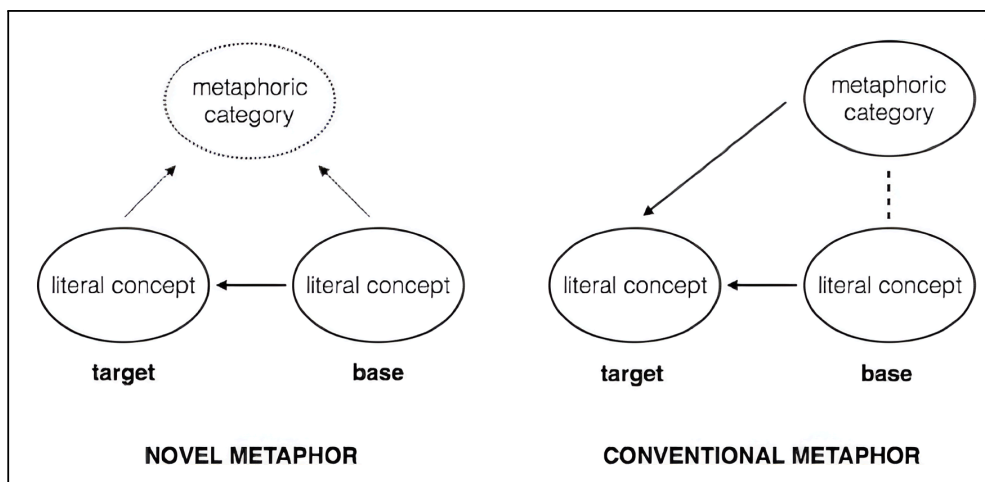
Currently, L1 (first language) and L2 (second language) MC research papers frequently investigate the differences in comprehension between *novel* (more creative) and *conventional* (less creative) metaphors, metaphors and similes, metaphors and anomalous sentences. The distinction between novel and conventional metaphors is of particular interest to me as well, as it has been seen to reliably predict processing times² (Gentner & Bowdle, 2005). Novel metaphors are

² In studies examining this distinction, processing times are typically measured through behavioral tasks such as lexical decision tasks, sensicality judgment tasks (SJT), or self-paced reading paradigms. For example, in SJT, participants are presented with metaphorical or literal

those that are unique, creative, and typically require more cognitive effort to interpret because they do not align with common, pre-existing associations between the vehicle term and the related figurative category (see Figure 3).

Figure 3

Novel and Conventional Metaphor Comprehension (adapted from Bowdle & Gentner, 1999)



For example, *the forest of time* is an unusual and creative construction, requiring the reader to engage in novel mappings between the domains of forest and time. Conventional metaphors, on the other hand, are widely recognized and have become fixed expressions or conceptual patterns that have been used so frequently that their figurative meaning is well understood. These metaphors tend to evoke familiar, conventional associations, such as “time is money” where “money” is commonly perceived as a vehicle domain representing the figurative category of a valuable resource, so time comes to be conceptualized in similar terms. Due to their

sentences and asked to quickly judge whether each sentence makes sense, with response times (RTs) recorded as an index of processing effort. Self-paced reading tasks, by measuring the time it takes participants to read each segment of a sentence, also help to identify points of increased cognitive load.

widespread usage, conventional metaphors are often processed more quickly and automatically than novel metaphors (Bowdle & Gentner, 2005).

Pursuing this line of inquiry, researchers attempt to gauge the degree of complexity of the semantic processes involved at both initial semantic access stages (N400³) and later meaning-integration stages (the Late Positive Complex⁴ (LPC)) as well as the comprehender's level of automaticity in accessing abstract metaphorical categories. Some metaphors are so entrenched in the language that they are processed almost as if they were literal statements, requiring little cognitive effort. For example, the metaphor *a bright idea* is understood immediately without actively reflecting on the connection between *brightness* and *intelligence*. This is due to the metaphor's *familiarity* and *conventionality* — it is widely used and has become almost automatic in its interpretation. The *similarity* between the vehicle (brightness) and the target (idea) enhances its *aptness*, making the metaphor both effective and easily comprehended. Over time, metaphors like this come to be interpreted almost instantaneously and with minimal awareness of their figurative nature. Researchers examine this automaticity to determine whether comprehenders can quickly access these abstract metaphorical categories or if additional cognitive effort is required.

The primary objective of the majority of the recent MC publications has been to test the effect of certain characteristics of metaphors on comprehension load:

³ The N400 component is a negative-going event-related potential (ERP) that typically peaks approximately 400 milliseconds after the presentation of a stimulus, and it is most prominent over centro-parietal regions of the scalp. In studies on metaphor comprehension, this neural response often varies in amplitude depending on the type of expression. Novel metaphors tend to elicit the strongest N400 responses, conventional metaphors produce intermediate responses, and literal statements evoke the weakest, indicating a gradation in processing demands (e.g., Arzouan et al., 2007; Coulson & Van Petten, 2002; Lai et al., 2009)

⁴ The Late Positive Component (LPC) is a positively oriented ERP waveform that typically peaks within the 500 to 900 millisecond window following stimulus presentation (Friedman & Johnson, 2000). This neural marker is frequently linked to processes that require reinterpretation or reevaluation of meaning, as well as increased engagement of working memory resources (e.g., Coulson & Van Petten, 2002, etc.)

metaphor familiarity, conventionality, aptness and other semantic dimensions. In this connection, gaining more experimentally supported knowledge about the correlation between certain qualities of metaphors and comprehension load will present a valuable contribution to our general understanding of language processing mechanisms. It is, therefore, my conviction that a more accurate and empirically supported understanding of metaphoric mapping will eventually enable researchers to form a comprehensive perspective on how knowledge about the surrounding world is represented in our mind and what the nature of the processing challenge is for the bilingual mind relative to monolingual, if any at all.

At the very core of metaphor comprehension discussion stand a whole host of theories and processing models approaching the issue of MC from various angles (e.g., The Comparison Model (e.g., Gentner & Clement, 1988; Gentner, Falkenhainer, & Skorstad, 1988); The Categorisation Model (Glucksberg et al., 1997), the Career of Metaphor model (Bowdle & Gentner, 2005), etc.). Noteworthy is the fact that all of these models have been proposed to explain single nominal metaphor comprehension (e.g., *this job is a jail*, or, more formally, X is a Y metaphors), clearly because of its structural simplicity. Metaphor research generally relies on simple X is Y metaphors, which occur infrequently in everyday discourse. Nonetheless, these metaphors serve as a foundation for many psycholinguistic theories of metaphor processing (Al-Azary & Katz, 2021; Gentner, 1983; Glucksberg & Keysar, 1990), and the mechanisms involved in processing X is Y metaphors are thought to generalize to other metaphorical forms, such as, for example, adjective-noun combinations as in *he gave her a sharp look* (Al-Azary et al., 2021).

Although there are more types of metaphors if we take a look at their formal typology (e.g., adjective metaphors or verbal metaphors (e.g., Neuman et al., 2013),

I am particularly interested in nominal *extended* metaphors, which are also referred to as *sustained* (Werth, 1994). Appearing very rarely in academic papers on figurative language processing, with the exception of a few (e.g., Ronderos & Falkum, 2023; Carston, 2010; Rubio Fernandez, 2007; Mathisen, 2023), sustained metaphors present special interest for research investigation because their length and structural characteristics might exert influence on processing times and accuracy when comprehending such utterances. Besides, to my knowledge, no research has been conducted on bilingual sustained metaphor processing, so it remains a largely understudied area of knowledge which, by all means, requires more analysis. My study is an attempt to fill the existing gap primarily by contributing new empirical findings into the domain of nominal extended metaphor processing – a form that has received little empirical attention compared to single nominal metaphors. By examining how their structural complexity and length influence comprehension — especially in bilingual speakers — it will provide the first known data on bilingual sustained metaphor processing and explore whether existing models, largely developed for simpler X is a Y forms, can account for the cognitive demands of extended figurative expressions.

2. Chapter two: top-down and bottom-up metaphor processing

The study of MC has witnessed significant theoretical development over the last 40 years since the publication of ‘Metaphors We Live By’ (Lakoff & Johnson, 1980), with two primary paradigms often serving as the framework for understanding how metaphors are processed: top-down approaches and bottom-up approaches. These paradigms represent contrasting but complementary perspectives on the mechanisms underlying metaphor processing. Top-down approaches emphasize the role of broader cognitive structures, such as conceptual mappings in shaping the

interpretation of metaphors. These models argue that metaphor comprehension is guided by pre-existing schemas, conceptual domains, or high-level associations that provide an interpretative framework (Lakoff & Johnson, 1980; Gibbs, 1994). By focusing on how these overarching structures influence the activation and selection of meaning, top-down theories, in their essence, assume that metaphor identification is based on predetermined conceptual metaphors (Krennmayr, 2013). Again, a well-known example is the conceptual metaphor TIME IS MONEY, which structures expressions like “*You’re wasting my time*” or “*I can’t afford to spend another hour on this,*” reflecting an underlying mapping between the abstract domain of time and the more concrete domain of economic value.

In contrast, bottom-up approaches, being rooted in word recognition, prioritize the local linguistic input and center around how metaphor comprehension emerges from the processing of the lexical, syntactic, and, most importantly, semantic features of the metaphor itself. In other words, metaphors are identified and processed as such before conceptual mappings are formulated (Krennmayr, 2013). Thus, while *semantics* originates from the bottom-up⁵ — built from the meaning of individual words and their combinations — *conceptual mappings* emerge top-down⁶, as higher-level structures that organize and give coherence to metaphorical interpretations once they are activated. Recent research has investigated the significance of such semantic characteristics as concreteness⁷ of metaphor and

⁵ In bottom-up models, semantics refers to the meanings activated by individual lexical items and their immediate syntactic context — this includes detecting selectional restriction violations or anomalous word pairings (e.g., “lawyer is a shark”), which may signal the presence of metaphor. In this sense, semantic processing operates locally and automatically, potentially triggering deeper interpretation.

⁶ Conversely, concepts are assumed to reside at a higher level of abstraction, involving structured knowledge domains and mappings (e.g., ANIMAL → PERSON) that are accessed after an initial semantic mismatch is detected.

⁷ *Concreteness* refers to the degree to which a word denotes a perceptible, tangible entity. In empirical studies, it is normally treated as a dichotomous variable: *concrete words* are those that represent physical, observable objects or experiences (e.g., *sword*), whereas *abstract words*

semantic neighbourhood density (SND) in metaphor comprehension – particularly their role as manipulating factors of comprehension difficulty (e.g., Kinstch, 2000; Xu, 2010; Al-Azary & Buchanan, 2017). SND refers to the degree of similarity between a word and its closest semantic associates — that is, how many meaningfully related words surround it in semantic space (Buchanan et al., 2001). A word from a dense semantic neighborhood shares high similarity with many nearby words (e.g., *dog*, which is closely related to *cat*, *pet*, *puppy*, *animal*), whereas a word from a sparse neighborhood has fewer closely related associates (e.g., *galaxy*, which has fewer high-overlap neighbors). Research shows that metaphors are often easier to comprehend when built around words from sparse neighborhoods, as these words are less semantically constrained and more open to novel figurative interpretations (Al-Azary & Buchanan, 2017; Al-Azary et al., 2019, 2021; Al-Azary & Katz, 2023).

The interplay between top-down and bottom-up approaches has shaped much of the theoretical discourse on metaphor processing. While top-down models have been particularly influential in explaining the role of conceptual metaphors and high-level cognitive mechanisms (e.g., schema activation, analogical mapping, or domain interaction), bottom-up models have shown how metaphor comprehension can be modulated not so much by the type of alignment between the domains⁸, but rather by purely linguistic and perceptual properties of words. Importantly, there is clear evidence supporting both top-down approaches (e.g., Gentner & Wolff, 1997;

refer to intangible concepts that lack direct sensory referents (e.g., *love*, *patience*) (Al-Azary & Buchanan, 2017).

⁸ Here, by “alignment between the domains,” I refer to the structural or relational similarity between the topic and vehicle domains in a metaphor. For example, in the conceptual metaphor ARGUMENT IS WAR, elements from the vehicle domain WAR such as *soldiers*, *attacks*, and *defenses* align systematically with elements in the topic domain ARGUMENT, such as *participants* (in a debate or argument), *criticisms*, and *justifications*. Top-down models emphasize how such mappings guide comprehension by projecting structured knowledge from one domain onto another, and by doing so they facilitate understanding through a certain type of alignment, be it analogical or categorical (the latter will be discussed further).

Bowdle & Gentner, 2005) and bottom-up approaches (Al-Azary & Buchanan, 2017) to metaphor processing. The particular interest in the latter approaches, which chronologically also emerged later, stems primarily from the observation of automatic⁹, near-instantaneous metaphor comprehension (Glucksberg et al., 1982). This rapid processing raises significant questions about whether traditional top-down, alignment-based models can fully account for the remarkable efficiency with which metaphors, mainly nominal ones (e.g., X is a Y), are often understood. At the same time, it is crucial to acknowledge that top-down approaches have amassed substantial empirical support across decades of research and remain foundational in theoretical discussions of metaphor comprehension. The robustness of these models lies in their ability to account for the role of entrenched metaphorical mappings and conventionality effects in facilitating comprehension. Therefore, it may be more productive to explore how they interact with bottom-up frameworks in a more integrated way that accommodates both rapid metaphor processing and takes into consideration the established conceptual structures and the proposed mappings of those structures.

The following section outlines major models of metaphor comprehension in an attempt to trace their developmental trajectory within the MC literature. This discussion begins with the foundational Comparison Models, which rely on feature-matching processes between the topic and vehicle, laying the groundwork for both top-down and bottom-up perspectives on metaphor interpretation.

⁹ The term “automatic” refers not to the absence of cognitive effort, but to the fact that the process unfolds without the speaker’s conscious control (Glucksberg et al., 1982; Cersosimo et al., 2024).

2.1 Top-down approaches to metaphor processing

My examination of MC literature starts with an overview of the major metaphor comprehension models in an attempt to show the developmental path of metaphor processing theory. The perhaps most rudimentary approach to metaphor comprehension rests on the assumption that metaphors are understood through pure matching, i.e. by matching identical elements constituting the topic and the vehicle of the metaphor. The goal is to find features that maximize the resemblance between them (Tversky, 1977). For example, in the metaphor *the sun is a lamp*, a pure matching model would identify shared features like *emitting light* or *producing warmth* — concrete, observable characteristics — as the basis for understanding the metaphor. However, this approach is limited because not all shared features are relevant to metaphor comprehension (e.g., in the metaphor *all surgeons are butchers*, both surgeons and butchers wear white coats, but these features are not part of our interpretation of that metaphor), and metaphoric meaning often depends on deeper relational or structural similarities rather than simple feature overlap.

Those ‘pure matching’ models, primarily because they merely propose a straightforward comparison mechanism, appear to disregard a whole host of core metaphoric principles of which we are now aware (Gentner & Wolff, 1997). In particular, matching models do not account for feature selection, metaphor asymmetry, and feature importation. Metaphors are indeed very asymmetrical in nature, and if we assume them to be only comprehended like literal similarity assertions¹⁰, we will encounter a problem when trying to reverse a metaphor as it will likely lose its original meaning or its comprehensibility entirely. *A surgeon is an artist*

¹⁰ Unlike metaphors, which are figurative and involve cross-domain mappings, *literal similarity assertions* stay within the same domain and refer to actual, non-metaphorical resemblances.

suggests that the surgeon possesses the skill and finesse associated with artists. However, *an artist is a surgeon* does not make the same kind of sense, even though the surface structure is similar. This exact asymmetry poses a challenge for models that treat metaphors as purely symmetrical similarity judgments.

The mapping of features between the topic and vehicle domains is not reciprocal: the vehicle domain's features are often transferred to the topic domain, but not the other way around. For example, in the metaphor *time is money*, *time* is understood through the lens of *money* and becomes characterised as a valuable resource. Thus, features that belong only to the metaphor vehicle (the second term in an "X is a Y" metaphor) are imported or projected onto the topic (X) during interpretation — even if these features were not known or applicable to the topic beforehand. Crucially, this mechanism of feature importation demonstrates how metaphors can *create meaning*, and not just highlight preexisting similarities.

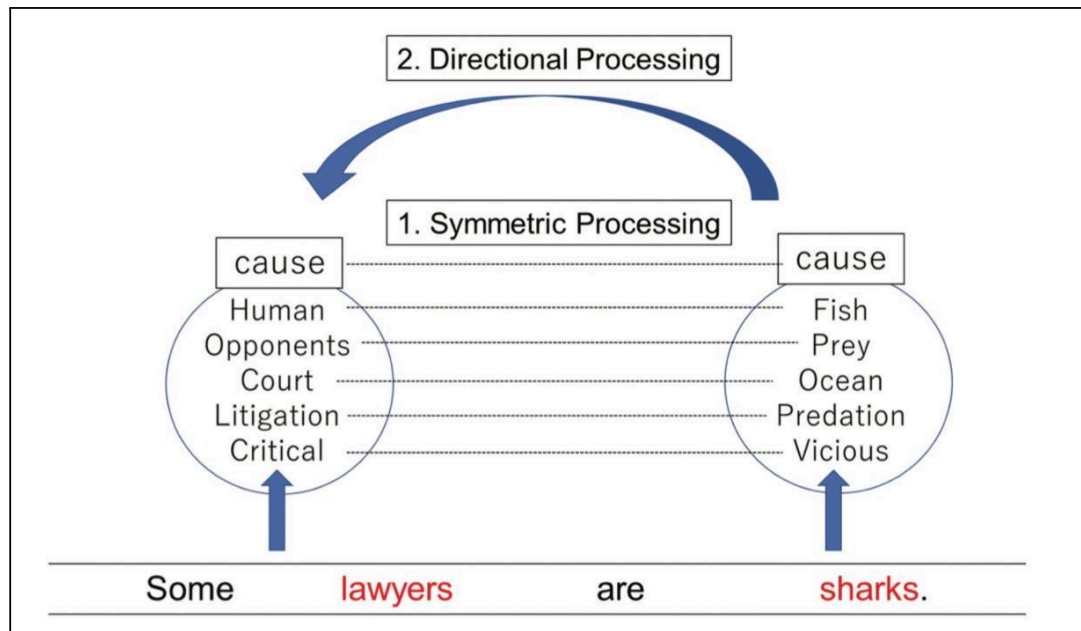
2.1.1 Analogical mapping approach

To tackle the above-mentioned theoretical shortcomings of the early similarity-based models – particularly their failure to explain metaphor asymmetry, selective feature rejection, and the importation of features – alternative approaches have emerged that offer more comprehensive accounts of metaphor processing. They incorporate mechanisms that account for the asymmetrical nature of metaphors, and the projection of salient vehicle-based features onto the topic, even when those features are not mutually shared. Among these are *alignment-based processing* and *categorization-based processing models*. One of such revised MC theories revolves around the principle of structural alignment and entails a feature-matching process between the topic and the vehicle of the metaphor (Gentner, 1983; Gentner & Wolff, 1997; Miller, 1993). Thus, the correct interpretation

of the metaphor, according to this model, involves identifying a horizontal property overlap between the properties of the topic and the vehicle in a fashion indicated in Figure 4 (Wolff & Gentner, 2011).

Figure 4

Visualized Structure Mapping Theory (Ikuta & Miwa, 2021)



Under this view, metaphor comprehension proceeds in two sequential phases. The first phase, feature overlap identification, involves (1) a relational process of property alignment, also referred to as *symmetric processing*. Here, the listener identifies structural parallels between the topic and the vehicle — treating both domains as co-equal in order to find shared relational patterns. This is followed by (2) a directional stage (i.e., *directional processing*), where specific features from the vehicle are projected onto the topic (Gentner & Wolff, 1997).

Take, for example, the metaphor "Lawyers are sharks." In the initial symmetric phase, the relationship between the topic *LAWYER* and the vehicle *SHARK* is evaluated by aligning their relational structures — such as the effect both exert on

their adversaries. One might align legal aggression with predatory aggression, or competitive court behavior with dominance in an ecosystem. These mappings may also include analogies between their environments (e.g., *courtroom* vs. *ocean*), and roles (*legal opponent* vs. *prey*).

In the subsequent directional phase, features unique to the vehicle — such as ruthlessness, relentlessness, or cold-blooded efficiency — are projected (or imported) onto the topic. Thus, the metaphor communicates that lawyers are understood not literally as sharks, but through the lens of shark-like qualities, inferred through this structured semantic extension.

2.1.2 Categorisation approach

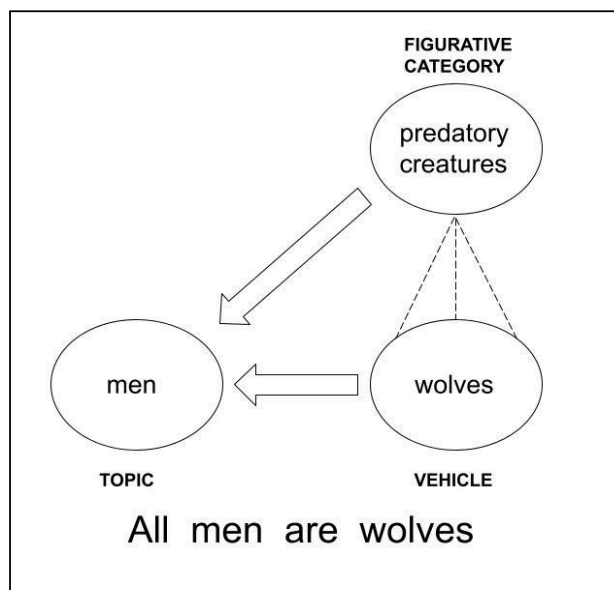
Another model of metaphor processing is the Categorization Model or Abstraction-first Processing Model, and it presupposes that whenever we operate with a metaphoric utterance, we access a generalized *figurative category* of the vehicle first rather than its *literal base* (Glucksberg, 2003). For example, in the metaphor *Men are wolves*, the term *wolves* is not initially understood in its literal sense (as furry, four-legged animals), but rather as a member of a broader figurative category — such as *predatory or dangerously seductive individuals*. The listener categorizes *men* as belonging to this abstract group, drawing on conventional associations. This model bypasses the need to activate and then suppress the literal meaning, offering a more efficient route to metaphor interpretation. The topic of the metaphor is only assumed to be involved minimally in this process, appearing at the final stage when the figurative category is projected onto the topic for verification¹¹

¹¹ Later interpretations of this processing model, however, assume that metaphor topic is actively involved initially in the dimension specification of the figurative category that is being derived. If we presume that for some metaphors the figurative category is already available and can be retrieved easily, it becomes theoretically plausible that this stage is optional and only occurs in novel metaphor comprehension.

(Gentner & Wolff, 1997). Interestingly, the vehicle of the metaphor WOLVES supposedly represents a more abstract category of *predatory creatures*. In this way, the vehicle acquires dual reference, for it is simultaneously representative of both the literal base and the figurative abstract category which is more generalized (Glucksberg et al., 1997; Bowdle & Gentner, 2005). Depending on how well-formed¹² the figurative category is and how promptly it can be accessed, metaphor interpretation will either involve a *derivation* of that figurative category on the spot and then applying it to the topic, or *accessing* (or retrieving) it directly without the need to generate it first (Gentner & Wolff, 1997).

Figure 5

Categorization-based Processing Model (All men are wolves)



¹² Figurative categories vary in how "well-formed" they are, and this affects how easily they are accessed during metaphor comprehension. A well-formed figurative category is one that is widely shared, conventionalized, and easily retrievable—such as *predator*ⁱ in the metaphor *men are wolves*ⁱ, or *valuable resource*ⁱⁱ in *time is money*ⁱⁱ. These categories are abstract but conventional, and so interpretation tends to be fast and automatic. In contrast, less well-formed figurative categories are novel, loosely defined, or context-dependent. For example, in the metaphor "*Ideas are seedlings*," the listener may need to construct the figurative category (e.g., "*things that require care and time to grow*") on the fly, based on shared features between *ideas* and *seedlings*. Here, the interpretation involves more inferencing and may take longer, as the figurative category is not readily available in memory.

2.1.3 Alignment-first or abstraction-first processing: career of metaphor

Bowdle and Gentner (2005) later proposed their own theoretical model of metaphor comprehension – The Career of Metaphor – which represents an integrative model, combining both the elements of analogical property alignment as a primary processing route in one type of metaphoric utterances and categorization-based processing patterns in another (see Figure 5). The Career of Metaphor proposes a diachronic model of metaphor development¹³, whereby each once novel metaphoric expression in due time experiences an increase of its *conventionality* and then, ultimately, becomes *dead* – that is, no longer metaphoric in its nature. For instance, the phrase *foot of the mountain* is no longer viewed as a metaphor as it has forfeited its figurative value – the relation between the base concept of the vehicle and the figurative category is no longer in place (Lakoff & Johnson, 1980). In contrast, consider a more actively metaphoric expression like *mountains swallowed the sun*, where the personification and image-based mapping remain salient to the reader or listener. The figurative connection between the mountain and the act of swallowing retains its figurative richness, unlike the *foot of the mountain*, which functions more like a lexicalized label than a metaphor (see Lakoff & Johnson, 1980 for continued discussion on how some parts of broader conceptual metaphors like *a metaphor is a person* become conventional while others may still maintain their figurative power).

Thus, every single developmental stage of metaphor existence presupposes its characteristic processing pattern in which the alignment of predicates in metaphor comprehension is subject to gradual change from horizontal¹⁴ (comparison-based) to

¹³ a framework describing how metaphorical expressions evolve over time

¹⁴ For example, when someone says “*Lawyers are sharks*,” we align the concept of a lawyer and that of a shark on a horizontal plane, looking across domains for features such as *aggression* or *relentlessness*. These shared features are drawn out through direct analogical mapping. The

vertical¹⁵ (categorization-based), with some intermediate phases¹⁶ of both types of alignments involved (Bowdle & Gentner, 2005). It is, in this respect, pertinent to note the gradual change of relationship between the literal vehicle concept and the corresponding figurative category, as in conventional metaphor processing the literal concept appears to acquire dual reference and thus becomes polysemous in nature (i.e., having more than one meaning) (Bolognesi & Werkmann Horvat, 2022). In the case of a dead metaphor, however, their relationship is switched to one of homonymy because the semantic association between the two concepts is very weak and not recognized by the speaker – the literal concept becomes less salient and, as a rule, is not activated in comprehension, thereby reducing the cognitive load (i.e., the reader/listener takes less time to compute the meaning of the metaphor). In this particular case, I define polysemy of metaphor vehicle as its ability to refer to two meanings, where one is the literal vehicle-related concept (e.g. in the case of *Lawyers are sharks*, we regard the literal meaning of *shark* as an instance of that) and the other is the derived, figurative vehicle-related concept (e.g. in the case of *Lawyers are sharks* we also regard *sharks* as *predatory creatures*). To use the exact terminology of Bowdle and Gentner, a polysemous base term has both a domain-specific meaning and a related domain-general meaning (Bowdle & Gentner,

metaphor is understood by *comparing* the two concepts across domains, hence the metaphorical “line” of reasoning moves horizontally — from one domain to another.

¹⁵ For instance, in *Jamie is a rock*, the vehicle rock no longer prompts comparisons to geological objects. Rather, *rock* has become a vertical category label — signifying someone who is emotionally stable, reliable, and grounded. The cognitive act now involves a vertical operation: classifying the topic as a *member* of a conceptually higher category represented by the vehicle. Hence, “vertical” reflects the taxonomic direction of metaphorical mapping — from abstract category (vehicle) to specific instance (topic).

¹⁶ Take the metaphor *time is a thief* for instance, where we might still compare time to a thief (both take something valuable without consent), but we may also begin to treat “thief” as a category that now includes abstract entities like time. At this stage, we gradually start to become aware of the figurative category associated with the vehicle and place the topic within it—while still maintaining some lateral (horizontal) relational comparison.

2005). Hence the notion of *the career of metaphor* – “evolution toward metaphoric polysemy” (ibid.).

A homonymous relationship between the literal base concept and the figurative category can be seen in the case of a dead metaphor. In general, it is natural for polysemous meanings to gradually turn into homonymous ones as secondary meanings over time become more well-formed and more prominent in the semantic structure of a polysemous word (Valera, 2020). Consider the term *arm* in the expression *arm of the chair*. Originally, *arm* refers to the upper limb of the human body. In the figurative use, *arm of the chair* describes the side support of the chair, resembling the function of an arm. As this metaphorical (or transferred) usage became widespread, the term *arm* in this context evolved into a dead metaphor, losing its original connection to the human limb and being understood simply as a part of the chair. Another well-known illustration of such gradual semantic change can be seen in the English verb *to fast*. Its older meaning — ‘to make firm’ — has, over time, become disconnected from its modern interpretation, ‘to abstain from food’, with the original link between steadfastness and self-denial now largely obscured (ibid.). A comparable development is found in Spanish, where *teniente* (‘lieutenant’) historically derives from the present participle of *tener* (‘to have’), reflecting a sense of someone holding a position (ibid.). These examples show us very well how meanings can evolve and drift apart over time. From this perspective, polysemy and homonymy are best viewed as points along a continuum rather than sharply divided categories.

This semantic shift demonstrates how frequent usage can turn a metaphor (or an instance of transferred use of a certain meaning) into a conventional term, making it less imaginative and more literal. This, in its turn, inevitably leads to disintegration

of once related meanings of a word into separate, heterogeneous lexical items, which retain the same lexical form. Such a process is called *split polysemy*, and it is considered one of the primary sources of homonymy in language (Arnold, 2014). Referring back to our *arm of the chair* example, polysemy is observed here in the original relationship between the human limb and the chair part, while homonymy is seen in the conventionalized term where *arm* in the context of furniture has become more of a fixed term independent of its original body-part semantics.

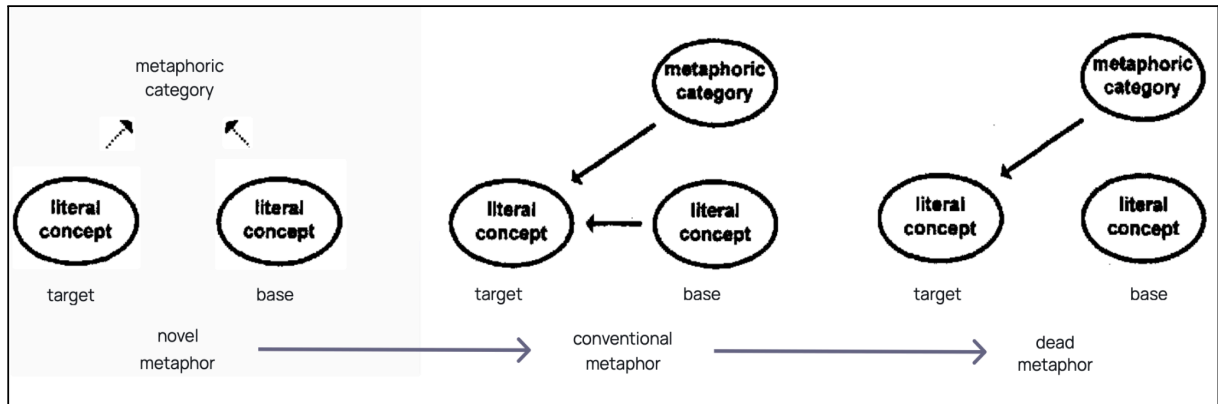
Bowdle and Gentner's empirical results support the proposed hypothesis, alongside indicating that, among other things, vertically aligned predicates (in Figure 6, the type of alignment we observe *a literal concept* and *a metaphoric category* to be in) are typically comprehended faster than those aligned horizontally (in Figure 6, the type of alignment literal concepts are in). This was determined by measuring the participants' comprehension time using an online Semantic Judgement Task. Participants pressed a key once they were certain they understood the figurative statement (e.g., *time is a river*), then typed their interpretation and pressed the enter key to continue to the next statement. The hypothesis concerning faster processing times associated with vertically aligned predicates (Bowdle & Gentner, 2005), such as in conventional metaphors, essentially derives from Bowdle and Gentner's theoretical argument about the structural content of metaphoric categories as opposed to literal base concepts. The alignment of the topic with the metaphoric category is less taxing from a computational perspective. This is because the predicates of metaphoric categories are typically more domain-general and limited in number, having undergone generalization as irrelevant properties are filtered out. Importantly, the crucial distinction between vertically aligned and horizontally aligned predicates lies in the 'maximal structural consistency' between the target and the

base in the former (ibid., p. 202). Thus, the assumption is that vertical alignment triggers less re-representation (i.e., rejection of the irrelevant features to avoid structural inconsistency) when matching predicates of the topic with those of the metaphoric category. In other words, due to high structural consistency, vertical alignment assumes fewer predicates will be rejected as inapplicable, for the majority of those that constitute the metaphoric category are relevant to the target. For instance, in the metaphor *time is a thief*, a predicate (or feature) like “*wears a mask*” — which might literally apply to a prototypical thief — would likely be rejected as inapplicable to the domain of *time*. Such features are too specific to the domain of the vehicle and fail to generalize across contexts. In contrast, predicates such as “*takes without permission*” or “*leaves you with less than you had before*” are more abstract and thus more relevant to the metaphorical understanding of time, making them more likely to be retained in the alignment process.

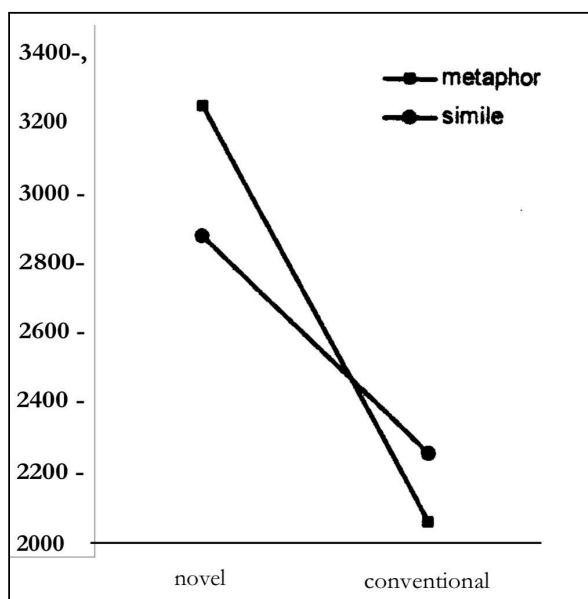
Reaction times were presented in support of this theory indicating that novel figurative phrases were comprehended more slowly than conventional ones (see Figure 6: comprehension times for novel [horizontally aligned] and conventional [vertically aligned] metaphors). On the whole, although Bowdle and Gentner do not explicitly state that in their paper, *The Career of Metaphor* describes a *qualitative* difference in the processing pattern. Referring back to the differentiation between vertical and horizontal alignment, these virtually represent two distinct processing models: categorization-based and comparison-based processing. In this regard, the contrast between novel and conventional metaphors manifests itself not only in the processing times, but rather in the different types of alignment and the respective differences in comprehension times that derive from that difference.

Figure 6

Career of Metaphor (adapted from Gentner & Bowdle, 2005)

**Figure 7**

Comprehension Times for Experiment 2 (from Gentner & Bowdle, 2005)



While there is clear evidence that *conventionality* modulates comprehension difficulty as is reflected in faster reaction times, an indeed interesting observation was made in a seminal study on the topic concerning the role of metaphor *similarity* in comprehension. Metaphor similarity – one of the semantic dimensions that metaphors are typically pre-tested for – refers to the extent to which the topic

concept inherently embodies the central abstraction associated with the vehicle.

While typically used to ensure stimulus quality in experimental design, this dimension also reflects a key factor in real-time comprehension: the more abstract overlap there is between topic and vehicle, the less conceptual restructuring (or re-representation) is required, which in turn facilitates interpretation.

This concept can be illustrated by comparing two metaphors: *that soldier is a pawn* and *that senator is a pawn* (Gentner & Wolff, 1997). In both cases, the vehicle term *pawn* conveys the notion of somebody being controlled by a higher authority with potentially hidden motives. This idea is already embedded in the central abstraction of *soldier* but is less likely to be a prominent constituent feature in the semantic representation of *senator*. As alignment-driven models suggest, metaphors with higher metaphorical similarity are predicted to be processed more quickly than those with lower similarity. As argued by Gentner & Wolff (1997), this is because a stronger abstract relational correspondence facilitates the alignment process. Having measured response times in a self-paced reading task, it appeared that conventionality does lead to faster processing on the condition that the similarity between the topic and the vehicle is high (ibid.). In cases when both metaphor conventionality and the similarity between the topic and the vehicle domains were high (e.g., *that argument is a war*, or *that sauna is an oven*), there was no indication of the participants processing the target metaphors through categorization.

MC research has seen a surge in publications since Bowdle and Gentner's seminal study in 2005, which demonstrated the facilitatory impact of conventionality on metaphor processing. Recent studies have extended these findings to L2 speakers. Although second language speakers generally show slower processing times compared to native speakers, the underlying cognitive mechanisms appear to

be similar. For instance, Lü et al. (2019) found that the underlying patterns of L1 and L2 metaphor acquisition are to a great extent of similar nature – namely, the initial activation of conceptually adjacent attributes¹⁷, followed by the suppression of those that are ultimately irrelevant to the metaphorical interpretation. The study found evidence of unfamiliar metaphors being more challenging to process for bilinguals as compared to familiar metaphors and discussed these findings within the framework of the *Predication Model* (Knitsch, 2000). L2 metaphor comprehension was found to be more complex and cognitively demanding due to an additional amount of linguistic and cultural information (greater density of semantic neighborhood) that is automatically retrieved in the bilingual mind. That is, understanding metaphors in L2 may necessitate not only filtering out irrelevant information but also suppressing non-target-language cues (Mercier et al., 2014). Additional language experiences, sociolinguistic and cultural information pose more of a strain on the executive control capacity of a bilingual speaker (Lü et al., 2019). Following the Predication Model, in second language metaphor of ‘X is a Y’ type, the semantic neighbours of the vehicle (Y) may be more weakly linked or located farther away within the L2 semantic network than in the native language, making its retrieval more demanding and requiring additional cognitive resources (Lü et al., 2019).

Building on this, it is important to consider how *conventionality* affects L2 metaphor processing — and here, a distinct pattern emerges. While conventionality plays a facilitative role for native speakers, its impact on non-native speakers appears more complex. Research reveals that the challenges L2 users face are not only related to the strength of associations in the semantic network but also to

¹⁷ By “adjacent attributes,” I refer to features that are closely associated with the metaphor vehicle in the mental lexicon or conceptual space, but that may or may not necessarily be relevant to the intended metaphorical meaning.

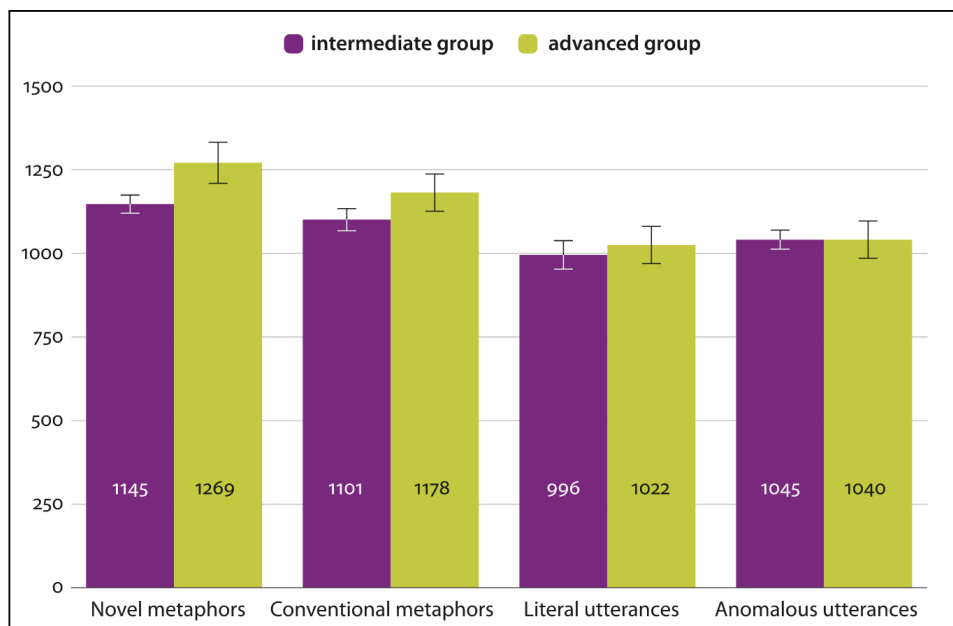
cross-linguistic differences in vehicle polysemy. As discussed above, vehicle polysemy in metaphors resulting from metaphorical and metonymic processes is a linguistic phenomenon that occurs across many languages and is characteristic for metaphor vehicles. The metaphorical meanings derived from such processes are often semantically transparent and, in some cases, shared across multiple languages. However, not all words exhibit consistent polysemous patterns. For instance, a mismatch may exist between a word's various meanings, or the word might not be polysemous in a speaker's L1 (Bolognesi & Werkmann Horvat, 2022). For example, in English, the word *carrot* is polysemous: it can refer to the literal vegetable, but also metaphorically to an incentive or reward used to motivate someone (ibid.). Russian shows a similar pattern — *морковь* (*morkov'*) can likewise be used metaphorically in political and everyday speech to represent a promised reward or enticement. However, this polysemous structure does not extend to Croatian, where *mrkva* retains only its literal, vegetable semantics, and the metaphorical extension is absent (ibid.).

This can naturally pose challenges for L2 speakers, as they may struggle to interpret the conventional meaning of a word when it appears out of context. The difficulty arises not only from the semantic opacity of the word but also from the L2 learner's potential lack of expectation that the word could have multiple meanings due to its monosemic nature in their L1. Unlike familiarity, which appears to predict faster processing across both native and non-native populations, metaphor conventionality, as follows from a number of studies on the topic (e.g., Jankowiak et al., 2017; Wang & Jankowiak, 2023), does not facilitate processing quantitatively for the non-native population. Behavioural findings, such as reaction times, indicated that processing metaphors in an L2, whether novel or conventional, requires

consistently more cognitive effort than processing literal sentences, suggesting that even proficient L2 speakers exhibit lower sensitivity to the degree of metaphor conventionality compared to L1 speakers (see Figure 8 and 9).

Figure 8

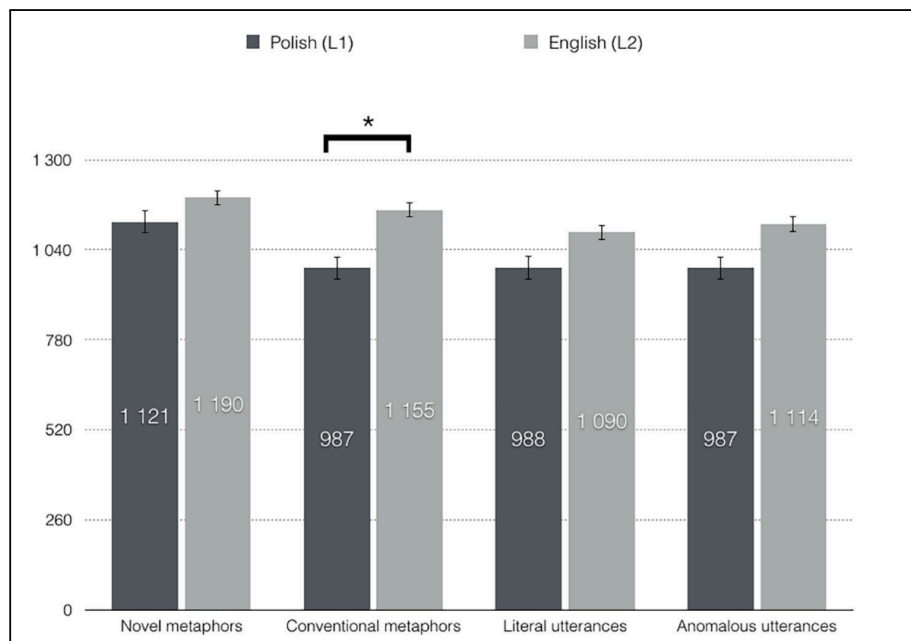
Reaction times (ms) to novel metaphors, conventional metaphors, literal utterances, and anomalous utterances among intermediate and advanced L2 speakers (Wang & Jankowiak, 2023)



As shown in Figure 8, there is no statistically significant difference between the mean reaction times for novel and conventional metaphors for both intermediate and advanced speakers of English as a second language (Wang & Jankowiak, 2023). In contrast, Figure 9, which compares native and non-native (ESL) populations, reveals that a statistically significant difference between novel and conventional metaphors emerges only for the L1 group, while no such difference is observed for the L2 speakers (Jankowiak et al., 2017).

Figure 9

Reaction times (ms) to novel metaphoric, conventional metaphoric, literal, and anomalous word dyads in Polish (dark grey) and English (light grey) (Jankowiak et al., 2017)



2.2 Literal and nonliteral routes to metaphoric meaning

Now that we have reviewed the three major MC models, it is pertinent to highlight that these are the classical metaphor processing models that were originally developed and since then have mostly been used in studies of monolinguals focusing on figurative language comprehension. However, the question of whether the literal meaning of a metaphor is always activated during processing, and whether this process differs between native and non-native speakers, remains an open issue. In addition to the classical models, two prominent approaches offer competing views on the matter.

The Direct Access Model, or *The Interactionist View*, puts forward the view that the nonliteral interpretation of a metaphoric statement may well be accessed

directly without having to first access the literal interpretation (Glucksberg, 2001). *The Indirect Processing Model* (Searle, 1979), quite in contrast, suggests that the computation of the literal interpretation is compulsory at the initial stage for it to be rejected for a certain reason so that a nonliteral interpretation might be constructed (Heredia & Muñoz, 2015). Both models, however, found it difficult to account for the empirical data collected by means of a cross-modal lexical decision task where the literal meaning was observed to remain activated even 300 ms after the offset of the metaphor regardless of familiarity and aptness (Blasko & Connine, 1993).

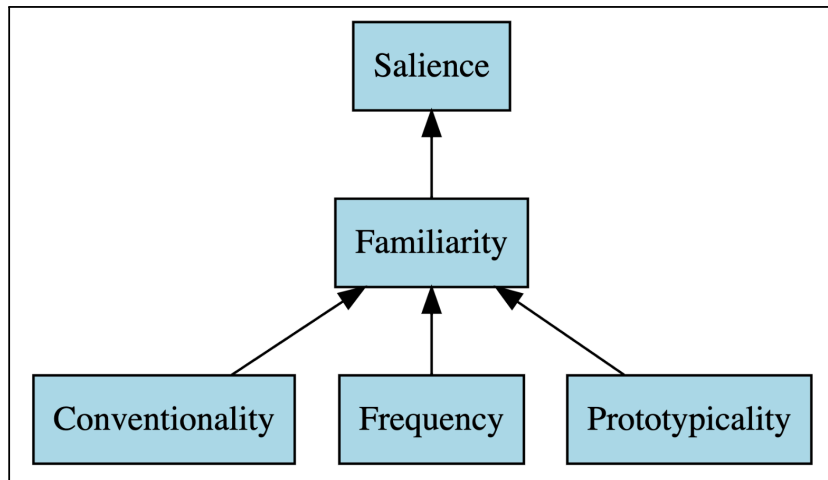
(1) *The belief that hard work is a ladder[1] is common [2] to this generation.*

(Blasko & Connine, 1993)

According to the *Indirect Processing Model*, the literal meaning is supposed to be inactive by position 2 [2] which is not the case, since in both positions the literal meaning remains active¹⁸, thereby pointing to the limitations of both models. To resolve this stalemate, a third model was introduced by Giora (2003) – the *Graded Salience Hypothesis*. Giora treats salience as a crucial factor in initial language comprehension, as a continuum index that is simultaneously modulated by “word frequency, familiarity, conventionality, and prototypicality” (p. 93).

¹⁸ In this study, while reading sentences, at specific positions within the sentences, participants were asked to perform a lexical decision task. In this task, they had to determine whether a string of letters presented to them was a real word or a nonword. The word targets they responded to were either related to the literal sense of the metaphor or the nonliteral sense, along with respective control words (Blasko & Connine, 1993). This setup allowed the researchers to assess how the activation of the literal and nonliteral meanings of metaphors unfolded.

Figure 10
Factors Constituting Semantic Salience



In particular, Giora attaches more significance to *familiarity* (see figure 10) as she does not treat the other factors – frequency, conventionality, and prototypicality – to consistently and necessarily predict familiarity (Gernsbacher, 1984; Wiley & Rayner, 2000). While conventionality, frequency, and prototypicality are often associated with increased salience, their effects are not automatic or direct. Instead, their contribution to salience is primarily mediated through familiarity. These factors enhance salience insofar as they increase the familiarity of a concept. However, if conventionality, frequency, or prototypicality are high but the concept remains unfamiliar, their ability to boost salience may be significantly diminished. In this sense, familiarity serves as a necessary conduit: without it, even highly conventional, frequent, or prototypical items may remain relatively non-salient. For example, a word may well be frequent but unfamiliar (Gernsbacher, 1984), and infrequent but familiar, as it is the case with obscene words, which are rarely used but are often high in familiarity (Wiley & Rayner, 2000). Similarly, although there are certain well-established (conventional) ways of referring to everyday phenomena (e.g. *it's*

raining cats and dogs), people do not necessarily use those frequently in their speech (Brown & Levinson, 1987). In this relation, Giora maintains that salient semantic units – those that through continuous use and frequent exposure appear ‘foremost on one’s mind’ and coded in one’s mental lexicon – are accessed faster than less salient ones, for the latter, as a rule, require additional effort to conduct inferential processing (Giora, 2003).

During real-time processing of unfamiliar metaphors, the literal meanings must often be revisited first, as they are more salient at that moment. This observation brings us to the logical premise that equally salient familiar metaphoric utterances, whose literal and nonliteral meanings are salient, are processed similarly at the early meaning activation stage. Conventional and familiar meanings are, in this way, automatically accessed, regardless of context. Primarily coming into play postlexically¹⁹, contextual information, in turn, is an effective tool for accelerated activation of contextually appropriate meanings, but is less powerful in helping to suppress irrelevant but salient meanings (Giora, 2003). More specifically, when the automatically accessed salient meanings are categorised as inappropriate, they are adjusted or redressed to fit the contextual information and authorial intent. This adjustment process occurs after the initial retrieval of meanings, making it post-lexical. In this regard, we should conclude that literal meanings even in a metaphorically priming context will still have a lower activation threshold if they are salient in the speaker’s mental lexicon. Here, by the term *activation threshold* I imply the level of cognitive effort or stimulation required to bring a particular piece of information or concept into conscious awareness or use. Lower activation thresholds

¹⁹ Studies employing *cross-modal lexical priming task* reveal that initial activation of meaning appears to occur automatically and independently of context, as both contextually appropriate and inappropriate meanings are accessed immediately upon encountering an ambiguous word, i.e., at zero interstimulus interval (Onifer & Swinney, 1981). Contextual effects emerge slightly later, selectively suppressing the irrelevant meaning while retaining the appropriate one (Giora, 2003).

indicate that the information is more easily and quickly retrieved, often because it is more familiar or frequently used (ibid.). Conversely, higher activation thresholds mean that more cognitive effort is needed to access the information, typically because it is less familiar or less frequently used. For instance, in a passage like "*The idea slithered through his mind before he could catch it,*" the context primes a metaphorical interpretation of "slithered" (*being elusive*). But the literal meaning of "slithered," (*moving in a smooth and winding way like a snake*), may still have a lower activation threshold and be quickly accessed if that meaning is particularly salient for the speaker.

Other studies that focus specifically on bilingual metaphor processing, such as the study by Johnson and Rosano (1993), show that bilinguals, in a manner similar to monolinguals, have direct access to the figurative "representation of metaphoric referential descriptions" particularly when they are highly proficient in their L2 or even dominant in it. It is still important to bear in mind that both literal and nonliteral meanings are equally accessible for activation (Heredia & Muñoz, 2015; Glucksberg, 2001; Onishi & Murphy, 1993). These observations appear of particular interest and relevance in our discussion of sustained metaphor comprehension which I will elaborate on in the next section.

2.3 Extended metaphor comprehension

As opposed to research in single nominal metaphor processing, sustained metaphors have rarely been investigated empirically. Crucially, sustained or extended metaphors differ from single metaphors in that they represent not a bicomponent figurative structure (e.g., *Men are wolves*) where *men* serves as a topic and *wolves* – as a vehicle, but a similar structure which is normally preceded by a

few more elements. I will borrow an illustration from a paper by Ronderos and Falkum (2023, p. 3) to clarify what an extended metaphor looks like:

*John doesn't like physical contact, and even his girlfriend finds it difficult to come close to him. She feels **pricked** by his **thorny** attitude every time he sees her. **John is a cactus.***

Here, apart from the single nominal metaphor *John is a cactus*, two more elements are added with the intention of creating a single enhanced image, that is, an image consisting of several interconnected and complementary elements (possibly also metaphors themselves, but not in this case) that enhance the coherence and internal consistency of the image by reconnecting all the two conceptual domains and ensuring their parallel functioning (Arnold, 2002). In this particular case, as well as in the case of all presently available research in the area of extended metaphor processing, we deal with additional elements complementing the primary vehicle component of the metaphor (*pricked* and *thorny* initially serve as relevant features of *cactus*, and only through analogical or categorical alignment are projected onto the topic [*John*]). Thus, I will consider those as additional vehicle elements. An expected question arises then: what will be the effect of such metaphor extension on the processing pattern?

2.3.1 'The lingering of the literal' and 'the lingering of the linguistic form'

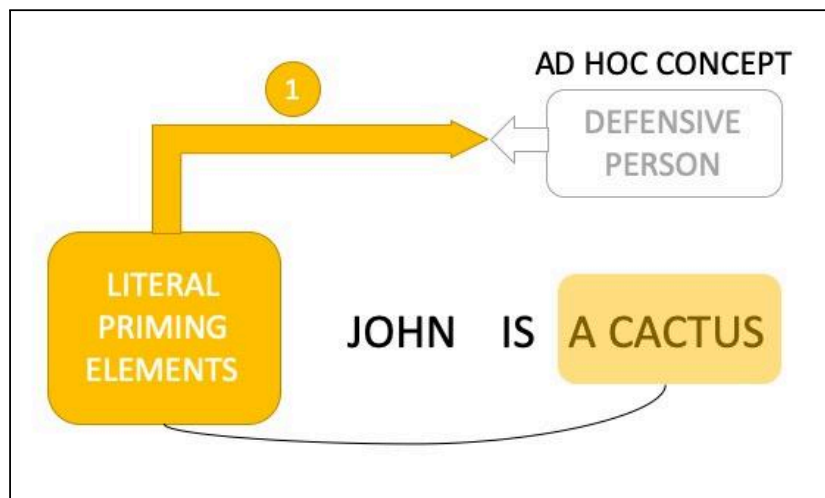
Carston (2010) put forward two different views on how the processing of such metaphorical utterances may proceed. She started to develop her theoretical framework within the Relevance Theory approach (Sperber & Wilson 1986, 1995, 2008; Carston 2002), according to which the interpretation of single metaphors

normally involves the adjustment of the encoded meaning by creating an *ad hoc concept* that is both broader and narrower than the encoded meaning at the same time (Wilson & Carston, 2006, 2007). For instance, in the metaphor “David is a rock,” the vehicle-related ad hoc concept “rock” broadens the meaning by encompassing the idea of strength, stability, and reliability that David embodies, thereby extending beyond the literal meaning of a physical rock. However, it also narrows the meaning by excluding other aspects of David’s personality or traits that may not align with the metaphorical image of a rock, such as his emotional vulnerability or creativity. In essence, this process resembles one in *the Categorization Model* where an abstract category is created from the literal base of the vehicle concept. As part of this modulation process, the relevant features become promoted, whereas others, irrelevant ones, become suppressed, the relevance index being determined by the topic of the metaphor.

In the case of extended metaphors, however, this processing model appears to be unsuccessful (Carston, 2010), as the preceding complementary vehicle elements amplify the vehicle’s literal meaning, creating a strong literal priming effect that interferes with the intended figurative interpretation (see Figure 11). As can be seen in Figure 11, the priming effect of the preceding literal elements precludes the formation and access of the ad hoc concept, thereby resulting in the literal interpretation of the whole utterance.

Figure 11

Proposed Novel Extended Metaphor Processing Model



In this connection, Carston argues, based on her theoretical framework, that relying on the relevance-theoretic ad hoc concept construction route would place an excessive cognitive burden on the comprehender (*ibid.*). Instead, comprehenders, in her view, are more likely to retain the initially activated *literal* meaning for the whole length of the comprehension process and only apply the *nonliteral* interpretation after the comprehension is complete. It is worth noting that the shift to the second mode occurs once a processing threshold is crossed — that is, when the cognitive effort required for constructing local ad hoc concepts becomes too high compared to the strong dominance and accessibility of the literal meaning (*ibid.*). This prolonged literal meaning activation is referred to by Carston as “lingering of the literal” account (*ibid.*, p. 305). Interestingly, this theoretical account predicts a comprehension time advantage (such as in reaction times) for the extended metaphor processing mode, as it is expected to be read similarly to a literal statement. As a result, it should be processed more quickly since it bypasses the lexical modulation stage. This processing pattern, as indicated by calculating the priming time of the

metaphor-related and literal-related targets, deviates from what is typically observed in single metaphor comprehension. From the results of a lexical decision task (priming paradigm), it is evident that in single metaphor comprehension the literal meaning is active at metaphor offset but diminishes in prominence and is replaced by the nonliteral interpretation by 1000 ms after metaphor offset, at least among L1 speakers (Ronderos & Falkum, 2023). Therefore, this model proposes a qualitatively different account of extended metaphor processing with a concomitant quantitative difference (ibid.).

Another theoretical perspective on this account suggests that extended metaphor processing is indeed distinct from single metaphors. However, instead of attributing this difference to “the lingering of the literal” (a *qualitative* and *quantitative* difference that fundamentally revises the processing model), it is attributed to “the lingering of the linguistic form” which is a *purely quantitative* difference (Wilson, 2018, p. 10). In this way, Wilson views faster processing times in extended metaphor comprehension as a result of low-level semantic priming which may be an effect of cumulative sequential same-domain activation of metaphor vehicles (e.g., *She feels pricked^a by his thorny^a attitude every time he sees her. John is a cactus^a*). In this context, “low-level” priming refers to the activation of related semantic information at a subconscious level. It suggests that the processing of extended metaphors may be facilitated by the activation of previously understood metaphors, which primes the semantic network and accelerates subsequent processing. For example, the mention of “pricked” and “thorny” in a previous context primes the semantic network, making the word “cactus” more quickly accessible in the subsequent context. This priming effect facilitates the comprehension of the extended metaphor by making related concepts more readily available in memory. Nonetheless, from Wilson’s perspective,

this is a mere difference in the degree of activation and not in the kind of processing (Ronderos & Falkum, 2023). While the predicted quantitative advantage for extended metaphors has been confirmed empirically (e.g., Keysar et al., 2000; Thibodeau and Durgin, 2008), studies that have directly investigated the possibility of a qualitative difference have found evidence for one (Ronderos & Falkum, 2023). This challenges Wilson's view and suggests that extended metaphor processing may involve qualitatively distinct mechanisms. In the following section, I will examine the major empirical studies that probe into the nature of these mechanisms (Rubio-Fernández et al., 2016; Ronderos & Falkum, 2023) and explore how the dynamics of literal and nonliteral meaning activation may differ depending on the structure and complexity of the metaphor. This discussion will help us build a fuller picture of how extended metaphors are processed and why they might challenge existing models of metaphor comprehension.

2.3.2 Empirical evidence for Carston's account: single & extended metaphors are processed in quantitatively and qualitatively different ways

2.3.2.1 An eye-tracking study by Rubio-Fernández et al.

In their research paper, Rubio-Fernández et al. (2016) set it as an objective for themselves to test Carston's hypothesis by conducting three experiments: self-paced reading, eye-tracking during natural reading, and cued recall. Sixteen English-speaking monolingual undergraduate students participated in the experiment. The materials included 21 extended metaphors and 21 single metaphors, with extended metaphors being presented in figurative passages (a single metaphor and a preceding figurative context) and single metaphors incorporated in literal contexts, as well as literal filler passages (see Figure 12).

Figure 12

Examples of Experimental Stimuli (Rubio-Ferdnández et al., 2016)

Extended metaphor: The school's pupils were notoriously unruly. The place was a zoo. New keepers were generally assigned to the safer animals: the half-domesticated goats, the placid giraffes, the comical penguins. Those animals you could handle **without protective gloves**. It was both the more intelligent and the more aggressive animals that required extra safety measures.

Single metaphor: The school's pupils were notoriously unruly. New teachers were generally assigned to the less troublesome classes: the younger kids, the more settled ones, and the more relaxed ones. They seldom made a lot of trouble, and you could work with them **without protective gloves**. It was both the more intelligent and the more aggressive pupils that required specialist care.

Literal control: The animals at the sanctuary had come from many different places. Some of them had been pets of owners who had died or had to move away. They were usually docile, and you could handle them **without protective gloves**. Other animals had been abused, or even trained to attack. They required special safety measures.

As part of the self-paced reading task, the participants were visually presented with various passages, one at a time, and were supposed to press the spacebar to move the text forward. Then, they took a short memory test²⁰, of which they had been informed in advance.

In the first experiment (self-paced reading), mean reading times across conditions were compared. Follow-up pairwise comparisons showed that there was a significant difference in the response times between literal and single metaphor passages (coefficient = 114.7, $SE = 28.6$, $t = 4.00$, $p < 0.001$), but no significant difference between passages with extended and single metaphors (coefficient = 38.4, $SE = 29.7$, $t = 1.30$, *n.s.*). According to Rubio-Ferdnández et al., “self-paced reading times might not be sensitive enough to reveal nuanced differences in the processing of single and extended metaphors” (p. 21).

Eye-tracking data, however, revealed a different picture – one that is more supportive of Carston's viewpoint. First fixation²¹ time results are indicative of longer fixations in a single-metaphor condition (see Table 1, first fixation: single-extended, p

²⁰ Since self-paced reading measures processing speed and comprehension, the memory test was crucial to ensure that participants actually engage with the content rather than just press the spacebar quickly without reading.

²¹ *First fixation duration* – the sum of all initial fixations in the critical region (i.e. the region of the passage which contained the metaphor target)

= 0,05), suggesting that single metaphors take significantly longer to be read compared to extended metaphors at least at the initial processing stage. Additionally, first fixation times for the extended metaphor and literal conditions were comparable. Regarding the later processing stage (i.e. first pass time²² and second pass time²³), the difference between the two metaphorical conditions became no longer significant, thus indicating a convergence in processing types later in time (see Figure 13, first pass time & second pass time: single-extended, $p > 0,05$). This suggests that at the earliest moment of comprehension, readers were likely engaging with the literal meanings of the extended metaphors, treating them much like literal language. However, as processing continued, differences began to emerge: at later stages (first pass time and second pass time), the processing times for extended metaphors started to approach those observed for single metaphors, which were consistently longer. This shift likely reflects the construction of non-literal interpretations, whereby readers gradually moved beyond the literal meanings activated initially and began integrating the metaphorical mappings required for full understanding. These observations are in line with Carston's predictions and evidence that the non-literal interpretation is constructed only later in time (Carston, 2010), with the primed literal meaning arguably facilitating early comprehension processes (Rubio-Ferdnández et al., 2016).

²² *First pass time* – the total time spent on fixations in the critical region before shifting attention beyond the critical region (including regressions back to earlier parts of the sentence before proceeding)

²³ *Second pass time* – the combined time spent on fixations in the whole passage after initially shifting attention beyond the critical region

Table 1

Pairwise t Values and p Values from Mixed Effects Regression (Rubio-Ferdnández et al., 2016)

Pairwise <i>t</i> values and <i>p</i> values from mixed effects regression.						
	Single-Literal		Single-Extended		Extended-Literal	
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
First fixation (ms)	1.58	0.12	1.96	0.05	-0.12	0.9
First pass time (ms)	0.99	0.33	0.28	0.78	0.67	0.51
Second pass time (ms)	1.39	0.17	1.22	0.23	0.18	0.86
Total time (ms)	3.00	0.007	1.68	0.09	2.2	0.03
Spillover first fixation (ms)	1.9	0.06	0.92	0.37	0.9	0.38
	Single-Literal		Single-Extended		Extended-Literal	
	<i>z</i>	<i>p</i>	<i>z</i>	<i>p</i>	<i>z</i>	<i>p</i>
Regressions in (%)	0.12	0.9	-1.75	0.08	-1.9	0.06
Regressions out (%)	-0.07	0.95	0.25	0.8	0.35	0.73

In the final experiment, the researchers seek to examine how processing disparities impact the recall of single versus extended metaphors in the final experiment (cued recall task). The task involved reading a passage containing either single or extended metaphors and then performing a distractor task to prevent short-term memory recall. Participants were then given a prompt related to the metaphor and asked to write down the first word that came to mind. The recall rates were measured by the number of correct target expressions participants remembered. If single metaphors entail local situational pragmatic shifts in lexical meaning (Sperber and Wilson, 2008), while extended metaphors do not, it is hypothesized that participants will recall target expressions more effectively in the single-metaphor condition than in the extended metaphor condition (Rubio-Ferdnández et al., 2016). By local pragmatic shift, we understand the process of lexical modulation whereby the literal base of the vehicle element of the metaphor transitions into a figurative category through generalization and specification of the encoded meaning (Wilson and Carston, 2006; Carston, 2010). In Carston's

terminology, this is known as the formation of an ad hoc concept. This shift involves a contextually driven change in meaning, specific to the instance of metaphorical use, which allows the denotation of the lexically encoded concept to acquire a more abstract, broader, and figurative sense within that context. For example, consider the metaphorical use of *bridge* in the sentence, “*Education is a bridge to success*”. In its literal sense, *bridge* denotes a physical structure designed to allow passage over an obstacle, like a river or a gap. In this metaphorical use, however, the meaning of *bridge* undergoes a contextually driven shift: it becomes *broader* because it extends beyond physical structures to include any means of overcoming an obstacle or achieving a goal. It also becomes *more abstract* because it no longer refers to a concrete, tangible object but instead to an intangible process — the role education plays in moving someone toward success (e.g., a facilitating linking element or means of transition). In this way, within the specific context of the utterance, the lexically encoded concept *bridge* is modulated to create a new ad hoc concept that captures the idea of facilitating progress or transition in a non-physical, figurative domain.

This mechanism of meaning modulation may help explain the pattern of results observed in the cued recall task. Specifically, cued recall task indicators showed better recall in single-metaphor contexts ($M = 6.5$; $SD = 2.3$) relative to the extended-metaphor context ($M = 5.2$, $SD = 2.5$), thereby pointing to a potential difference in processing models ($t = 3.10$, $df = 27$, $p < 0.01$). From this, Rubio-Ferdnández et al. deduced that these comparably higher recall rates for single metaphors might signal that there is indeed some qualitative difference in the way extended are processed, that is, with considerable reliance on the literal meaning, which reduces the distinctiveness of the metaphor in the mind of the comprehender,

making it less memorable (Rubio-Ferdnández et al., 2016: experiment 3). By measuring quantitative recall data it became possible to infer differences in memorability, and, hence, make conclusions regarding metaphor distinctiveness and the cognitive operations that had a positive effect on the participants' memory. As previously discussed, examples of such cognitive operations are local pragmatic adjustments that the reader/listener has to make as part of the lexical modulation process.

2.3.2.2 A study by Ronderos and Falkum: a cross-modal lexical priming paradigm

Another recent attempt to investigate the issue of extended metaphor processing, which also produced supportive evidence for Carston's "lingering of the literal" account, was a paper by Ronderos and Falkum (2023). Their initial research goal was to test Carston's hypothesis and to observe more closely how the suppression of the literal meaning was manifested at different stages of meaning processing if it was at all. The design of the experiment was similar to that of Rubio-Ferdnández et al. (2016) with minor modifications. The changes to the experimental design included presenting both primes and targets visually instead of using a cross-modal paradigm, using a subset of items to shorten the experiment, and reducing the number of Inter-Stimulus Intervals (ISI) tested from three to two to minimize experimental conditions. Different ISIs are crucial because they allow researchers to capture how the time available between prime and target presentation influences meaning activation. In the original experiment, three ISIs were tested: 0 ms, 400 ms, and 1000 ms. Ronderos and Falkm used only two: 0 ms and 1000 ms. So by doing, they simplified the experimental design and focused on the key intervals of interest, as two ISIs are sufficient for capturing the trends in meaning

activation over time and allowing for a clear distinction between *initial*, automatic activation and *later*, contextually mediated processing. An example of a critical trial can be seen below (see Figure 13).

Figure 13

Example of a Critical Trial in All Conditions (Ronderos & Falkum, 2023)

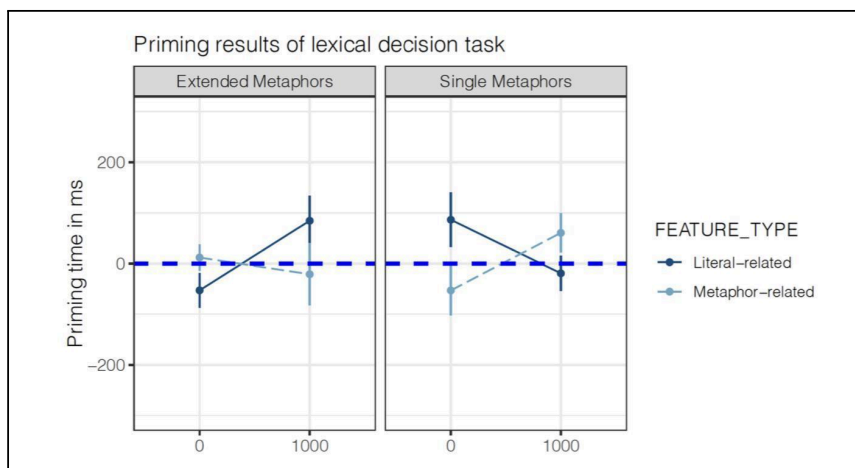
Factor 1: CONTEXT, level: 'mismatch'		Factor 1: CONTEXT, level: 'match'	
Factor 2: METAPHOR TYPE, level: 'Single Metaphor'		Factor 2: METAPHOR TYPE, level: 'Single Metaphor'	
<i>John loved paddling his canoe through the steep canyon and enjoyed rolling over in the white water of the rapids. The river poured down cascading in great amounts.</i>		<i>John doesn't like physical contact, and even his girlfriend finds it difficult to come close to him. She feels rejected by his distant attitude every time he sees her.</i>	
Factor 2: METAPHOR TYPE, level: 'Extended Metaphor'		Factor 2: METAPHOR TYPE, level: 'Extended Metaphor'	
<i>John loved paddling his canoe through the steep canyon and enjoyed rolling over in the white water of the rapids. The river poured down cascading in a foaming fizz.</i>		<i>John doesn't like physical contact, and even his girlfriend finds it difficult to come close to him. She feels pricked by his thorny attitude every time he sees her.</i>	
Metaphor prime		Metaphor prime	
<i>The river was champagne.</i>		<i>John is a cactus.</i>	
Factor 3: Inter-Stimulus Interval (ISI)		Factor 3: Inter-Stimulus Interval (ISI)	
0 Milliseconds	1000 Milliseconds	0 Milliseconds	1000 Milliseconds
Factor 4: FEATURE TYPE		Factor 4: FEATURE TYPE	
Target Word		Target Word	
Metaphor-related	Literal-related	Metaphor-related	Literal-related
<i>Spike</i>	<i>Plant</i>	<i>Spike</i>	<i>Plant</i>

The results of the analysis, counter to the relevance-theoretic view, appear to illustrate once again that the case of extended metaphor processing is noticeably different from that of single-metaphor contexts. By calculating the priming time in milliseconds for each target, literal-related and metaphor-related, Ronderos and Falkum (2023) report that literal-related features of sustained metaphors remain highly active even 1000 ms after the metaphor offset, sometimes just as active as the metaphoric meaning which is supposedly constructed by that time (see Figure 14: extended metaphors, priming time of literal-related and metaphor related features after 1000 ms interstimulus interval). As shown in Figure 14, the trends in the activation of the metaphor-related meaning differ notably between conditions. For single metaphors, activation of the metaphor-related meaning increases from 0 ms to

1000 ms, suggesting a construction of the nonliteral interpretation over time. In contrast, for extended metaphors, metaphor-related activation decreases over the same interval, indicating the absence of ad hoc concept construction. Meanwhile, activation trends for the literal-related meaning follow the opposite pattern, with a decrease over time for single metaphors and an increase for extended metaphors. It is, in this respect, concluded by the researchers that comprehenders are apparently unable to suppress the promoted literal meaning in extended metaphors, which may well serve as a substantial reason to consider Craston's model to be a highly plausible one, reflecting qualitatively dissimilar processing patterns in prolonged metaphorical utterances.

Figure 14

Activation Pattern of Literal-related and Metaphor-related Features (Ronderos & Falkum, 2023, p. 7)



2.4 Predictions about extended metaphor processing: bilingualism and anticipated effects of conventionality

2.4.1 Bilingual metaphor processing

Mastering figurative language, and metaphor in particular, is a central part of achieving communicative competence in a second language. However, research has repeatedly shown that for L2 speakers, metaphors can present persistent challenges, even when they are highly conventionalized within the target language (Bolognesi & Werkmann Horvat, 2022). Studies by Littlemore (2001) and Littlemore et al. (2011) reveal that metaphoric expressions, despite their frequency in everyday English, are often misunderstood by learners. In academic contexts especially, misinterpretations of seemingly straightforward metaphors, such as *skirting around an issue*, can disrupt comprehension—sometimes without learners even realizing a misunderstanding has occurred (Bolognesi & Werkmann Horvat, 2022). This raises important questions about the relationship between metaphorical competence and broader language proficiency. While it might seem intuitive to assume that better general L2 skills would lead to better metaphor comprehension, findings suggest otherwise. Evidence indicates that an individual's ability to use and interpret metaphors in their L2 is closely tied to their metaphorical competence in their L1 (Johnson, 1996; Littlemore, 2010). That is, a solid foundation in figurative thinking in one's native language appears to transfer to the second language, even when overall L2 proficiency is limited.

Building on these insights about the cross-linguistic foundations of metaphorical competence, it becomes natural to ask how bilinguals process more complex forms of figurative language, such as extended metaphors. Undoubtedly, the theoretical and empirical accounts investigated in the work of Rubio-Ferdnández

et al. (2016) and Ronderos & Falkum (2023) that were discussed in the previous section give rise to many questions, including questions about bilingual extended metaphor processing – a domain of figurative language processing research that has not been studied yet, to the best of my knowledge. However, it seems possible to build predictions and test hypotheses about what processing path a bilingual mind is likely to take when exposed to a sustained metaphor context.

Existing studies in bilingual single metaphor processing reveal that bilinguals indeed, depending on their level of language proficiency and language dominance, might at times be able to access the non-literal interpretations of metaphoric utterances directly. Heredia and Muñoz (2015) observed highly proficient bilinguals to be more likely to process the intended meaning of metaphors early in comprehension, with the literal meaning remaining “a possibility for bilinguals even 1000 ms after they have accurately resolved the linguistic ambiguity” (p. 103); that is, even after participants accurately discerned the intended metaphoric meaning. This is, however, a different pattern from the one found in monolinguals. A group of L1-dominant bilinguals (Spanish-English bilinguals) were found to have activated both the literal and the non-literal (intended) meaning 300 ms after the offset of the metaphor reference. Clearly, what we observe is a growing prominence of the literal meaning and, consequently, automatic access to that meaning in bilinguals which negatively correlates with language dominance (self-rated through Language History Questionnaire for Bilingual Participants) and, potentially, language proficiency. When bilinguals exhibit lower proficiency or dominance in a language, they often show stronger automatic access to the literal meaning. Bilingual competence involves two closely related but distinct concepts: *language dominance* and *proficiency* (Montrul, 2015). Language dominance refers to the pattern in which one of a bilingual's

languages is used more frequently in certain contexts and is typically processed with greater ease (ibid.). Proficiency, on the other hand, concerns the level of grammatical knowledge and fluency an individual has in a language (ibid.). While research often treats these two notions as interchangeable, they are not the same. Proficiency can be considered one component of language dominance, but the two are not fully equivalent — high proficiency often accompanies dominance, yet dominance encompasses more than just linguistic skill (ibid.). Heredia and Muñoz (2015) consider language dominance as well as language proficiency to be factors that condition *salience*²⁴ – an important modulator of metaphoric processing that, as has been discussed previously, is characterized by individual variation and primarily rests on meaning familiarity. Crucially, it is a reflection of a certain meaning in the mental lexicon of the comprehender that is subject to individual variability. If bilinguals are more dominant in their L2, they are likely to access the non-literal interpretations faster due to their considerable salience. Their general language proficiency and more frequent exposure to L2, which certainly varies from one individual to another, results in their increased *metaphoric fluency* – that is, their ability to come up with as many possible interpretations of a metaphoric utterance as possible. And this is exactly the type of ability that appears most necessary when dealing with metaphors that are novel and unfamiliar. Thus, the general distinction between monolinguals and bilinguals would still reside in perhaps qualitatively different approaches to metaphoric processing: native speakers, largely due to their considerable linguistic experience with figurative expressions, prioritize the figurative

²⁴ It is important to bear in mind here that salient meanings are generally accessed more quickly, and salience itself is shaped by factors such as familiarity, frequency, conventionality, and prototypicality. Heredia and Muñoz's claim about the predictive value of L2 proficiency and dominance essentially arises because proficiency and dominance are closely linked to increased exposure and use, which in turn enhance familiarity, conventionality and the rest of the semantic factors. As a result, for highly proficient and L2-dominant bilinguals, lexical-semantic properties associated with salience naturally acquire greater strength. I should admit it appears quite difficult to disentangle the effects of language proficiency and dominance from the inherent lexical-semantic features that contribute to salience.

meaning (which is still more salient than it is in the bilingual mind), whereas bilinguals are inclined to approach metaphors analytically, attending to the literal meaning, which holds more salience, as well as to the figurative meaning (Kecskes, 2006). This leaves bilinguals, on aggregate, with more meanings to suppress in order to arrive at the intended (figurative) meaning.

Recent models of bilingual language control highlight how the recurrent demands of managing two languages shape bilinguals' cognitive systems. According to Green and Abutalebi's Adaptive Control Hypothesis (2013), the executive control processes that bilinguals recruit during language use are not static; instead, they dynamically adapt to the demands of different interactional contexts — such as switching between languages with different speakers or code-switching within conversations. They propose that bilinguals fine-tune cognitive mechanisms like interference suppression, task disengagement, and goal maintenance depending on how often and under what conditions they use both languages. This framework is particularly relevant for metaphor processing in L2, where bilinguals must manage the competing activation of literal and figurative meanings under heightened cognitive load (Lü et al., 2019). Individual suppressing abilities in bilingual metaphor processing were studied by Lü et al. (2019), and their research revealed a significant effect of executive control abilities (measured by a Stroop Task) on familiar metaphor comprehension. Evidently, L2 learners not only have to suppress irrelevant information similar to native speakers but also must contend with factors like varied language experiences and cultural backgrounds, which heighten the demand on executive control (*ibid.*).

At first glance, this additional cognitive burden would seem to place L2 speakers at a significant disadvantage when it comes to interpreting figurative

language, particularly in cases where meanings deviate substantially from literal norms. In the context of extended metaphor comprehension, this challenge may be even more pronounced: faced with a strong priming of literal meanings and limited automatic access to conventional figurative interpretations, L2 speakers might initially default to literal readings, with few immediate cues to suggest an alternative route. Because of such a strong literal priming effect, bilingual speakers may not be able to access the non-literal interpretation given that even monolinguals struggle to suppress the literal-related features present in additional vehicle elements. There is sufficient evidence to believe that bilinguals will not resort to the relevance-theoretic model of lexical adjustment (ad hoc concept construction) (Kecskes, 2006) and will immediately proceed to the literal interpretation of the extended metaphor, just as monolinguals, presumably with no significant difference in processing pace.

However, extended metaphors also offer a potential counterbalance to these difficulties. Unlike isolated or brief metaphorical expressions, extended metaphors unfold gradually across a broader linguistic and conceptual landscape, providing repeated, mutually reinforcing cues that can scaffold the construction of nonliteral meanings. This elaboration may serve to ease the executive control load typically required for suppressing literal interpretations, allowing bilingual speakers to shift progressively toward the intended figurative meaning without relying solely on rapid, automatic inferencing. Thus, rather than uniformly exacerbating comprehension difficulties, extended metaphor contexts might, under certain conditions, mitigate them by offering cognitive support precisely where bilingual readers need it most. Interestingly enough, if both L1 and L2 speakers/listeners supposedly avoid the ad hoc concept construction stage and process the entire passage literally, L2 speakers are likely to appear in a similar environment – where they are unable to suppress

the literal-related associates in the passage – and not have any significant differences in processing compared to L1 speakers. I am hypothesising that this scenario will with great likelihood hold true for contexts when monolinguals are compared with highly proficient bilinguals. However, more empirical data is necessary to be able to make such conclusions with more confidence. My empirical investigation will probe into these questions; however, it will do so more from the perspective of quantitative measurements of processing, such as reaction times.

2.4.2 Conventinality effects on extended metaphor processing

Elaborating on the speculations of Rubio-Fernandez et al. (2016), I would argue that the effect of conventionality indeed seems to be an issue worth investigating within the framework of sustained metaphor processing. Referring back to the Career of Metaphor Model, I will investigate whether the processing model will change depending on the degree of conventionality of the extended metaphor (e.g., [*<metaphoric context> John is a cactus*] which is a novel extended metaphor, as opposed to [*<metaphoric context> Argument is war*] which is a conventional extended metaphor). As was discussed earlier, categorization-driven processing normally takes less time than processing of novel metaphorical utterances as no structural alignment needs to be established. It appears theoretically convoluted to try and develop a new model of conventional extended metaphor processing with the framework of Relevance Theory because relevance theory assumes abstraction-first processing for novel metaphors as well. Rather, for the sake of maintaining a clearer theoretical distinction between the two processing modes, I propose it might be optimal to assume that the facilitatory effect observed in the processing of extended novel metaphors can in fact be theorized as stemming from the priming of the vehicle domain (see Wilson, 2018). This perspective aligns with the notion that when

novel metaphors are encountered, both the topic and vehicle domains are activated, allowing for structural alignment to occur more fluidly. The priming of the vehicle domain can enhance the processing of the metaphor, as the features of the vehicle domain are activated and can facilitate the selection of appropriate features within that conceptual domain. This is particularly significant in novel metaphors, where the lack of established figurative categories necessitates a more active engagement with both domains in order to derive meaning.

In contrast, conventional metaphors often involve a more immediate access to established figurative categories, which can lead to quicker processing times. However, the reactivation of the vehicle domain in these cases can also play a role in metaphor comprehension. In particular, when the vehicle domain is reactivated, it may encourage a deeper structural alignment which is otherwise not required as the literal base of the vehicle domain is not attended to (Werkmann Horvat et al., 2023). For example, in a conventional metaphor like *time is money*, listeners usually access the figurative meaning directly, without consciously reflecting on the literal attributes of *money*. However, if the metaphor is embedded in a context that brings specific features of *money* back into focus — such as in *invest your time wisely or you'll go bankrupt* — the vehicle domain *money* is reactivated in a more detailed way. This reactivation can prompt the listener to engage in a richer structural mapping between the concepts of time and money, reinforcing the deeper relational parallels between the two domains rather than relying solely on the pre-established figurative shortcut that categorization-based processing offers us. This makes the metaphor more deliberate and potentially slows down processing as the listener navigates through the revitalised featural spectrum of the vehicle domain.

Within the framework of Deliberate Metaphor Theory (DMT), deliberateness distinguishes metaphors that are specifically employed to influence the reader's or listener's perspective on a subject, from those that do not aim for such an outcome (Steen, 2016). A metaphor is considered *deliberate* when it is intentionally used to prompt the listener or reader to adopt an unfamiliar perspective on a given subject, encouraging them to interpret that subject through the lens of this alternative viewpoint (Steen, 2009). In Steen's words, deliberate metaphors serve as "perspective changers," meaning they are used intentionally to offer a new or external viewpoint on a particular topic (see the example from earlier: *investⁱ your timeⁱⁱ wisely or you'll go bankruptⁱ*). By doing so, these metaphors encourage the audience to focus on the vehicle domain—the conceptual area from which the metaphor is drawn—thereby prompting them to reconsider their understanding of the topic domain (the concept or topic being discussed).

Thus, the vehicle term's prior fragmented exposure via lexically represented elements can facilitate meaning computation for novel metaphors but may also introduce complexity for conventional metaphors, as the baseline condition – a nominal metaphor in isolation – assumes different processing procedures for the two metaphor types. The change of the condition to that of an extended metaphor will likely entail a change in the type of alignment for conventional metaphors and no such fundamental change for novel metaphors (see Figures 15 and 16).

Figure 15

Processing Procedures for the Two Metaphor Types: Novel and Conventional

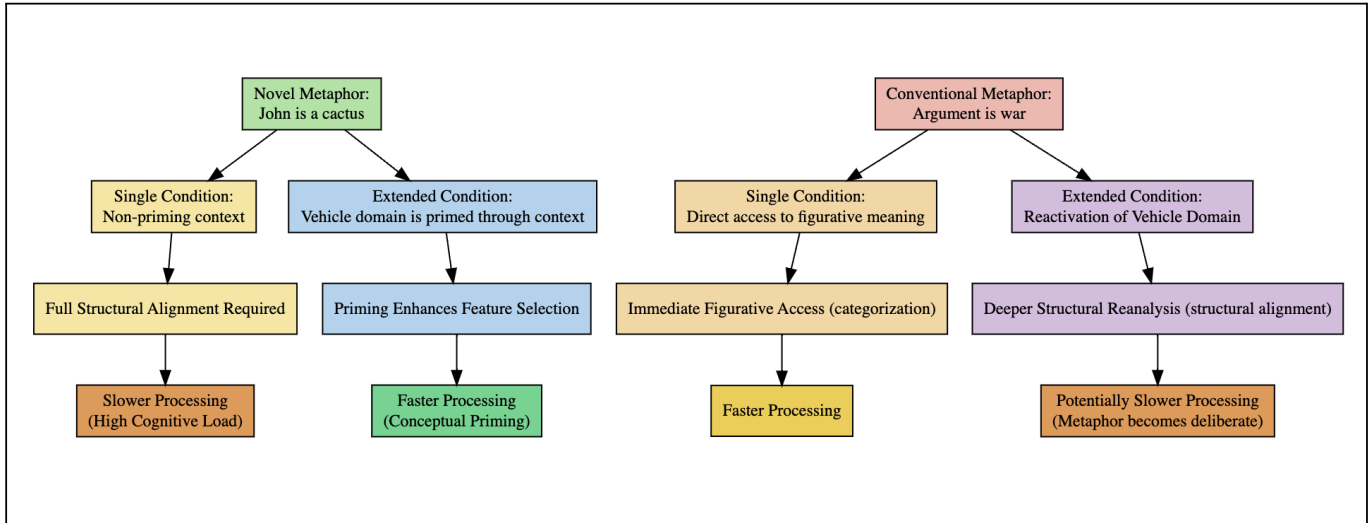
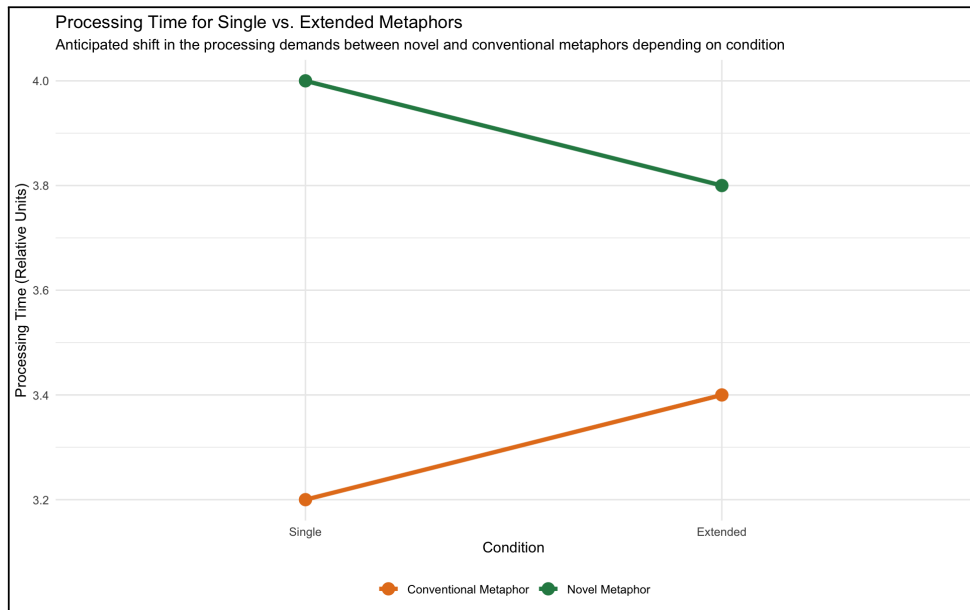


Figure 16

Anticipated Trends in Processing Times: Single and Extended Metaphors



Following this line of argument, in the case of conventional metaphors, features related to the vehicle in the context help to reactivate the literal meaning of the vehicle. This communicative function prompts the reader to reconsider their interpretation of the source domain, activating its literal meaning and engaging in a cross-domain comparison. Consequently, the contextual cues that prime the vehicle or source domain contribute to the deliberateness of the metaphor. The impact of these cues on processing demands, however, remains an open question given the scarcity of empirical studies investigating this phenomenon.

Werkmann Horvat et al. (2023), using eye-tracking and a forced-choice semantic relatedness task, attempted to examine whether introducing linguistic elements related to the vehicle domain (e.g., *father cut^t the budget like grass* which is essentially the representation of the underlying conceptual metaphor REDUCTION IS CUTTING) could shift readers' focus toward the source domain, and through this, encourage them to interpret conventional metaphors through cross-domain mapping. The results showed that when the source domain was activated, participants spent significantly more time fixating on target object (budget) during both the second reading and subsequent regressions, compared to when no source domain activation occurred, which confirmed their predictions. Longer fixations for the activated condition reflect that if the metaphor vehicle is reactivated through contextual vehicle-related cues, the reader requires more cognitive resources to process such a metaphor. This was also interpreted by the researchers as an indication of the shift to alignment-based processing, which is considered more cognitively demanding as we now know from our review of the metaphor processing models in previous sections. The study used verbal metaphors as stimuli, with the condition intended to prime the vehicle domain including only a single word that

followed the metaphor, acting as the priming element. However, I'm not entirely convinced this approach is optimal. In the reactivated condition, not only is the source domain primed, but the entire sentence is transformed into a simile (e.g., *the farther cut the budget like grass*) which is figuratively meaningful. Research on simile processing suggests that the explicit comparison in similes can hinder the retrieval of long-term memory information needed for meaning comprehension, which may be reflected in reduced N400 amplitudes for similes compared to nominal metaphors (Jankowiak et al., 2021). Also, earlier research on monolingual novel metaphor comprehension produced plenty of evidence that novel similes, by triggering comparison mechanisms, are processed more quickly and easily (Bowdle & Gentner, 2005). In that respect, changing the paradigm of stimuli presentation to avoid the potential confounding effect of the collateral activation of comparison mechanisms might have resulted in a more pronounced increase in comprehension difficulty.

2.4.3 Conventinality effects within the framework of Relevance Theory

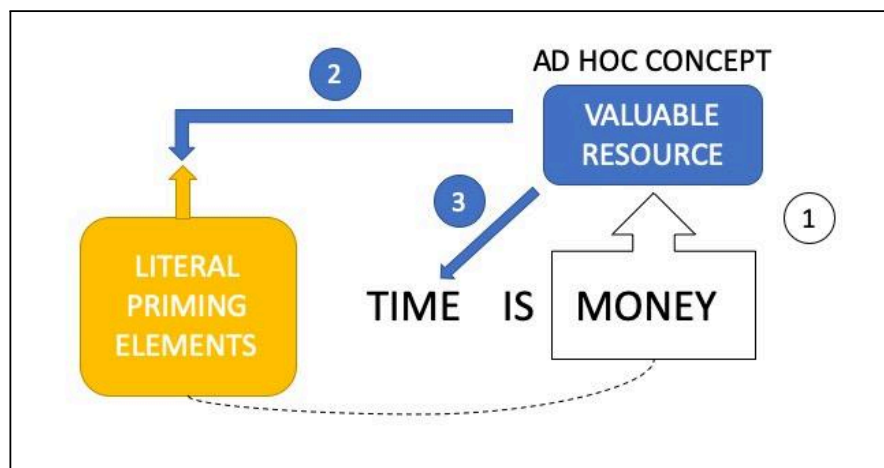
Up to this point, Relevance Theory has primarily been discussed in relation to the processing of novel metaphors, particularly in terms of the pragmatic adjustments required to infer nonliteral meanings. However, within the broader discussion of conventionality effects on extended metaphor comprehension, it becomes apparent that a relevance-theoretic perspective can also illuminate how conventional metaphors are processed when embedded in extended discourse. As we already know, Carston argues that in the case of extended metaphors, the literal meaning is not merely suppressed but remains salient, affecting the overall interpretative process (Carston, 2010). On the other hand, the structural alignment approach, as articulated in the work of Gentner and Bowdle, suggests that metaphor comprehension involves a structural mapping between the topic and vehicle

domains. This model emphasizes that metaphors are initially processed through comparisons that align the two domains, which can later evolve into more abstract categorizations as the metaphor becomes conventionalized.

If we were to hypothesise about the effects of conventionality relying on the assumptions of Relevance Theory, the necessity to construct or access an ad hoc concept will be conditioned by its prominence in the mental lexicon as it has been conventionalised overtime (Bowdle & Gentner, 2005). In the strong version of this hypothesis – that is, if we assume that conventional figurative categories will remain readily accessible even with the vehicle domain being primed by the context – this might hinder the reactivation of the literal vehicle domain of the metaphor, the entire utterance still being processed through categorization with insignificant differences²⁵ in reaction times (see Figure 17 and refer to the footnotes for a more detailed description of the processes involved).

Figure 17

*Alternative Conventional Extended Metaphor Processing Model (the strong version)*²⁶



²⁵ The differences in reaction times between two conditions: single conventional and extended conventional metaphors

²⁶ 1 – The ad hoc concept of the metaphor vehicle is accessed due to its salience.

2 – The semantic salience of the ad hoc concept overrides the priming effect of the preceding literal elements.

3 – The interpretation of the metaphor proceeds via the process of vertical alignment between the ad hoc category *valuable resource* and the metaphor topic *time*.

It also appears plausible that the preceding vehicle-related priming elements might demand additional comprehension time necessary to inhibit these features first. This model will minimize the priming effect of the literal antecedent features and will turn them into irrelevant properties, thereby triggering additional cycles of spreading activation as part of the construction-integration process (Knitsch, 2000).

2.5 Research gap

Although I have to admit that the nature of extended metaphor processing remains a very novel area of research, some progress has been made to unveil the differences in sustained metaphor comprehension relative to single nominal metaphor comprehension. Unlike single metaphors with a bicomponent structure, sustained metaphors also contain complementary elements that reinforce the literal base of one of the metaphoric elements, thus triggering the literal meaning to persist and, indeed, predominate in the comprehension process. This more complex structure questions the applicability of all previously proposed processing models that imply the activation of the nonliteral meaning at least at some stage of complex meaning comprehension (lexico-semantic access or meaning integration) both in monolinguals and bilinguals. It then follows that while the literal meaning appears prevalent because it is metarepresented as a whole, the figurative meaning should be calculated later in time and will derive from this holistic literal representation, i.e. from an accumulation of all literal vehicle-related elements present in the sentence. By *metarepresentation*, Carston refers to the conceptual representation that encompasses the literal meaning, which is not considered inherently factual. Instead, it is held for further analysis of its conceptual properties (Carston, 2010).

Currently, available empirical data do not allow us to make supported conclusions as to how exactly the construction of this figurative meaning proceeds, and this should be analyzed in further research. More insights into the specificity of posterior (delayed) figurative meaning formation may be gained through real-time analysis (eye-tracking and electrophysiological experiments) so that the delayed meaning-integration processes could be better measured and analyzed. And this is what I have every intention to do in my further research; however, it is beyond the scope of the current empirical project.

Existing studies in bilingual single metaphor comprehension revealed that language proficiency and language dominance appear to be important determinants of how quickly bilinguals can access the nonliteral interpretation (e.g., Heredia & Altarriba, 2001). Apparently, this will not be the case in sustained metaphor processing as bilinguals are likely to comprehend such metaphors just as quickly as monolinguals by only accessing the literal route. Nonetheless, my prediction would be that there might be a potential shift in the processing mode once the metaphor becomes more conventional²⁷ in society (a term more applicable to monolinguals) and more salient²⁸ (a more precise term for bilinguals). I would expect categorical processing (ad hoc concept construction or ad hoc concept access) to take over or at least to confront literal processing to the extent that both literal and nonliteral meaning will be simultaneously activated and the literal meaning will gradually recede (or become inhibited), in line with the relevance-theoretic view. In this case, bilinguals will most likely demonstrate differences in that they will struggle to

²⁷In this paper I adopt the definition of *conventional* meaning as an established meaning that is widely shared among L1 speakers of a certain language (Werkmann Horvat, 2022)

²⁸As the notion of salience is essentially hypernymic in nature because it rests on a number of other characteristics such as familiarity, conventionality, prototypicality, and frequency, its measure derives from the measurable indexes of all of these characteristics. It can be obtained by combining the scores from familiarity, conventionality, prototypicality, and frequency to create a composite salience index.

suppress the literal-related features more than monolinguals due to lack of experience with figurative language and, consequently, attenuated salience of metaphoric meanings relative to salient literal meanings.

2.6 Research Objectives

In my research I investigate the contrasts between single and extended metaphor processing involving reaction times and accuracy rates. The major theoretical foundations for my research have been developed by Carston (2010) in her “lingering of the literal” account as well as by Gentner and Bowdle’s (2005) Career of Metaphor Model. My study explores the differences between L1 and L2 extended metaphor processing, as no research has yet examined the performance of bilingual speakers when exposed to a sustained metaphor context, nor has any study so far employed conventional extended metaphors in their experimental paradigm to investigate processing demands. The research will be guided by the following research questions:

- (1) Do advanced non-native English speakers show similar processing patterns to native English speakers when comprehending novel extended metaphors?
- (2) If so, are their response times significantly different from those of native English-speaking participants?
- (3) What is the effect of conventionality in extended metaphors on processing times?

2.7 Empirical Predictions

The following empirical predictions outline the anticipated behavioral patterns in response times and accuracy rates, grounded in the body of theoretical work published so far on metaphor processing. Prior research has highlighted the role of factors such as conventionality, pragmatic context, and language nativeness in

shaping how metaphors are understood and evaluated. These predictions build on existing findings, considering how utterance structure (single vs. extended) and the conventionality of figurative categories may influence processing efficiency.

1. *Both L1 and L2 populations are expected to take longer to judge single metaphors as comprehensible compared to extended metaphors. Extended metaphors will be interpreted literally, at least initially, which will facilitate faster processing.* This prediction follows from the assumption that single metaphors offer fewer contextual cues to support interpretation, increasing inferential load, whereas extended metaphors provide scaffolding that favors a literal (and thus easier) initial interpretation (Carston, 2010; Rubio-Fernández et al., 2016).
2. *Extended novel metaphors are anticipated to elicit response times comparable to those of literal utterances, reflecting their literal interpretation.* Because extended metaphors maintain strong literal salience across their structure, participants are likely to rely on literal readings during initial processing, leading to faster response times (Rubio-Fernández, 2007).
3. *L2 participants are expected to take longer to respond to novel single metaphors compared to L1 participants. However, I expect to see no statistically significant difference between the two groups in the extended novel metaphor condition.* This prediction reflects prior findings that L2 speakers struggle more with novel metaphor interpretation due to lower figurative competence (Littlemore, 2001; Jankowiak, 2017), but extended contexts can mitigate this difficulty by promoting literal construal.
4. *Potentially, L1 participants, because they more readily retrieve conventional meanings, will process conventional metaphors faster than novel metaphors.*

It is likely that L2 participants will not have such a processing advantage and will struggle equally with single novel and single conventional metaphors, consistent with the results from previous studies (Jankowiak, 2017, 2023).

5. *For conventional metaphors, shifting from a single metaphor to an extended metaphor condition will result in slower reaction times.* This is due to the shift in processing models, likely involving a deeper structural alignment and re-engagement with the vehicle's featural spectrum compared to single metaphors, which do not require such alignment (Werkmann Horvat et al., 2023).

3. Chapter three: the current study

3.1 Experiment 1: pre-test of stimuli

3.1.1 Methods

Fifteen undergraduate and graduate students from the University of Victoria participated in the rating study. The primary goal of this rating study is to establish baseline perceptual data on a set of target stimuli by collecting participant judgments across four key semantic dimensions: comprehensibility, metaphoricity, felt familiarity, and conventionality (Jankowiak et al., 2017). By gathering these ratings, the study aims to characterize how participants perceive the target phrases. These perceptual measures will serve as a foundation for selecting and categorizing stimuli in subsequent experimental tasks, alongside helping to ensure that target phrases vary systematically along relevant semantic dimensions. All participants were native speakers of English and completed the task in a controlled setting on an iMac computer in a lab at the University of Victoria. The experiment was designed and conducted using PsychoPy v2024.2.4 – a widely used open-source software for psycholinguistic experiments.

The stimuli consisted of a set of sentences classified into four categories: 80 metaphors, 40 literal sentences, and 40 anomalous sentences. Of these, 40 metaphors, 40 literal sentences and 40 anomalous sentences were sourced from an existing normed stimulus set²⁹ (Jankowiak, 2020). However, to maintain a balanced representation of conventional and novel metaphors, an additional 40 new metaphors were introduced. Each participant was presented with all of the stimuli in a random order, one at a time, and asked to rate them according to specific criteria.

- *Comprehensibility* measured how easily a sentence could be understood. A rating of 1 indicated that the sentence was totally meaningless, making no sense and being impossible to interpret, whereas a rating of 10 indicated that the sentence was completely meaningful, requiring no effort to comprehend.
- *Metaphoricity* assessed the extent to which the presented phrase was perceived as figurative. A rating of 1 represented a very literal statement, where the meaning was conveyed directly without requiring any figurative interpretation. In contrast, a rating of 10 signified that the sentence was highly metaphorical and that some meaning adjustment was necessary in order to arrive at a possible interpretation.
- *Familiarity* captured the extent to which participants had encountered the ideas expressed in the phrase before (1 – very unfamiliar, 10 – very familiar).

²⁹ The materials used in the norming tasks included 120 novel nominal metaphors (e.g., *Viruses are travellers*), 120 novel similes (e.g., *Viruses are like travellers*), 120 literal sentences (e.g., *These people are travellers*), and 120 anomalous sentences (e.g., *Tables are travellers*). Cloze probability tests ensured that sentence-final target words were not predictable from context. Familiarity ratings assessed how often participants encountered each sentence type; metaphoricity ratings evaluated how metaphorical or literal the items were; and meaningfulness ratings judged how interpretable the sentences were (Jankowiak, 2020).

- *Conventionality* measured the strength of association between the metaphor's vehicle (i.e., the source domain of the metaphor) and its figurative meaning. A rating of 1 indicated that the metaphor was highly uncommon, meaning that the figurative association was rare or unlikely. A rating of 10, on the other hand, suggested that the metaphor was widely recognized and had a strong, well-established figurative meaning. To illustrate this concept, participants were provided with two examples:
 - ❖ *Her room is **Norway*** (Novel): When we say that something is Norway, how common is it to use this concept to convey the idea of 'a place where it is very cold'?
 - ❖ *Time is **money*** (Conventional): When we say that something is money, how common is it to use this concept to convey the idea of 'something that is valuable'?

Participants were instructed to use the provided Likert scale (1-10) to indicate their ratings for each semantic dimension. They were also informed that they could revise their answers before proceeding to the next trial. Once they had read and understood the instructions, they were asked to click anywhere on the screen to continue to the practice trials, where they could familiarize themselves with the task before beginning the main experiment. Self-paced breaks were provided every 20 trials, during which participants could pause before continuing to the next block of sentences.

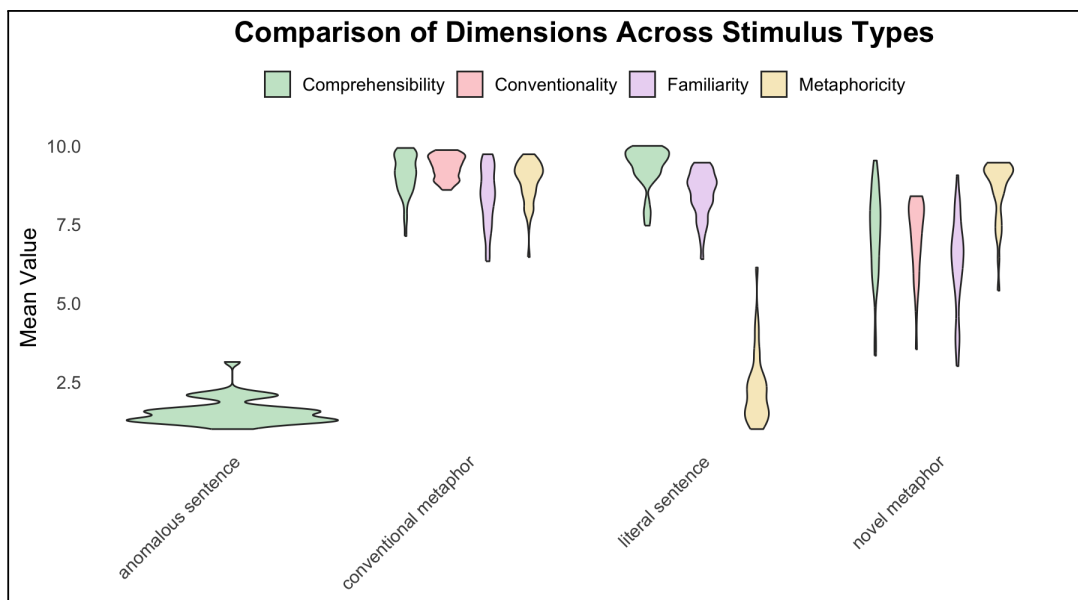
3.1.2 Results and discussion

A one-way ANOVA was conducted to examine the effect of stimulus type on various semantic dimensions, followed by Tukey's Honest Significant Difference

(HSD) test for pairwise comparisons between conditions. All statistical analyses were performed in RStudio. The ANOVA results revealed significant effects of stimulus type on all dimensions of interest (comprehensibility, metaphoricity, familiarity, and conventionality) (see Figure 18).

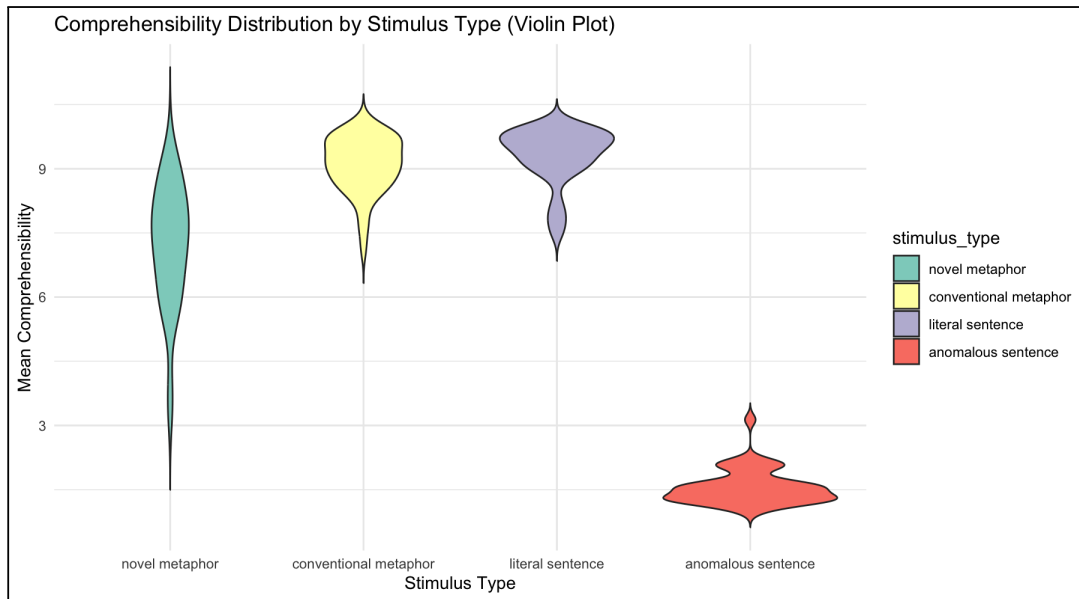
Figure 18

Comparison of Dimensions across Stimulus Types



Comprehensibility

The results from the post-hoc comparisons using Tukey's HSD test offer a fairly clear picture of how participants rated the different sentence types. Conventional metaphors stood out as significantly more comprehensible than novel metaphors, with a mean difference of 1.89 (95% CI: 1.40, 2.38, $p < 0.001$), which is not surprising since conventional metaphors are likely more familiar to most people. Likewise, literal sentences were rated significantly higher than novel metaphors, with a mean difference of 2.11 (95% CI: 1.60, 2.61, $p < 0.001$), indicating that the literal sentences were more straightforward and easier to understand (see Figure 19).

Figure 19*Comprehensibility Distribution by Stimulus Type*

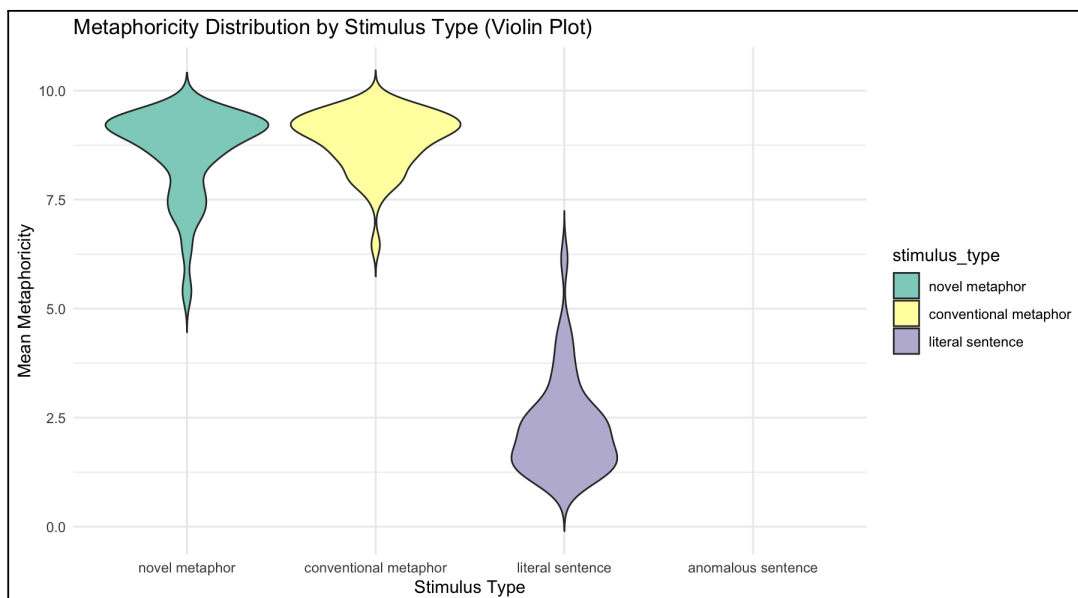
Anomalous sentences were rated much lower than novel metaphors, with a mean difference of -5.65 (95% CI: -6.16, -5.14, $p < 0.001$). This result is expected because anomalous sentences would likely be perceived as nearly entirely nonsensical. Interestingly enough, there was no significant difference between literal sentences and conventional metaphors, with a very small mean difference of 0.21 (95% CI: -0.26, 0.68, $p = 0.645$). This lack of a clear distinction could be attributed to the emerging polysemy which is characteristic of conventional metaphors. In this case, conventional metaphors may leverage both their literal and figurative meanings in a way that allows them to be processed in a similar manner to literal sentences, making them equally accessible to participants in terms of ease of understanding. Anomalous sentences were just as well rated significantly lower than both conventional metaphors (-7.54, 95% CI: -8.01, -7.07, $p < 0.001$) and literal sentences (-7.76, 95% CI: -8.24, -7.27, $p < 0.001$), confirming the anticipated idea that participants would find them particularly difficult to interpret.

Metaphoricity

When examining the metaphoricity ratings, the results indicated a clear effect of utterance type (see Figure 20). Specifically, there was no significant difference in metaphoricity between conventional metaphors and novel metaphors, with a mean difference of 0.27 (95% CI: -0.22, 0.77, $p = 0.386$). This suggests that both categories were perceived similarly in terms of their metaphorical nature and participants did not rate one as more metaphorical than the other.

Figure 20

Metaphoricity Distribution by Stimulus Type



However, when comparing literal sentences to novel metaphors, the difference was substantial, with a mean difference of -6.32 (95% CI: -6.83, -5.82, $p < 0.001$). Similarly, when comparing literal sentences to conventional metaphors, the difference was again significant, with a mean difference of -6.60 (95% CI: -7.07, -6.13, $p < 0.001$). This further emphasizes that literal sentences are consistently rated as less metaphorical than both novel and conventional metaphors, thus

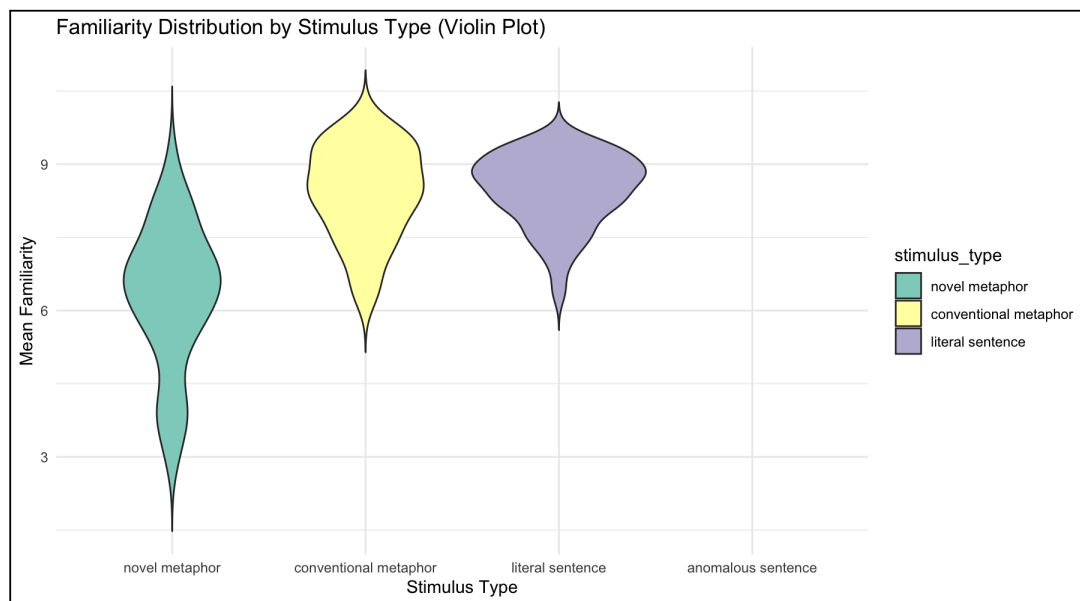
reinforcing the distinctiveness of literal language from figurative expressions independent of their level of conventionality.

Felt familiarity

The familiarity ratings, above all, revealed conventional metaphors were rated significantly more familiar than novel metaphors, with a mean difference of 2.10 (95% CI: 1.53, 2.67, $p < 0.001$) (see Figure 21). This suggests that conventional metaphors are more familiar to participants, likely because they are more established in everyday language use. Similarly, literal sentences were also rated as more familiar than novel metaphors, with a mean difference of 2.16 (95% CI: 1.57, 2.75, $p < 0.001$). Again, this suggests that participants found literal sentences to be more familiar compared to novel metaphorical expressions.

Figure 21

Familiarity Distribution by Stimulus Type



However, when comparing conventional metaphors to literal sentences, no significant difference was observed in terms of familiarity. The mean difference was just 0.06 (95% CI: -0.49, 0.60, $p = 0.965$), indicating that literal sentences and

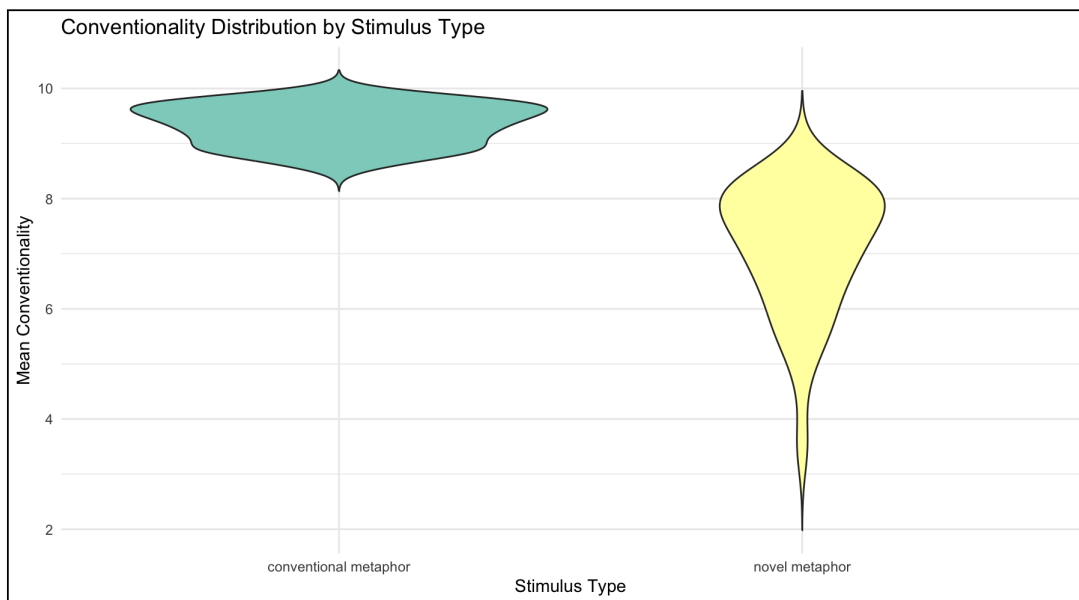
conventional metaphors were rated similarly in terms of how familiar they felt to the participants. This finding aligns with the idea that conventional metaphors, though figurative, may be so commonly used that they become almost as familiar as literal expressions.

Conventionality

The analysis of conventionality, as anticipated, revealed a significant difference between conventional and novel metaphors (see Figure 22). Conventional metaphors were rated as significantly more conventional than novel metaphors, with a mean difference of 2.22 (95% CI: 1.86, 2.59, $p < 0.001$). This indicates that conventional metaphors are generally considered more entrenched in language and that the association between the vehicle domain and the corresponding figurative category is very strong. The results support the notion that conventional metaphors, due to their frequent usage and widespread recognition, are perceived as more typical and common in comparison to novel metaphors.

Figure 22

Conventionality Distribution by Stimulus Type

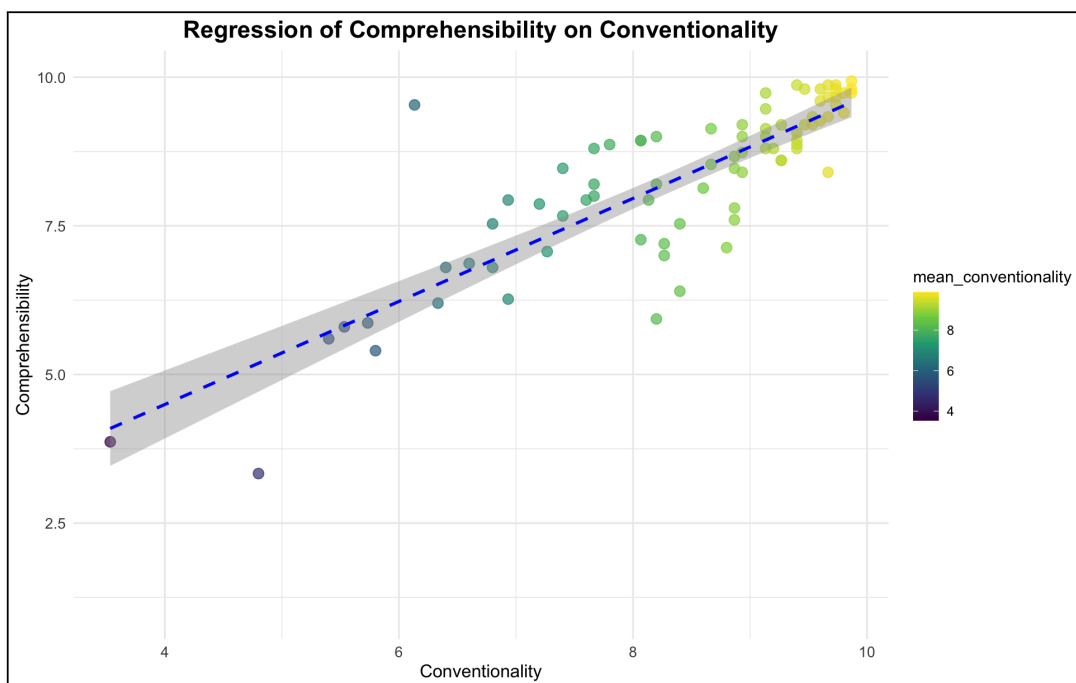


the dividing line between what is considered conventional and what is novel is not always straightforward. The purpose of the current rating study was, among other things, to address this issue by establishing a continuum of conventionality. This continuum allows us to move beyond the oversimplified dichotomy and better capture the nuanced gradations of metaphor conventionality. By creating this spectrum, we are now in a position to run linear regression analyses that can more precisely examine the relationships between conventionality and other factors, rather than relying on the arbitrary distinction between two categories.

The linear regression model had mean comprehensibility as the dependent variable, while the independent variable was mean conventionality. The model aimed to determine whether an increase in conventionality predicted an increase in comprehensibility (see Figure 24).

Figure 24

Regression of Comprehensibility on Conventionality



The regression model revealed a significant positive relationship between conventionality and comprehensibility ($\beta = 0.866$, $t(78) = 13.86$, $p < 2e-16$), indicating that for every unit increase in conventionality, comprehensibility increased by approximately 0.87 units. The model explained 71.1% of the variance in comprehensibility ($R^2 = 0.7111$), which essentially is indicative of a strong predictive relationship between the two variables. The residual standard error was 0.7598, indicating that the model fit reasonably well. The p-value was highly significant, thereby suggesting that conventionality is a reliable predictor of metaphor comprehensibility in the dataset. The Intercept was marginally significant ($p = 0.056$), but it does not hold as strong an effect as the main predictor.

3.2 Experiment 2: validation study

3.2.1 Methods

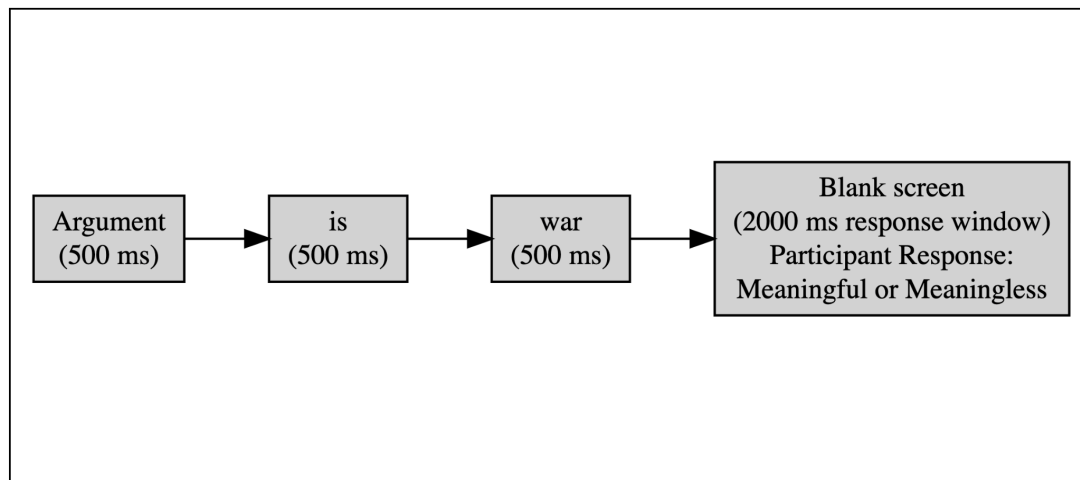
In this experiment, participants engaged in a semantic judgment task (SJT) in which they read target stimuli presented one word at a time and classified each phrase as either meaningful or meaningless. The same set of stimuli from Experiment 1 was used. Eleven participants — all native speakers of English and either undergraduate or graduate students at the University of Victoria — took part in the study.

Before beginning the main task, participants received detailed instructions regarding the response key assignments, which were counterbalanced to minimize response biases. Those with an even participant number were instructed to press “D” for “meaningful” and “K” for “meaningless,” whereas those with an odd participant number received the opposite assignment to control for handedness bias. To familiarize them with the experimental procedure, a short practice phase consisting of 8 target phrases was included. Participants were informed that words would be

displayed sequentially on the screen, followed by a two-second response window. If they failed to respond within this time, a message appeared reminding them to respond more quickly in subsequent trials and their response for that trial was not recorded.

Figure 25

Visual Timeline of an Experimental Trial (Experiment 2)

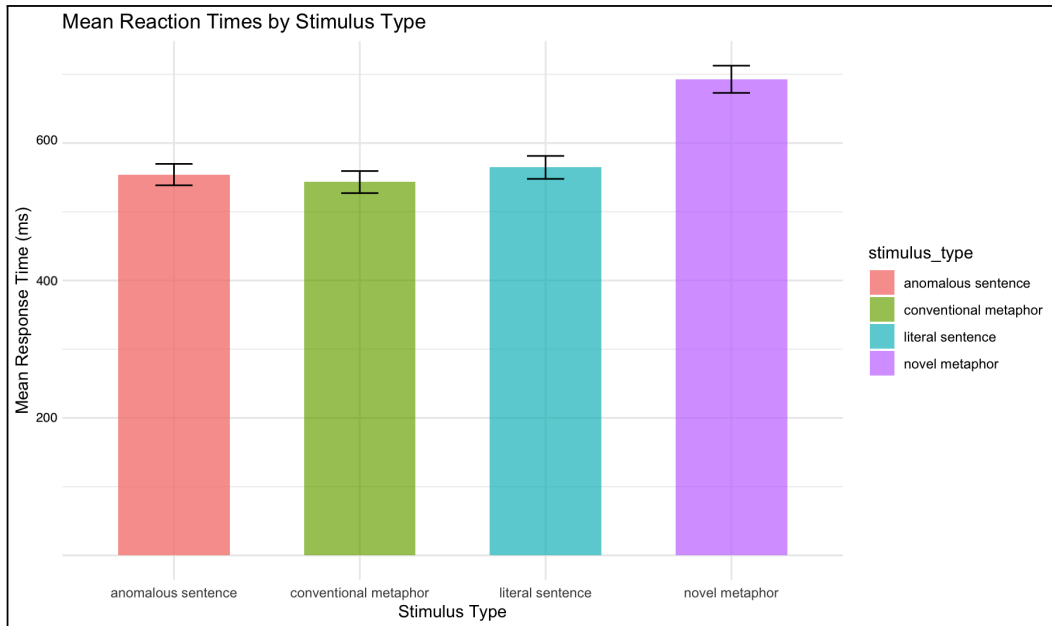
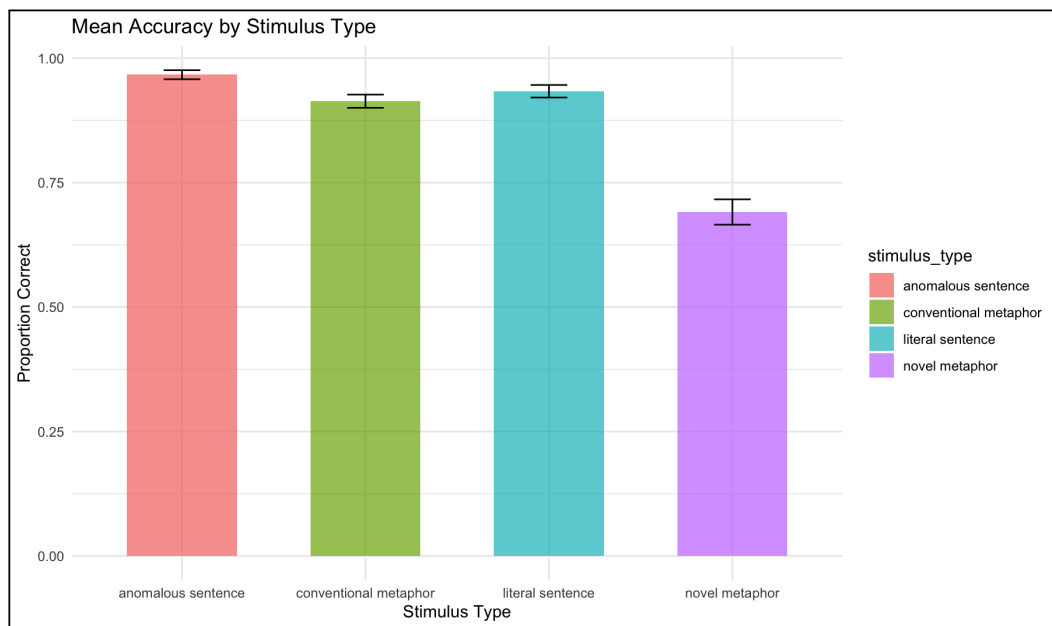


The main experiment followed the same structure as the practice phase but included a larger set of stimuli presented in a randomized order. Each word within a sentence appeared for 500 milliseconds before being replaced by the next word (see Figure 25). Following the final word, the screen turned blank, signaling the start of the response window. Participants had up to two seconds to categorize the phrase as meaningful or meaningless. Their reaction times and accuracy rates were recorded for analysis, with the reaction times being measured from the offset of the last nominal element in the target phrases. To reduce the potential confound associated with the speed-accuracy tradeoff often observed in behavioral studies (Salthouse & Hedden, 2002), only correct responses were included in the reaction time analysis.

To prevent participants from adopting a rhythmic response pattern, an inter-trial interval jittering mechanism was implemented. This interval varied randomly between 1.5 and 3.0 seconds, ensuring that participants could not anticipate the exact onset of the next trial. Such jittering helps to reduce predictability, maintain attentional engagement, and prevent habituation effects that could influence reaction times (Breiter et al., 1996). Additionally, to minimize fatigue and sustain accuracy, participants were given short breaks after every ten trials.

3.2.2 Results and discussion

To examine whether utterance type affected participants' performance, we conducted analyses of variance (ANOVAs) on both reaction time (RT) and accuracy data. Linear and generalized linear mixed-effects models were fitted using the *lme4* package in R, with stimulus type (novel metaphor, conventional metaphor, literal sentence, and anomalous sentence) as a fixed effect and participant as a random intercept to account for repeated measures. Type III Wald chi-square tests were used to assess the significance of the fixed effect. The analysis revealed a significant main effect of stimulus type on both reaction times, $\chi^2(3) = 43.89, p < .001$, and accuracy, $\chi^2(3) = 152.02, p < .001$, which indicates that the type of utterance influenced how quickly and accurately participants responded.

Figure 26*Mean Reaction Times by Stimulus Type***Figure 27***Mean Accuracy by Stimulus Type*

Post-hoc comparisons for reaction time, using Bonferroni correction, revealed several significant differences between stimulus types (see Table 2). Specifically, participants responded significantly slower to novel metaphors compared to anomalous sentences ($p < 0.001$), conventional metaphors ($p < 0.001$), and literal sentences ($p < 0.001$). In contrast, no significant differences were observed between anomalous sentences and conventional metaphors, nor between anomalous sentences and literal sentences (both $p > 0.05$).

Table 2

Post-hoc Comparisons for Reaction Time (Bonferroni Corrected)

Post-hoc Comparisons for Reaction Time (Bonferroni Corrected)						
contrast	estimate	SE	df	t.ratio	p-value	Significance
anomalous sentence - conventional metaphor	0.0494050	0.0193225	1653.364	2.5568658	n.s.	
anomalous sentence - literal sentence	0.0195864	0.0199881	1653.338	0.9799026	n.s.	
anomalous sentence - novel metaphor	-0.0956251	0.0227197	1655.302	-4.2089075	$p < 0.001$	***
conventional metaphor - literal sentence	-0.0298186	0.0195510	1653.273	-1.5251679	n.s.	
conventional metaphor - novel metaphor	-0.1450300	0.0223131	1654.524	-6.4997826	$p < 0.001$	***
literal sentence - novel metaphor	-0.1152115	0.0228992	1654.688	-5.0312450	$p < 0.001$	***

These results suggest that novel metaphors elicited slower reaction times compared to other stimulus types, most probably due to necessity to reach maximal structural consistency between the metaphor predicates. Structural alignment is part of the process of analogical reasoning. When processing a novel metaphor, the brain must engage in deeper levels of reasoning to find similarities between the topic and vehicle domains, especially because those connections are not as readily apparent.

Table 3*Post-hoc Comparisons for Accuracy (Bonferroni Corrected)*

Post-hoc Comparisons for Accuracy (Bonferroni Corrected)						
contrast	estimate	SE	df	z.ratio	p-value	Significance
anomalous sentence - conventional metaphor	1.0418679	0.3251217	Inf	3.2045477	p < 0.01	**
anomalous sentence - literal sentence	0.8549250	0.3397362	Inf	2.5164378	n.s.	
anomalous sentence - novel metaphor	2.7270377	0.3029940	Inf	9.0003019	p < 0.001	***
conventional metaphor - literal sentence	-0.1869428	0.2528019	Inf	-0.7394835	n.s.	
conventional metaphor - novel metaphor	1.6851699	0.2003741	Inf	8.4101164	p < 0.001	***
literal sentence - novel metaphor	1.8721127	0.2233918	Inf	8.3804016	p < 0.001	***

The accuracy results align with the reaction time findings, with novel metaphors being rated the least accurately. Specifically, novel metaphors were significantly less accurate than anomalous sentences (estimate = -2.73, $p < 0.001$), conventional metaphors (estimate = -1.69, $p < 0.001$), and literal sentences (estimate = -1.87, $p < 0.001$). Anomalous sentences showed higher accuracy compared to novel metaphors, with a significant contrast against conventional metaphors (estimate = 1.04, $p < 0.01$), but no significant difference compared to literal sentences. Conventional metaphors did not differ significantly from literal sentences in terms of accuracy ($p = n.s.$), suggesting that both were processed with similar levels of accuracy.

Conventionality effects***Reaction times***

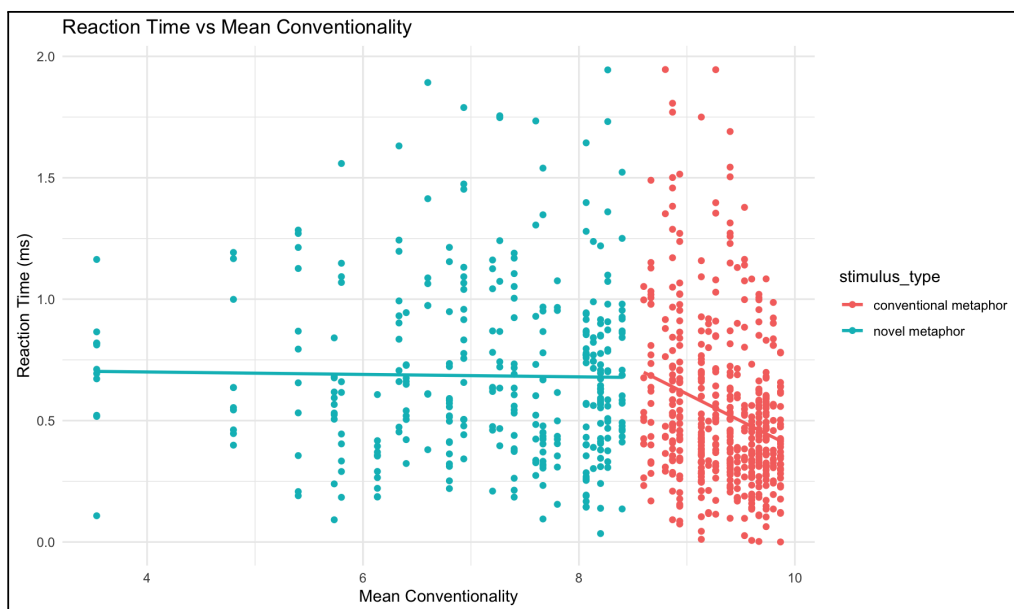
A linear regression analysis was conducted to examine the relationship between participants' reaction times and the mean conventionality of the metaphors used. The analysis included data from 11 participants and aimed to determine whether the conventionality of a metaphor influences the time taken to process it.

The dependent variable in this analysis was reaction time, while the independent variable was the mean conventionality score.

The results indicate a significant negative relationship between mean conventionality and response time ($\beta = -0.05581$, $SE = 0.00880$, $t = -6.342$, $p < 0.001$), suggesting that as the conventionality of a metaphor increases, the time required to process it decreases (see Figure 28).

Figure 28

Reaction Times vs Mean Conventionality



The intercept ($\beta = 1.06738$, $SE = 0.07521$, $p < 0.001$) represents the estimated response time when mean conventionality is at zero. The model accounts for approximately 4.59% of the variance in response times ($R^2 = 0.04585$, adjusted $R^2 = 0.04471$), indicating a small but statistically significant effect. The overall model fit was significant, $F(1, 837) = 40.22$, $p < 0.001$, so the relationship between conventionality and reaction time is statistically significant and warrants further investigation in future studies.

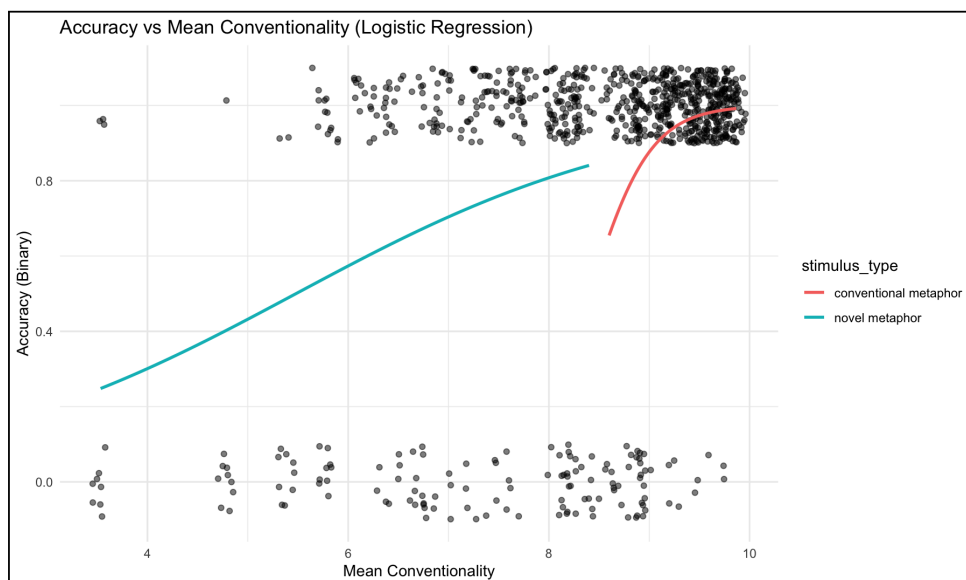
Accuracy rates

The relationship between participants' accuracy in identifying metaphors as meaningful and the mean conventionality of the stimuli was examined using a logistic regression model, as accuracy was treated as a binary outcome (correct/incorrect). The model aimed to determine the influence of mean conventionality on the likelihood of participants correctly categorizing metaphors. The results of the logistic regression revealed a statistically significant relationship between mean conventionality and accuracy (see Figure 29). The coefficient for mean conventionality was 0.6981, with a corresponding z-value of 9.901 ($p < 2e-16$). This indicates that for each unit increase in the conventionality of the metaphor, the odds of a correct response increased significantly, suggesting that more conventional metaphors were more likely to be correctly identified by participants.

Further model diagnostics revealed a null deviance of 753.35 and a residual deviance of 639.59. These statistics suggest that the logistic regression model provided a good fit to the data. Overall, the findings underscore the importance of conventionality in metaphor processing, as participants were more accurate in identifying metaphors with higher conventionality.

Figure 29

Accuracy vs Mean Conventionality (Logistic Regression)



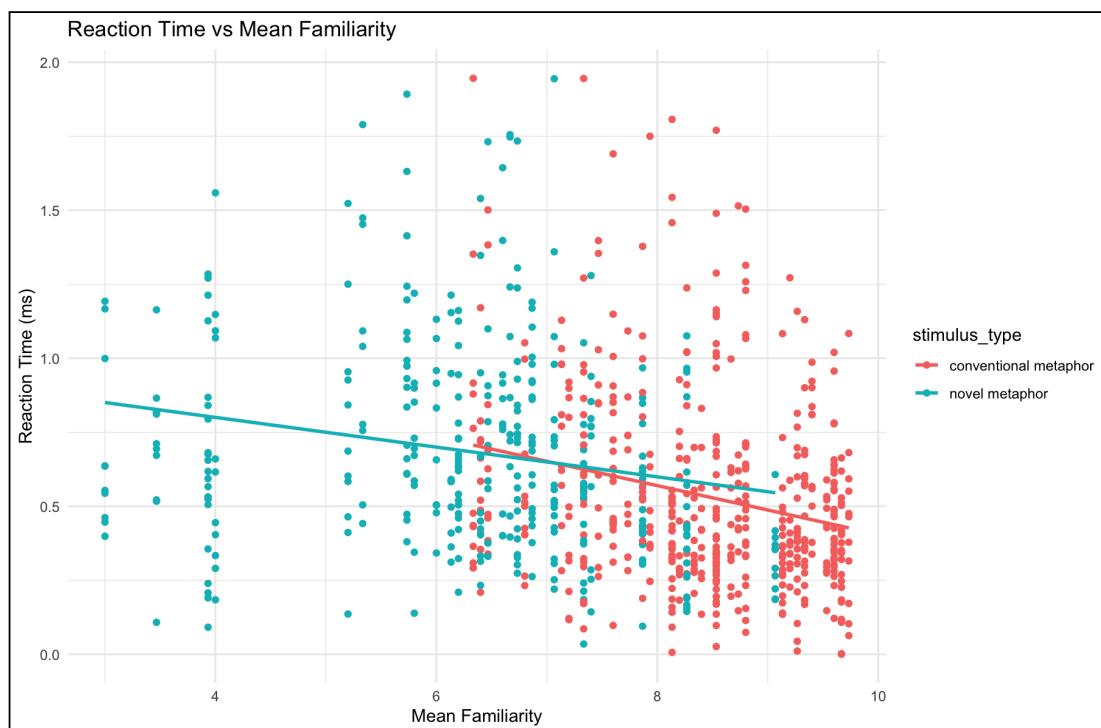
Familiarity effects

Reaction times

To analyse the effect of metaphor familiarity on reaction times, a linear regression analysis was conducted. The dependent variable, reaction time, was modeled as a function of the independent variable, mean familiarity. The model revealed that the intercept, with an estimate of 1.096 and a standard error of 0.057, was significantly different from zero ($p < 2e-16$). This suggests that when the mean familiarity is zero, the predicted reaction time is approximately 1.096 ms.

Figure 30

Reaction Time vs Mean Familiarity



The coefficient for mean familiarity was estimated at -0.066, with a standard error of 0.007. This negative coefficient indicates that for each unit increase in mean familiarity, the reaction time decreases by 0.066 ms. The relationship between mean

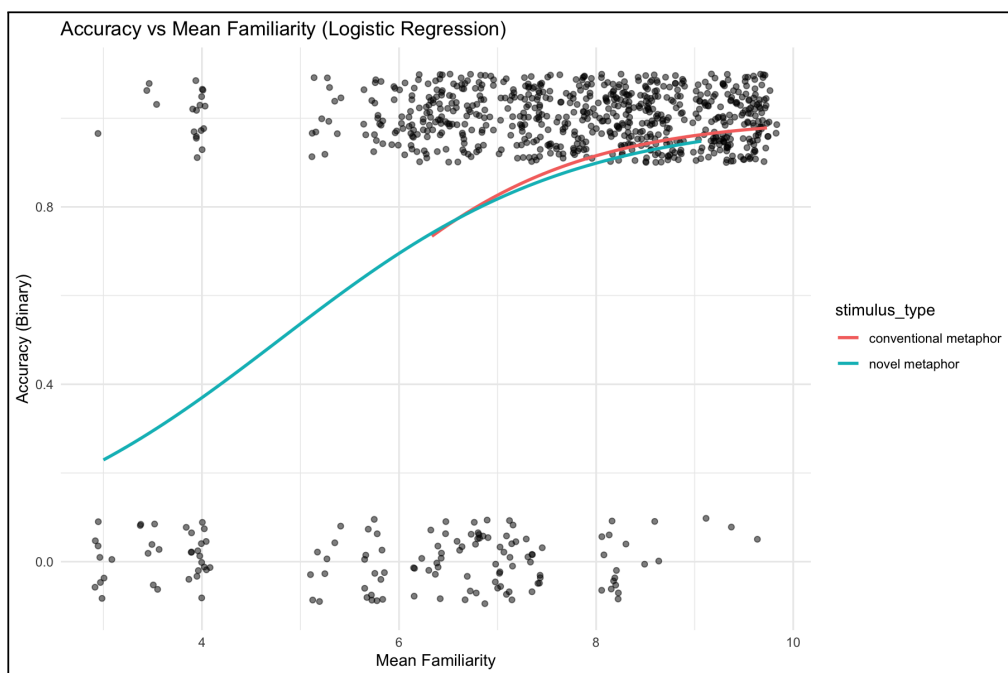
familiarity and reaction time was highly significant ($p < 2e-16$), pointing to a robust inverse association between these variables. The multiple R-squared value was 0.088, suggesting that approximately 8.8% of the variance in reaction time could be explained by mean familiarity. Although this represents a modest proportion of variance, it still reflects a meaningful relationship. The adjusted R-squared value, which accounts for the number of predictors in the model, was 0.087, indicating minimal overfitting. The F-statistic of 80.8, with a p-value less than $2.2e-16$, confirmed that the overall model was statistically significant and provided strong evidence for the relationship between reaction time and mean familiarity.

Accuracy rates

A logistic regression analysis was conducted to examine the relationship between accuracy and mean familiarity. The model estimated the log-odds of a correct response as a function of mean familiarity.

Figure 31

Accuracy vs Mean Familiarity (Logistic Regression)



The results revealed that the intercept was significant ($\beta = -3.55$, $SE = 0.47$, $z = -7.58$, $p < .001$), indicating that when mean familiarity is at zero, the log-odds of a correct response are negative, meaning accuracy is low. The coefficient for mean familiarity was also highly significant ($\beta = 0.73$, $SE = 0.07$, $z = 10.73$, $p < .001$), suggesting that as familiarity increases, the probability of a correct response also increases. This positive relationship indicates that more familiar items were judged with greater accuracy.

Discussion

The observation that reaction times only begin to decrease within the category of conventional metaphors, and not across the entire spectrum of metaphor conventionality, provides compelling support for the idea that categorical processing only emerges for highly conventionalized metaphors (see Figure 28). As we can clearly see, for the majority of the metaphor spectrum, including novel metaphors of varying degrees of conventionality, processing remains dominated by structural alignment and cross-domain mapping mechanisms (Gentner & Wolff, 1997; Bowdle & Gentner, 2005). This finding appears to align with Alignment-First Processing (Gentner & Bowdle, 2001), where metaphor comprehension initially relies on analogical mapping between conceptual domains rather than on direct category membership retrieval. In this view, novel metaphors—regardless of their degree of conventionality—require an effortful process of comparing relational structures between source and target domains. This explains why increasing the conventionality of a novel metaphor does not necessarily lead to faster reaction times, as participants still engage in relational reasoning and structural alignment rather than automatic retrieval.

However, for highly conventional metaphors, reaction times do begin to decrease, suggesting a shift toward categorical processing. This transition aligns with the Career of Metaphor Theory (Bowdle & Gentner, 2005), which posits that metaphors initially understood through comparison gradually become stored as lexicalized category members. I feel inclined to interpret this finding as an indication that only when a metaphor reaches a certain threshold of conventionality does its processing switch from analogical alignment to direct retrieval from memory, resulting in faster comprehension. Once again, the observed nonlinear effect of conventionality on reaction times suggests that alignment-based processing remains dominant across much of the spectrum. It is only for highly conventionalized metaphors that a more automatic, categorical processing mechanism takes over. This distinction underscores the idea that metaphor comprehension exists along a continuum, where structural alignment remains the primary processing strategy until a metaphor becomes entrenched enough to undergo a categorical shift.

3.2.3 Limitations

These findings should be interpreted with caution due to the relatively small sample size of participants ($N = 11$) and the within-subjects nature of the data, which introduces dependencies that are not accounted for in a simple linear and logistic regression model. While the number of observations is substantial, the responses are not fully independent, as multiple trials were collected from the same individuals. However, it is important to bear in mind that this experiment, in the first place, was designed as an initial investigation to identify potential trends in the relationship between conventionality and reaction times as well as accuracy rates, rather than to draw definitive conclusions. The primary goal was to assess whether these stimuli exhibit meaningful variation in comprehension and processing, which would justify

their inclusion in the next phase of the study and support the validation of the experimental instrument and entire paradigm.

3.3 Experiment 3

This curious pattern of results — where only highly conventional metaphors benefit from expedited processing — raises important questions about the extent to which figurative language processing mechanisms are sensitive not only to conventionality but also to broader contextual and linguistic factors. If alignment-based processing dominates early stages of metaphorical understanding and only gradually gives way to categorical retrieval, then it becomes crucial to explore how this trajectory might be modulated by the speaker's linguistic background and the nature of the discourse context. Specifically, we must ask: Do these processing shifts manifest similarly for non-native speakers, who may not share the same entrenched metaphorical categories as native speakers? And does the presence of extended context influence how metaphors are judged as meaningful in terms of accuracy of judgement and speed of that judgement? To address these questions, Experiment 3 builds on the findings of Experiment 2 by examining how metaphorical expressions are processed in both figuratively extended and figuratively minimal (neutral) contexts, and how these effects interact with participants' language background.

3.3.1 Methods

3.3.1.1 Design

This experiment employed a semantic judgment task (SJT) to investigate how native (L1) and non-native (L2) English speakers process metaphorical expressions in an *extended* and *single* condition (see Figure 32). Participants were presented with short contextual paragraphs followed by a target phrase, and they were asked

to judge whether the phrase was *meaningful* or *meaningless*. The task was designed and implemented using PsychoPy.

Figure 32

Examples of Stimuli for Novel Single and Novel Extended Metaphors

Novel Single Metaphor condition	
Context	Target Expression
No matter how hard Julia tried to focus on the future, her previous experiences and decisions continually influenced her present choices.	The past is an anchor
When Lucas greeted everyone with a warm smile, it noticeably lifted the mood in the room and made interactions more pleasant.	A smile is a sweetener

Neutral context
(No metaphor-related elements)
Novel metaphor

Novel Extended Metaphor condition	
Context	Target Expression
John doesn't like physical contact, and even his girlfriend finds it difficult to come close to him. She feels pricked by his thorny attitude every time he sees her.	John is a cactus
When he announced his choice, it shot through the room with sudden force, cutting through any previous uncertainties and leaving an immediate impact.	Decision is a bullet

Metaphorical context
(Contains metaphor-related elements)
Novel metaphor

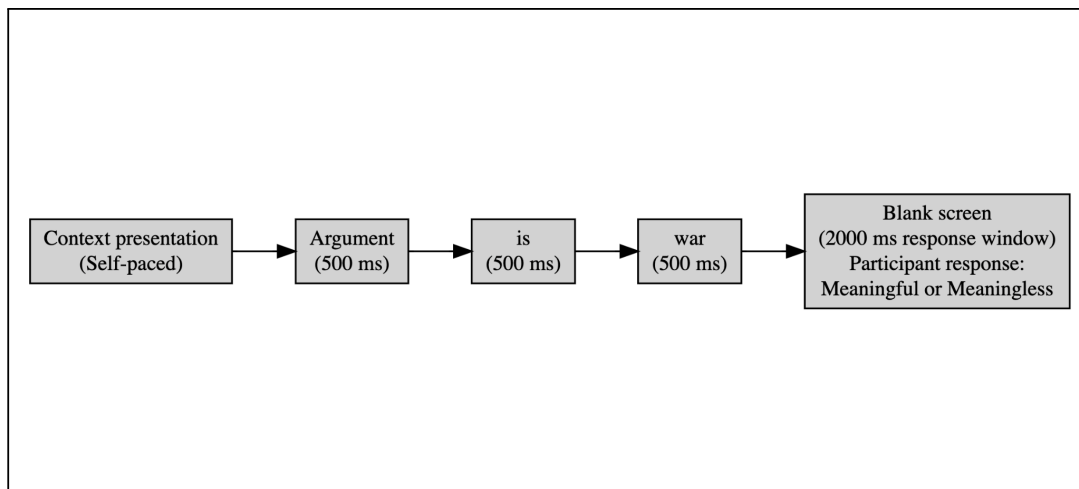
3.3.1.2 Materials and procedure

The experimental stimuli were drawn from a pre-validated set of metaphorical and non-metaphorical expressions. The procedure was the same as in experiment 2, with one modification: each trial also contained a short context paragraph followed by a target sentence (see Figure 33). Participants were allowed to read the context passages for as long as needed in order to ensure full comprehension of the passage before proceeding. This self-paced presentation was intended to encourage deeper semantic integration of the contextual information and to ensure that the additional vehicle-related metaphor primes in the extended condition would be attended to properly by the readers. Just as in Experiment 2, the participants were given self-paced breaks after every ten trials to minimize fatigue and sustain accuracy and attention span.

Stimuli were divided into the following categories: novel metaphors, conventional metaphors, literal expressions, and anomalous expressions. In addition, all metaphors appeared in two conditions: in the single condition (neutral) and in the extended condition (metaphorically priming), while literal and anomalous phrases appeared only once in the neutral condition.

Figure 33

Visual Timeline of an Experimental Trial (Experiment 3)



All metaphors and literal phrases were coded as meaningful, and anomalous phrases were coded meaningless. A set of practice items was used at the beginning of the experiment to familiarize participants with the task. Each word in the target phrase was presented sequentially for 500 ms each, followed by a response time window (2000 ms). Responses were made via keyboard input. Reaction times and accuracy were recorded for each trial. A progress bar was displayed to keep participants informed of their progress.

To control for order effects and manage semantic priming between related stimuli, the experiment implemented a structured randomization procedure. Specifically, target items that shared the same metaphorical source (e.g., two conditions using the same vehicle term such as *anchor* in the metaphor *past is an anchor*) were treated as pairs and were deliberately separated by a buffer of at least 20 unrelated filler trials (other metaphor or non-metaphor targets). This enforced space and time buffer was used to reduce potential carryover or repetition effects in sentence comprehension. The fillers used as buffers were randomly selected from the pool of unrelated stimuli to maintain unpredictability. This trial structure was

implemented through the code, such that one item from each critical pair was inserted into the trial list, followed by a randomized set of 20 filler trials, before the second item from the pair was introduced. After all critical pairs were placed with appropriate spacing, the remaining fillers were randomly distributed across the remaining trial sequence. This approach ensured that no two semantically related metaphor trials occurred in close succession and helped to maintain the independence of individual trials.

3.3.1.3 Participants

In total, 20 L1 English speakers ($M_{age} = 21.4$, $SD = 2.26$) and 23 L2 English speakers ($M_{age} = 25.39$, $SD = 4.41$) participated in the study. Participants were assigned to one of two groups based on self-reported language background³⁰: L1 English speakers and L2 English speakers. In line with the approach taken by Wang and Jankowiak (2023), participant selection for the L2 group involved measuring their vocabulary size, which was assessed using the LexTale test (Lemhöfer & Broersma, 2012). It was administered prior to formally inviting individuals to participate in the main experiment, and the selected performance threshold was 75% ($M_{LexTale\ score} = 82.64\%$, $SD = 7.59$, range: 75-96.25). Upon launching the task, participants were asked to indicate whether English was their first (L1) or second (L2) language. The participant number was automatically recorded, and responses were saved along with language group, stimulus type, response time, and accuracy. All data were anonymized and stored in a structured CSV file.

³⁰ Their language status (L1 or L2) was determined based on order of acquisition.

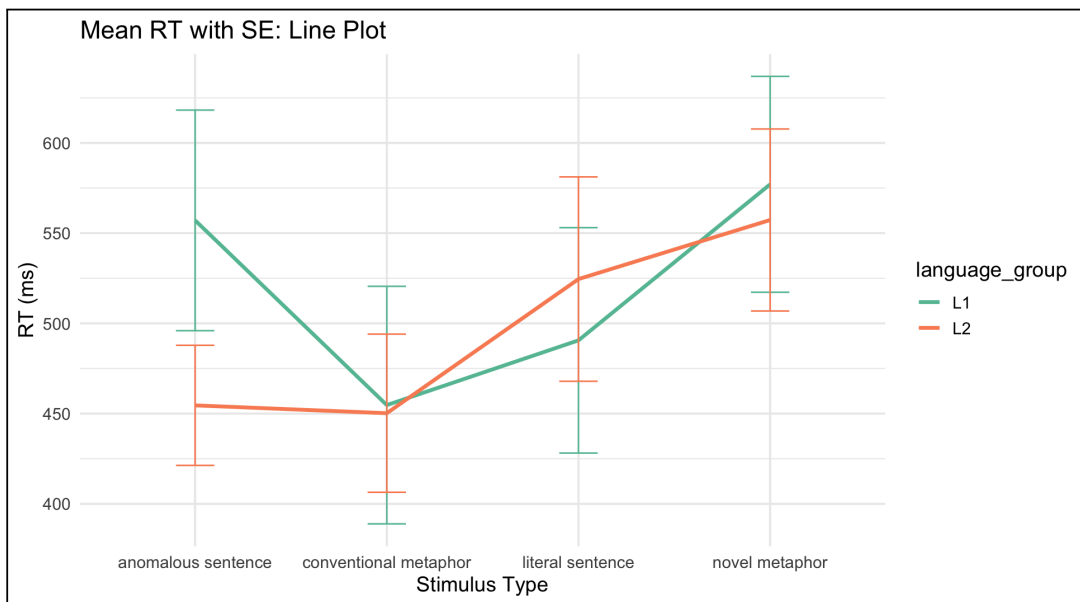
3.3.2 Results and discussion

3.3.2.1 Reaction time results across all utterance types in the single condition

To investigate how metaphor processing is influenced by language background and stimulus type, a mixed-design ANOVA was conducted on reaction time data from the single-phrase condition. The model included language group (L1 vs. L2) as a between-subjects factor and stimulus type (i.e., novel metaphors, conventional metaphors, literal sentences, and anomalous sentences) as a within-subjects factor. Only correct responses were included in the analysis. This analysis was chosen to assess whether reaction times varied by stimulus type and whether this effect interacted with participants' language background (see Figure 34).

Figure 34

Mean RT with SE: Line Plot



The results revealed a significant main effect of stimulus type, $F(2.70, 110.71) = 4.45$, $p = .027$, generalized eta squared (η^2) = .027, indicating that participants'

reaction times significantly differed depending on the type of stimulus (see Table 4). There was no main effect of language group, $F(1, 41) = 0.12, p = .729$, indicating that L1 and L2 speakers did not differ in their overall reaction times in this condition. Moreover, the interaction between language group and stimulus type was also non-significant, $F(2.70, 110.71) = 1.67, p = .198$. This is indicative of the fact that the impact of stimulus type on reaction times was comparable across both language groups — that is, L1 and highly proficient L2 speakers were similarly sensitive to the differences in the semantic qualities across all utterances.

Table 4

ANOVA: RT (Single Condition)

ANOVA: RT (Single Condition)					
Effect	df	MSE	F	ges	p.value
language_group	1, 41	189075.19	0.12	.002	.729
stimulus_type	2.70, 110.71	23475.93	4.45 **	.027	**007**
language_group:stimulus_type	2.70, 110.71	23475.93	1.67	.010	.181

To follow up on the significant main effect of stimulus type observed in the ANOVA for the single condition, pairwise comparisons were conducted separately within each language group and between groups for each utterance type. For all analyses, p -values for pairwise comparisons were adjusted using the Bonferroni correction to account for multiple comparisons. The goal of these comparisons was to identify whether participants responded differently to various stimuli, and whether these effects differed between L1 and L2 speakers (see Table 5 and 6).

Table 5*Post-hoc: RT (Within-Group)*

Posthoc: RT (Within-Group)							
contrast	language_group	estimate	SE	df	t.ratio	p.value	
anomalous.sentence - conventional.metaphor	L1	102.372161	52.87760	41	1.93602118	0.3587	
anomalous.sentence - literal.sentence	L1	66.495575	50.70744	41	1.31135746	1.0000	
anomalous.sentence - novel.metaphor	L1	-19.983080	47.88524	41	-0.41731186	1.0000	
conventional.metaphor - literal.sentence	L1	-35.876586	44.56746	41	-0.80499504	1.0000	
conventional.metaphor - novel.metaphor	L1	-122.355241	37.95457	41	-3.22372916	**0.0149**	
literal.sentence - novel.metaphor	L1	-86.478655	39.88879	41	-2.16799372	0.2161	
anomalous.sentence - conventional.metaphor	L2	4.346679	49.30862	41	0.08815253	1.0000	
anomalous.sentence - literal.sentence	L2	-69.988589	47.28493	41	-1.48014583	0.8789	
anomalous.sentence - novel.metaphor	L2	-102.729512	44.65322	41	-2.30060707	0.1594	
conventional.metaphor - literal.sentence	L2	-74.335269	41.55937	41	-1.78865234	0.4864	
conventional.metaphor - novel.metaphor	L2	-107.076192	35.39282	41	-3.02536521	**0.0256**	
literal.sentence - novel.metaphor	L2	-32.740923	37.19649	41	-0.88021533	1.0000	

Among native speakers, a significant difference was found between novel and conventional metaphors, with novel metaphors eliciting slower reaction times (*estimate* = -122.36 ms, $p = .015$). This suggests that conventional metaphors, which rely on more entrenched conceptual mappings, are processed more fluently than novel metaphors that require the construction of new metaphorical associations. Curiously enough, a trend was also observed for novel metaphors being slower than literal sentences (*estimate* = -86.48 ms), though this difference did not reach statistical significance after correction ($p = .216$). No other pairwise contrasts were significant, indicating that literal and conventional metaphors were processed with comparable ease in this group.

L2 speakers showed a similar pattern, with significantly slower reaction times for novel metaphors compared to conventional ones (*estimate* = -107.08 ms, $p = .026$). This parallel effect reinforces the role of metaphor conventionality in reducing processing demands, even for non-native speakers, provided they are highly

proficient. A non-significant trend was also observed for slower responses to novel metaphors compared to anomalous sentences (*estimate* = -102.73 ms, $p = .159$), though again, this did not hold after correction. Other contrasts showed no reliable differences ($p > .4$).

Table 6

Post-hoc: RT (Between-Group)

Posthoc: RT (Between-Group)						
contrast	stimulus_type	estimate	SE	df	t.ratio	p.value
L1 - L2	anomalous.sentence	102.522346	67.19119	41	1.52583014	0.1347
L1 - L2	conventional.metaphor	4.496864	77.26996	41	0.05819679	0.9539
L1 - L2	literal.sentence	-33.961819	84.14222	41	-0.40362398	0.6886
L1 - L2	novel.metaphor	19.775913	77.68452	41	0.25456696	0.8003

Pairwise comparisons of reaction times between L1 and L2 speakers for each stimulus type revealed no statistically significant differences. While L1 participants were, on average, 102 ms slower than L2 participants in responding to anomalous sentences, this difference did not reach significance ($p = .135$). It is my assumption that L1 speakers might engage in more effortful or deeper processing, even for anomalous stimuli, attempting to extract meaning where none exists. L2 speakers, in contrast, might more quickly dismiss nonsensical input, resulting in faster but more superficial processing. This might have to do with the nature of the task where a lot of original linguistic stimuli are presented and overtime it becomes more difficult to disentangle sensicality from nonsensicality. Native speakers may, in fact, be more cautious or methodical in their judgments, especially when a sentence seems off but not obviously wrong — a “double-checking” strategy. L2 participants may make quicker, more decisive judgments based on surface cues or perceived semantic

mismatch, which is also reflected in their lower accuracy for novel metaphors. L2 speakers appear to be rejecting ambiguity more readily, which results in their numerically higher reaction times for anomalous sentences and significantly lower accuracy for novel metaphors, compared to native speakers.

For the remaining stimulus types — conventional metaphors ($p = .954$), literal sentences ($p = .689$), and novel metaphors ($p = .800$) — group differences were minimal and also non-significant. These results suggest that, across all sentence types, L1 and L2 speakers responded with broadly similar timing, and there is no evidence of a consistent speed advantage for either group at the level of individual stimulus types.

3.3.2.2 Accuracy rates across all utterance types in the single condition

A mixed-design ANOVA was conducted to examine how sentence type and participant language group influenced accuracy in judging the meaningfulness of single-sentence utterances. The analysis revealed a significant main effect of the language group, $F(1, 41) = 5.24$, $p = .027$, indicating that L1 and L2 participants differed in their overall accuracy (see Figure 35).

Figure 35

Accuracy by Stimulus Type (Single Condition)

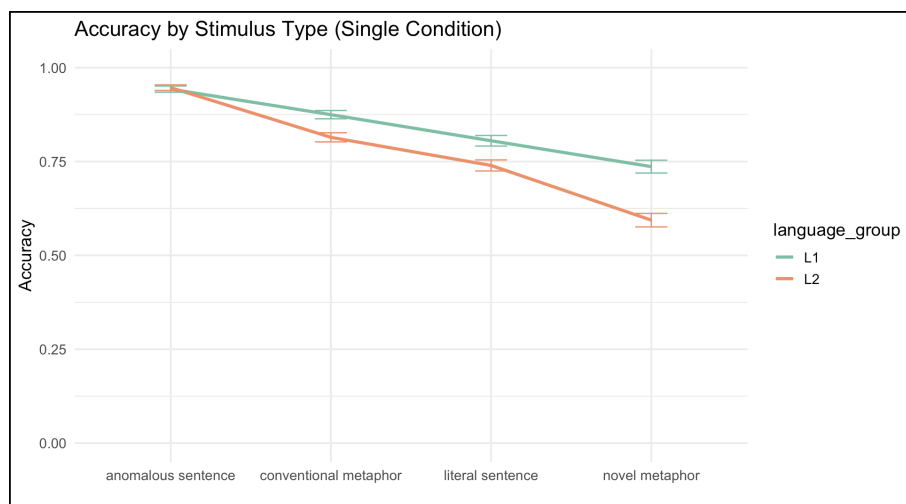


Table 7*ANOVA: Accuracy (Single Condition)*

ANOVA: Accuracy (Single Condition)						
Effect	df	MSE	F	ges	p.value	
language_group	1, 41	0.04	5.24 *	.052	.027	
stimulus_type	2.31, 94.68	0.02	37.76 ***	.345	<.001	
language_group:stimulus_type	2.31, 94.68	0.02	2.51 +	.034	.078	

A highly significant main effect of stimulus type was also observed, $F(2.31, 94.68) = 37.76$, $p < .001$, showing that accuracy varied substantially depending on the type of sentence participants judged. Although the interaction between language group and sentence type did not reach conventional significance, $F = 2.51$, $p = .078$, it indicated a potential trend toward different patterns of performance across groups.

Table 8*Post-hoc: Accuracy within Groups*

Posthoc: Accuracy within groups						
contrast	language_group	estimate	SE	df	t.ratio	p.value
anomalous.sentence - conventional.metaphor	L1	0.06931620	0.03527235	41	1.965171	0.3372
anomalous.sentence - literal.sentence	L1	0.13996421	0.04655099	41	3.006686	**0.0270**
anomalous.sentence - novel.metaphor	L1	0.20662656	0.04814643	41	4.291628	**0.0006**
conventional.metaphor - literal.sentence	L1	0.07064801	0.04110488	41	1.718726	0.5592
conventional.metaphor - novel.metaphor	L1	0.13731035	0.02655889	41	5.170033	**0.0000**
literal.sentence - novel.metaphor	L1	0.06666235	0.03826300	41	1.742214	0.5338
anomalous.sentence - conventional.metaphor	L2	0.13125226	0.03289164	41	3.990444	**0.0016**
anomalous.sentence - literal.sentence	L2	0.20823249	0.04340902	41	4.796986	**0.0001**
anomalous.sentence - novel.metaphor	L2	0.35597977	0.04489678	41	7.928849	**0.0000**
conventional.metaphor - literal.sentence	L2	0.07698023	0.03833050	41	2.008329	0.3074
conventional.metaphor - novel.metaphor	L2	0.22472751	0.02476629	41	9.073925	**0.0000**
literal.sentence - novel.metaphor	L2	0.14774728	0.03568043	41	4.140849	**0.0010**

For the L1 group, several contrasts revealed significant differences.

Participants were significantly more accurate in judging anomalous sentences than literal ones ($p = .027$) with no significant difference between anomalous sentences and conventional metaphors. Accuracy was also significantly higher for anomalous sentences compared to novel metaphors ($p < .001$), indicating that novel metaphors may have been misclassified or interpreted with greater difficulty. Importantly, conventional metaphors were judged significantly more accurately than novel ones ($p < .001$), validating central the idea of the Career of Metaphor hypothesis that conventional categories enhance comprehension. Other contrasts, such as between literal and novel or anomalous and conventional sentences, were not statistically significant, indicating that the nature of the downward trend in accuracy from anomalous sentences to novel metaphors is rather smooth.

In the L2 group, the pattern was generally more pronounced. Similarly, accuracy was significantly higher for anomalous than literal sentences ($p < .001$), and markedly higher for anomalous than novel metaphors ($p < .001$). In fact, novel metaphors were judged less accurately than all other sentence types. As we can see on the visual (see Figure 35), the drops of accuracy (from anomalous sentences to conventional, and from literal sentences to novel metaphors) are more significant. Conventional metaphors also yielded significantly higher accuracy than novel ones ($p < .001$), and anomalous sentences were judged more accurately than conventional metaphors ($p = .002$), thus underscoring the difficulties L2 speakers experience in accurately judging creative figurative language as meaningful.

Table 9*Post-hoc: Accuracy between Groups*

Posthoc: Accuracy between groups						
contrast	stimulus_type	estimate	SE	df	t.ratio	p.value
L1 - L2	anomalous.sentence	-0.003787014	0.02104019	41	-0.1799896	0.8580
L1 - L2	conventional.metaphor	0.058149042	0.03762961	41	1.5453002	0.1300
L1 - L2	literal.sentence	0.064481269	0.05714719	41	1.1283366	0.2657
L1 - L2	novel.metaphor	0.145566199	0.05190535	41	2.8044545	**0.0077**

Post-hoc comparisons between L1 and L2 groups across sentence types revealed a selective pattern of group-level differences in accuracy. Notably, a significant disparity was observed only for novel metaphors ($p < 0.05$), with L1 participants demonstrating superior performance relative to their L2 counterparts. In contrast, no statistically significant differences emerged between groups for conventional metaphors, literal sentences, or anomalous items. This suggests that while L2 speakers are comparable to native speakers in their ability to identify literal and conventional meanings — and to reject semantically anomalous input — they encounter pronounced difficulties when required to interpret novel figurative language. These results underscore the heightened processing demands associated with novel metaphor comprehension, particularly for individuals operating in a non-native language. Such metaphors require greater cognitive flexibility, deeper inferencing, and more robust integration of context-independent semantic cues, all of which pose increased challenges in a second language context.

3.3.2.3 Reaction time results: metaphor type x language group x condition

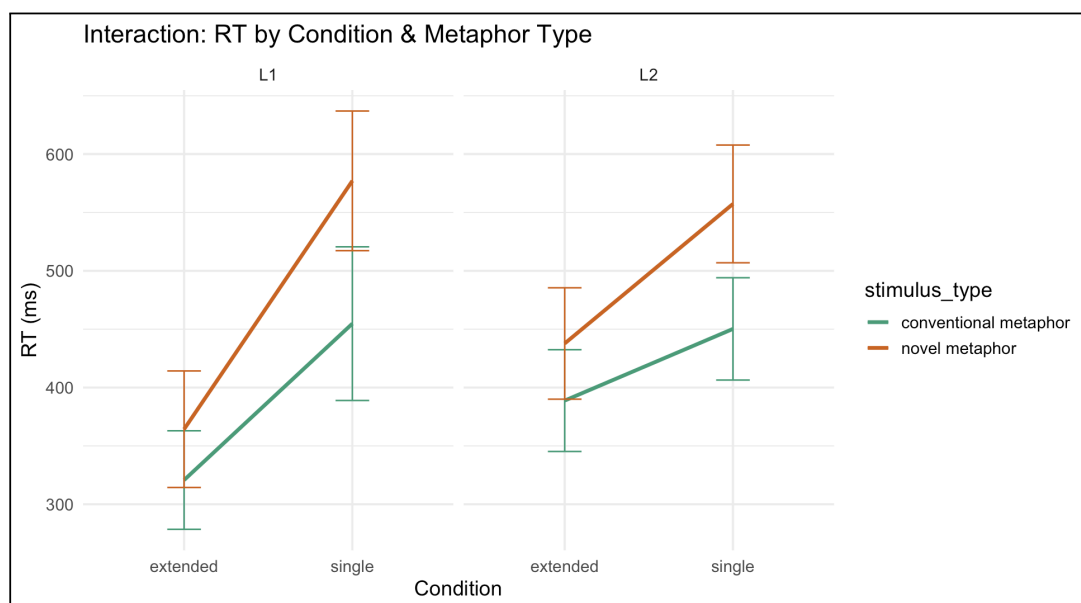
To examine the effects of *metaphor type*, *condition*, and *language group* on reaction times, I conducted a three-way mixed-design repeated measures ANOVA using the `av_ ez()` function from the `afex` package in R. This analysis included two

within-subject factors — stimulus type (novel vs. conventional metaphors) and condition (single vs. extended metaphors)—and one between-subject factor, language group (L1 vs. L2). The dependent variable was reaction time (ms), calculated from correct trials only. This approach was selected to simultaneously assess the independent and interactive effects of metaphor novelty, context, and nativeness on processing speed, and to determine whether the patterns differed across participant groups.

A significant main effect of stimulus type was observed, $F(1, 24) = 22.48, p < .001$, which essentially tells us that novel metaphors elicited longer reaction times than conventional metaphors across all participants. There was also a significant main effect of condition, $F(1, 24) = 43.95, p < .001$, such that single metaphors were processed more slowly than extended metaphors. In contrast, the main effect of the language group was not significant, $F(1, 24) = 0.21, p = .652$, so we conclude that overall reaction times did not differ between L1 and L2 participants.

Figure 36

Interaction: RT by Condition & Metaphor Type



In terms of interactions, a significant *stimulus type* \times *condition* interaction was found, $F(1, 24) = 4.19$, $p = .047$, reflecting that the effect of metaphor conventionality on reaction time differed depending on whether the metaphor was presented in a metaphorically priming (extended) or metaphorically non-priming (single) context (see Table 10). Additionally, a significant *language group* \times *condition* interaction emerged, $F(1, 24) = 4.34$, $p = .044$, indicating that L1 and L2 participants differ in how much the extended- single distinction affects processing times. No other interactions reached significance, including the *language group* \times *stimulus type* interaction, $F(1, 24) = 0.02$, $p = .885$, and the three-way interaction, $F(1, 24) = 0.10$, $p = .759$.

Table 10

ANOVA: RT (Metaphors Only, Both Conditions)

ANOVA: RT (Metaphors Only, Both Conditions)					
Effect	df	MSE	F	ges	p.value
language_group	1, 41	178377.45	0.21	.004	.652
stimulus_type	1, 41	12327.95	22.48 ***	.030	<.001
language_group:stimulus_type	1, 41	12327.95	0.02	<.001	.885
condition	1, 41	16953.26	43.95 ***	.076	<.001
language_group:condition	1, 41	16953.26	4.34 *	.008	.044
stimulus_type:condition	1, 41	11965.74	4.19 *	.006	.047
language_group:stimulus_type:condition	1, 41	11965.74	0.10	<.001	.759

Within L1 group post-hoc pairwise comparisons revealed a consistent pattern: *extended metaphors, regardless of novelty, were generally processed faster than their single-sentence counterparts* (see Table 11). Specifically, extended conventional metaphors were processed significantly faster than single conventional

metaphors (estimate = -134 ms, $p = 0.0076$), and even more so compared to novel single metaphors (-256 ms, $p < .0001$). Similarly, extended novel metaphors were significantly faster than novel single metaphors (-213 ms, $p < .0001$). No significant difference was found between extended conventional and extended novel metaphors, suggesting that metaphoric context may reduce the processing cost typically associated with novelty.

For L2 speakers, the same general trend was observed, though with less statistical robustness. Extended novel metaphors were processed significantly faster than novel single ones as well (-120 ms, $p = 0.0082$). Conventional single metaphors were also processed faster than novel single metaphors (-107 ms, $p = 0.0256$). However, contrasts between extended and single conventional metaphors were not significant, possibly reflecting different processing strategies in L2 participants, which I will discuss further.

Table 11

Post-hoc: RT (Metaphors Only, within Groups)

Posthoc: RT (Metaphors Only, Within Groups)						
contrast	language_group	estimate	SE	df	t.ratio	p.value
conventional.metaphor extended - novel.metaphor extended	L1	-43.55447	31.44550	41	-1.3850780	1.0000
conventional.metaphor extended - conventional.metaphor single	L1	-134.00616	38.69455	41	-3.4631795	**0.0076**
conventional.metaphor extended - novel.metaphor single	L1	-256.36140	28.66777	41	-8.9424940	**0.0000**
novel.metaphor extended - conventional.metaphor single	L1	-90.45169	45.89422	41	-1.9708734	0.3331
novel.metaphor extended - novel.metaphor single	L1	-212.80693	37.34476	41	-5.6984409	**0.0000**
conventional.metaphor single - novel.metaphor single	L1	-122.35524	37.95457	41	-3.2237292	**0.0149**
conventional.metaphor extended - novel.metaphor extended	L2	-48.92456	29.32308	41	-1.6684660	0.6171
conventional.metaphor extended - conventional.metaphor single	L2	-61.42283	36.08285	41	-1.7022722	0.5777
conventional.metaphor extended - novel.metaphor single	L2	-168.49903	26.73284	41	-6.3030729	**0.0000**
novel.metaphor extended - conventional.metaphor single	L2	-12.49827	42.79658	41	-0.2920391	1.0000
novel.metaphor extended - novel.metaphor single	L2	-119.57446	34.82417	41	-3.4336628	**0.0082**
conventional.metaphor single - novel.metaphor single	L2	-107.07619	35.39282	41	-3.0253652	**0.0256**

Pairwise comparisons of RTs between groups (L1 vs. L2) for each condition did not yield statistically significant differences. For extended conventional and extended novel metaphors, L1s were slightly faster than L2s (-68 ms and -73 ms respectively), but these differences were non-significant ($ps > 0.27$).

Table 12

Post-hoc: RT (Metaphors Only, between Groups)

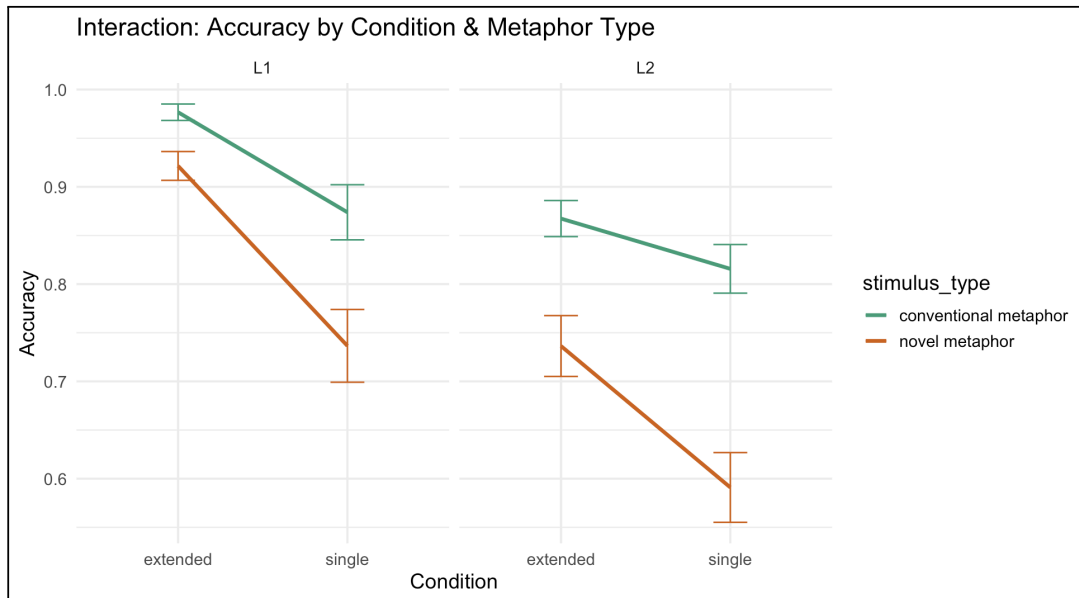
Posthoc: RT (Metaphors Only, Between Groups)							
contrast	stimulus_type	condition	estimate	SE	df	t.ratio	p.value
L1 - L2	conventional.metaphor	extended	-68.086464	61.13813	41	-1.11364984	0.2719
L1 - L2	novel.metaphor	extended	-73.456556	69.18600	41	-1.06172565	0.2946
L1 - L2	conventional.metaphor	single	4.496864	77.26996	41	0.05819679	0.9539
L1 - L2	novel.metaphor	single	19.775913	77.68452	41	0.25456696	0.8003

3.3.2.4 Accuracy rates: metaphor type x language group x condition

ANOVA analyses of the accuracy data revealed a strong main effect of *language group*, $F(1,41) = 18.36$, $p < .001$, with L1 speakers significantly outperforming L2 speakers in overall accuracy. The effect size was moderate to large, indicating a reliable group-level difference. A significant main effect of *metaphor type type* was also observed, $F(1,41) = 91.77$, $p < .001$. Participants were more accurate in interpreting conventional metaphors compared to novel ones, in line with the findings from our previous analyses. In addition, a main effect of *condition* emerged, $F(1,41) = 43.50$, $p < .001$, indicating that, as predicted, metaphors presented in the extended condition were comprehended more accurately than those presented in the single condition.

Figure 37

Interaction: Accuracy by Condition & Metaphor Type



A notable interaction *language group x stimulus type* was found, $F(1,41) = 8.14$, $p = .007$, $\eta^2 = .207$. This suggests that the advantage of conventional over novel metaphors differed by group, L2 speakers struggling more with novel metaphors both in the single and extended condition, compared to L1 participants. In contrast, there was no significant interaction *language group x condition*, $F(1,41) = 1.52$, $p = .224$. This indicates that both L1 and L2 groups benefited similarly from extended metaphor context, which, in my interpretation, proves that contextual facilitation appears to be a shared processing advantage, independent of language background.

Table 13
ANOVA: Accuracy (Metaphors Only)

ANOVA: Accuracy (Metaphors Only)						
Effect	df	MSE	F	ges	p.value	
language_group	1, 41	0.04	18.36 ***	.204	<.001	
stimulus_type	1, 41	0.01	91.77 ***	.237	<.001	
language_group:stimulus_type	1, 41	0.01	8.14 **	.027	.007	
condition	1, 41	0.01	43.50 ***	.196	<.001	
language_group:condition	1, 41	0.01	1.52	.008	.224	
stimulus_type:condition	1, 41	0.00	22.87 ***	.031	<.001	
language_group:stimulus_type:condition	1, 41	0.00	0.10	<.001	.754	

A significant interaction *stimulus type x condition* was also detected, $F(1,41) = 22.87$, $p < .001$. This pattern suggests that the benefit of contextual extension was not uniform across metaphor types: specifically, novel metaphors benefited more from being embedded in a supportive extended context, likely because they require more inferential work to interpret when presented in a neutral context. Finally, the three-way interaction between language group, stimulus type, and condition was not significant, $F(1,41) = 0.10$, $p = .754$.

Table 14
Post-hoc: Accuracy (Metaphors Only, within Groups)

Posthoc: Accuracy (Metaphors Only, Within Groups)							
contrast	language_group	estimate	SE	df	t.ratio	p.value	
conventional.metaphor extended - novel.metaphor extended	L1	0.05519821	0.02306499	41	2.393160	0.1282	
conventional.metaphor extended - conventional.metaphor single	L1	0.10284807	0.02616958	41	3.930061	**0.0019**	
conventional.metaphor extended - novel.metaphor single	L1	0.24015842	0.03677499	41	6.530483	**0.0000**	
novel.metaphor extended - conventional.metaphor single	L1	0.04764986	0.03112794	41	1.530775	0.8010	
novel.metaphor extended - novel.metaphor single	L1	0.18496021	0.03349426	41	5.522147	**0.0000**	
conventional.metaphor single - novel.metaphor single	L1	0.13731035	0.02655889	41	5.170033	**0.0000**	
conventional.metaphor extended - novel.metaphor extended	L2	0.13103410	0.02150822	41	6.092281	**0.0000**	
conventional.metaphor extended - conventional.metaphor single	L2	0.05168513	0.02440326	41	2.117960	0.2417	
conventional.metaphor extended - novel.metaphor single	L2	0.27641264	0.03429285	41	8.060357	**0.0000**	
novel.metaphor extended - conventional.metaphor single	L2	-0.07934897	0.02902695	41	-2.733631	0.0552	
novel.metaphor extended - novel.metaphor single	L2	0.14537854	0.03123356	41	4.654562	**0.0002**	
conventional.metaphor single - novel.metaphor single	L2	0.22472751	0.02476629	41	9.073925	**0.0000**	

Post-Hoc tests revealed that, for L1 speakers, both metaphor conventionality and extended context significantly improved comprehension. Accuracy was significantly higher for extended metaphors compared to single ones, regardless of whether the metaphor was novel or conventional. Specifically, accuracy was higher for conventional metaphors in the extended condition ($p = .0019$), and a similarly strong benefit was observed for novel metaphors ($p < .0001$). L1 speakers also understood conventional metaphors significantly better than novel ones when both were presented in the single-sentence condition ($p < .0001$), confirming the expected processing advantage for conventionalized expressions.

Interestingly, there was no significant difference in comprehension between novel metaphors in the extended condition and conventional metaphors in the single ($p = .8010$). This finding suggests that for native speakers, supportive discourse can compensate for a lack of strong association between the metaphor vehicle and its corresponding figurative category: novel metaphors, when elaborated in context, become just as comprehensible as conventional metaphors with no priming context.

In contrast, L2 speakers showed a more constrained pattern of effects. Like their L1 counterparts, they benefited from additional context when interpreting novel metaphors, with significantly higher accuracy in the extended condition compared to the single condition ($p = .0002$). However, in line with the reaction time findings, L2 speakers did not significantly benefit from the extension of conventional metaphors ($p = .2417$), suggesting that the addition of metaphorically priming context did not meaningfully facilitate accurate categorisation of expressions that should, in principle, be conventional to them.

Moreover, although novel metaphors presented in the extended form were somewhat more difficult than conventional metaphors in single form, the difference was only marginally significant ($p = .0552$). This result implies that, for L2 speakers, extended context helps to narrow the comprehension gap between novel and conventional metaphors, but conventional expressions are not processed with the same ease as they are by native speakers in the extended condition.

Table 15

Post-hoc: Accuracy (Metaphors Only, between Groups)

Posthoc: Accuracy (Metaphors Only, Between Groups)							
contrast	stimulus_type	condition	estimate	SE	df	t.ratio	p.value
L1 - L2	conventional.metaphor	extended	0.10931198	0.02141841	41	5.103646	**0.0000**
L1 - L2	novel.metaphor	extended	0.18514787	0.03631652	41	5.098172	**0.0000**
L1 - L2	conventional.metaphor	single	0.05814904	0.03762961	41	1.545300	0.1300
L1 - L2	novel.metaphor	single	0.14556620	0.05190535	41	2.804454	**0.0077**

Post-hoc comparisons examining accuracy between L1 and L2 speakers reveal significant group differences for most conditions, particularly when metaphors are presented with extended context. L1 speakers significantly outperformed L2 speakers on extended conventional metaphors ($p < .00001$) and extended novel metaphors ($p < .00001$), showing robust advantages in both cases. The difference was especially pronounced for novel metaphors, where the L1-L2 accuracy gap was largest. A smaller, yet statistically significant, group difference was also observed for novel metaphors in the single-sentence condition ($p = .0077$), again with L1 speakers showing higher accuracy. However, for conventional metaphors presented in single form, the group difference did not reach significance ($p = .13$), suggesting that L2 speakers were as successful as L1 participants in categorising non-primed conventional metaphors as meaningful.

3.3.2.5 Effects of conventionality as a continuous variable

Because metaphor conventionality exists on a continuum rather than as a dichotomy between novel and conventional types (see the discussion of this in Experiment 2), we conducted a regression analysis to examine how differences in conventionality ratings predict reaction times across conditions and participant groups. Rather than relying on a categorical classification, this approach allowed us to assess whether more conventional metaphors, as judged by participants, are processed more efficiently, and whether this effect is modulated by condition (single vs. extended) or language background (L1 vs. L2).

The results revealed a significant main effect of conventionality: metaphors rated as more conventional were associated with faster reaction times. Specifically, each one-point increase in rated conventionality corresponded to a 33 millisecond decrease in response latency. This aligns with prior work suggesting that conventional metaphors are more entrenched in memory and are thus accessed and integrated more fluently.

Figure 38
Reaction Time vs Conventionality: Linear Regression

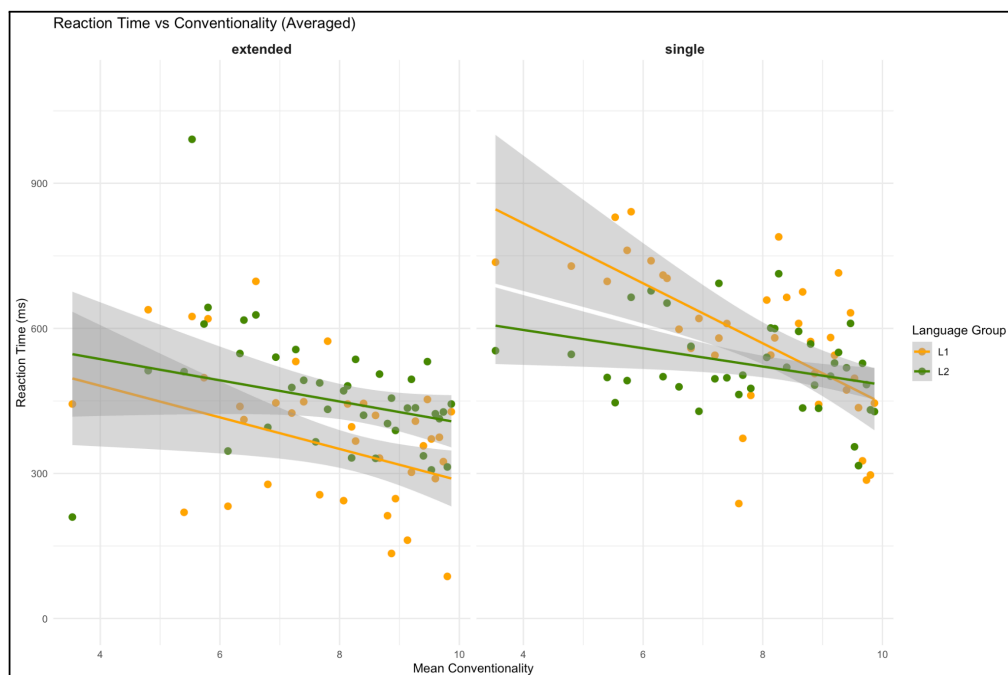


Table 16*Linear Regression: Reaction Time ~ Conventinality*

Linear Regression: Reaction Time ~ Conventinality				
term	estimate	std.error	statistic	p.value
(Intercept)	611.98711	112.23287	5.45283297	< 0.001
mean_conventinality	-32.66344	13.33435	-2.44957197	0.01433
conditionsingle	453.59717	158.60758	2.85987079	0.00425
language_groupL2	12.15444	154.30151	0.07877069	0.93722
mean_conventinality:conditionsingle	-29.38140	18.83555	-1.55989119	0.11884
mean_conventinality:language_groupL2	10.75215	18.33074	0.58656371	0.55752
conditionsingle:language_groupL2	-405.34708	217.81039	-1.86100895	0.06279
mean_conventinality:conditionsingle:language_groupL2	32.39363	25.86390	1.25246469	0.21045

Utterance condition also significantly influenced processing: metaphors presented in neutral context (single condition) elicited substantially slower responses than those embedded in an extended metaphorical context. The size of this effect — over 450 milliseconds — demonstrates the powerful role of contextual support in facilitating metaphor comprehension, particularly by providing discourse cues that constrain meaning interpretation.

Interestingly, there was no significant main effect of language group, indicating that, once conventionality and condition were controlled for, L1 and L2 speakers did not differ in their overall reaction times. However, a borderline interaction between utterance condition and language group suggested a more noteworthy pattern: L1 participants showed a much greater benefit from extended metaphorical context than their L2 counterparts. Although this interaction did not reach statistical significance, it is still indicative of a certain trend: this trend implies that while L2 speakers do process metaphorical meaning with comparable speed on average, they may rely less on discourse-level cohesion or may struggle to integrate cues across multiple

sentences as efficiently as native speakers. Crucially, there were no significant interactions between conventionality and either condition or group, indicating that the facilitative effect of conventionality was stable across both structural formats and across L1 and L2 speakers. This finding, along with the findings from our previous analyses, challenges prior claims that L2 speakers are insensitive to conventionality; instead, our data suggest that both groups benefit similarly from metaphor conventionality, even if they differ in how effectively they use extended context to support interpretation.

To complement the reaction time findings, we next examined participants' accuracy using a logistic regression model with conventionality, condition, and language group as predictors, along with all possible interactions.

Figure 39

Accuracy vs Conventionality: Logistic Regression

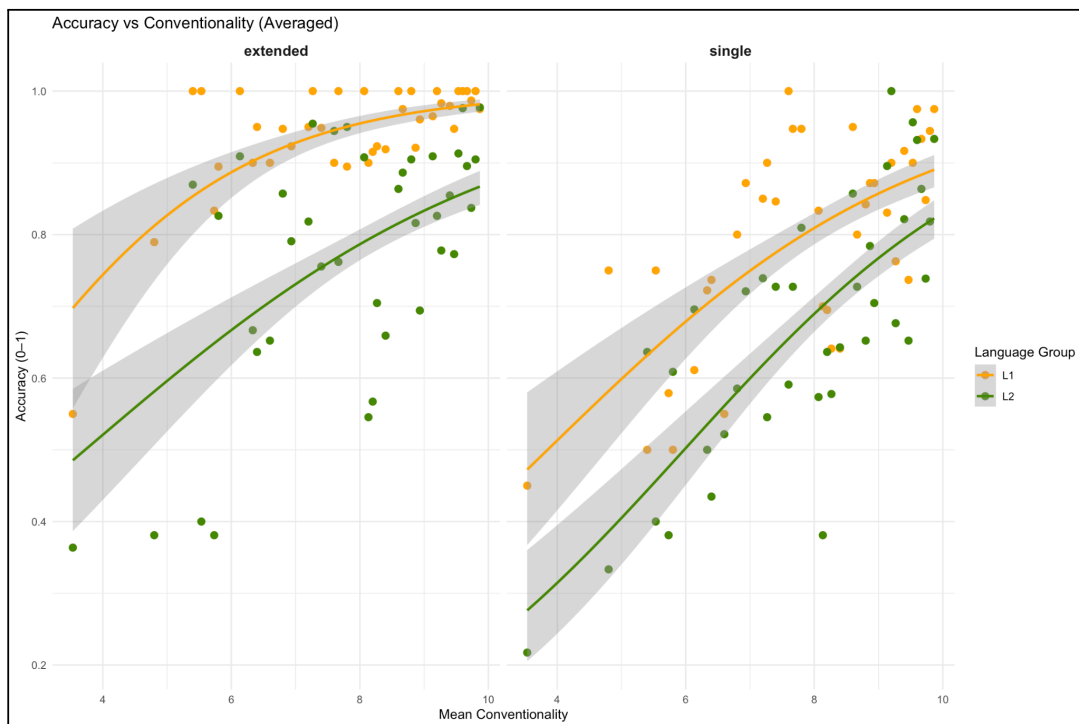
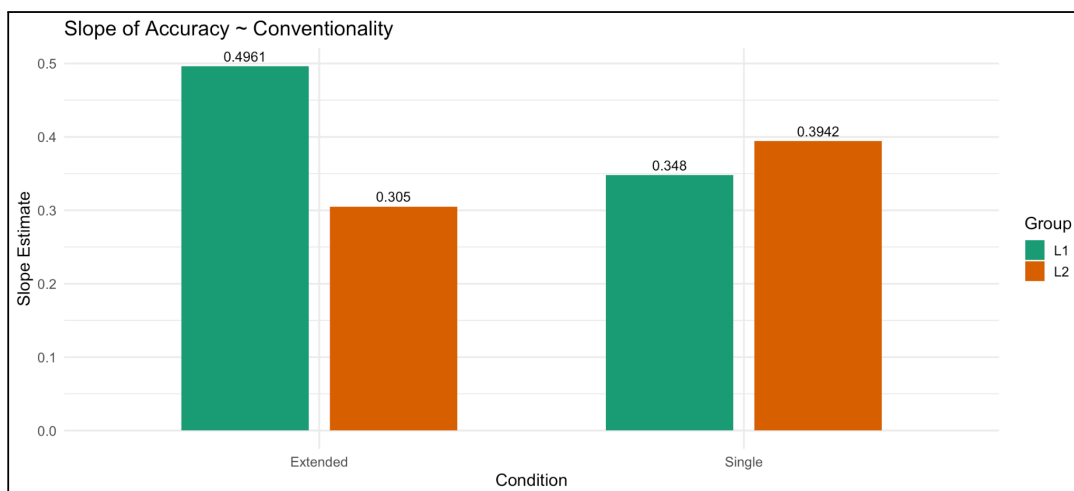


Table 17*Logistic Regression: Accuracy ~ Conventinality*

Logistic Regression: Accuracy ~ Conventinality				
term	estimate	std.error	statistic	p.value
(Intercept)	-0.9177593	0.55700023	-1.6476821	0.0994
mean_conventinality	0.4961394	0.07441316	6.6673604	<0.001
conditionsingle	-0.4225401	0.67631890	-0.6247647	0.5321
language_groupL2	-0.2188291	0.66034562	-0.3313857	0.7404
mean_conventinality:conditionsingle	-0.1480717	0.08828073	-1.6772823	0.0935
mean_conventinality:language_groupL2	-0.1910615	0.08624266	-2.2153941	0.0267
conditionsingle:language_groupL2	-0.7975706	0.83636228	-0.9536186	0.3403
mean_conventinality:conditionsingle:language_groupL2	0.2373405	0.10682242	2.2218231	0.0263

Unsurprisingly, higher conventionality was associated with significantly greater accuracy across the board ($\beta = 0.496$, $p < .001$). However, this benefit was not uniform across groups. A significant interaction between conventionality and language group ($\beta = -0.191$, $p = .027$) showed that L2 speakers benefited less from conventionality than their L1 counterparts. While both groups were more accurate when metaphors were more conventional, the slope of improvement was steeper for L1s. Crucially, a significant three-way interaction ($\beta = 0.237$, $p = .026$) qualified this pattern: for L2 speakers, the decline in conventionality-based accuracy in the single condition was attenuated. In fact, L2s appeared to benefit more from conventionality in the single condition than in the extended one — an unexpected but revealing result (see Figure 40).

Figure 40*Slope of Accuracy ~ Conventionality*

Taken together, these findings suggest that while conventionality robustly supports metaphor comprehension, it is exactly this combination of an increase in conventionality and metaphor extension that appear to result in a very marginal and attenuated facilitatory effect on metaphor processing for L2 speakers. One possible interpretation is that L2 speakers may find extended metaphorical contexts difficult to integrate, diminishing their ability to use surrounding information effectively. In contrast, when the metaphor appears in isolation (i.e., in a neutral context), those same conventional cues may be more salient or easier to apply without the added cognitive demands of integrating multiple metaphorical cues in the passage. L1 speakers may flexibly integrate vehicle-priming elements into an enriched, dynamically constructed metaphorical frame — allowing for both literal and abstract features to co-exist temporarily in working memory — L2 speakers may rely more heavily on stable, pre-stored figurative categories that are already decontextualized and very formulaic in nature. These conventional metaphor categories are often abstract and generalized, and as such, may lack the literal-specific features that

extended context attempts to reactivate. Without sufficient cognitive resources to reconcile the incoming literal features with the abstract category, L2 speakers may either ignore these features or experience interference, ultimately resulting in no net facilitation.

4. Chapter four: general discussion

The present study examined how metaphor type (novel vs. conventional) and language background (L1 vs. highly proficient L2) influence the processing of extended and single metaphors, with a focus on reaction times and accuracy. Several major findings emerged, some of which validating already existing theories in the domain of metaphor processing, and some providing new insight into processing peculiarities in metaphor comprehension on a behavioural level.

Across all analyses, a consistent main effect of metaphor type was observed: novel metaphors were processed more slowly and less accurately than conventional metaphors, for both L1 and L2 speakers. This is in strong agreement with the Career of Metaphor Hypothesis (Bowdle & Gentner, 2005), which posits that novel metaphors initially require structural alignment and analogical mapping between domains, whereas conventional metaphors are processed more categorically due to their entrenched status in the mental lexicon. Our results also align with the Graded Salience Hypothesis (Giora, 2003), which proposes that more salient meanings (e.g., conventional metaphors) are accessed faster than less salient ones (e.g., novel metaphors). The slower reaction times and lower accuracy for novel metaphors likely reflect the cognitive effort needed to construct new mappings and achieve relational consistency between the invoked domains, as opposed to retrieving already available figurative categories when interpreting conventional expressions.

The fact that both L1 and L2 speakers showed this effect serves as an indication of the universality of these cognitive demands, supporting the idea that the basic mechanisms of metaphor processing are shared across language backgrounds, although proficiency most likely modulates their efficiency. Contrary to some previous findings (Jankowiak et al., 2017; Werkmann Horvat et al., 2021), the parallel reaction time pattern observed in both L1 and L2 participants suggests that sensitivity to metaphor conventionality is not exclusive to native speakers. Rather, conventionality appears to afford similar processing benefits across linguistic backgrounds — at least in the non-priming, single metaphor condition.

4.1. Contextual facilitation: extended metaphors reduce processing costs

A crucial finding of the study was the processing advantage for extended metaphors compared to single metaphors, across both L1 and L2 groups. Reaction times were faster and accuracy was higher when metaphors were embedded within a supportive discourse context which contained vehicle-related literal priming elements. This provides empirical support for both Wilson's and Carston's theoretical accounts of extended metaphor comprehension, as they both assume a quantitative advantage for extended metaphor processing (Wilson, 2018; Carston, 2010). Interestingly, the *stimulus type* × *condition* interaction showed that novel metaphors benefited disproportionately more from an extended context than conventional metaphors. As novel mappings are initially difficult to establish, repeated metaphorical primes or supporting semantic frames help reduce processing demands by narrowing possible interpretations and reinforcing intended mappings. Context therefore acts as a scaffolding mechanism that promotes the successful interpretation of otherwise ambiguous or demanding figurative language. This facilitatory effect was evident for both L1 and L2 participants, suggesting that

contextual support is a powerful processing aid, although overall L2 participants still showed lower accuracy, especially for novel metaphors.

4.2. Language background & conventionality effects: strategic processing differences

While overall reaction times did not differ significantly between L1 and L2 speakers, accuracy rates were consistently lower in L2 speakers, particularly for novel metaphors. In our task, it seems that L2 participants appeared to adopt a less cautious and surface-based processing strategy, particularly when encountering novel metaphors and anomalous sentences. Their tendency to reject ambiguity more quickly (faster reaction times to anomalous stimuli, though not statistically significant) coupled with lower accuracy for novel metaphors suggests that L2 speakers may prioritize efficiency over depth, leading to quicker but sometimes less accurate judgments.

Conversely, L1 participants may have engaged in deeper inferential processes, attempting to recover plausible meanings even in ambiguous or anomalous cases — what might be seen as a "double-checking" strategy. This echoes findings from previous research (e.g., Kintsch, 2000) showing that native speakers are more likely to engage in coherence-building efforts during language comprehension. Thus, while both groups benefited from extended metaphor context, L2 participants were particularly vulnerable to the challenges posed by novel figurative constructions in terms of accuracy.

One of the more intriguing patterns in the data was the differential effect of context on conventional metaphor comprehension. While L1 speakers benefited from contextual support across both conventional and novel metaphors, L2 speakers only showed facilitation for the novel items. One possible explanation is that L2

participants process conventional metaphors as lexicalized chunks or familiar idioms (non-deliberate metaphors). That is, they may retrieve a stored figurative meaning without engaging in deeper integration with the surrounding discourse. If the meaning is already accessed in a formulaic manner, the metaphor extension provided by the context may be underutilized or even bypassed altogether. In this view, the lack of context effects may reflect a ceiling effect in processing: the L2 participant has retrieved what they perceive to be the intended meaning, and additional metaphorical elaboration does not noticeably enhance comprehension or decision-making efficiency. L2 speakers may rely on more piecemeal or locally constrained strategies for interpreting figurative language, attending less to discourse-level metaphor cohesion than their L1 counterparts. This might explain why context proves useful for novel metaphors, where meaning is uncertain and inferential support is needed, but has limited impact for conventional metaphors, which are already understood quite rapidly in figuratively minimal contexts.

Contrary to what was initially anticipated, the introduction of a metaphorically rich context for conventional metaphors did not result in a shift toward alignment-based processing, or at least there was no indication of that in the data collected in our experiment. The working hypothesis was that such context would reactivate the otherwise unattended-to vehicle domain, thereby triggering deeper structural mapping between topic and vehicle domains — akin to the kind of processing observed with novel metaphors. This reactivation was expected to increase cognitive demands and manifest in longer processing times, as participants would engage in more effortful cross-domain alignment rather than relying on pre-established figurative categories to be retrieved rapidly (Werkmann Horvat et al., 2023). However, the data did not support this prediction. Instead of a shift in

processing mode, the presence of metaphoric context facilitated processing. For L1 speakers, this facilitation was statistically significant, indicating that the context enhanced speed of interpretation, but it did so without altering the underlying processing architecture. In L2 speakers, a similar trend was observed, but it did not reach significance, which may reflect a less efficient or more selective use of contextual information depending on metaphor type rather than a qualitative difference in processing strategy.

Crucially, the absence of any reversal or slowdown in processing times suggests that the addition of metaphoric context did not prompt a re-evaluation of the metaphor's literal features or initiate a structural alignment process. Rather, it appears that the context simply bolstered the salience of the already conventionalized figurative meaning, making retrieval more efficient for L1s and, to a lesser extent, for L2s. This implies that, for conventional metaphors, contextual elaboration serves a supportive rather than transformative role in comprehension — reinforcing existing figurative mappings rather than restructuring them fundamentally.

4.3 Limitations and future research

While the current study provides robust evidence for how metaphor type and context interact during comprehension across L1 and L2 speakers, several limitations should be acknowledged. First, although the target metaphors were pre-rated for factors such as familiarity, metaphoricity comprehensibility, and conventionality (see results of Experiment 1), these ratings were collected in isolation, without the context sentences in which the metaphors were ultimately embedded in Experiment 3. As such, the influence of contextual scaffolding on perceived metaphoricity and comprehensibility was not systematically assessed. Future studies should consider collecting holistic ratings of the full context-plus-target

stimulus combinations to better understand how surrounding discourse modulates the interpretive salience of metaphorical content.

Second, although participants were grouped by language background, the L2 group may have been heterogeneous in terms of their L1 and cultural background which were not systematically measured or accounted for due to time constraints and general difficulty to recruit L2 participants with similar L1 backgrounds.

Third, while response times and accuracy rates offer useful indicators of overall processing difficulty, they are coarse measures that do not capture the dynamics of meaning activation and integration over time. To further look into this, future work could adopt complementary experimental paradigms such as cross-modal or regular lexical priming paradigms (e.g., Ronderos and Falkum, 2023). These would allow for the examination of how literal and metaphor-related features are activated or suppressed at various temporal stages of comprehension. Crucially, electrophysiological methods like electroencephalography (EEG) could offer additional insight into the neurocognitive processing of metaphors, and it is my intention to employ EEG in future to continue investigating extended metaphor comprehension. ERP components such as the N400 and the Late Positive Complex (LPC) are known to index semantic access and meaning integration difficulty, respectively, and could help disentangle the early versus late stages of metaphor comprehension.

While limitations remain — particularly in the need to collect context-sensitive ratings and to employ temporally sensitive methodologies — the current findings with no doubt lay valuable groundwork for future investigations. By incorporating methodologies such as EEG, cross-modal lexical priming paradigms, subsequent studies can further illuminate the time course and representational mechanisms

involved in metaphor understanding. Ultimately, such research will reveal more to us about figurative language processing, and particularly so in such a novel field as extended metaphor processing, and clarify how metaphor comprehension unfolds in real time across diverse speaker populations with more precise temporal tracking.

References

- Al-Azary, H., & Buchanan, L. (2017). Novel metaphor comprehension: Semantic neighbourhood density interacts with concreteness. *Memory & Cognition*, 45(2), 296–307. <https://doi.org/10.3758/s13421-016-0650-7>
- 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*, 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(3), 557–570. <https://doi.org/10.3758/s13421-020-01109-2>
- Al-Azary, H., & Katz, A. N. (2023). On choosing the vehicles of metaphors 2.0: The interactive effects of semantic neighborhood density and body-object interaction on metaphor production. *Frontiers in Psychology*, 14, 1216561. <https://doi.org/10.3389/fpsyg.2023.1216561>
- Al-Azary, H., McAuley, T., Buchanan, L., & Katz, A. N. (2019). Semantic processing of metaphor: A case-study of deep dyslexia. *Journal of Neurolinguistics*, 51, 297–308. <https://doi.org/10.1016/j.jneuroling.2019.04.003>
- Ardila, A., Cieślicka, A. B., Heredia, R. R., & Rosselli, M. (2017). *Psychology of bilingualism: The cognitive and emotional world of bilinguals* (Vol. 5). Springer International Publishing. <https://doi.org/10.1007/978-3-319-64099-0>
- Arnold, I. (2002). *Stilistika sovremennogo angliyskogo yazyka* [Contemporary English stylistics]. Moscow: Prosveshenye.
- Arnold, I. (2014). *Leksikologiya sovremennogo angliyskogo yazyka* [Lexicology of modern English] (3rd ed., rev.). Moscow: Flinta: Nauka.

- Arzouan, Y., Goldstein, A., & Faust, M. (2007). Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Research, 1160*, 69–81. <https://doi.org/10.1016/j.brainres.2007.05.034>
- Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M., & Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin, 86*(3), 203–216. <https://doi.org/10.1016/j.brainresbull.2011.07.015>
- Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(2), 295–308. <https://doi.org/10.1037/0278-7393.19.2.295>
- Bowdle, B. F., & Gentner, D. (1999). Metaphor comprehension: From comparison to categorization. *Proceedings of the Annual Meeting of the Cognitive Science Society, 21*, 90–95. <https://escholarship.org/uc/item/8qs8n92v>
- Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*(1), 193–216. <https://doi.org/10.1037/0033-295X.112.1.193>
- Bolognesi, M., & Werkmann Horvat, A. (2022). *The metaphor compass: Directions for metaphor research in language, Cognition, Communication, and Creativity* (1st ed.). Routledge. <https://doi.org/10.4324/9781003041221>
- Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., Strauss, M. M., Hyman, S. E., & Rosen, B. R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17*(5), 875–887. [https://doi.org/10.1016/S0896-6273\(00\)80219-6](https://doi.org/10.1016/S0896-6273(00)80219-6)

- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, 8(3), 531–544. <https://doi.org/10.3758/BF03196189>
- Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Blackwell Publishing.
- Carston, R. (2010). Metaphor: Ad hoc concepts, literal meaning and mental images. *Proceedings of the Aristotelian Society*, 110, 295–321.
- Cersosimo, R., Domaneschi, F., & Al-Azary, H. (2024). Automatic metaphor processing in developmental dyslexia. *Journal of Communication Disorders*, 111, 106448. <https://doi.org/10.1016/j.jcomdis.2024.106448>
- Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, 30(6), 958–968. <https://doi.org/10.3758/bf03195780>
- Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.
- Friedman, D., & Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research and Technique*, 51(1), 6–28. [https://doi.org/10.1002/1097-0029\(20001001\)51:1<6::AID-JEMT2>3.0.CO;2-R](https://doi.org/10.1002/1097-0029(20001001)51:1<6::AID-JEMT2>3.0.CO;2-R)
- Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, 113(2), 256–281. <https://doi.org/10.1037/0096-3445.113.2.256>

- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2), 155–170.
https://doi.org/10.1207/s15516709cog0702_3
- Gentner, D., & Clement, C. (1988). Evidence for relational selectivity in the interpretation of analogy and metaphor. *Psychology of Learning and Motivation*, 22, 307–358. [https://doi.org/10.1016/s0079-7421\(08\)60044-4](https://doi.org/10.1016/s0079-7421(08)60044-4)
- Gentner, D., Falkenhainer, B., & Skorstad, J. (1988). Viewing metaphor as analogy. In D. H. Helman (Ed.), *Analogical reasoning: Perspectives of artificial intelligence, cognitive science, and philosophy* (pp. 171–177). Springer Netherlands. https://doi.org/10.1007/978-94-015-7811-0_8
- Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language*, 37(3), 331–355.
<https://doi.org/10.1006/jmla.1997.2527>
- Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.
- Giora, R. (2003). *On our mind: Salience, context, and figurative language*. Oxford University Press.
- Glucksberg, S. (2001). *Understanding figurative language: From metaphor to idioms* (Vol. 36). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780195111095.001.0001>
- Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, 7(2), 92–96. [https://doi.org/10.1016/S1364-6613\(02\)00040-2](https://doi.org/10.1016/S1364-6613(02)00040-2)
- 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., Gildea, P., & Bookin, H. B. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, 21(1), 85–98. [https://doi.org/10.1016/S0022-5371\(82\)90467-4](https://doi.org/10.1016/S0022-5371(82)90467-4)
- Glucksberg, S., McGlone, M. S., & Manfredi, D. (1997). Property attribution in metaphor comprehension. *Journal of Memory and Language*, 36(1), 50–67. <https://doi.org/10.1006/jmla.1996.2479>
- Glucksberg, S., & McGlone, M. S. (2001). *Understanding figurative language: From metaphors to idioms*. Oxford University Press.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5), 515–530. <https://doi.org/10.1080/20445911.2013.796377>
- Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science*, 10(5), 164–168. <https://doi.org/10.1111/1467-8721.00140>
- Heredia, R. R., & Cieślicka, A. B. (Eds.). (2015). *Bilingual figurative language processing*. Cambridge University Press.
- Heredia, R. R., & Muñoz, M. E. (2015). Metaphoric reference: A real-time analysis. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 89–116). chapter, Cambridge: Cambridge University Press.
- Ikuta, M., & Miwa, K. (2021). Structure Mapping in Second-Language Metaphor Processing. *Metaphor and Symbol*, 36(4), 288–310. <https://doi.org/10.1080/10926488.2021.1941971>
- Jankowiak, K. (2020). Normative data for novel nominal metaphors, novel similes, literal, and anomalous utterances in Polish and English. *Journal of*

Psycholinguistic Research, 49(4), 541–569.

<https://doi.org/10.1007/s10936-020-09695-7>

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), e0175578. <https://doi.org/10.1371/journal.pone.0175578>

Jensen, D. (2006). Metaphors as a bridge to understanding educational and social contexts. *International Journal of Qualitative Methods*, 5(1), 36–54.

<https://doi.org/10.1177/160940690600500104>

Johnson, J., & Rosano, T. (1993). Relation of cognitive style to metaphor interpretation and second language proficiency. *Applied Psycholinguistics*, 14(2), 159–175. <https://doi.org/10.1017/S014271640000953X>

Johnson, J. M. (1996). Metaphor interpretations by second language learners: Children and adults. *The Canadian Modern Language Review*, 53(1), 219–241. <https://doi.org/10.3138/cmlr.53.1.219>

Keckes, I. (2006). On my mind: Thoughts about salience, context and figurative language from a second language perspective. *Second Language Research*, 22(2), 219–237. <https://doi.org/10.1191/0267658306sr266ra>

Keysar, B., Shen, Y., Glucksberg, S., & Horton, W. S. (2000). Conventional language: How metaphorical is it? *Journal of Memory and Language*, 43(4), 576–593. <https://doi.org/10.1006/jmla.2000.2711>

- Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, 7(2), 257–266. <https://doi.org/10.3758/BF03212981>
- Krennmayr, T. (2013). Top-down versus bottom-up approaches to the identification of metaphor in discourse. *Metaphorik.de*, 24, 7–36.
- Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, 1284, 145–155.
<https://doi.org/10.1016/j.brainres.2009.05.088>
- Lakoff, G., & Johnson, M. (1980). Conceptual metaphor in everyday language. *The Journal of Philosophy*, 77(8), 453–486. <https://doi.org/10.2307/2025464>
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Littlemore, J. (2001). The use of metaphor in university lectures and the problems that it causes for overseas students. *Teaching in Higher Education*, 6(3), 333–349. <https://doi.org/10.1080/13562510120061205>
- Littlemore, J. (2010). Metaphoric competence in the first and second language: Similarities and differences. In *Cognitive Processing in Second Language Acquisition* (Vol. 13, pp. 293–316). John Benjamins Publishing Company.
<https://doi.org/10.1075/celcr.13.20lit>
- 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>
- Lü, J., Liang, L., & Chen, B. (2019). The effect of executive control ability on the comprehension of second language metaphor. *International Journal of Bilingualism*, 23(1), 87–101. <https://doi.org/10.1177/1367006917709096>

- López, B. G., Vaid, J., Tosun, S., & Rao, C. (2017). Bilinguals' plausibility judgments for phrases with a literal vs. non-literal meaning: The influence of language brokering experience. *Frontiers in Psychology*, *8*, 1661. <https://doi.org/10.3389/fpsyg.2017.01661>
- Mathisen, E. K. (2023). Processing differences between single and extended metaphors and similes [Master's thesis, University of Oslo]. <https://www.duo.uio.no/handle/10852/104394>
- Mercier, J., Pivneva, I., & Titone, D. (2014). Individual differences in inhibitory control relate to bilingual spoken word processing. *Bilingualism: Language and Cognition*, *17*(1), 89–117. <https://doi.org/10.1017/S1366728913000084>
- Miller, G. A. (1993). Images and models, similes and metaphors. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 357–400). chapter, Cambridge: Cambridge University Press.
- Montrul, S. (2015). Dominance and proficiency in early and late bilingualism. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals* (1st ed., pp. 15–35). Cambridge University Press. <https://doi.org/10.1017/CBO9781107375345.002>
- Neuman, Y., Assaf, D., Cohen, Y., Last, M., Argamon, S., Howard, N., & Frieder, O. (2013). Metaphor identification in large texts corpora. *PLoS ONE*, *8*(4), e62343. <https://doi.org/10.1371/journal.pone.0062343>
- Neves, E. P., Cravo, A. M., & Carthery-Goulart, M. T. (2021). Functionality of the N400 component and its application in studies of figurative language processing: A systematic review. *Revista Linguagem Em Foco*, *13*(4), Article 4. <https://doi.org/10.46230/2674-8266-13-7263>

- Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, 9(3), 225–236. <https://doi.org/10.3758/BF03196957>
- Onishi, K. H., & Murphy, G. L. (1993). Metaphoric reference: When metaphors are not understood as easily as literal expressions. *Memory & Cognition*, 21(6), 763–772. <https://doi.org/10.3758/BF03202744>
- Peng, Z., & Khatin-Zadeh, O. (2023). Research on metaphor processing during the past five decades: A bibliometric analysis. *Humanities and Social Sciences Communications*, 10(1), 928. <https://doi.org/10.1057/s41599-023-02465-5>
- Pynte, J., Besson, M., Robichon, F.-H., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain and Language*, 55(3), 293–316. <https://doi.org/10.1006/brln.1996.0107>
- Recanati, F. (2004). *Literal meaning*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511615382>
- Ronderos, C. R., & Falkum, I. L. (2023). Suppression of literal meaning in single and extended metaphors. *Frontiers in Psychology*, 14, 1135129. <https://doi.org/10.3389/fpsyg.2023.1135129>
- Rubio-Fernandez, P. (2007). Suppression in metaphor interpretation: Differences between meaning selection and meaning construction. *Journal of Semantics*, 24(4), 345–371. <https://doi.org/10.1093/jos/ffm006>
- Rubio-Fernández, P., Cummins, C., & Tian, Y. (2016). Are single and extended metaphors processed differently? A test of two Relevance-Theoretic accounts. *Journal of Pragmatics*, 94, 15–28. <https://doi.org/10.1016/j.pragma.2016.01.005>

- Salthouse, T. A., & Hedden, T. (2002). Interpreting reaction time measures in between-group comparisons. *Journal of Clinical and Experimental Neuropsychology*, 24(7), 858–872. <https://doi.org/10.1076/jcen.24.7.858.8392>
- Searle, J. R. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 92–123). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511609213.006>
- Sperber, D., & Wilson, D. (1986). IX—Loose talk. *Proceedings of the Aristotelian Society*, 86(1), 153–171. <https://doi.org/10.1093/aristotelian/86.1.153>
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Blackwell.
- Sperber, D., & Wilson, D. (2008). A deflationary account of metaphors. In *The Cambridge handbook of metaphor and thought* (pp. 84–106). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816802.007>
- Steen, G. (2009). Deliberate metaphor affords conscious metaphorical cognition. *Cognitive Semiotics*, 5(1–2), 179–197.
<https://doi.org/10.1515/cogsem.2013.5.12.179>
- Steen, G. J. (2016). Mixed metaphor is a question of deliberateness. In *Mixing metaphor* (Vol. 6, pp. 113–132). John Benjamins Publishing Company.
<https://doi.org/10.1075/milcc.6.06ste>
- Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18(5), 523–534.
[https://doi.org/10.1016/S0022-5371\(79\)90284-6](https://doi.org/10.1016/S0022-5371(79)90284-6)
- Thibodeau, P. H., & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PLoS ONE*, 6(2), e16782.
<https://doi.org/10.1371/journal.pone.0016782>

- Thibodeau, P., & Durgin, F. H. (2008). Productive figurative communication: Conventional metaphors facilitate the comprehension of related novel metaphors. *Journal of Memory and Language*, 58(2), 521–540.
<https://doi.org/10.1016/j.jml.2007.05.001>
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327–352.
<https://doi.org/10.1037/0033-295X.84.4.327>
- Valera, S. (2020). Polysemy versus homonymy. Oxford University Press.
<https://doi.org/10.1093/acrefore/9780199384655.013.617>
- Wang, X., & Jankowiak, K. (2023). Electrophysiological insights into the role of proficiency in bilingual novel and conventional metaphor processing. *Linguistic Approaches to Bilingualism*, 13(2), 163–189.
<https://doi.org/10.1075/lab.20051.wan>
- Werkmann Horvat, A., Bolognesi, M., & Althaus, N. (2023). Attention to the source domain of conventional metaphorical expressions: Evidence from an eye tracking study. *Journal of Pragmatics*, 215, 131–144.
<https://doi.org/10.1016/j.pragma.2023.07.011>
- Werkmann Horvat, A., Bolognesi, M., & Lahiri, A. (2021). Processing of literal and metaphorical meanings in polysemous verbs: An experiment and its methodological implications. *Journal of Pragmatics*, 171, 131–146.
<https://doi.org/10.1016/j.pragma.2020.10.007>
- Werkmann Horvat, A., Bolognesi, M., Littlemore, J., & Barnden, J. (2022). Comprehension of different types of novel metaphors in monolinguals and multilinguals. *Language and Cognition*, 14(3), 401–436.
<https://doi.org/10.1017/langcog.2022.8>

- Werth, P. (1994). Extended metaphor—A text-world account. *Language and Literature: International Journal of Stylistics*, 3(2), 79–103.
<https://doi.org/10.1177/096394709400300201>
- Wiley, J., & Rayner, K. (2000). Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements. *Memory & Cognition*, 28(6), 1011–1021. <https://doi.org/10.3758/BF03209349>
- Wilson, D. (2018). Relevance theory and literary interpretation. In T. Cave & D. Wilson (Eds.), *Reading Beyond the Code: Literature and Relevance Theory* (pp. 185-204). Oxford University Press.
<https://doi.org/10.1093/oso/9780198794776.003.0011>
- Wilson, D., & Carston, R. (2006). Metaphor, relevance and the “emergent property” issue. *Mind & Language*, 21(3), 404–433.
<https://doi.org/10.1111/j.1468-0017.2006.00284.x>
- Wilson, D., & Carston, R. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 230–259). Palgrave Macmillan.
- Wolff, P., & Gentner, D. (2011). Structure-mapping in metaphor comprehension. *Cognitive Science*, 35(8), 1456–1488.
<https://doi.org/10.1111/j.1551-6709.2011.01194.x>
- Xu, X. (2010). Interpreting metaphorical statements. *Journal of Pragmatics*, 42(6), 1622–1636. <https://doi.org/10.1016/j.pragma.2009.11.005>