The Metaphor Awakening Effect:
A Time-Course Investigation of the Literal Meaning during Metaphor Comprehension

Laura Pissani

A Thesis

In the Department of

Psychology

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Psychology)

at Concordia University

Montreal, Quebec, Canada

July 2023

CONCORDIA UNIVERSITY SCHOOL
OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: **Laura Pissani**

Entitled: **The metaphor awakening effect: A time-course investigation of the literal meaning during metaphor comprehension**

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

DOCTOR OF PHILOSOPHY (Psychology)

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

Signed by the final examining committee:

_______________________________________ Chair
Dr. A. Bale

_______________________________________ External Examiner
Dr. E. Smith

_______________________________________ External to Program
Dr. P. Trofimovich

_______________________________________ Examiner
Dr. D. Poulin Dubois

_______________________________________ Examiner
Dr. N. Segalowitz

_______________________________________ Supervisor
Dr. R. G. de Almeida

Approved by ____
Dr. Andrew Chapman, Graduate Program Director

September 6th, 2023 ____
Dr. Pascale Sicotte, Dean
Faculty of Arts and Science

Abstract

The metaphor awakening effect: A time-course investigation of the literal meaning during metaphor comprehension

Laura Pissani, Ph.D.
Concordia University, 2023

Metaphors have been an object of fascination and a matter of debate since ancient times. What has attracted researchers is how metaphors are so seamlessly understood when their literal meaning differs from what they convey metaphorically. Some scholars have proposed that listeners attain the metaphorical content serially, where the literal interpretation is initially derived, combined with pragmatic information, and then the metaphorical content is determined. Other scholars, however, have argued that the efficiency with which metaphors are understood does not allow for the literal meaning to be derived first and then rejected in favour of the metaphorical content. Rather, they contend that most conventional metaphors are directly retrieved from semantic memory without the need for any inferential work. This thesis presents three manuscripts that investigated two-word metaphors such as *broken heart* and *sharp tongue*. The first manuscript reported a norming study to serve as an open source of materials required to run experiments such as those in the current thesis. The second manuscript examined whether the literal meaning of conventional metaphors was available, and could be recovered, immediately after the metaphorical content had been attained. In a maze task, participants read sentences word by word in a self-paced manner and then choose which of two words correctly continues the sentence, where the distractor words were either related or unrelated to the metaphorical content of the sentence. The results of this study yielded a significant awakening effect, whereby longer response times and lower accuracy rates were obtained in trials in which the literal meaning was cued immediately after the metaphor had been processed. This pattern of results suggests that the literal meaning was awakened during sentence processing. The third manuscript examined whether the awakening effect could be found further away from the metaphorical expression. The results of this study also yielded a significant awakening effect. However, it was weaker when compared to the original maze. Lastly, in Appendices A and B, the effects of familiarity and aptness on the awakening effect were analyzed. The results of these analyses indicated that, when the literal meaning is cued immediately after the metaphorical expression has been processed, familiarity and aptness do not have an overall effect. However, further downstream, the awakening effect is indeed modulated by familiarity and aptness. Altogether, the results from the series of studies presented in the current thesis provide compelling evidence in support of the literal-first approach.

Acknowledgements

This endeavor would not have been possible without Roberto G. de Almeida. He mentored me and supported me throughout my years at Concordia University. I am grateful for his thoughtful supervision, advice, and lessons during my studies. He was encouraging and trusting, which boosted my confidence when I needed it the most. I am also grateful for his sense of humour, sociability, and flexibility, which made my Ph.D. journey a more fulfilling and enjoyable experience.

I also could not have undertaken this journey without my thesis committee, Norman Segalowitz and Diane Poulin-Dubois, who generously provided knowledge and expertise. Additionally, this endeavour would not have been possible without the support from the Fonds de recherche du Québec, who financed my doctoral program.

I am also fortunate to have an incredible lab team. Thanks to Caitlyn Antal, Kyan Salehi, and Tobias Ungerer for their encouragement, feedback, and top-notch advice. Most importantly, thanks for cheering me up before conference talks. Thanks to all the volunteers in the lab who helped me recruit participants and code data. I would also like to recognize Jessie Gratton for being my right hand, always willing to assist, and always bringing positive energy.

Special thanks to my partner and colleague, Christopher Genovesi, for his constant support, his wisdom, and his encouragement. Thanks for inspiring me, for taking the time to listen to me, for motivating me every day to keep going, and for always being there for me.

I could not be where I am today without my support system, my family and friends. Thanks to my mom for always believing in me, for her unconditional love, and for her daily support from afar. Thanks to my brother because I know I can always count on him. Thanks to my grandfather for his knowledge, his lessons, and his stories. Thanks to my best friends, Ruth Yancce, Catarina Borges, and Nada Hafez. Thanks for hearing me vent, for boosting my mood, and for adding balance to my life.

Big thanks to Hamad Al-Azary, without whom I would not be doing my Ph.D. in Canada today. Thanks for his mentorship, his friendship, and his words of wisdom at every step of the way, especially during the toughest times.

To every single one of you, thank you, you did for me way more than you know.

Contribution of Authors

For all presented manuscripts, Laura Pissani was the lead author and was closely supervised by Dr. Roberto G. de Almeida through the progression of the studies. Dr. de Almeida was greatly involved in the writing of the introduction and discussion of all manuscripts and he is the second author of all papers herein.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Aristotle wrote that mastering the *use* of metaphor was a sign of genius. Since, according to Aristotle, this is an innate ability, I will settle for mastering the *study* of metaphor.

From ancient times, metaphor has allured academics in philosophy, psychology, linguistics, and other domains. This may be so because of its pervasiveness in language or because of its power to convey meaning in a way that literal statements do not—and perhaps, cannot. Or, most convincingly, it may be so because even though most metaphors are blatant lies, they are taken sincerely, and seamlessly understood by interlocutors. A metaphor is a rhetorical device whereby one thing is presented in terms of another. The thing being described is the *topic*, while the thing used to convey its metaphorical sense is the *vehicle*. Consider Romeo, for instance, when he declares that *Juliet is the sun*. His fervent love for Juliet (the topic) inspired him to represent her as the sun (the vehicle). And, knowing Juliet is not a large, celestial body, one can promptly understand that he was talking about her beauty and all the great qualities in her that one can associate with the sun.

The basic idea is that, when speaking metaphorically, there is a divergence between what the words mean literally and what they convey metaphorically. For instance, when someone confesses having a *broken heart*, competent speakers understand that the person is referring to a state of extreme grief and despair rather than to a ruptured organ. In this instance, the intended meaning differs from the meaning of its components. Instead, the topic of the metaphor is an emotional state—conveyed by *heart*—and the vehicle is *broken*, which predicates the commotion of this emotional state. Nonetheless, competent speakers reach the intended meaning rather fast and efficiently. This mystery has been at the *heart* of research in metaphor: To wit, how do speakers infer the intended meaning of a statement when it differs from the literal one?

Some headway has been made since the first experimental study addressing this question (Clark & Lucy, 1975). However, the idea that the literal meaning is crucial during metaphor comprehension has been rejected by certain scholars (see Holyoak & Stamenković, 2018 for a review), claiming instead that metaphorical meaning is attained directly without any inferential work. What would then be the difference between metaphor and polysemy? The current thesis aims to address the premature resolution that the three-stage model has been "quite conclusively rejected" (p. 644), which has been made based on a few studies unsuited to tap into the early stages of metaphor processing, where, by hypothesis, the literal meaning takes place.

The direct-indirect access debate

Two main models have attempted to describe how metaphorical meaning is derived and whether it differs from how literal meaning is retrieved. The first, historically, is the indirect model. It assumes that during conversation, speaker and listener are in mutual agreement to cooperate (Grice, 1975). This cooperative behaviour is guided by subordinate maxims of conversations. Thus, if one of the maxims of conversation is flouted, the listener remains under the assumption that the overall cooperative principle is being observed. So, the listener believes that the violation is taking place only at the level of *what is said*, but it is still observed at the level of *what it is intended* (Grice, 1975). Further, Searle (1979) proposes that when the listener recognizes that an utterance is not meant literally, s/he is prompted to search for an alternative meaning. Accordingly, the indirect model proposes a serial account of metaphor interpretation, where the literal meaning is first accessed and subsequently rejected if it results unsuitable in the given context. Then, combined with rules of conversation, the listener searches for an alternative non-literal interpretation, which triggers a third stage (Grice, 1975; Janus & Bever, 1985; Searle, 1979). Based on these observations, Clark and Lucy (1975) outlined three stages that the listener undergoes to attain the intended meaning of a figurative expression. In the first stage, a proposition that corresponds to the literal meaning of the sentence is formed. In the second stage, this literal proposition is tested to determine whether it is plausible and appropriate in context. If that is the case, the literal interpretation is accepted as the intended meaning. Otherwise, the literal interpretation is rejected as the intended meaning, in which case, an additional stage is prompted. In the third stage, the literal meaning is combined with pragmatic information in order to derive the intended, figurative meaning.

The three-stage model outlined by Clark and Lucy (1975) has three predictions:

> Prediction 1: The listener should show evidence that he had come to the literal interpretation of a sentence before he had come to its conveyed interpretation. For example, if one sentence is positive and another negative and yet both have the same conveyed meaning, we might expect the negative to take longer to comprehend, since previous work on negatives indicates this should happen for literal meaning.
>
> Prediction 2: The listener should take longer whenever the intended meaning is different from the literal meaning, all other things being equal. This prediction is based on the assumption that deducing the conveyed meaning from the literal meaning takes time. Thus, *It's stuffy in here, Jeeves*, should take less time to comprehend, if all else is equal, when taken as a comment about stale air than when taken as a request to open the window. The stickler in Prediction 2 is the condition "all other things being equal." The listener attempting to understand *It's stuffy in here, Jeeves*, by [this] model, has to register the context, and by definition, the context will not be the same under the first and second interpretations. So there is always the possibility that the context takes longer to register in one case than in the other, thereby confounding the time difference of Prediction 2. Despite this difficulty, however, one can bring plausible, even if not conclusive, evidence to bear on Prediction 2.
>
> Prediction 3: The listener should show evidence that his final representation of a sentence is its intended meaning. So when the intended meaning differs from the literal meaning, as in conveyed requests, he should be using the conveyed meaning, not the literal meaning, in all further uses of the interpretation of the sentence (Clark & Lucy, 1975, pp. 58-59).

The outline of the three-stage model of metaphor comprehension allowed previous theoretical postulates to be tested empirically. The first prediction assumes that, when two utterances convey the same meaning, the one that does so indirectly will take longer to process. In line with this prediction, Clark and Lucy (1975) found that positive requests were judged true faster than negative ones, while negative requests were judged false faster than positives, and positive requests were judged faster than negative requests. The second prediction entails that metaphorical statements take longer to process than literal ones. This prediction has been the most widely tested, with evidence for and against depending on the methods employed, some of which are discussed in the current and next sections. The third prediction refers to whether the intended meaning is indeed attained by the listener. This prediction has been considered by most researchers by including sensicality judgments or probe questions in their experimental studies. In sum, Clark and Lucy’s model suggests that figurative statements behave differently than literal ones. In particular, it suggests that figurative expressions are harder to process than literal ones, given that figurative content is derived only after the literal interpretation has been built.

Contrary to the three-stage model, the direct access view contends that metaphor comprehension occurs automatically and effortlessly. That is, the literal meaning of metaphorical expressions is not accessed and subsequently rejected. Instead, the metaphorical content is directly retrieved. Proponents of the direct access view claim that the literal meaning is bypassed during metaphor comprehension and that comprehension of literal and metaphorical statements require the use of the same cognitive strategies and mechanisms (Glucksberg et al., 1982). Thus, the direct access view challenges the three-stage model on both empirical and theoretical grounds.

For instance, Ortony et al. (1978) argued that the three-stage model is restricted to certain cases, such as expressions presented in isolation. To that effect, Ortony and colleagues examined the effects of context on metaphor comprehension. The authors measured reading time to a target sentence (e.g., *Regardless of the danger, the troops marched on*), which could be interpreted literally or metaphorically based on the preceding context, which also varied in length (i.e., short and long) and type (i.e., biasing towards a literal or metaphorical interpretation of the target sentence). Their results indicated that literal target sentences yielded longer reading times than their metaphorical counterparts only when the preceding context was short. Thus, they argued that supportive context—rather than the literalness of the expression—determines the time it takes for an expression to be processed. In turn, Glucksberg et al. (1982) tested the three-stage model's tenet that metaphorical meaning is only attained after the literal interpretation of the expression has failed in context (Searle, 1979). In a series of experiments, participants were asked to verify whether sentences were literally true or false. Glucksberg and colleagues found a large metaphor interference effect, wherein metaphors (e.g., *Some jobs are jails*) took significantly longer than scrambled metaphors (e.g., *Some jobs are birds*) to be judged false. The authors claimed that participants had generated the metaphorical interpretation of the expressions even in cases where the literal meaning was sufficient. Therefore, they concluded that the interpretation of the metaphorical meaning is automatic. Similarly, Keysar (1989) employed a verification task but included a follow-up experiment, in which participants were only required to read the sentences (without making judgements) while their reading time to the full sentence was measured. Keysar (1989) replicated Glucksberg et al.'s findings. Thus, their results were consistent with the conclusion that access to metaphorical meaning is automatic, even in cases where the literal meaning is appropriate.

Against these claims, Janus and Bever (1985) argued that to test the predictions of the three-stage model properly, two conditions must be met. First, reading time measurements should be restricted to the target phrase rather than to the full sentence. This allows one to capture the precise moment at which either metaphorical or literal interpretation happens and avoids extra processing costs at the end of the sentence. Second, metaphor comprehension must be evaluated in context. This is so because, as specified in the second stage, the derived meaning is tested for suitability *in context*. To account for these conditions, Janus and Bever employed Ortony et al.'s materials but rather than measuring reading time at the offset of the full sentence (e.g., *the fabric had begun to fray*), they measured it at the offset of the constituent boundary (e.g., *the fabric).*

Their results showed that metaphorical target phrases were read slower than literal ones after supportive context. In a separate rating study, target phrases were found to be more predictable following literal rather than metaphorical biasing context. This may have affected reading time—although Janus and Bever argued that the predictability of the target phrase did not correlate with reading time. Overall, this pattern of results is compatible with the idea that metaphors are processed serially, wherein the literal meaning is derived first, contrasted in context, and then rejected in favour of a metaphorical interpretation. Moreover, Inhoff et al. (1984) replicated Ortony et al.'s findings by employing eye tracking. However, they measured reading time as the total duration of fixations made to the full sentence rather than to the target phrase. Interestingly, they found that metaphorical sentences yielded a higher number of regressions in comparison to literal sentences, which—consistent with the three-stage model—was interpreted as participants disputing their initial interpretation.

The issue of familiarity and aptness

One of the crucial issues in the dispute between direct and indirect models is the familiarity and aptness of a metaphor. Blasko and Conine (1993) found evidence that metaphors rated highly familiar were accessed as rapidly as literal statements. Interestingly, low familiar metaphors were also accessed rapidly when they were rated highly apt. These findings suggest that metaphorical expressions can be processed at the same speed as literal statements provided the former are either highly familiar or highly apt.

In this sense, familiarity constitutes an important variable for understanding the nature of metaphor comprehension because conventional metaphors are usually associated with a particular alternative content. Conventional metaphors, in this regard, differ from novel metaphors, whose metaphorical content is indeterminate, or less salient, and needs to be inferred by the listener in real time.

The career-of-metaphor hypothesis (Bowdle & Gentner, 2005) poses that novel metaphors are understood via comparison when they are first encountered. In this manner, the metaphorical content is interpreted by aligning the representations of the topic and vehicle and importing predicates from the latter to the former. The aligned predicates then become more strongly activated with frequent use, while irrelevant predicates are suppressed. Over time, this process yields a metaphorical category, which—if found to provide useful information—becomes productive and it is paired with other topics. For instance, in the novel metaphor *nuclear anger,* the vehicle *nuclear* is used as “excessive”, but this sense has not yet been generalized to other topics (e.g., *nuclear love*, *nuclear thirst, nuclear pain*). Conversely, conventional metaphors involve polysemous vehicles, with both literal and metaphorical senses already associated. For instance, in *bright screen*, the vehicle *bright* is used as “smart” (as in *bright student*), which is a metaphor category, and it is paired with other topics (e.g., *bright idea*, *bright memory*, *bright professor*). Therefore, Bowdle and Gentner argue that different mechanisms are involved in understanding novel and conventional metaphors. On the one hand, novel metaphors are understood via analogy, comparison, or feature mapping, where access to the literal meaning is required. On the other hand, most conventional metaphors are understood via categorization, where the metaphorical content is recovered directly from semantic memory thus bypassing the literal meaning.

Most scholars agree that novel and conventional metaphors are processed differently such that the former “invite sense creation” while the latter “invite sense retrieval” (Bowdle & Gentner, 2005, p. 199). In particular, it has been argued that most conventional metaphors are stored as lexicalized, noncompositional expressions and, as such, do not require further inferential operations to derive the conventional content associated with them (Glucksberg, 2003; Keysar et al., 2000).

Crucial to this issue, however, is the time course of events and whether the methods employed are able to account for it. In particular, offline tasks used to examine the difference in comprehension between literal and metaphorical statements do not tap into the early moments of metaphor comprehension (e.g., Glucksberg et al., 1982; Harris, 1976; Keysar, 1989; McElree & Nordlie, 1999; Ortony et al., 1978; Pollio et al., 1984). That is, offline methods do not capture the moment in which the listener uses the initial input to calculate the speaker's intentions. For instance, in the initial stages of metaphor processing, one may rely on an early parsing that is based on linguistic (viz., syntactic-semantic) predicate-argument relations, which are then further interpreted through non-linguistic (viz., pragmatic) inferential processes. In contrast, online methods, which tap into the early moments of metaphor comprehension, have demonstrated that literal meaning plays a more central role than previously thought. For example, a number of studies employing online methods show differences between metaphors and literal expressions in reading time (Brisard et al., 2001; Janus & Bever, 1985; Shibata et al., 2012), priming effects (Rubio Fernández, 2007; Patalas & de Almeida, 2019), ERPs (De Grauwe et al., 2010; Lai et al., 2009; Weiland et al., 2014), and eye movements (Ashby et al., 2018; Inhoff et al., 1984; Olkoniemi et al., 2016).

In line with the three-stage model, results from Chapters 3 and 4 of this thesis demonstrate that the literal meaning is available—even in the case of highly conventional metaphors—and it can be detected early on during real-time sentence processing. Afterward, the metaphorical content is rapidly attained. Against the three-stage model, these results suggest that the literal meaning is not rejected but rather used to derive the metaphorical content. In that case, as discussed in Chapter 5, then, the literal meaning lingers shortly after the metaphorical content has been attained and, thus, it can be recovered. If it is not recovered, it wanes over time.

The case of two-word metaphor combinations

In contrast to previous studies, the present set of studies uses two-word metaphor combinations. Most of the research to date has focused on copular metaphors such as *My lawyer is a shark* and *Juliet is the sun*. Thus, a major part of the present thesis is to endeavour beyond the *X is Y* structure. Two-word metaphor combinations represent particularly interesting cases for research because the relationship between the first and the second word of these expressions is complex. Namely, they vary in syntactic type, vehicle position, and literalness. First, two-word metaphor combinations include a variety of syntactic structures such as adjective-noun as in *bright student*, noun-noun as in *night owl*, and adjective-adjective as in *icy clean*. Second, they vary in the position of the vehicle, which refers to whether the metaphorical content is carried by the modifier such as *brilliant* in *brilliant idea*, by the head such as *owl* in *night owl*, or by both constituents combined such as *red flag*. Third, some expressions are meaningful only when interpreted metaphorically, whereas others are meaningful both metaphorically and literally. For instance, while expressions such as *broken heart* can only be interpreted metaphorically, the interpretation of expressions such as *cold feet* is determined by context. Overall, the interplay between these factors may be informative of the nature of metaphor comprehension. This is where my story begins.

The present thesis

The series of studies in the present thesis was aimed at investigating the metaphor *awakening* effect. I hypothesized that the literal meaning of a metaphor is initially accessed during metaphor comprehension but remains dormant unless a subsequent cue in the sentence *awakens* it. Accordingly, the aim of this thesis is three-fold. First, to examine whether the literal meaning of conventional metaphors can be recovered after the metaphorical content has been attained. If so, this would indicate that the literal meaning is available during metaphor comprehension. Second, to trace the time course of the literal meaning at different points after the metaphorical interpretation has been derived. That is, whether the availability of the literal meaning fluctuates downstream. Third, to determine whether properties of the expressions themselves—familiarity and aptness—can affect the availability of the literal meaning during sentence comprehension. The outcome of these studies will inform the direct-indirect access debate of whether metaphorical content is realized via literal meaning as well as specify the time course of the literal meaning during metaphor comprehension.

Outline of the Chapters

Chapter 2 provides norms for 309 two-word metaphor combinations. As explained above, this type of metaphor consists of a modifier followed by a head—such as *broken heart* and *early bird*—and they are the object of study throughout the present thesis. For these metaphors, norms include subjective ratings of familiarity and aptness—which are integral factors in metaphor comprehension—along with other linguistic variables of interest. I also examined the effect of context on perceived familiarity and aptness by presenting all metaphors in sentence context as well as in isolation. In Chapter 2, I also compared the collected ratings with frequency scores to examine to what extent familiarity and aptness are a function of the usage of the expression and, then, with each other to examine to what extent these constructs differ from one another. This norming study was developed in parallel to all experiments herein presented and, thus, the way in which familiarity and aptness influence metaphor comprehension was examined in our data afterward and included as Appendices A and B.

In Chapters 3 and 4, I reported three experiments aimed at tracing the early and late stages of metaphor comprehension. In Chapter 3, I examined whether the literal meaning of highly conventional metaphors could be triggered by a cue presented subsequently. To that effect, I conducted a separate norming study to select 24 highly conventional metaphors. These expressions were then embedded in carrier sentences (e.g., *John is an early bird so he can attend morning classes*) and two distractors were created for the word selection task. The related distractor was associated with the literal meaning of the metaphor (e.g., *fly*), while the unrelated distractor was not associated with either the literal or metaphorical meaning of the expression (e.g., *cry*). In a maze task, participants read each word of the sentence at their own pace and, 1 to 3 words after the metaphor, a lexical choice appeared on the screen. During the lexical choice, participants were presented with the most appropriate word to continue the sentence (e.g., *attend*) paired with one of the two distractors. Participants were instructed to select the most meaningful word to continue the sentence. For instance, in the first part of the sentence *John is an early bird so he can,* the most meaningful word to continue the sentence was *attend*—regardless of whether it was paired with *cry* or *fly.* A meaningfulness rating study was conducted to ensure that the correct word was, in fact, the most meaningful word to continue the sentence, while both distractors were equally anomalous.

I reasoned that, if conventional metaphors such as *early bird* are indeed lexicalized, trials in which the correct word was paired with the related distractor *fly* would not interfere with the interpretation of the metaphor and, thus, would yield response times and accuracy rates comparable to trials with the unrelated distractor *cry*. Conversely, if conventional metaphors are understood via literal meaning, trials with *fly* as a distractor would interfere with interpretation as they would conflict with—or *awaken*—the literal meaning of *early bird* and, possibly, elicit a reinterpretation of the metaphor in real time, consequently yielding longer response times and lower accuracy rates.

In Chapter 4, I conducted two experiments that concentrated on the late stages of metaphor processing. These experiments examined whether the literal meaning of highly conventional metaphors was also available, and could be recovered, further downstream during sentence comprehension. To that end, I adapted the materials from Chapter 3 to create a medium (i.e., with 6 to 8 words between the metaphor and the lexical choice) and a large maze (i.e., with 10 to 13 words between the metaphor and the lexical choice). At the end of Chapter 4, I compared response times and accuracy rates across all three maze experiments to trace the availability of the literal meaning at different time points after the metaphorical content had been attained.

Lastly, I examined whether familiarity and aptness influenced the metaphor awakening effect in the original maze (Appendix A) and in the medium and large mazes (Appendix B). I reasoned that the literal meaning of highly familiar and highly apt metaphors is less salient and, therefore, it would be more difficult to recover—or *awaken*.

Altogether, the results provide compelling evidence supporting that the literal meaning is available during conventional metaphor comprehension, which will be further discussed in Chapter 5.

Chapter 2: Norms for 309 two-word metaphor combinations[1]

Laura Pissani and Roberto G. de Almeida

Concordia University

[1] This manuscript is currently in preparation:

Pissani, L., & de Almeida, R. G. (In preparation). Aptness, familiarity, and other linguistic variables for 309 two-word metaphor combinations in context and isolation

Abstract

Norms for 309 two-word English metaphorical expressions (e.g., *broken heart*, *early bird*) in sentence context and in isolation were obtained from 100 participants. These norms serve as materials for research in metaphor comprehension and include ratings of familiarity and aptness. Familiarity was conceived as the extent to which participants had previously heard or read that expression. Aptness was conceived as the extent to which the vehicle captured important features of the topic. In addition to these two main variables, which are known to play a key role in metaphor comprehension, we obtained frequency scores for the whole expression as well as for each constituent separately from the Corpus of Contemporary American English (COCA) database. We found that expressions presented in context were considered more familiar compared to when they were presented in isolation. However, aptness ratings remained consistent in both conditions. Our findings showed that familiarity and frequency have a moderate positive correlation, while aptness and frequency have a low correlation. This supports the idea that familiarity, unlike aptness, is determined by the frequency of its use rather than by the relationship between the topic and vehicle. Lastly, we found a high positive correlation between familiarity and aptness. This is compatible with previous studies suggesting that more apt metaphors are more frequently used in conversation, which in turn increases their familiarity.

Aptness, familiarity, and other linguistic variables for 309 two-word metaphor compounds in context and isolation

Metaphorical expressions such as *My lawyer is a shark* or *Mary is an early bird* jazz up our language use. What makes them particularly appealing to empirical investigation is that the contents they convey ultimately differ from what the expressions mean literally. Given that the interpretation of metaphorical expressions appeals to cognitive resources that go beyond linguistic denotation, their investigation in isolation and utterance contexts constitutes a window into how the brain composes meaning and how language interfaces with other cognitive systems. Research on metaphors is thus at the confluence of a variety of issues bearing on the nature of human cognitive architecture.

In this research context, metaphorical norms serve two main goals. The first goal is to enable experimental studies to rely on more robust sets of materials than those devised within a specific laboratory or designed for a specific experiment conducted with a relatively small set of materials and sample size. Norming studies are important methodologically because they allow laboratories and research groups to investigate different hypotheses about the processing, representation, and neuronal implementation of metaphors, while relying on the same source of materials, thus facilitating cross-experimental comparisons. In addition, norms allow for better replication of experiments—which is key to the validity of purported psychological and neurological mechanisms—by contributing to the standardization of materials (Nosek et al., 2015). Besides serving as an open source of experimental materials, a study on norms further contributes to understanding how metaphorical expressions are interpreted and how their processing is implemented in the brain. Regarding this second goal, a norming study can also be considered an "offline" (that is, not in real time) processing experiment, one in which properties of expressions can be tested for different factors and under different methods and interpretive conditions.

The present study has these two main goals. First, we provide norms for 309 two-word metaphor combinations—such as *broken heart* and *early bird*—presented in sentence context and in isolation. These norms include the linguistic properties of the expressions (i.e., their linguistic structure and semantic opacity), and ratings about subjective familiarity and aptness. Second, these two presentation formats also enabled us to test our main variables of interest—familiarity and aptness—which are well known to play a key role in metaphor comprehension (Blasko & Connine, 1993). Familiarity is an important variable in metaphor research because it reflects the extent to which participants have heard or read the expression in the past. The assumption is that the more familiar the expression is, the easier it is to recover its content from memory or to compose its meaning during comprehension (Gentner & Bowdle, 2008). Aptness reflects the extent to which one word captures important properties of another in order to convey figurative content. In copular metaphors, such as *My lawyer is a shark*, the vehicle *shark* is used to predicate some property of *lawyer*—perhaps that of being ruthless, sneaky, or aggressive. In the two-word expressions employed in the present study, such as in *broken heart*, the relation between “topic” and “vehicle” is less evident. In the case of *broken heart*, *broken* may be used to convey the rupture of an emotional state conveyed by *heart* (usually referring to love). Therefore, in *broken heart* the vehicle *broken* is predicating some state of rupture of the heart—or some love gone awry. Aptness in the present context, then, means how appropriate it is to predicate this sense of rupture of an emotional state such as love. We further discuss the linguistic properties of these expressions below. Suffice it to say that, as several studies have shown, aptness plays a central role in the process of attaining the metaphorical content of a given expression (Chiappe & Kennedy, 1999; Chiappe et al., 2003; Roncero & de Almeida, 2014)—a role that is often taken to be greater than familiarity (Roncero & de Almeida, 2014). Thus, while familiarity reflects the usage properties of a given expression, aptness reflects the degree of semantic appropriateness of an expression, independent of how familiar it may be. In the present study, familiarity and aptness were investigated both as baseline conditions (in isolation) and when the expressions were embedded in metaphorical contexts. Given that metaphor interpretations are largely contextually driven, the same word combinations can be interpreted literally or metaphorically depending on the utterance context. For instance, *cold feet* could refer to either actual feet being cold due to weather or to an emotional state such as a lack of confidence. Thus, deviations in sensitivity to familiarity or aptness can be taken as the effect of context on an expression's interpretation.

Thus far, there have been relatively few large metaphor norming studies, with most employing the more traditional and productive copular form (*x is y*), with their differences being only on the nature of the *x* and *y* constituents. The majority of Katz et al.'s (1988) items, for instance, includes long topic and vehicle phrases (e.g., *The creative mind is a kettle on the stove*, *Thunderclouds are wild horses galloping across the sky*; see also Campbell & Raney, 2015). In Roncero and de Almeida's (2015) norms, all expressions had simple noun phrases in the topic and vehicle positions for both metaphors and simile forms (e.g., *Roads are snakes* / *Roads are like snakes*). Other studies (Cardillo et al., 2010; Cardillo et al., 2017) have provided norms for different metaphorical constructions such as verb phases (e.g., *The insults hopped on her tongue*) but also complex copular ones (e.g., *The editorial was a brass-knuckle punch*) together with literal sentences of the same form.

What is unique about the present metaphor norms is that they are the first to be obtained for two-word metaphorical expressions such as *broken heart*, *early bird*, and *sharp mind* in isolation and in sentence context. Although the relationship between the two words in these expressions is, on the surface, relatively simple, on closer inspection they carry many semantic intricacies. To begin with, these expressions include a variety of syntactic categories, but predominantly adjective-noun (Adj-N) such as *broken heart* and noun-noun (N-N) such as *drug mule*. These expressions are also complex in how the two words are used metaphorically. We will use the term “vehicle” to refer to the word that carries most of the metaphorical content (e.g., *sharp*) and “topic” to refer to the constituent that is predicated on (e.g., *mind*). It can also be the case that both constituents work together as the vehicle to describe a topic that remains implicit, as it is the case of *smoking gun* used to refer to a type of evidence that is incontrovertible, not to a type of gun. In most cases, however, the vehicle is the first word of the pair (e.g., *dark*), and it is followed by the topic, which can maintain its literal meaning (e.g., *personality*) or bear another figure of speech, such as metonymy (e.g., *mind*). Thus, contrary to the topic and vehicle of copular metaphors, in which the topic (e.g., *lawyer*) precedes the word used for predication (e.g., *shark*), in most two-word metaphors the vehicle precedes the topic. In this way, these expressions behave as English compound words, whose modifier is the first constituent, and the head is the second (e.g., *blueberry*, *driveway*). This is typically the case with very productive adjectives such as *bright* which is conventionally used to refer to someone or something intelligent (*bright student*, *bright idea*, etc.). However, the relation between the two words is not always clear. For instance, *sharp mind* predicates “sharpness” of mind, but *mind* is used metonymically to refer to someone’s intellectual ability, with only *sharp* being used metaphorically to express the idea of someone being able to think fast and clearly. In the present norms, the distinction between metaphorical and metonymic uses of a word is often blurred given that both are considered figurative forms (i.e., their meanings do not come from the object that they refer to—such as in *mind*—but from how the word is used to refer or predicate of something or someone). In other cases, only the second word, the vehicle, is metaphorical. For instance, in *early bird*, *bird* is used to refer to a person who does things in the morning or sooner rather than later.

A third way in which the expressions we studied vary is whether the expression has an equivalent literal use—such as *cold feet*—or if it can only be interpreted figuratively, as in *broken heart*. The advantage of embedding these expressions in context is, thus, that we can ensure that the ratings provided in a biasing context refer to the metaphorical content, rather than to the literal meaning. Further, context has been shown to play a role in the way metaphors are processed. For instance, Bambini et al. (2016) examined the role of context during metaphor comprehension by contrasting literal and metaphorical expressions in both minimal and supportive contexts (i.e., where the ground—the relationship between the topic and vehicle—of the metaphor was made explicit). They found that, when expressions were presented in minimal context, metaphors yielded greater N400 and P600 amplitudes in comparison to their literal counterparts. In contrast, when expressions were presented in supportive context, metaphors showed a reduced N400 amplitude thus matching their literal counterparts.

Before we present our norms, it is important to note that, to our knowledge, there have been few psycholinguistic studies investigating the nature of two-word metaphorical expressions (Al-Azary et al., 2021; Arzouan et al., 2007; Gagné, 2002; Goldstein et al., 2012; Forgács et al., 2015; Park, 2020; Pissani & de Almeida, 2021, 2023). For instance, Gagné (2002) studied exclusively noun-noun combinations (e.g., *closet heart, trumpet voice*) adapted from copular metaphors (e.g., *Heart are closets, some voices are trumpets)*, and found that factors influencing the comprehension of copular metaphors (particularly aptness, salience, and expectancy) also affected two-word metaphor combinations. In turn, Forgács et al. (2015) examined novel adjective-noun combinations where the same head was preceded by either an abstract (e.g., *conditional schedule*), concrete (e.g., *printed schedule*), or metaphorical modifier (e.g., *thin* *schedule*). Their results suggest that the modifier affects how the whole expression is processed, such that reading the same noun may yield a larger concreteness effect following a concrete adjective rather than an abstract one. Interestingly, metaphorical word pairs rated more abstract yielded a larger N400 effect on the noun when compared to the noun of word pairs rated more concrete. More recently, Al-Azary et al. (2021) examined modifier-noun phrases (e.g., *shark lawyer*), and found that, similarly to copular metaphors (e.g., *My lawyer is a shark*), they were subject to the metaphor interference effect (Glucksberg et al. 1982), which states that the processing of the metaphorical content is automatic. That is, it cannot be suppressed. It is important, however, to highlight the heterogeneity of the materials employed in these studies with most being adapted and rated *ad hoc*.

It is also important to note that the investigation of these metaphorical expressions plays a key role in understanding compositionality. Crucially, compositionality bears on how an expression's meaning is a function of its constituents and how they are structured—for example, whether *cold feet* is a function of the meaning of *cold* and *feet* and their predication relation, such that *cold* is predicated on *feet*. Compositionality is one of the fundamental characteristics of human cognitive architecture and it is said to underly the productivity of our linguistic and conceptual systems (Frege, 1884; Fodor, 1975, 2001; Grandy, 1990; Partee, 1995). But clearly, not all linguistic expressions appear to compose the same way: while *cold feet* may be compositional on its own when it is inserted in a metaphorical context, its sense differs from the meanings of its constituents and how they combine. Compositionality of linguistic expressions is then crucially linked to the literal/metaphorical distinction (Gibbs & Colston, 2012). Two recent studies (Pissani & de Almeida, 2021, 2023 [see Chapters 3 and 4]) have shown that even in sentences such as *Most people agree that a broken heart can be difficult to overcome,* where the metaphorical interpretation is biased, the literal meaning of the expression remains active for several seconds after the expression has been processed, suggesting that literal interpretation—and, thus, composition—is attained even in metaphorical contexts. That is, these expressions require a local composition besides the content that the two-word expression contributes to its carrier sentence.

The present norms

We collected norms for 309 metaphorical expressions in isolation and sentence contexts (see Appendix F). All expressions in our set are of the form YX, with words representing two main grammatical combinations, Adj + Noun (N = 189) and Noun + Noun (N = 117) with a few Adj + Adj (N = 3). From these, 63 expressions carry their metaphorical content in the first word (e.g., *bright student*), 156 expressions in the second word (e.g., *old flame*), and 90 expressions in both words (e.g., *smoking gun*). Another dimension of our set is that some of these expressions are always interpreted metaphorically (e.g., *emotional rollercoaster*), while others may also be interpreted literally (e.g., *cold turkey*). For all the expressions, we collected familiarity and aptness ratings along with analyses of the effect of context and vehicle position. We also included COCA frequency scores for the co-occurrence of both words that constitute the combination as well as scores for the individual constituents. Further, we have examined the relationship between these variables by calculating the correlation coefficient between familiarity, aptness, and frequency. All norms are available at https://osf.io/y9g3s.

Method

Participants

We recruited 134 native speakers of English between the ages of 23 and 66 years (M = 37.51, SD = 10.45; 89 M) with self-reported normal or corrected-to-normal vision and without any history of reading or hearing disability. After data preparation, we removed 34 participants (detailed in Data Preparation). Thus, these norms are based on 100 participants. All participants were recruited using Amazon Mechanical Turk (MTurk, for short). Recruitment was restricted to countries where the official or dominant language is English (e.g., Canada, US, UK) and to MTurk workers with an approval rate of 80% or higher. Each participant was compensated with CAD$12 for their participation.

Materials

We collected a total of 309 two-word metaphor combinations. From these expressions, 260 were obtained from everyday conversations and diverse media sources such as newspapers, blog posts, streaming services, video-sharing platforms, and social networks. The remaining 49 expressions were obtained from Briner et al. (2018), where local norms were compiled for 100 metaphorical and literal word pairs. For all expressions, we provided scores on diverse linguistic variables and collected ratings for familiarity and aptness with and without context (see Table 2A for a summary of the variables).

Grammatical structure. All expressions are composed of two words, where the first word (modifier) is either an adjective (e.g., *red flag*) or a noun (e.g., *drug mule*), while the second word (head) is always a noun, except for three cases in which both words are used predominantly as adjectives (e.g., *boiling mad*).

Metaphorical content carrier (vehicle). The metaphorical content of the expressions can be carried by the modifier, the head, or both. For instance, in expressions such as *bright student* the modifier carries the metaphorical content and can be paired with different nouns (*bright student*, *bright author*, *bright idea*, etc.). In this case, all instances of *bright* play the same role, that is, predicating some form of "cleverness" to the head noun referent. For those cases, we selected only one occurrence (*bright student*). However, we included expressions with the same modifier only if they contributed unique semantic content to the expression. For instance, in *bright side*, *bright* is used as "hopeful" or "promising", while in *bright student, bright* is used as "intelligent" or "clever". Conversely, some expressions carry the metaphorical content in the head, but the modifier can be interpreted literally. For instance, in expressions such as *old flame*, the modifier *old* can be literally interpreted as "former", while the head *flame* can be figuratively interpreted as "lover". In some cases, metaphorical heads can be paired with different modifiers as is the case of *old flame* and *new flame* or *red flag* and *green flag,* in which cases, we selected the most frequently used. Finally, other expressions carry the metaphorical content in both words. For instance, the expression *smoking gun* is used together to refer to conclusive evidence.

Literalness. Most of the expressions can be understood only metaphorically (e.g., *emotional rollercoaster*). However, some can be interpreted both metaphorically and literally. In some of these expressions, the metaphorical meaning may be more salient (e.g., *cold turkey*), while in others the literal meaning may be more salient (e.g., *hot water*). It is important to note that participants were aware that they were rating metaphorical English expressions. Thus, even in the case of metaphors that can also be understood literally, ratings are meant to reflect their metaphorical interpretation.

COCA frequency score. We included frequency scores obtained from the Corpus of Contemporary American English (COCA; Davies, 2008) for each word separately as well as for the co-occurrence of both words that compose the two-word combination. COCA is a large, long-established English corpus and contains over one billion words. It is important to mention that the COCA database does not distinguish between metaphorical and literal uses, thus expressions like *hot water* (5124) and *big brother* (3766) had the highest scores, while the most frequent exclusively metaphorical expression was *movie star* (2687). Expressions with the lowest frequency scores included *penguin huddle* (1) and *burning moment* (1), while 38 expressions did not appear in the database and were marked with a frequency score of zero. Amongst the latter, some were contemporary expressions such as *online scrub* (0), whereas others were domain-specific expressions such as *violin graph* (0) or *spider plank* (0).

Table 2A: Summary of all the variables present in the norms

Variable	Abbreviation	Description
Syntactic type	type	describes the grammatical structure of the metaphorical combination, whether Adj-Noun (AN), Noun-Noun (NN), or Adj-Adj (AA).
Vehicle position	vehicle_position	indicates whether the metaphorical content, carried by the vehicle, is in the first, second, or both words of the expression
Vehicle	vehicle	names the metaphorical term used to describe the subject of the metaphor.
Topic	topic	names the subject of the metaphor being described by the expression.
Literalness	literalness	marks whether the expression can be interpreted literally (yes) or only metaphorically (no).
COCA frequency score	COCA_expression	provides the COCA frequency score for the co-occurrence of both constituents of the expression.
	COCA_modifier COCA_head	provides the COCA frequency score for either the first (modifier) or second (head) of the expression.
Familiarity	FAM-C FAM-I	provides subjective familiarity ratings separately for expressions embedded in context (C) or presented in isolation (I).
Aptness	APT-C APT-I	provides subjective aptness ratings separately for expressions embedded in context (C) or presented in isolation (I).

Procedure

The study was programmed in Psychopy3 (Version 2021.2.3; Peirce et al., 2022), with stimuli presentation and data collection via the Pavlovia online platform. Upon registration, participants were assigned to one of four lists in order of registration. After receiving a web link to our study, participants were directed to a virtual consent form followed by a demographic form. During the rating task, participants saw either a sentence with the metaphorical expression in upper case (e.g., *The EARLY BIRD always gets the best seat at the movie theatre*.) or the metaphorical expression in isolation (e.g., *EARLY BIRD*). After 3 seconds, a 1-to-7 numerical scale appeared below the sentence or the expression, and participants were required to rate the expression in uppercase, even if it was presented embedded in a sentence. Further, we included additional trials as attention checks. These trials required participants to press a specific number on the rating scale (e.g., *For this sentence, press the number five so we know you are paying attention.*), and these were randomly presented one time per block.

Familiarity rating collection. We asked participants to rate metaphorical expressions with and without context for familiarity. Participants were encouraged to use the full scale, with not-so-well-known expressions rated more towards the middle (2, 3, 4, 5, 6), reserving 1 for truly not familiar expressions and 7 for very familiar ones. In the instructions (see Appendix D), familiarity was defined as the extent to which participants had heard or read each expression in the past. In addition, two detailed examples including a familiar and a less familiar expression were provided. For instance, if the expression *feeling blue* was a well-known expression to participants, they were advised to give it a high rating (perhaps 6 or 7), whereas if *crying wolf* was not as well-known, they were advised to provide a lower rating (perhaps 2 or 3).

Further, familiarity ratings were collected in two blocks with context as a within-subjects factor, which was counterbalanced in lists one and two. In list one, participants rated expressions from 1 to 155 in context and expressions from 156 to 309 without context. Conversely, in list two, different participants rated expressions from 1 to 155 without context and expressions from 156 to 309 in context. Participants were only assigned to one list and encouraged to take a break between blocks.

Aptness rating collection. We asked participants to rate metaphorical expressions with or without context for aptness. In the instructions (see Appendix E), aptness was defined as the extent to which the vehicle captures important features of the topic and six different examples including apt and inapt expressions with different vehicle positions were provided. For instance, the expression *silky hair* could be considered a highly apt expression because *silky* captures important features of the hair (namely, shininess and smoothness). On the other hand, *silky sunset* may be considered a less apt expression since it is less common for sunsets to be both shiny and smooth. Therefore, *silky hair* would receive a high rating (perhaps 6 or 7), whereas *silky sunset* would receive a lower rating (perhaps 3 or 4). In addition, the same use of the scale for familiarity was encouraged for aptness.

In contrast to familiarity, aptness ratings were collected in three blocks with the vehicle position as a within-subjects factor and context as a between-subjects factor, both of which were counterbalanced in lists three and four. In both lists, expressions were divided into three blocks (A, B, and C) according to the position of the vehicle to facilitate the rating task. Thus, block A contained 156 expressions where only the modifier was metaphorical (e.g., *bright student*), block B contained 63 expressions where only the head was metaphorical (e.g., *early bird*), while block C contained 90 expressions where both words were metaphorical (e.g., *red flag*). In list three, all expressions were shown in context and blocks appeared in consecutive order (i.e., A, B, C). In list four, all expressions were presented without context and the order of the blocks differed from list three (i.e., B, C, A).

Data preparation

For our analyses, we only included data from eligible participants who correctly followed the instructions. Thus, we removed 11 participants for not using the full rating scale (e.g., using only 5 to rate all expressions), nine participants for incorrectly responding to at least one attention check, two participants for having previously participated in a different version of the study, and 12 participants for not completing the task.

Results

The present norms were obtained from ratings on familiarity and aptness provided by 100 participants. We examined the effect of context on both familiarity and aptness ratings. For aptness ratings, we also examined the effect of vehicle position. Further, we calculated the correlation coefficient to assess the relationship between familiarity ratings and COCA frequency, aptness ratings and COCA frequency, and familiarity ratings and aptness ratings.

Familiarity

The familiarity ratings for metaphors presented in context ranged from 2.96 to 6.2, with a mean of 4.89 (SD = 1.85). For metaphors presented in isolation, familiarity ratings ranged from 2.48 to 6.4, with a mean of 4.74 (SD = 1.93). We used R (R Core Team, 2020) and lme4 (Bates et al., 2015) to perform a linear mixed-effects model to analyze the effect of context on familiarity ratings. As fixed effects, we entered context into our full model. As random effects, we had by-subject and by-item random intercepts. We obtained p-values by likelihood ratio tests of the full model against the null model including only random effects. Our full model was a significantly better fit to the data than the null model, $\chi^2(1) = 33.57$, $p < .001$. There was a main effect of context on familiarity ratings, with an increase of 0.15, 95% CI [0.10, 0.20] for metaphorical expressions in context compared to the same expressions presented in isolation. Afterward, we visually inspected residual plots and found no evident deviation from normality.

Further, we computed Pearson correlation coefficients to assess the linear relationship between familiarity ratings and COCA frequency scores. We found a significant correlation between frequency and familiarity in context, r (307) = .44, $p < .001$, and between frequency and familiarity in isolation, r (307) = .45, $p < .001$).

Aptness

The aptness ratings for metaphors presented in context ranged from 2.88 to 6.12, with a mean of 4.72 (*SD* = 1.76). For metaphors presented in isolation, aptness ratings ranged from 2.88 to 6.4, with a mean of 4.45 (*SD* = 1.75). We also performed a linear mixed-effects model to analyze the effect of context on aptness ratings. We entered context and vehicle position into our full model as fixed effects. As random effects, we had by-subject and by-item random intercepts. We also obtained p-values by likelihood ratio tests of the full model against the null model. Our full model was a significantly better fit to the data than the null model, $\chi^2(2) = 33.95$, $p < .001$. There was no main effect of context on aptness ratings, $\chi^2(1) = 2.43$, $p = .119$. There was, however, an interaction between context and vehicle position, whereby the effect of context was significant for expressions in which the vehicle was the first word (e.g., *bright student*) increasing aptness ratings by .42 ($p = .016$), while the opposite was true for the same expressions presented in isolation, in which aptness ratings decreased by .26 ($p = .007$).

We also computed Pearson correlation coefficient to assess the linear relationship between aptness ratings and COCA frequency scores. We found a significant correlation between frequency and aptness in context, r (307) = .30, $p < .001$) and between frequency and aptness in isolation (r (307) = .33, $p < .001$).

Familiarity versus aptness

Finally, we computed Pearson correlation coefficients to assess the linear relationship between familiarity and aptness ratings. We found a significant correlation between familiarity and aptness, both in context, r (307) = .76, $p < .001$, and in isolation, r (307) = .80, $p < .001$. These results are discussed next considering their potential contribution to future studies and what they inform us about the nature of two-word metaphor comprehension.

Discussion

We collected norms for 309 metaphor combinations such as *broken heart* presented in isolation and embedded in context (e.g., *She was left with a broken heart after the split with her partner*).

We had two main goals with the present norms. First, we sought to provide a robust set of experimental materials to support the execution of experiments investigating the representation, processing, and neuronal implementation of metaphors. As an open source, our norms promote the standardization of materials and allow for comparison across studies. Conversely, local norms that are collected *ad hoc* may be limited to a smaller set of materials, sample size, and time frame. Second, we aimed to contribute to the understanding of metaphor interpretation and implementation in the brain by treating our norms as offline data. It is important to note that several metaphor processing studies have also relied on "offline" methods to test the nature of metaphor comprehension. Among these offline methods are semantic judgement (McElree & Nordlie, 1999), sentence verification (Glucksberg et al., 1982), ratings and fill-in-the-blank tasks (Bowdle & Gentner, 2005), as well as usage data from the internet (Roncero et al., 2016). In the same vein, our norms provide data obtained offline—for instance, the contrast between familiarity ratings in context versus in isolation—to inform about the nature of metaphor interpretation. While this was not our main goal, we argue that contextual information is crucial for understanding metaphorical expressions of the kind we used here.

We collected ratings for familiarity and aptness, which are integral factors in metaphor comprehension (Blasko & Conine, 1993; Chiappe & Kennedy, 1999; Chiappe, Kennedy, & Chiappe, 2003; Bowdle & Gentner, 2005; Jones & Estes, 2006; Roncero & de Almeida, 2014; Roncero et al., 2016). All ratings were collected with and without support from context, which has also been shown to modulate metaphor interpretation (Janus & Bever, 1985; Bambini et al., 2016). First, we examined the effect of context on both familiarity and aptness. We found a main effect of context for familiarity, whereby metaphors in context were rated significantly more familiar than the same expressions in isolation. These findings are partially compatible with the idea that subjective ratings of familiarity and aptness are confounded with processing fluency (Thibodeau & Durgin, 2011; Thibodeau et al., 2018). That is, participants may base their ratings on how easily they understand the sentences. Thus, the biasing context in which our expressions were embedded may have facilitated their comprehension, which in turn yielded higher familiarity ratings. However, this was not the case for aptness, which was not improved by context. Rather, aptness appears to be an internal property of the expression, which is determined by the relationship between the topic and the vehicle of the metaphor regardless of context. Interestingly, we found an interaction between context and vehicle position, whereby aptness ratings for expressions in context with the vehicle as the first word (e.g., *She was a bright student* *who learned quickly*) were significantly higher than expressions with the vehicle as the second or both words. Conversely, the same expressions in isolation (e.g., *bright student*) obtained lower ratings compared to those with the vehicle as the second or both words. Although unexpected, these results may indicate that biasing context combined with the first word being metaphorical facilitated the comprehension of these particular expressions; which, according to the processing fluency account (Thibodeau & Durgin, 2011; Thibodeau et al., 2018), yields higher aptness ratings.

We found three significant correlations in our norms. First, we examined the correlation between familiarity ratings and COCA frequency scores. We found a moderate positive correlation between familiarity and frequency regardless of context. These results, although not surprising, support the idea that the familiarity of an expression is a function of the frequency of its usage, which is in line with most of the literature comparing familiarity and different forms of frequency (e.g., Senaldi et al., 2022; Thibodeau & Durgin, 2011; Wisniewski & Murphy, 2005). We then examined the correlation between aptness ratings and COCA frequency scores. We found a low positive correlation between aptness and frequency. This finding suggests that, unlike familiarity, aptness is not purely obtained by the repetition of an expression, but rather expresses the appropriateness —or quality—of a metaphor. Finally, we assessed the correlation between familiarity and aptness ratings both in context and in isolation. We found a high positive correlation between familiarity and aptness regardless of context. We argue that, even though each variable expresses different dimensions of a metaphor, metaphors that are perceived to be more apt may be more frequently used in conversation and, therefore, those may become more familiar (Thibodeau & Durgin, 2011; Roncero & de Almeida, 2014).

Metaphors occur in various forms, including copular (e.g., *My lawyer is a shark*), verbal (e.g., *Maria devoured the paper*), and two-word metaphor combinations (e.g., *brilliant idea*).

However, most of the research to date has focused on the former. A major contribution of the present norming study is to endeavour beyond the *X is Y* structure. Two-word metaphor combinations represent particularly interesting cases for research because the relationship between the first and the second word of the expression is complex. First, two-word metaphor combinations include a variety of syntactic types (i.e., adjective-noun, noun-noun, and adjective-adjective). Second, they can carry the metaphorical content (or *vehicle*) in the modifier, the head, or both constituents. Third, the combination of both constituents can yield an expression that is meaningful only metaphorically or both metaphorically and literally. Together—the range of syntactic structures, vehicle position, and literalness—can be informative of how metaphorical meaning is attained. For instance, Pissani and de Almeida (2023) found that the literal meaning was available further during sentence comprehension only for expressions in which the first or second word was metaphorical, but not both. This suggested that, after the metaphorical content has been attained, the individual concepts decay, which is expedited when both constituents of the expression are metaphorical.

The present norms are a valuable resource for researchers interested in the nature of metaphor interpretation, with a focus on two-word metaphor combinations. Crucially, our materials are designed to investigate current theories of metaphor processing while considering integral factors known to affect the comprehension process.

Chapter 3: Examining the awakening effect during the early stages of metaphor interpretation

The previous study compiled and normed an extensive number of two-word metaphor combinations to serve as an open-source resource of materials, and to provide familiarity and aptness data to feed the series of studies in this thesis. However, to advance data collection for the present experiment, several metaphors were taken from this compilation and normed for familiarity separately. We then selected the 24 expressions rated the most familiar to employ in the present experiment, whose object of study was conventional metaphors.

In the current chapter, we examined whether the literal meaning of conventional metaphors was accessed during metaphor comprehension. To that effect, we used a version of the maze task, in which sentences (e.g., *John is an early bird so he can*) were presented word by word on the screen. At the maze juncture, a word selection task was presented, where the correct word to continue the sentence (e.g., *attend*) was paired with a distractor that could be either related to the literal meaning of the metaphor (e.g., *cry*) or unrelated (e.g., *fly*). Participants were then instructed to select the most meaningful word to continue the sentence, while their response time and accuracy were measured.

If the three-stage model is right in that the literal meaning is accessed during metaphor comprehension, then responses in trials with the related cue (e.g., *fly*) would be delayed and less accurate than those with the unrelated cue (e.g., *cry*). This would be an indication of conflict between the literal meaning and the figurative interpretation of conventional metaphors. In contrast, if the direct-access model is right in that conventional metaphors function as lexicalized expressions and, as such, are derived directly from semantic memory, then responses in trials with the related cue would not differ in response time or accuracy from those with the unrelated cue.

This manuscript was published in August 2021:

Pissani, L., & de Almeida, R. G. (2021). Can you mend a broken heart? Awakening conventional metaphors in the maze. *Psychonomic Bulletin & Review*, *29*(1), 253-261.

Can you mend a broken heart? Awakening conventional metaphors in the maze

Laura Pissani and Roberto G. de Almeida

Concordia University

Abstract

Conventional metaphors such as *broken heart* are interpreted rather fast and efficiently. This is because they might be stored as lexicalized, noncompositional expressions. If so, they require sense retrieval rather than sense creation. But can their literal meanings be recovered or "awakened"? We examined whether the literal meaning of a conventional metaphor could be triggered by a later cue. In a maze task, participants (N = 40) read sentences word by word (e.g., *John is an early bird so he can . . .)* and were presented with a two-word choice. Participants took longer and were less accurate when the correct word (*attend*) was paired with a literally-related distractor (*fly*) rather than an unrelated one (*cry*). This suggests that the literal meaning of a conventional metaphor is not circumvented, nor that metaphors simply involve sense retrieval. The metaphor awakening effect suggests that the mechanisms employed to process conventional metaphors are dynamic with both metaphorical sense and literal meaning being available.

Can you mend a broken heart? Awakening conventional metaphors in the maze

In the song by the BeeGees, *How can you mend a broken heart?* the lyrics invite us to attend to the literal sense of *broken heart*—or, instead, to *mend* it metaphorically. Either way the verse goes, it leads us to awaken the literal meaning of a conventional metaphor. Metaphorical expressions pepper our language use, and their literal interpretations appear to go unnoticed because they are usually linked to a conventional metaphorical—thus non-literal—content. But how are these expressions represented and processed in real time? Are they represented in lexicalized form—thus linked to a concept that is non-compositional, that is, one that is not a function of the meanings of its constituent elements and how they combine? If they are represented as independent forms, can their literal meanings be recovered or "awakened" as the BeeGees' verse invite us to do? If so, what does this process reveal about metaphorical processing? We were interested in understanding what conventional metaphor processing can tell us about the interface between arguably two different levels of comprehension: semantic (based on what is actually said) and pragmatic (what is intended by the speaker in a particular context).

Recent research has focused on how we calculate the content that a metaphor carries—whether *directly*, relying on what has been stored based on use, or *indirectly*, via a literal form of interpretation. While *direct* interpretation may rely primarily on the retrieval of stored representations (Gibbs, 1994; Gibbs & Colston, 2012), the *indirect* form relies on a proposition that is faithful to the literal interpretation—which, when deemed false, triggers the search for an alternate meaning in line with context and as an *approximation* to the speaker's intentions (Grice, 1975; Searle, 1979).

A key issue in the dispute between direct and indirect theories is the conventionality of a metaphor. In the case of copular metaphors—those with the form *X is Y*—such as *My lawyer is a shark*, conventionality is taken to refer to the degree to which the vehicle *shark* is used to refer to a particular alternative sense (e.g., *aggressive*, *shrewd*), predicating this sense to the topic *lawyer*. The conventionality of *shark*, in this case, comes from its frequent use contributing that particular alternative sense to numerous expressions (*the banker is a shark*, *your friend is a shark*, etc.). Conventionality here is then distinct from but interacts with familiarity, which refers to how well-known a full expression such as *My lawyer is a shark* is (Bowdle & Gentner, 2005; Roncero & de Almeida, 2015).

Conversely, the conventionality of metaphors of the form *Adjective-Noun* can come from the modifier, the head, or both. For instance, in expressions such as *bright student* only the modifier *bright* is metaphorical, and it is conventionally used to refer to someone or something smart—such as in *bright idea*, *bright scientist*. In some expressions, however, only the head, but not the modifier, can be metaphorical: an *old flame* refers to an ex-lover, where *old* maintains its literal meaning as 'former', while the head *flame* is used metaphorically to refer to a lover. In other expressions, both terms are used together to form a metaphor, as in the case of *broken heart*, where both *broken* and *heart* have productive, metaphorical uses. Here, we refer to all these *Adjective-Noun* types of metaphors as conventional, even when only one of the constituents is used metaphorically. We do so because metaphoricity comes from the combination of both terms, which are conventionally used together to refer to something figuratively.

Conventionality constitutes an important variable for understanding the nature of meaning processing because while conventional metaphors are blatantly false, like all other forms of metaphors, *qua* linguistic expressions they are usually associated with a particular alternative content. Conventional metaphors, in this regard, are in stark contrast with novel metaphors, whose metaphorical contents are indeterminate and need to be calculated by the listener or reader. This contrast has been termed "stored" vs. "fresh" (Morgan, 1979) or "lexicalized" vs. "novel" (Blank, 1988), which Bowdle and Gentner (2005, p. 199) claim "invite sense retrieval" (conventional) in opposition to "sense creation" (novel). As Morgan (1979, p. 129) put it, "recognizing the phrase [*…is a pig*] one knows immediately what is intended. It's an institutionalized metaphor, and knowledge that the phrase 'a pig' is used this way short circuits the process of figuring out the metaphor from literal meaning…". Explicit in this contrast, is the idea that there are two processes involved in understanding metaphors, depending on whether the metaphor is novel or conventional, with conventional ones requiring the retrieval of its associated content.

We examined whether conventional metaphor processing in fact bypasses or "short circuits" the literal meaning. In particular, we investigated whether or not the literal meaning could be recovered or "awakened" by a later cue. This shift in interpretation—from metaphorical to literal—to our knowledge has been investigated only by Goldstein et al. (2012), but with a radically different method. They induced thoughtful semantic processing of two-word metaphors by asking participants to make a semantic judgment and to interpret a subset of the materials. ERPs were recorded during a later task in which the subset of conventional metaphors that had been explained in terms of their literal meaning elicited a higher N400 component and a lower late positive component in contrast to the subset that had been left unexplained. According to Goldstein et al., this pattern suggests that explaining a conventional metaphor might create new structural relations between the two words. Although these results might indicate a shift in processing mode, this shift happened over time—from the exposure to the test phase—and was triggered by asking participants to explain a conventional metaphor rather than relying on moment-by-moment, real-time processes of interpretation.

We reasoned that if conventional metaphors indeed "bypass" the literal interpretation but can be "shifted" back to the literal meaning by subsequent cues, this literal meaning might in fact be available during initial processing. Thus, rather than engaging participants in a thinking task, we investigated the potential awakening of the literal meaning of *Adj+N* metaphors by employing a maze task (Forster et al., 2009; Forster, 2010). This task is taken to provide an accurate estimate of the processing cost of a target word in the presence of a given distractor during word-by-word self-paced reading. The maze task ensures that consecutive words in a sentence are integrated with its previous elements, as participants must continuously and rapidly understand unfolding segments of the sentence to be able to make an accurate lexical choice. Hence, the task yields a measure of word-by-word incremental interpretation (Gallant & Libben, 2020), with findings comparable to those obtained in eye-tracking studies (Boyce et al., 2020; Forster et al., 2009).

To our knowledge, no other study has investigated the potential metaphor-to-literal shift in real time. We hypothesized that if conventional metaphors are lexicalized, involving only meaning retrieval, meaning composition should be insulated from a cue that refers to the literal meaning of the metaphor. That is, the proposition that readers form should consist of the metaphorical content obtained by sense retrieval. However, if the literal meaning is available even in cases of conventional metaphors, a subsequent cue might cause an *awakening* effect by making the literal content of a conventional metaphor to be (re-)accessed in real time. Specifically, we predicted that the presence of a word such as *mend* after a metaphor such as *broken heart* would engender processing costs compatible with the hypothesis that the literal meaning of *broken heart* has been awakened.

Experiment 1: The original maze

Method

Participants

Participants were 40 native speakers of English between the ages of 18 and 47 years old (*M* = 22.20, *SD* = 4.91; 39 F) with normal or corrected-to-normal vision. The sample size was based on similar studies that employed the same technique and investigated related phenomena (e.g., Gallant & Libben, 2020; Witzel & Forster, 2014). Participants were recruited via the Concordia University online participant pool and were compensated with course credit.

Materials

Materials consisted of 24 experimental sentences containing highly conventional two-word metaphorical expressions obtained from the combination of adjectives and nouns (e.g., *broken heart, early bird*) plus 48 filler sentences. The metaphorical expressions were chosen based on their familiarity ratings because familiarity is a good predictor of conventionality (Roncero & de Almeida, 2015). In a separate rating study, 10 participants rated 48 metaphors such as *early bird* and 48 other non-metaphorical expressions (e.g., *red wine*, *handsome woman*) which were presented in upper case and embedded in simple carrier sentences. Their task was to rate how well they knew the expression on a scale of 1 to 7, with 7 being the most familiar. We then selected 24 metaphor combinations that were rated above 4 (*M* = 6.36, *SD* = 0.77). Based on those expressions, 24 sentences were created for the maze task (e.g., *John is an early bird so he can attend morning classes*; see Appendix). For each sentence, we selected two types of distractor words, which were to be entered as choices in the maze portion of the task: (1) a related distractor (e.g., *fly*), which was semantically associated with the literal meaning of the metaphor combination, and (2) an unrelated distractor (e.g., *cry*), which was not semantically associated with the metaphor. Both types of distractors were matched to each other in overall frequency, length, and syntactic category based on data from the MRC database (Coltheart, 1981). Both distractors were grammatically equivalent, but none of them was semantically appropriate to continue the sentence.

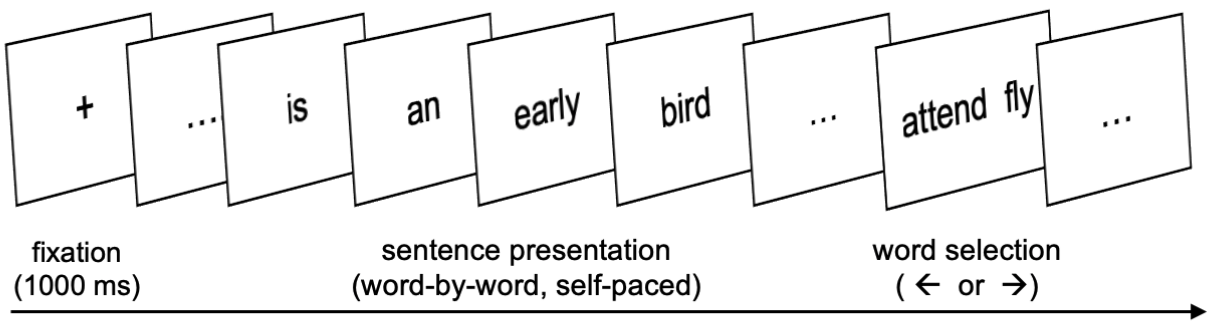
Based on the two conditions, two mixed lists were created so that participants would see only one version of each sentence, but both distractor conditions. Thus, participants were presented with half of the items in the related-distractor condition, while the other half of the items was presented in the unrelated-distractor condition. In addition, for each sentence, the position of the distractor was counterbalanced for each list, so that half of the participants read the distractor on the right, and the other half read the same distractor on the left (e.g., *attend/fly* or *fly/attend*).

Procedure

The experiment was programmed using PsychoPy3 version 3.2.4 (Peirce & MacAskill, 2018). Participants were tested via the Pavlovia online platform (pavlovia.org), which allows data collection remotely with high response time (RT) accuracy (Bridges, Pitiot, MacAskill, & Peirce, 2020; Grootswagers, in press). Participants were randomly assigned to one of two lists when they registered for the experiment. After receiving a web link to access the experiment, participants were directed to a virtual consent form, which was followed by a demographics form. During the experiment, participants were presented with sentences word by word, in a self-paced manner (see Figure 3A). At any point during the sentence (1 to 3 words after the metaphorical expression; *M* = 1.96; *SD* = 0.91), they were presented with two words to choose from so that the sentence could continue. Only one word was semantically appropriate to continue the sentence.

Participants were required to select the appropriate word by pressing the left or right arrow according to the chosen word's position on the screen. In about 25% of the trials, a comprehension question was presented. Participants were instructed to answer the question by pressing the Y key for YES and the N key to for NO to answer the question. It was emphasized that they should move as quickly as possible through the sentence and make their decision as fast and as accurately as possible because their time was being recorded. Participants were presented with 8 practice trials before the experimental phase (24 experimental trials + 48 fillers, randomly distributed). The experimental session lasted approximately 15 minutes.

Figure 3A: The procedure for the maze task



Each trial in the maze task started with a 1000 ms fixation cross, followed by the first word of the sentence. Participants were instructed to press the up arrow on the keyboard to indicate that they had read the word and move on to the next word in the sentence. The metaphor (*early bird*) was followed by one to three words (*so he can* …), followed by the word selection task—a maze juncture—consisting of the target (*attend*) presented together with a distractor that was semantically related (*fly*) to the literal meaning of the metaphor or an unrelated control (*cry*). Participants had to select which word would continue the sentence by pressing the corresponding left or right arrow.

Data Preparation

Analyses of RT and accuracy were restricted to participants who achieved 75% or higher in both the lexical choice and comprehension question tasks. Overall accuracy scores were calculated for each participant. All participants achieved 89% or higher in the lexical choice task and 83% or higher when responding to the comprehension question. Based on these criteria, all 40 participants were included in the analysis. Incorrect trials (i.e., those where participants either selected the incorrect word or did not respond correctly to the probe question) were removed from the RT analyses only (12 % of all data points).

All RTs above 3 seconds were trimmed prior to data analysis (3% of all data points). Further, outliers were defined as more than 2.5 standard deviations from the mean and calculated per participant to preserve individual variability. Outliers were replaced with 2.5 standard deviations above the mean for the upper tail (2 % of all data points).

Results

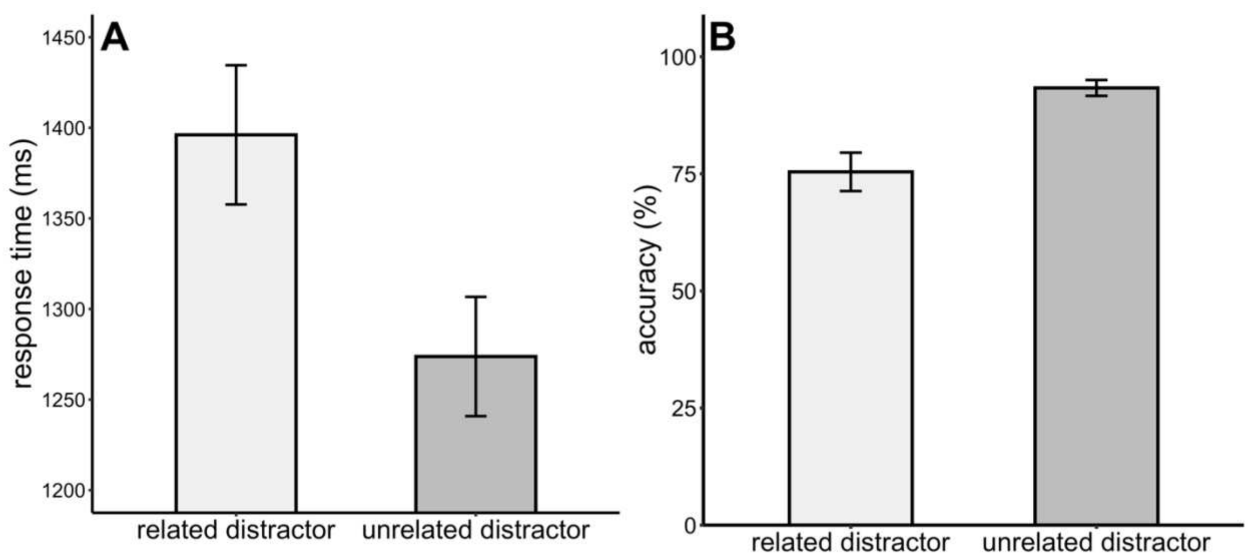
Response times

We used R (R Core Team, 2012) and *lme*4 (Bates et al., 2015) to perform a linear mixed effects analysis of RTs to word selection. As fixed effects, we entered the type of distractor into our first model. As random effects, we had by-subject and by-item random intercepts. In a second model, we added the distance (in number of words) between the metaphor combination and the lexical choice as a fixed effect. P-values were obtained by likelihood ratio tests of the full model against the null model including only random effects (subject and item). Our first model was a significantly better fit to the data than the null model ($\chi^2(1) = 17, p < .001$). The type of distractor affected the response time, increasing it by 117 ms, 95% CI [61.44 – 171.77]. Moreover, the model with distance as a second predictor was not significantly better than the first model. There was no main effect ($\chi^2(1) = 1.90, p = .16$) or interaction ($\chi^2(1) = 0.39, p = .53$). Thus, the distance between the metaphorical expression and the maze juncture, which ranged between 1 and 3 words, did not affect RTs. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality.

As predicted, participants took longer to select the correct word (*attend*) when it was paired with the literally-related distractor (*fly*; M = 1396 ms, SD = 188) rather than the unrelated one (*cry*; M = 1274 ms, SD = 161), as shown in Figure 3B (panel A).

Additionally, due to the variation of position of the metaphorical word (i.e., first, second, or both) within the two-word combinations, we included position as a factor in our model. Results indicated no significant main effect ($\chi^2(2) = 0.59$, $p = .74$) or interaction ($\chi^2(2) = 1.75$, $p = .41$). Thus, the position of the metaphorical word, whether it was the first word (e.g., *warm welcome*), the second word (e.g., *early bird*), or both words (e.g., *red flag*), did not affect response times.

Figure 3B: RT and accuracy for the related and unrelated distractors



A: Response time to the correct word as a function of the related and unrelated distractors. B: Accuracy selecting the correct word as a function of the related and unrelated distractors. Error bars are standard errors.

Accuracy

For the accuracy analyses, we used the *glm* function (R Core Team, 2012) to perform logistic regression by modeling the probability of observing a correct word selection as a function of the type of distractor. Results indicated a reliable effect of type of distractor (logit difference: +1.96, SE = 0.29, z = 6.78, $p < .001$). We then calculated the probabilities of the log odds, which predicted that the probability of observing a correct word selection was 81% when the correct answer (*attend*) was paired with a related distractor (*fly*), while the probability of observing a correct word selection increased to 97% when the correct answer (*attend*) was paired with an unrelated distractor (*cry*).

As predicted, participants were less accurate when selecting the correct word (*attend*) when it was paired with the literally-related distractor (*fly*; *M* = 15.1, *SD* = 4.02) rather than the unrelated one (*cry; M* = 18.6, *SD* = 1.64).

Meaningfulness ratings

After conducting our main experiment, we conducted a meaningfulness rating study to ensure that the correct answer (*attend*) had a higher chance of being selected than both distractors (*fly* and *cry*) and, further, to determine whether both distractors were equally anomalous to continue the sentence. For this task, we recruited 21 additional participants from the Concordia community (*M* = 23.86, *SD* = 3.8; 17 F), none of whom had participated in the main experiment. The study was conducted online, using the Pavlovia (pavlovia.org) platform, with all 24 experimental sentences and 24 fillers. Each trial consisted of an individual sentence segment presented in isolation (e.g., *John / is an / early / bird / so he can / attend / morning classes*). Participants were instructed to determine how meaningful each segment was as a continuation of the sentence by pressing a number between 1 (very bad) and 5 (very good) on the computer keyboard. Participants were encouraged to use the full scale with examples of what would constitute bad, good, and not-so-good/bad continuations. After each segment rating, participants were immediately presented with the next segment, and so on until the end of the sentence. We avoided presenting syntactic units (e.g., *an early bird*) in order to obtain ratings for individual constituents of the metaphor expression and, more importantly, for the segment that contained either the correct maze choice (*attend*), the related distractor (*fly*), or the unrelated distractor (*cry*). Three lists were created, each one containing eight items of each kind (correct, related, or unrelated), so that participants only saw one of the versions of the same sentence.

We used R (R Core Team, 2012) and *lme*4 (Bates et al., 2015) to perform a linear mixed effects analysis of meaningfulness ratings, taking into account ratings provided at the key maze choice word. As fixed effects, we entered the type of answer into our full model. As random effects, we had by-subject and by-item random intercepts. P-values were obtained by likelihood ratio tests of the full model against the null model including only random effects (subject and item). The full model was compared to a null model consisting of only random effects and was found to provide a statistically significant better fit to the data, $\chi^2(2) = 160.15$, $p < .001$, which also indicated an overall significant main effect of answer type. Planned comparisons between answer types showed that correct answers yielded significantly higher mean meaningfulness ratings than both related, $t(462) = 7.38$, $p < .001$, and unrelated distractors, $t(462) = 13.81$, $p < .001$. That is, regardless of whether the correct answer was paired with a related or an unrelated distractor, the correct answer was always the most meaningful (correct: $M = 4.36$, $SD = 0.93$; related: $M = 3.39$, $SD = 1.54$; unrelated: $M = 2.54$, $SD = 1.47$). In addition, the related distractor yielded significantly higher mean meaningfulness ratings than the unrelated distractor, $t(462) = 6.41$, $p < .001$. Therefore, to guarantee that both related and unrelated distractors were equally anomalous as maze choices, we removed 7 items where the difference in mean meaningfulness ratings between the related and unrelated distractors was the highest. Afterwards, planned comparisons to the new dataset including 17 items indicated that correct answers remained significantly higher than both related and unrelated distractors (correct: $M = 4.30$, $SD = 0.97$;

related: *M* = 2.99, *SD* = 1.58; unrelated: *M* = 2.96, *SD* = 1.49), while meaningfulness ratings for both related and unrelated distractors were not significantly different, $t(322) = 2.09$, p = .09.

Based on the results of this meaningfulness rating task, we conducted further RT and accuracy analyses restricted to the 17 items for which the correct answer was more meaningful than both related and unrelated distractors, while related and unrelated distractors were rated equally anomalous. The rating scores for all materials as well as the items removed for the purposes of re-analyses are shown in the Appendix.

RT re-analyses

For response time re-analyses, we also used *lme*4 (Bates et al., 2015) to perform a linear mixed effects analysis of RTs to word selection, using the same data-analytic procedures employed with the full dataset. As in our main analyses, the full model was a significantly better fit to the data than the null model, $\chi^2(1) = 14.07$, $p < .001$. The type of distractor affected the response time, increasing it by 122 ms, 95% CI [58.53 – 185.31]. Further, the model with distance as a second predictor was not significantly better than the first model. That is, there was no main effect, $\chi^2(1) = 2.28$, $p = .13$, or interaction, $\chi^2(1) = 1.09$, $p = .29$.

Similar to our results with the full dataset, participants took longer to select the correct word (*attend*) when it was paired with the literally-related distractor (*fly*; *M* = 1397 ms, *SD* = 186) rather than the unrelated one (*cry; M* = 1264 ms, *SD* = 168). With regards to the variation in position of the metaphorical word within the two-word expression (first, second, or both), again results indicated no significant main effect, $\chi^2(2) = 2.15$, $p = .34$, or interaction ,$\chi^2(2) = 0.35$, $p = .83$.

Accuracy re-analyses

For the accuracy re-analyses, we again used the *glm* function (R Core Team, 2012) to perform logistic regression by modeling the probability of observing a correct word selection as a function of the type of distractor. Results, again, indicated a reliable effect of type of distractor (logit difference: +1.73, SE = 0.34, z = 5.06, $p < .001$). We then calculated the probabilities of the log odds, which predicted that the probability of observing a correct word selection was 84% when the correct answer (*attend*) was paired with a related distractor (*fly*), while the probability of observing a correct word selection increased to 97% when the correct answer (*attend*) was paired with an unrelated distractor (*cry*). As obtained in the main analyses with the full dataset, participants were less accurate when selecting the correct word (*attend*) when it was paired with the literally-related distractor (*fly*; M = 15.9, SD = 4.09) rather than the unrelated one (*cry; M* = 18.6, SD = 1.66).

Discussion

We employed a maze task to examine whether the literal meaning of a conventional metaphor could be triggered or "awakened" by a subsequent cue. This would indicate that the literal interpretation was available during conventional metaphor processing. Results support our awakening hypothesis, showing a significant increase in RTs to the correct alternative when it was paired with a literally-related (*fly*) rather than an unrelated distractor (*cry*). Furthermore, accuracy decreased significantly in the literally-related condition in contrast to the unrelated one. In a maze task, an increase in response time and a decrease in accuracy are indicators of a higher processing cost to a target word as a function of the type of distractor. As Gallant and Libben (2020, p. 7) suggested, the assumption is that at the maze juncture "it is impossible for the participant to suppress the activation of the distractor word or the consequences of its activation."

It is important to note that the maze task forces semantic composition, with each word's meaning being integrated into the ongoing proposition. At the time the juncture is presented, the two-word metaphor (*early bird*) has already been interpreted according to its conventional content, which is triggered only by the individual content of the lexical items that are accessed. That is, in order to obtain the pragmatic content *EARLY RISER* from the conventional metaphor *early bird* the two concepts *EARLY* and *RISER* need to be composed. Hence, the awakening effect cannot be accounted for by the "activation" of lexical items through simple association (e.g., *bird* → *fly*). Such associations are determined over the words' conceptual representations—thus, they are established via the *content* that each word yields. Figure 3C depicts the model we propose for the awakening effect. Crucially, our proposal is that the concepts obtained by lexical access may also form a proposition that is faithful to the input expression with the cue (e.g., *fly*) working to enhance the literal proposition built out of the lexical denotations of the token sentences.[2]

Several other studies have suggested that competing interpretations for sentences linger in working memory (see Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Reyna, Corbin, Weldon & Brainerd, 2016). Data from Christianson et al., for instance, suggest that, for temporary ambiguous sentences such as *While the man hunted the deer ran into the woods*, subjects retain two propositions, a false one, compatible with the idea that [*the man hunted the deer*], which is derived from misparsing the sentence, and a true one, compatible with the idea that [*the deer ran into the woods*]. However, if only *one* proposition about the sentence was constructed during incremental interpretation, then these two propositions would be mutually

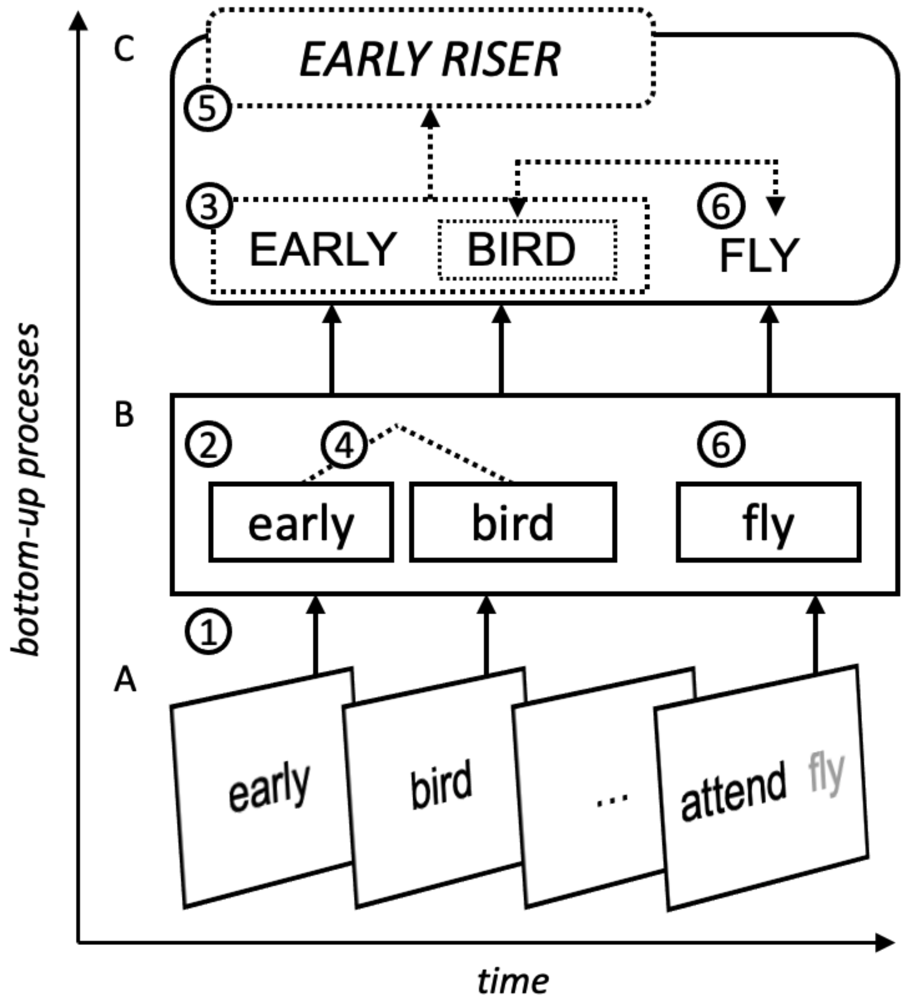
---

[2] While we have assumed that lexical concepts engage in classical composition (see, e.g., de Almeida, 2018, for discussion), our results and model (Figure 3C) are also compatible with different forms of compositionality (see, e.g., Pelletier, 2017, for a wide range of views on the nature of compositionality).

exclusive.[3] In the present case, we suggest that both literal and metaphorical interpretations linger, with the less salient literal content remaining a viable—if a dormant one—interpretation even for conventionalized metaphors. Although our study does not provide evidence for the nature of these interpretations—whether they are full propositions or only partial ones—our suggestion is that both literal and metaphorical interpretations are available simultaneously, compatible with theories of propositional memory representation that postulate "true" and "false" memory traces for sentences and events (see Reyna et al., 2016, for a review).

[3] As suggested by an anonymous reviewer, our results are compatible with an incremental, constraint-satisfaction model of sentence interpretation, which takes into account several sources of information (e.g., lexical, syntactic, semantic and even pragmatic) as every new lexical item is processed. However, it is also possible to assume that incremental interpretation is driven by structural principles (e.g., argument structure), largely shielded from pragmatics and world knowledge (see Ferreira & Nye, 2018, for discussion).

Figure 3C: The model for the awakening effect



The model entails two main levels of representation and processing after (A) lexical input: one for (B) word recognition and linguistic-structural operations, and one for (C) conceptual processes. During incremental interpretation, (1) input tokens (2) are recognized (lowercase words) and their conceptual representations (uppercase) (3) are accessed and incrementally composed into propositions. Then, the next incoming token items (4) are combined via linguistic operations, such as phrase structuring. In turn, this triggers a (5) stored, conventional representation associated with the full phrase interpretation. At the same time, the concepts that are accessed during lexical input also trigger (6) associated concepts. Finally, the relation among the related concepts is enhanced by the later cue, thus "awakening" the concepts accessed during literal interpretation. We suggest that a proposition related to the "lower" literal interpretation is available together with one built with the "higher" pragmatic information.

It could be argued, however, that the literally-related distractor *fly* may be working as an extension of the metaphor *early bird.* According to the conceptual metaphor theory (Lakoff & Johnson, 1980), conceptual metaphors (e.g., *LOVE IS A JOURNEY*) govern our thought and elicit multiple metaphorical expressions (e.g., *we are going too fast, it has been a bumpy road, this relationship is not going anywhere*). Thus, when arriving at the distractor *fly*, readers might assume that it is a continuation of the metaphor, which in turn might delay their response. Crucially, our experiment measures online responses, which do not allow the participant to reflect on the relationship between *early bird* and *fly*. Yet, if the reader took *fly* as a metaphorical extension, that would entail that the literal meaning of *early bird* was being accessed in real time to link *fly* and *early bird*, which indeed supports our hypothesis. Conversely, if the reader took *fly* literally and still longer RTs were obtained, it would follow that the meaning of *fly* triggers the literal meaning of *early bird*, which ultimately supports our hypothesis that the literal meaning of conventional metaphors had been initially available.

Our results suggest that the literal meaning of a conventional metaphor is not necessarily "short circuited" (Morgan, 1979), nor that conventional metaphors simply involve the retrieval of an associated meaning. The results rather suggest that the conventional content may take priority but the literal meaning is available and may be awakened by a subsequent cue. This cue might yield a conflict between the metaphorical content and the literal meaning, which may force a reinterpretation of the conventional expression. If so, this reinterpretation may involve a process that is similar to that of a novel metaphor. In this regard, our results are partially in line with those of Goldstein et al. (2012) who found that after explaining conventional metaphors, participants' pattern of activation for these conventional metaphors resembled that of novel metaphors. Crucially, our study demonstrates that a shift from conventional to literal meaning may occur in real time, given a triggering cue. In our study, the distance between the conventional metaphor and the trigger was between 1 and 3 words, which is estimated to be in the 250 – 750 ms range (Forster et al., 2009), suggesting that the awakening effect occurs fast and automatically. Our accuracy data also supports this conclusion: although the literally-related distractor fits syntactically, it renders the sentence semantically anomalous. Yet, in almost one-fourth of the trials, participants chose the literally-related distractor. These results were obtained with the full dataset and with a subset of materials that showed greater contrast between correct choice and the two maze distractors. Overall, this pattern can be further interpreted as a conflict between the dominant metaphorical content and the availability of the literal meaning, which lingers briefly until it is awakened by the literal cue.

The classical dichotomy between direct and indirect metaphor interpretation rests on the assumption that metaphors are either accessed directly or via literal interpretation. The effect we obtained suggests that a different process might be at work. Word constituents quickly give rise to metaphorical content as they attempt to semantically compose, but their literal meanings remain available and can be enhanced by further information in the context. This suggests that literal meanings are always accessed, but their availability may be inhibited by conventional content.

The view we espouse is not far from what other theories propose. For instance, for Bowdle and Gentner (2005), conventional metaphors can access literal meanings but with associated figurative senses having primacy over literal meaning during comprehension. This model indeed proposes that conventional vehicles of copular metaphors (e.g., *shark* in *My lawyer is a shark*) are polysemous, for they "refer both to a literal concept and to an associated metaphorical category" (p. 199)—which meaning (or sense) wins depends on numerous factors "including the context of the metaphor and the relative salience of each meaning of the [vehicle] term" (p, 199).

In addition, our results—and model—are compatible with Giora's (2003) graded salience hypothesis. In her model, if metaphorical content is most salient, it is accessed during comprehension, with salience being determined by conventionality and other stimulus properties.

Accordingly, the less salient meaning—possibly the literal meaning—"may not reach sufficient levels of activation to be visible" (p. 11). Beyond salience, the awakening effect might shed light on how literal meaning and pragmatic content might interface in the course of sentence interpretation. It suggests that pragmatic content is quickly computed, without however impeding on underlying processes of semantic—i.e., literal—access and composition. As words compose into propositions, there should not be a need to re-interpret the content of a conventional metaphor if all that is happening is simply *retrieval* of a conventional content. The awakening cue signals that information about the literal interpretation of a conventional metaphor lingers during interpretation processes. To wit, it suggests that broken hearts can be literally mended.

Chapter 4: Examining the awakening effect during the late stages of metaphor interpretation

The results from the previous study indicated a significant awakening effect. The awakening effect refers to the recovery of the literal meaning after the conventional metaphor has been processed. In the case of the previous experiment, the awakening effect was demonstrated by longer response times and lower accuracy rates in trials with the related cue in comparison to those with the unrelated cue. This pattern of results suggests that the literal meaning was accessed during metaphor comprehension, and the related cue was able to prompt it when presented immediately afterward, which in turn caused conflict with the figurative content of the metaphor.

Yet, the path of the literal meaning during metaphor comprehension remained unclear. In Chapter 4, we adapted the number of words between the offset of the metaphor and the maze juncture to examine whether the literal meaning could be recovered at two later time points. In the previous study, the distance between the metaphor and the maze juncture was between 1 and 3 words. This distance was increased to 6 to 8 words in the medium maze (Experiment 1) and 10 to 13 words in the large maze (Experiment 2).

This manuscript was accepted for publication in June 2023:

Pissani, L., & de Almeida, R. G. (In Press) Early birds can fly: Awakening the literal meaning of conventional metaphors further downstream. *Metaphor and Symbol*

Early birds can fly: Awakening the literal meaning of conventional metaphors further downstream

Laura Pissani and Roberto G. de Almeida

Concordia University

Abstract

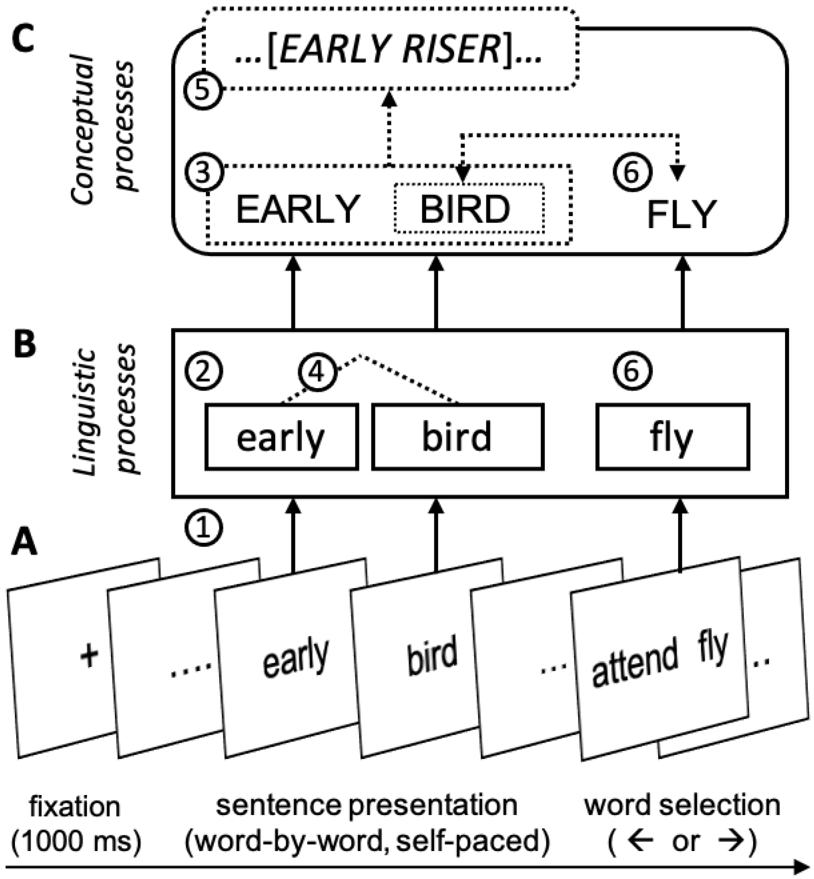
Conventional metaphors such as *early bird* are interpreted rather fast and efficiently. This is so because they might be stored as lexicalized, non-compositional expressions. In a previous study, employing a maze task, we showed that, after reading metaphors (*John is an early bird so he can*…), participants took longer and were less accurate in selecting the appropriate word (*attend*) when it was paired with a literally-related distractor (*fly*) rather than an unrelated one (*cry*). This suggests that the literal meaning of conventional metaphors is awakened or made available immediately after their metaphorical interpretation. But does the literal meaning remain available further downstream, during sentence comprehension? In two experiments also employing a maze task, we examined whether the awakening effect can be obtained when there is a medium (6 to 8 words) and a large (11 to 13 words) distance between the metaphor and lexical choice. Results indicated that the metaphor awakening effect persists but decreases as word distance increases. An analysis of our data based on a GPT model showed that our maze effect could not be attributed to target predictability. Overall, our results suggest that the literal meaning of a metaphor is accessed and remains available for about three seconds, fading as the sentence unfolds over time. The results support a model of metaphor comprehension that postulates the availability of both literal and metaphoric content, in the course of sentence processing.

Early birds can fly: Awakening the literal meaning of conventional metaphors further downstream

When someone says, *John is an early bird…* and continues *so he can fly to the morning classes,* attention is being called to the literal meaning of the conventional metaphor *early bird—perhaps as a pun or as a way to extend the metaphor*. But what does this tell us about the way a metaphor is processed in real time? In a recent study (Pissani & de Almeida, 2021), we examined the metaphor *awakening* hypothesis, which postulates that processing a metaphor (e.g., *early bird*) followed by a cue related to its literal meaning (e.g., *fly*) can cause the literal meaning of the metaphor to be enhanced or "awakened". The awakening hypothesis predicts that the literal meaning of a metaphor may be initially processed but remains dormant until the subsequent cue in the sentence "awakens" it.

In our previous study, we employed a maze task (e.g., Forster et al., 2009; Forster, 2010), wherein participants read sentences word by word in a self-paced manner and were presented with a two-word choice at a maze juncture. Each sentence contained a metaphorical combination (noun-noun [*NN*] or adjective-noun [*AdjN*]; e.g., *early bird*) followed by one to three words (e.g., *John is an early bird so he can …*) after which, participants had to choose the most suitable word out of two alternatives provided in order to continue the sentence (see Figure 4A for sample procedure). At the maze juncture, the appropriate word (e.g., *attend*) was accompanied by one distractor that could be either related (e.g., *fly*) or unrelated (e.g., *cry*) to the literal meaning of the metaphor. Participants took significantly longer and were significantly less accurate when the appropriate word was paired with the literally-related distractor rather than the unrelated one. That is, when *attend* was presented with *fly* as the distractor, participants took longer to select *attend* and made more errors (thus, selecting *fly*) compared to when the maze contained the unrelated word *cry* as the distractor.

Figure 4A: The expanded model for the awakening effect



(A) The maze procedure employed in the present Experiments 1 and 2 as well as in Pissani and de Almeida (2021); the difference between the experiments was the distance between the maze juncture (word selection) and the metaphorical expression (early bird). (B) Schematic representation of the proposed linguistic processes involved in the early processing of conventional metaphors. Those include word recognition and syntactic structuring of incoming phrases. (C) Schematic representation of conceptual processes encompassing the access of token words' meanings (concepts) and their composition into propositions, including further pragmatic processes such as access to the metaphorical content associated with words and conjoined phrases. Words written in lowercase letters stand for lexical items; words in uppercase letters stand for concepts. The "awakening" effect is represented in (6): access to a concept related to one of the concepts that constitutes the metaphorical expression enhances the activation of its related concept (thus FLY causes BIRD to be awakened or enhanced). See text for further details.

In the present study, we examined whether the metaphor awakening effect can be obtained further downstream, that is, whether the literal meaning of a conventional metaphor can be recovered when there is a longer distance between the metaphor and the literally-related cue. To that effect, we adapted our materials to increase the time interval between the metaphor and the juncture to 6-8 words for the *medium* maze (Experiment 1), and to 11-13 words for the *large* maze (Experiment 2). This manipulation was motivated by two goals: first, we were interested in tracing the time course of the literal meaning of a metaphor—that is, whether or not the literal meaning remains as a viable interpretation even long after the conventional metaphorical content has been accessed and incorporated into ongoing processes of interpretation. Second, more broadly, we wanted to inform models of *real-time* metaphor comprehension, shedding light on how the composition of metaphorical sentences might be attained. We start by briefly situating our study in the context of metaphor models and the variables of interest in the present study.

Models of metaphor comprehension

Metaphor comprehension models have ranged between those that assume direct access to stored metaphorical content (so-called *direct* models; e.g., Gibbs, 1994; Gibbs & Colston, 2012) and those that propose a literal-first interpretation, with the metaphorical interpretation being further obtained by pragmatic inferences (*indirect* or *pragmatic* models; e.g., Grice, 1975; Searle, 1979). Direct models propose that literal meaning can often be bypassed, either because there is no proper *literal* meaning, given that words may conventionally carry metaphorical senses, or because context determines the interpretation of a figurative expression without the need to resort to a literal interpretation (see, e.g., Gibbs & Colston, 2012, for a review). Indirect models, on the other hand, assume that literal meaning is first accessed, then rejected—because it flouts conversational principles—finally giving rise to a metaphorical interpretation. It is important to note that one of the main points of contention between these models is the idea that a literal interpretation might be attained and then "rejected" (Clark & Lucy, 1975; Searle, 1979). Direct models do not deny that some degree of access to a more literal interpretation of words in the sentence may occur (Gibbs & Colston, 2012). But they dispute the idea that a literal interpretation of the *whole* expression might be constructed and only then *rejected* before an alternative, contextually appropriate interpretation is sought out. Thus far, most studies taken as evidence for the direct access model have employed offline methods such as paraphrasing (Harris, 1976), sentence classification (Pollio et al., 1984), semantic judgement (McElree & Nordlie, 1999), and sentence verification (Glucksberg et al., 1982), neither of which examines the time-course of metaphor interpretation. The present study bears on the *direct-indirect* debate by cueing the literal meaning of highly conventional metaphors using an online method that targets two different time points during metaphor interpretation.

We should note that metaphor theories are not exhausted by the *direct-indirect* axis. Numerous other variables enter into the determination of when and how a metaphorical expression is interpreted—chiefly among them is *conventionality.* For instance, the conceptual metaphor theory (CMT; Lakoff & Johnson, 1980) proposes that conventional, seemingly superordinate conceptual metaphors such as LOVE IS A JOURNEY give rise to everyday expressions such as *Are we going too fast in our relationship?* or *It has been a long, bumpy road* by mapping properties of the vehicle (JOURNEY) onto the topic (LOVE). Moreover, conceptual metaphors are purported to be productive and to elicit novel linguistic expressions that function as metaphorical extensions within the same metaphor family. For instance, in her song *Getaway Car*, Taylor Swift describes a doomed romantic relationship by introducing novel expressions such as *You were driving the getaway car* and *there were sirens in the beat of your heart* which, under the conceptual metaphor theory, could be described as stemming from the conceptual metaphor LOVE IS A JOURNEY. Lakoff (1993) suggests that extensions of conceptual metaphors are readily interpretable since cross-domain mappings belong to our conceptual system, which enables us to draw metaphorical inferences in real time. In this sense, CMT is compatible with direct access theories of metaphor processing, which predict that metaphor interpretation occurs automatically. More recently, Thibodeau and Durgin (2008) found that the comprehension of novel metaphorical expressions was facilitated by a series of preceding sentences seemingly accessing a related conventional conceptual metaphor. In their series of experiments, participants were instructed to press a button after having comprehended each sentence. They found that target expressions (e.g., *She loved to gamble*) were facilitated after having read scenarios involving either conventional (e.g., […] *Joan decided to take her chances and have the operation*) or novel expressions (e.g., […] *Joan decided to ante up and have the operation*) as long as they stemmed from supposedly the same conceptual metaphor (e.g., LIFE IS A GAMBLING GAME). Crucially, they suggested that conventional metaphors are productive and can facilitate novel metaphors within the same metaphor 'family' as proposed by Lakoff and Johnson (1980).

Contrary to CMT, the metaphor awakening effect (Pissani & de Almeida, 2021) is in line with indirect theories of metaphor processing that propose that metaphor interpretation is mediated by access to the literal meaning. Accordingly, we suggest that the literal interpretation of an expression deemed metaphorical can be quickly recovered by any cue that is related to its literal meaning[4]. Also, while CMT proposes that novel extensions of conventional metaphors can be understood directly given pre-existing conceptual mappings, the metaphor awakening effect suggests that novel extensions prompt access to the literal meaning of conventional metaphors. Thus, the metaphor awakening effect and CMT converge in that conventional metaphors may

[4] We note, however, that the present study was not designed to investigate the conceptual metaphor theory (Lakoff and Johnson, 1980; Lakoff, 1993), nor were our materials suited for such a task. Although some of the cues we employed may be taken as novel extensions of conventional metaphors, we cannot assert that (a) all of our cues are metaphorical extensions, as some may be literal cues. For instance, we anticipate that the literal meaning of *cold feet* can be triggered by the cue *warm* regardless of whether the latter is used metaphorically (e.g., *warm welcome*) or literally (e.g., *warm weather*). Nor can we assert that (b) all extensions belong to the same metaphor family, for we may not have the theoretical grounds to establish all cases in which a metaphor belongs to one or another metaphor family. For instance, it is not obvious whether *warm blood, warm gesture, hot take, hot minute*, *cold glance*, *cold turkey, cool cat,* and *cool head* belong to the same metaphor family.

facilitate access to their novel extensions. Further, we have suggested that the literal interpretation—initially rejected or not—is available even when the metaphor is highly conventional. But, assuming that the interpretation of a metaphorical expression, during real-time sentence processing, occurs soon after the expression is processed, how is it possible for the literal meaning to still be available? And how long does the literal meaning remain as a viable interpretation for the metaphorical expression? These questions are posed on the assumption that the *proposition* that the metaphorical sentence expresses carries among its constituents the metaphorical content, *not* the literal one. If, however, the literal meaning can still be "awakened", this suggests that the metaphorical expression can be *re-interpreted*—or *shifted* from the ongoing metaphorical content back to the literal meaning.

The alleged shift from novel to conventional

As mentioned above, a key factor in the way a metaphor is processed is its degree of conventionality. This is so because metaphor use is taken to strengthen a particular relationship (e.g., between particular topics and vehicles) or, in Lakoff and Johnson's (1980) system, it emerges from a conventionalized mapping between different conceptual domains. For some, the meaning of a conventional metaphor is "stored" (Morgan, 1979) or "lexicalized" (Blank, 1988) or invites "sense retrieval" (Bowdle & Gentner, 2005), contrary to novel metaphors, which involve "sense creation". Indeed, this contrast cuts across models of metaphor interpretation—whether the metaphorical content is accessed directly or indirectly. In Bowdle and Gentner's (2005) proposal, for instance, the process of metaphor interpretation may vary, depending on how novel or conventional it may be. When a metaphor of the form *X is Y* is novel (e.g., *My boyfriend is a capybara*), it is interpreted as a comparison (Gentner, 1983), with its interpretation involving "structural alignment" between predicates related to the vehicle (*capybara*) and the topic (*boyfriend*). But as the metaphor becomes conventionalized, there is a "shift" in processing mode, from comparison to categorization, with the vehicle obtaining a new sense, one that corresponds to a more abstract category related to the figurative sense.

More relevant to the present study is how metaphors of the form *XY* (*NN* or *AdjN*) might be interpreted. We assume that a novel expression such as *tent dress* would be interpreted by comparing (or "structurally aligning") the vehicle *tent* to the literal topic *dress*—thus possibly yielding an interpretation such as that of a dress with a certain shape or extremely loose. The conventionalization of metaphors is also taken to involve structural alignment, with supposedly common predicates between topic and vehicle becoming increasingly salient over time, leading to the creation of a novel abstract category. For instance, for the conventional expression *bright student*, *student* (topic) is included in the category *bright* (vehicle), which refers to people or objects that are smart, and it is used productively in other expressions such as bright idea, bright author, etc. And because conventionalization creates polysemy, both literal (e.g., bright as in emitting or reflecting a lot of light) and figurative (as in smart) senses of the conventional vehicle may be accessed during comprehension. Thus, following the career of metaphor, conventional metaphors carry ambiguous or polysemic terms, and their senses can be available, thus potentially allowing for a shift in interpretation.

An interesting hypothesis of Bowdle and Gentner's (2005) theory, then, is the "shift" from comparison to categorization that may occur over time. In their Experiment 3, they investigated this hypothesis using a sentence completion task (study phase) and a rating task (test phase). In each trial of the study phase, participants were exposed to two expressions with the same vehicle but different topics in either the simile or metaphor form (*A bee is* [*like*] *a butterfly*; *A moth is* [*like*] *a butterfly*) and had to provide a topic to a third statement containing the same vehicle (e.g., ______ *is* [*like*] *a butterfly*). In the test phase, participants were presented with a scale—1 to 10—to indicate which form (simile or metaphor) they preferred, with the two expressions (e.g., *A ballerina is like a butterfly* and *A ballerina is a butterfly*) constituting the extreme points of the scale. Bowdle and Gentner found that exposure to novel expressions (i.e., those containing novel vehicles) in similes led to a greater preference for the metaphor form. The claim is that the lab manipulation (study phase) affected the conventionality of the expression, supporting the processing shift hypothesis.

Goldstein et al. (2012) examined this shift from novel to conventional in an ERP study that first required participants to explain novel metaphors. Goldstein et al. reasoned that by having participants explain novel metaphors, these would become conventionalized thus affecting how they were processed—supposedly "shifting" them from novel to conventional. Results showed that the subset of novel metaphors that participants were asked to explain prior to the ERP study yielded the same ERP patterns as those of conventional metaphors, namely, a smaller N400 and a greater late positive complex (LPC) amplitude. Moreover, Goldstein et al. examined the opposite over-time shift, from conventional to novel. They found that explaining conventional metaphors had the reverse effect compared to novel metaphors: the subset of conventional metaphors that was explained yielded the same ERP patterns as a novel metaphor, in particular, a greater N400 and a reduced LPC amplitude.

The present study

To our knowledge, our previous study (Pissani & de Almeida, 2021) was the first to examine a processing shift from conventional to "novel"—that is, to a *literal* interpretation—in real time. Contrary to a shift over time, by hypothesis, a real-time shift involves automatic access to the literal meaning and possibly a reinterpretation of the expression—not metaphorically but literally. The awakening effect we proposed postulates that as a conventional metaphor is processed, the metaphorical content may be accessed only after the literal meaning is incrementally composed. The model in Figure 4A illustrates the sequence of events bearing on the incremental interpretation of a metaphorical expression and its awakening by a subsequent cue. Linguistic processes (Panel B) operate on the perceptual input (step 1 in Panel A) of lexical tokens (step 2), which incrementally access their conceptual representations (step 3 in Panel C). While further linguistic processes structurally merge incoming lexical tokens (step 4), the proposition that the sentence expresses is quickly formed with the conventional content (step 5). The availability of the literal meaning of the metaphor (*early bird*) is obtained by a subsequent cue (*fly*) related to the literal meaning of a constituent (*bird*) of the conventional metaphor (step 6). Notice that, by hypothesis, the proposition that the sentence expresses may have already incorporated the figurative sense of the metaphor (as in step 5) and by awakening or enhancing the literal sense, the proposition may engender a brief shift from a metaphorical interpretation *back* to a literal one. Thus, by engaging participants in a fast-paced rather than an off-line thinking task (viz., overtly explaining the meaning of a novel metaphor as in Goldstein et al., 2012), results may reflect the moment-by-moment processes of metaphor interpretation—and in particular the availability of the literal meaning. This suggests the possibility of a real-time shift in interpretation, from metaphorical to literal, which is brought to bear by the literally-related cue.

The present study aimed to examine to what extent the literal meaning remained available—whether or not it could be awakened further downstream. That is, the goal was to trace the availability of the literal meaning—thus, the extent of the awakening phenomenon—after many words and phrases have been interpreted. Since the proposition that the sentence expresses might be formed incrementally, it is expected that the greater the lag between the conventional metaphor content (step 5) and the appearance of the awakening cue (step 6), the harder it should be for the awakening effect to be obtained. But it could also be possible for an alternative proposition—one with the literal content—to remain active, in parallel with, though perhaps not with the same strength of the proposition carrying the figurative content.

Experiment 1: The medium maze

Method

Participants

Participants were 40 native speakers of English between the ages of 19 and 35 years old (*M* = 23.1, *SD* = 3.65; 32 F) with normal or corrected-to-normal vision. The sample size was based on Pissani and de Almeida (2021; *N* = 40), which used the same design to examine the metaphor awakening effect. Participants were recruited via the Concordia University online participant pool and were compensated with course credit.

Materials

Materials were based on Pissani and de Almeida's set of 24 experimental sentences, which contained highly conventional two-word metaphor combinations (e.g., *broken heart*). In that study, the distance from the metaphor (*early bird*) to the maze juncture (*attend/fly*, *attend/cry)* was between one and three words (e.g., *John is an early bird so he can attend morning classes*). In the current study, we increased the number of words such that there were six to eight words (*M* = 7.17, *SD* = .81) between the metaphor and the maze juncture (e.g., *John is an early bird so he can eat a healthy breakfast and attend morning classes*). For each sentence, we employed the original two types of distractors, each of which accompanied the correct answer in the maze juncture: (1) a related distractor (e.g., *fly*), semantically associated with the literal meaning of the metaphorical expression, and (2) an unrelated distractor (e.g., *cry*), not semantically associated with the literal meaning of the metaphorical expression. Additionally, both types of distractors were matched to each other in overall frequency, length, and part of speech[5]. It is important to note that the original set of 24 items, including both distractors, was rated afterward for meaningfulness to ensure that both distractors were equally anomalous to continue the sentence while making sure that the correct word was distinctively the best fit. In doing so, seven items were removed resulting in a reduced set of 17 items. Here, we present both results, starting with the full set of 24 items followed by the re-analyses with the reduced set of 17 items in order to compare the present results with those of our previous study.

The full set of materials was distributed in two mixed lists, wherein the type of distractor was counterbalanced within subjects so that participants were exposed to both conditions, but only presented with one version of each item. Thus, participants saw a half of the sentences in the related-distractor condition and the remainder in the unrelated-distractor condition. Overall, participants were presented with 24 experimental trials plus 48 filler sentences, 25% of which were followed by a closed-ended comprehension question (e.g., *Does John prefer the evening classes?).* Additionally, the position of the distractor was counterbalanced so that half of the participants read the distractor on the right, and the other half read the same distractor on the left (e.g., *attend/fly* or *fly/attend*).

[5] Frequency scores were obtained from the Medical Research Council (MRC; Coltheart, 1981) psycholinguistic database, and their logged values were compared between conditions, which yielded no significant differences, $t(23) = 1.41$, $p = .171$. Length was verified using the LEN() function on Excel, and part of speech was verified by the authors.

Table 4A: Sample materials employed in all mazes

	Example of stimuli per experiment
The original maze (Pissani & de Almeida, 2021)	*John is an early bird so he can (attend - fly/cry) morning classes.* 1 2 3
Experiment 1: The medium maze	*John is an early bird so he can eat a healthy breakfast and (attend - fly/cry)* 1 2 3 4 5 6 7 8 *morning classes*.
Experiment 2: The large maze	*John is an early bird so he can go to the gym and exercise and then* 1 2 3 4 5 6 7 8 9 10 11 *(attend - fly/cry) morning classes.*

Sample materials employed in the present study (Experiments 1 and 2) and in our previous study (Pissani & de Almeida, 2021). Participants saw word by word in a self-paced manner and were required to make a word choice at the maze juncture (shown here in parenthesis). The only difference between the medium and the large mazes was the number of words between the metaphor (*early bird*) and the maze juncture (*attend - fly/cry*).

Procedure

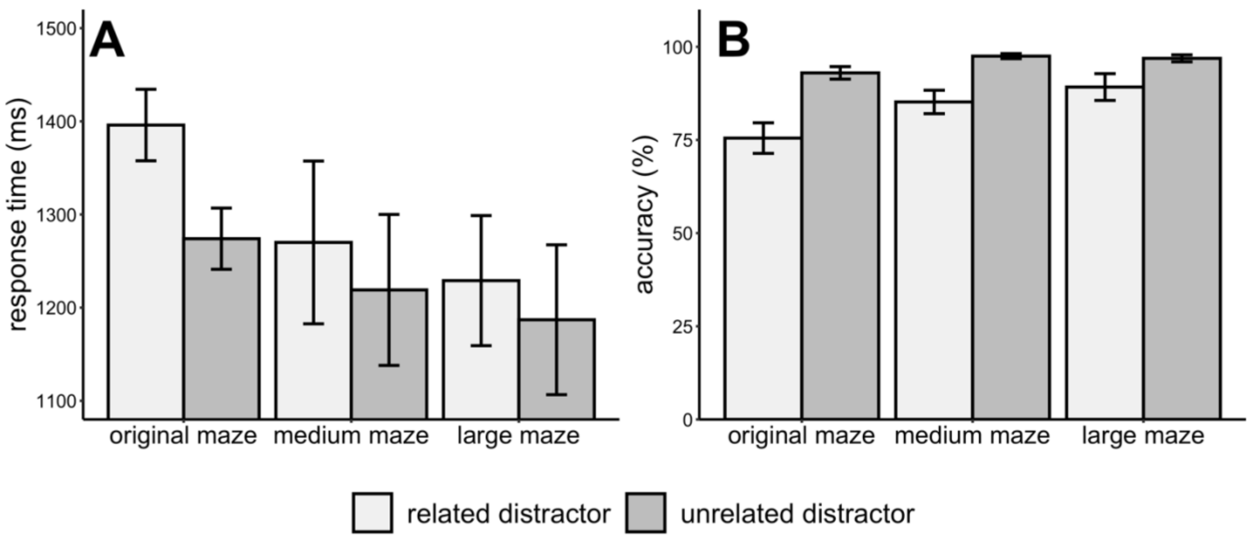
The experiment was run using the Pavlovia online platform (pavlovia.org). Upon registration, all participants signed a consent form, after which they were assigned to one of two lists following the order of registration. We used a version of the maze task (Forster et al., 2009; Forster, 2010) in which each trial started with a 1000 ms fixation cross followed by one word at a time (see Figure 4A). Participants were instructed to press the downward arrow on their keyboard (↓) to move to the next word of the sentence. At the maze juncture, participants saw two words in parallel and had to press either the left (←) or right (→) arrow to select the word that best fits the sentence. During word selection, the correct answer was paired with one of two distractors. Thus, half of the participants saw *attend* versus *fly* (literally-related distractor), whereas the other half saw *attend* versus *cry* (literally-unrelated distractor).

Data preparation

All analyses were restricted to participants who reached 75% or higher in both the lexical choice and comprehension portions of the maze task. Overall, the accuracy scores for the lexical choice ranged between 83 and 100 % (*M* = 92.75, *SD* = 3.49), while for the comprehension section, scores ranged between 78 and 100 % (*M* = 94.2, *SD* = 5.75). Thus, no participant was removed from the analyses. Further, incorrect trials in either the lexical choice task (8.65 % of all data points) or the comprehension section (0.73 % of all data points) were not included in the RT analyses. Further, given the nature of the task (i.e., to capture real-time metaphor processing), RTs larger than 3000 ms were removed prior to data analyses (1.98 % of all data points) assuming that values above this magnitude increase the probability that responses will be affected by strategic effects (viz., the cognitive penetrability of the task; Pylyshyn, 1984). Lastly, remaining outliers (i.e., data points larger than 2.5 *SD* from the mean) were calculated per participant to preserve individual variability and replaced with 2.5 *SD* above the mean for the upper tail (1.88 % of all data points)[6].

[6] We employed a more conservative approach following Forster et al.' (2009) procedures (i.e., removing RTs longer than 1500 ms and replacing remaining outliers per participants with 2 *SD* above the mean).

Figure 4B: RT and accuracy across all mazes



Response times to the correct word continuation as a function of the related and unrelated distractors for the 24-item dataset in the original study (Pissani & de Almeida, 2021) and the present study's medium (Experiment 1) and large mazes (Experiment 2). B: Accuracy selecting the correct word as a function of the related and unrelated distractors for the 24-item datasets.

Results

Response times (24 items)

Figure 4B depicts the response times (RT) and accuracy results. We used R (R Core Team, 2012) and *lme*4 (Bates et al., 2015) to perform linear mixed-effects analyses of RTs to the word selection portion of the maze task. As fixed effects, we entered the type of distractor in the full model. As random effects, we had by-subject and by-item random intercepts. We obtained p-values by likelihood ratio tests of the full model against the reduced model. The full model was a significantly better fit to the data than the null model, $\chi^2(1) = 4.39$, $p = .036$. The type of distractor affected the RT, increasing it by 51 ms, 95% CI [3.27, 98.35]. We visually inspected residual plots and did not find any obvious deviations from homoscedasticity or normality. As predicted, participants were slower to select the correct word (e.g., *attend*) when it was paired with the literally-related distractor (*fly*; *M* = 1270 ms, *SD* = 428) compared to the unrelated one (*cry*; *M* = 1219 ms, *SD* = 397).

Further, to control for the position of the metaphorical vehicle, whether it was in the first (e.g., *warm welcome*), second (e.g., *early bird*), or both words (e.g., *red flag*), we included the position of the word as a factor in our model. Our analyses showed no main effect of position, $\chi^2(2) = 3$, $p = .223$, or interaction between the position of the vehicle and the type of distractor, $\chi^2(2) = 3.84$, $p = .147$. Thus, the position of the metaphorical vehicle did not affect RT.

Accuracy (24 items)

For the accuracy analyses, we used the *glm* function (R Core Team, 2012) to perform logistic regression following the steps recommended by Winter (2019). First, we modeled the probability of observing a correct word selection as a function of the type of distractor. Results indicated a reliable effect of the type of distractor (logit difference: +1.91, SE =0.32, z =5.99, $p <$ .001). We then calculated the probabilities of the log odds, which predicted that the probability of observing a correct word selection was 85% when the correct answer (*attend*) was paired with a literally-related distractor (*fly*) in the juncture, while the probability of observing a correct word selection increased to 98% when the correct answer was paired with a literally-unrelated distractor (*cry*). As predicted, participants were less accurate when selecting the correct word when it was paired with the literally-related distractor (*fly*; *M* = 17, *SD* = 3.07) rather than the unrelated one (*cry*; *M* =19.5, *SD* = 0.66).

Response times (17 items)

For RT re-analyses, we also performed linear mixed-effects analyses of RTs to word selection using the same data-analytic procedures employed with the full data set. For this reduced data set, the full model was not significantly better than the null model, $\chi^2(1) = 0.92$, $p = .335$. Note, however, that our full model estimated a slope of 29 ms, 95% CI [-29.79, 87.3] in the literally-related distractor condition. Thus, for the 17-item data set, participants tended to be slower to select the correct word when it was paired with the literally-related distractor (*fly*; M = 1264 ms, SD = 401) rather than the unrelated one (*cry*; M = 1234 ms, SD = 420), but this difference was not statistically significant.

Again, to control for the position of the metaphorical vehicle, we included it as a factor in our model. Our analyses showed no main effect, $\chi^2(2) = 2.28$, $p = .32$, or interaction, $\chi^2(2) = 1.86$, $p = .395$. Thus, the position of the metaphorical vehicle did not predict RT.

Accuracy (17 items)

For the accuracy re-analyses, we also modeled the probability of observing a correct word selection as a function of the type of distractor using logistic regression. Results indicated a reliable effect of the type of distractor (logit difference: +1.82, SE = 0.37, z = 4.9, $p < .001$). We then calculated the probabilities of the log odds, which predicted that the probability of observing a correct word selection was 86% when the correct answer was paired with a literally-related distractor (*fly*), which increased to 97% when it was paired with a literally-unrelated distractor (*cry*). Similar to the full set, participants were less accurate when the choice involved the literally-related distractor (*fly*; M = 16.2, SD = 4.93) in comparison to the unrelated one (*cry*; M =19.5, SD = 0.62).

Experiment 2: The large maze

Method

Participants

Participants were 40 native speakers of English between the ages of 19 and 36 years old (*M* = 23.7, *SD* = 4.96; 36 F) with normal or corrected-to-normal vision. Recruitment procedures followed those of Experiment 1. None of the participants in this experiment had participated in Experiment 1.

Materials

For this experiment, materials were also based on the original 24 core sentences from Pissani and de Almeida (2021). In Experiment 2, however, we increased the number of words between the metaphor and the maze juncture such that there were between 11 and 13 words (*M* = 11.75, *SD* = .79) (e.g., *John is an early bird so he can go to the gym and exercise and then attend morning classes*; see Table 4A).

Procedure

The procedure was identical to Experiment 1.

Data preparation

Data preparation procedures were identical to those in Experiment 1. Accuracy scores for the lexical choice ranged between 78 and 100 % (*M* = 93.65, *SD* = 4.56), while for the comprehension section, scores ranged between 72 and 100 % (*M* = 93.61, S*D* = 5.7). Incorrect trials in either the lexical choice task (6.04 % of all data points) or the comprehension section (1.46 % of all data points) were not included in the RT analyses.

In addition, RTs larger than 3000 ms were trimmed prior to data analyses (2.19 % of all data points), while outliers were replaced with 2.5 *SD* above the mean for the upper tail (1.35 % of all data points).

Results

Response times (24 items)

We also performed linear mixed-effects analyses of RTs to word selection as a function of the type of distractor. The full model was a significantly better fit to the data than the null model, $\chi^2(1) = 8.14$, $p = .004$. The type of distractor affected RT, increasing it by 55 ms, 95% CI [17.38, 93.22]. Similar to Experiment 1, participants were slower to select the correct word when it was paired with the literally-related distractor (*fly*; *M* = 1229 ms, *SD* = 342) rather than the unrelated one (*cry*; *M* = 1187 ms, *SD* = 349).

Lastly, we added the position of the metaphorical vehicle as a factor in our model. Our analyses showed no main effect, $\chi^2(2) = 1.24$, $p = .538$. Surprisingly, however, there was a significant interaction, $\chi^2(2) = 313.87$, $p < .001$, whereby the type of distractor affected only metaphor combinations when the vehicle was the first (by 149.89 ms, 95% CI [59.23, 240.55], $p < .001$) or second word (by 177.99 ms, 95% CI [55.38, 300.60], $p < .001$), but not when the vehicle comprised both words (-18.25 ms, 95% CI [-72.30, 35.81], $p = .508$).

Accuracy (24 items)

We also performed logistic regression by modeling the probability of observing a correct word selection as a function of the type of distractor. Results indicated a reliable effect of the type of distractor (logit difference: +1.33, SE = 0.3, z = 4.41, $p < .001$). In addition, the probabilities of the log odds predicted that the probability of observing a correct word selection was 89% when the correct answer was paired with a literally-related distractor (*fly*), which increased to 97% when the correct answer (*attend*) was paired with a literally-unrelated distractor (*cry*). Similar to Experiment 1, participants were less accurate when the correct word was paired with the literally-related distractor (fly; $M = 17.8$, $SD = 3.52$) rather than the unrelated one (cry; $M =19.4$, $SD = 0.92$).

Response times (17 items)

For RT re-analyses, we used the same data-analytic procedures employed with the full data set. For this reduced data set, the full model was also a significantly better fit to the data than the null model, $\chi^2(1) = 8.22$, $p = .004$. The type of distractor affected RT, increasing it by 69 ms, 95% CI [22.07, 116.56]. Similar to the full data set, participants were slower when the correct answer was paired with the literally-related distractor (*fly*; $M = 1226$ ms, $SD = 348$) rather than the unrelated one (*cry*; $M = 1208$ ms, $SD = 362$).

Again, we added the position of the metaphorical vehicle as a factor in our model. Our analyses showed no main effect, $\chi^2(2) = 1.71$, $p = .425$, or interaction ,$\chi^2(2) = 4.67$, $p = .097$. Thus, the position of the metaphorical vehicle did not predict RT.

Accuracy (17 items)

For the accuracy re-analyses, we also modeled the probability of observing a correct word selection as a function of the type of distractor using logistic regression. Results indicated a reliable effect of type of distractor (logit difference: +0.96, SE = 0.34, z = 2.72, *p* = .006). In addition, the probabilities of the log odds predicted that the probability of observing a correct word selection was 91% in the literally-related condition (*fly*), which increased to 96% in the unrelated condition (*cry*). Similar to the full data set, participants were less accurate when the correct word was paired with the literally-related distractor (*fly*; *M* = 18.2, *SD* = 4.02) rather than the unrelated one (*cry*; *M* = 19.2, *SD* = 1.03).

Comparison across maze experiments

RT analyses across maze experiments

In order to examine the change in the strength of the awakening effect as a function of the distance between the metaphorical expression (*early bird*) and the maze juncture, we compared RTs from the full data set for the related and unrelated distractors separately. First, we performed a linear mixed-effects analysis of RTs to word selection for the related distractor only. As fixed effects, we entered the maze distance (original, as in Pissani and de Almeida, 2021, medium [present Experiment 1], and large [Experiment 2]) in the full model. As random effects, we had by-subject and by-item random intercepts. We obtained p-values by likelihood ratio tests of the full model against the reduced model. The full model was a significantly better fit to the data than the null model, $\chi^2(2) = 9.312$, $p = .01$. Maze distance affected the RT when the related distractor was present, significantly decreasing it by 108.48 ms, 95% CI [14.9, 202.36], from the original to the medium maze, and by 141.52 ms, 95% CI [48.53, 234.51], from the original to the large maze. There was, however, no significant change from the medium to the large maze.

We then performed a linear mixed-effects analysis of RT to word selection for the unrelated distractor only. The full model was not significantly better than the null model, $\chi^2(2) = 2.829$, $p = .241$. Maze distance did not affect RT for the unrelated distractor.

As predicted, participants' responses in trials with the literal cue were significantly slower in the original maze (M = 1396 ms) in comparison to both medium (M = 1270 ms) and large mazes (M = 1229 ms). Conversely, in trials with the unrelated cue, there was no significant change in RT across mazes.

Accuracy analyses across maze experiments

We also examined the change in the strength of the awakening effect across mazes by comparing the accuracy of the responses from the full data set for the related and unrelated distractors separately. First, we performed logistic regression by modeling the probability of observing a correct response as a function of the maze distance for the related distractor only. Results indicate no effect of maze distance between the original and the medium maze (logit difference: +0.28, SE =0.16, z =1.78, p = .076). The probability of observing a correct word selection for the trials that contained the literal cue was 89% in the original maze and 91% in the medium maze. There was, however, a reliable effect of maze distance between the original and the large maze (logit difference: +0.51, SE =0.17, z =3.09, p = .002). The probability of observing a correct word selection for the trials that contained the literal cue was 89% in the original maze, which significantly increased to 93% in the large maze. But there was no significant change from the medium to the large maze.

Finally, we performed logistic regression for the unrelated distractor only. Results indicated no effect of maze distance between the original and the medium maze (logit difference: +0.27, SE =0.4, z =0.69, p = .493), nor between the original and the large maze (logit difference: +0.04, SE =0.37, z =0.11, *p* =.96). The probability of observing a correct word selection for the trials that contained the unrelated cue was 97% in the original maze, 98% in the medium maze, and 97% in the large maze.

Participants' responses in trials with the literal cue were significantly more accurate in the large maze in comparison to both the original and medium mazes. However, in trials with the unrelated cue, there was no significant change in accuracy.

Expressions' predictability

We used the GPT-2[7] model (Radford et al. 2019a, Radford et al. 2019b) to calculate the probability that each sentence (e.g., *John is an early bird so he can* [juncture]) would be completed with the correct word (e.g., *attend*), related distractor (e.g., *fly*), and unrelated distractor (e.g., *cry*). These cloze probability scores are taken to capture the predictability of a word as a function of the preceding words. We then used the negative log probability to derive surprisal scores, which are a measure of the degree of difficulty for each target word (Levy, 2008).

Considering that, as mentioned above, our dependent variables were RT and accuracy to perform a lexical choice, we reasoned that the difference between the surprisal scores for the correct word and each distractor would capture the difficulty, due to surprise, to select the correct word at the maze juncture. The obtained surprisal difference scores were significantly different between conditions, $t(23) = 4.42$, $p < .001$, which indicates that choosing the correct word was

[7] A language model such as GPT-2 can perform language tasks such as reading comprehension, summarization, translation, and question answering. In addition, the GPT-2 model yields reliable estimates of cloze probabilities as it has been trained on approximately 8M webpages to predict the next word given the previous ones (Radford et al., 2019a). To obtain our surprisal scores, we used the large version of the GPT-2 model, which contains 762M parameters and 36 layers (Radford et al., 2019b).

more difficult when it was paired with a related distractor (correct versus related distractor; *M* = -2.59, *SD* = 3.74) rather than an unrelated one (correct versus unrelated distractor; *M* = -6.13, *SD* = 4.02). A greater difference in processing difficulty engendered by the "surprisal" effect may be associated with a slower RT and a lower accuracy rate due to increased competition between word choices. Hence, we included the difference in surprisal scores for all 24 items in our LME models to examine whether they were sufficient to explain our observed effects.

RT as a function of surprisal difference scores

We used the same data-analytic procedures employed in our previous analyses to model the effect of the surprisal difference scores (i.e., the surprisal score for the correct word minus the surprisal score for the distractor with which it was paired) in both the medium (Experiment 1) and large (Experiment 2) mazes. The full models, including the type of distractor and surprisal difference scores as fixed factors, were compared to their respective reduced models without the effect of the surprisal difference scores. Neither of the full models provided a better fit to the data when compared to the reduced models. Thus, there was no main effect of surprisal difference scores for the medium maze: $\chi^2(1) = 1.08$, $p = .298$, $R^2 = 0.28$, 95% CI [-15.37, 5.08] or large maze: $\chi^2(1) = 2.41$, $p = .12$, $R^2 = 0.38$, 95% CI [-16.26, 1.95]). There was also no interaction between the type of distractor and the surprisal difference scores for the medium, $\chi^2(1) = 0.06$, $p = .804$, $R^2 = 0.28$, 95% CI [-16.88, 8.14] or the large maze, $\chi^2(1) = 2.39$, $p = .12$, $R^2 = 0.39$, 95% CI [-23.51, -1.14].

Accuracy as a function of surprisal difference scores

In order to follow up our accuracy analyses, we incorporated the surprisal difference scores as a second predictor in our generalized linear model, as computed above, to examine how this difference affects the probability of observing a correct word selection. For the medium maze, results indicated no reliable effect of surprisal difference scores (logit difference: -0.009, SE = 0.032, z = -0.281, p = .778). For the large maze, results also indicated no reliable effect of surprisal difference scores (logit difference: -0.008, SE = 0.035, z = -0.239, p = .811). Thus, the probabilities of the log odds predicted in sections 2.2.2 and 3.2.2 remain unaltered for both the medium and large mazes, respectively.

Discussion

We investigated the extent to which the literal meaning of conventional metaphorical expressions such as *early bird* are available during sentence comprehension, aiming to further understand the *awakening* effect we proposed (Pissani & de Almeida, 2021). In our previous study, also employing a maze task, we found that in trials where the metaphor was immediately followed by a literal cue (e.g., *fly*), participants' choice for the correct continuation of the sentence at the maze juncture was slower and less accurate in comparison to trials with an unrelated cue. It has been a common assumption in metaphor research that the content of a conventional metaphor is stored, rather than built incrementally during real-time processing. Thus, by hypothesis, once this metaphorical content is accessed, it takes its place in the proposition incrementally built by the comprehender. Our effect was obtained when the cue appeared 1 to 3 words after the second word of a Noun-Noun or Adjective-Noun combination, suggesting that the literal meaning was immediately available—though, perhaps, "dormant"—for further processing.

The present study extended those results, employing the same technique and similar materials, but with a greater distance between the metaphorical expression and the maze juncture aiming to trace the time course of the availability of the literal meaning. The results—relying on both RT and accuracy patterns—suggest that the literal meaning is available beyond the 1-3 words we measured before, remaining active despite the intervening processing of words and phrases. We estimated that the literal interpretation was available for up to 13 words (or about 3250 ms) after the metaphor had been processed. The availability of literal meaning, however, seems to decay at that point, as shown by the differences in RT and accuracy between the short maze (1-3 words) of our original study and both the medium (Experiment 1) and large (Experiment 2) mazes investigated in the present study. Further, our analyses relying on GPT-2, suggest that the differences we obtained may not be due to the greater predictability of each target as a possible continuation for the sentence. The locus of our effect—extending our previous results (Pissani & de Almeida, 2021)—seems to be in the availability of the literal content of the metaphorical expression beyond its metaphorical interpretation.

What is perhaps most surprising about what we called "awakening" is that the literal meaning seems to be available even when a *conventional* metaphor is processed. Virtually all theories of metaphor processing—even the ones that assume indirect access to metaphorical content—take conventionality to be a key factor determining the direct access to metaphorical content (Searle, 1979; Giora, 2003), by either bypassing the literal meaning or by accessing more salient representations. The supposed consequence of this process is that the proposition that the comprehender builds ought to include the metaphorical content among its constituents. But what happens to the literal meaning at this point, when the proposition that is built incrementally has among its constituents the conventional metaphorical content? The standard answer to this question is that there is no literal meaning: it has been rejected, or it has given priority to a more salient metaphorical content. What we demonstrate here is that the literal meaning *is* available and that it remains so up to about three seconds during the processing of the sentence, possibly waning soon after.

What is not clear from our results is whether the literal meaning replaces the original interpretation—the proposition built with the conventional content—or whether it remains available, perhaps as an alternative proposition built in parallel. Yet, a third alternative is that concepts related to the multiple meanings of words remain active though with different degrees of activation reflecting their strength or preference for a given utterance context. Our study cannot dissociate between the first two alternatives—that is, whether the interpretation changes or whether an alternative (literal) interpretation is built in parallel to the metaphorical one. Many studies have suggested that propositions can be built in parallel during sentence comprehension. For instance, Christianson et al. (2001) showed evidence that participants retain propositions that are consistent with multiple interpretations of temporarily ambiguous (garden-path) sentences such as *While Bill hunted the deer ran into the woods*. Antal and de Almeida (2021) have also suggested that when indeterminate sentences such as *Lisa began the book* are presented in discourse (e.g., contexts about *Lisa* wanting to read a book), propositions consistent with both the enriched (*Lisa began reading the book*) and the original sentences compete in memory (see also Riven & de Almeida, 2021; Sachs, 1967). Along the same lines, evidence from false memory research also points to the idea that both true and false propositions are retained (e.g., Reyna et al. 2016; see de Almeida, 2018, for review). These studies are compatible with the hypothesis of two propositions, a literal one and one carrying the enriched, metaphorical content brought about by the conventional metaphor.

However, the possibility that two propositions may be built during sentence processing—one faithful to the literal meaning and another one carrying the metaphorical content—does not imply a complete shift in interpretation. As we reviewed in the introduction, for instance, Bowdle and Gentner (2005) suggested that a shift—from comparison to categorization—is obtained even after a few trials of repeated exposure to a vehicle. A similar effect was obtained by Goldstein et al. (2012) in an ERP study relying on participants first explaining the meaning of a novel metaphor, with results suggesting that the metaphor was later processed similarly to conventional ones. Interestingly, Goldstein et al. also obtained the reverse effect: after explaining a conventional metaphor, ERP results for conventional metaphors showed a pattern similar to novel ones, suggesting a shift in interpretation from conventional back to novel. While we have no direct evidence for this shift, the availability of the literal meaning suggests that the awakening effect might be consistent with the idea of a temporary shift in the interpretation—one that becomes briefly more salient than the metaphorical content itself.

It is possible, however, that the effect we obtained simply reflects the availability—possibly due to priming—of *single concepts*, those that are compatible with the denotations of the words in the sentence. This effect is consistent with several views of metaphor representation and processing, in particular, those that are explicit about some form of lexical ambiguity or polysemy associated with items commonly used metaphorically (e.g., Bowdle & Gentner, 2005; Giora, 2003). If that is the case, the awakening effect is limited to specific concepts accessed as the sentence unfolds in real time. But even under this more restricted interpretation, the awakening effect suggests that the literal meaning is not discarded or rejected as soon as the lexical item makes contact with its conceptual representation, but that it works at an intermediary stage in the composition of the pragmatic content that the sentence conveys. The model we propose takes the literal meaning of word constituents, then, to be accessed first (step 4 in Figure 4A), within the first milliseconds post-recognition. The availability of literal meaning is determined by this initial mapping between words/morphemes and their conceptual representations. The composition of the two-word metaphorical expression, then, retrieves an expression that conventionally corresponds to the two concepts (viz., *early bird* → EARLY & BIRD → 'EARLY RISER'). Notice that given the incremental nature of sentence processing—and, in particular, under the conditions of a self-paced reading task—access to a potentially metaphorical content associated with the constituents of the two-word metaphorical expression can only be obtained after the literal meaning is accessed.[8] For instance, even if a metaphor such as *early bird* is conventional, there is no evidence for *early* to have an associated metaphorical content. It is only when the two items compose that the metaphorical content can be retrieved. Interestingly, metaphor combinations where both words are interpreted metaphorically (e.g., *red flag*) were not affected by the awakening cue in the large maze, though they were in the short and medium versions of the task. We hypothesize that this type of combination is faster to retrieve the metaphorical content in comparison to combinations where only the first (e.g., *warm welcome*) or second (e.g., *early bird*) word is metaphorical. This is yet another piece of the puzzle of how sentences carrying metaphors of the form *XY* compose as they unfold in real time. We suggest that, once the meaning of the metaphorical expression has been retrieved and integrated into the ongoing proposition, the individual concepts RED and FLAG decay—and they do so faster than the items for which one member of the pair is initially taken to be literal. This decay, although faster, was only noticeable after 2500 – 3250 ms of reading the metaphor, which further illustrates the real-time path of the literal meaning of conventional expressions.

There remains, however, the possibility that the effects we are tapping into result from a spread of activation of lexical associations with one of the words of the expression, which are picked up by the cue, thus yielding a lexical priming effect—that is, with the activation of the

[8] We note that single-word recognition times are in the order of milliseconds, with some classical RSVP studies suggesting that, with about 60 ms of exposure, words can be recognized and integrated into an ongoing propositional representation of the sentence (see, e.g., Forster, 1970; and Potter, 2018, for a review).

word *bird* in the expression *early bird* facilitating the processing of the word *fly* when it appears as a distractor in the maze juncture. There are, however, a few points that preclude us from adopting a pure lexical-priming explanation of our effects. One is that for a simple lexical priming explanation to hold, we would also have to assume that the lexical token *bird*—*qua* word—remains active yielding a form of intra-lexical priming effect with the related distractor *fly* when it is first recognized. Under this explanatory framework, token words remain active *qua words* concomitantly with their meanings or senses while also yielding a metaphorical sense. According to this view, then, words, meanings, and their pragmatic senses are all active, but facilitation would be attributed to the lexical level, on the assumption that *bird* and *fly* are *lexically* related similar to classic intra-lexical semantic associations such as *salt-pepper* (see, e.g., Neely, 1991, for review). Besides an intra-lexical form of association, we are left with semantic or conceptual relations. We have assumed that our task is taken to pick up conceptual representations between, say, *bird* and *fly* on the assumption that these relations are obtained beyond their lexical/morphological representations. Only in this sense, then, what we obtained is a type of priming effect. But notice that, according to our proposal, the priming effect (represented as the dotted line between FLY and BIRD in Figure 4A) is obtained after the words are understood, that is during the encoding of the meaning of the sentence, not during the early stages of word recognition (e.g., at short SOAs with single word presentations). And, if so, they are representative of a literal meaning that lingers. There are two further reasons for not simply adopting a lexical-priming view of our results. One is that there is no account of what happens to word tokens once their meanings become part of a propositional representation of their carrier sentence. Indeed, classic (e.g., Sachs, 1967, 1974; Kintsch, 1974) as well as recent studies (e.g., Riven & de Almeida, 2021; Antal & de Almeida, 2021) have suggested that verbatim representations of words and sentences are quickly lost, giving rise to their propositional representations (or “gist”), yielding false memories of the actual linguistic input (see also Reyna & Brainerd, 2016). A second one is theoretical and bears on the status of associations in the process of language comprehension: if associations are how our mental states are constituted, our thoughts, as Fodor (2008, p.98) suggests, "would be forever getting derailed” by associations. This of course would hold for each and every word in the sentence, with each and every word activating multiple senses in the course of sentence comprehension. While the relations obtained between the metaphorical expressions and the literally related distractor cannot be resolved without further research, our proposal is that the maze cue signals the endurance of the literal representation of the metaphorical expression.

Our results are not entirely compatible with any extant model of metaphor processing. Against direct models, our data suggest that literal representations are entertained in the process of comprehending conventional metaphors and that they remain viable interpretations of these expressions. And against the traditional pragmatic model (e.g., Clark & Lucy, 1975; Searle, 1979), we suggest that this literal interpretation is *not rejected*, but rather held in working memory, together with the conventional interpretation retrieved from memory in the course of metaphor processing. Our results, thus, point to a third alternative, beyond the classical direct/indirect dichotomy: both literal and metaphorical representations are held in memory concomitantly in the early stages of metaphor processing. This view is compatible with many studies suggesting that multiple propositions about a sentence may remain available in memory, in particular when sentences carry fleeting syntactic ambiguities (e.g., Christianson et al., 2001), yield false memories (Reyna et al., 2016), or are embedded in biasing contexts (Sachs, 1967; Antal & de Almeida, 2021; Riven & de Almeida, 2021). All these studies suggest that the propositional content computed about a sentence might take into account multiple alternatives, possibly compatible with different meanings or uses of the sentence's lexical constituents and how they combine.

In summary, our results provide further evidence for the awakening effect together with the time course of activation (and decay) of the literal meanings of conventional metaphors. We demonstrate that not only does the literal meaning of a metaphorical expression is available immediately after its comprehension, but also that the literal interpretation remains available further downstream. We have previously (Pissani & de Almeida, 2021) shown that the literal meaning was available soon (1 to 3 words; 250 to 750 ms) after metaphor processing. In the present study, we extend this effect, demonstrating that the literal meaning is still available, but faded, 6 to 8 words (1500 to 2000 ms) and 10 to 13 words (2500 to 3250 ms) after the metaphorical expression has been processed. Overall, the present study suggests that conventional metaphor processing may involve entertaining both, the literal content associated with the denotational meanings of the words and the conventional content, both of which are retrieved as soon as the metaphorical expression unfolds. These remain available for interpretation—and in fact, it is one possible way in which a metaphor can be extended or shifted into literal content.

Chapter 5: Final Discussion

The present thesis examined the time course of the literal meaning during metaphor comprehension. Chapter 2 included a norming study where 309 two-word metaphorical expressions were rated for familiarity and aptness in context and in isolation. Chapter 3 employed a maze task to investigate whether the literal meaning of highly conventional metaphors could be recovered immediately after the metaphorical content had been attained. Chapter 4 reported two experiments that examined whether the literal meaning could be recovered further downstream during sentence comprehension. The central question of my thesis is whether the literal meaning is available during metaphor comprehension and, if so, whether it can be triggered after the metaphorical expression has been processed.

In the following sections, I discuss the awakening effect in light of the results obtained in Chapters 3 and 4. Then, I integrate the results presented in Appendices A and B to argue how familiarity and aptness can modulate the awakening effect. Next, I discuss these effects in the context of models of metaphor processing with emphasis on the direct-indirect debate. Lastly, I present a minimalist model of metaphor processing that accounts for the evidence herein presented.

The awakening effect

The awakening effect, as I propose it, refers to the phenomenon that the literal meaning of conventional metaphors can be recovered by a subsequent cue that is related to the literal interpretation of the metaphor. In the current thesis, I have presented evidence in favour of the awakening effect from three main experiments, which have revealed a delay in response time and a decrease in accuracy rate only in trials where the cue is related—rather than *unrelated*—to the literal meaning of the metaphor.

In Chapter 3, I examined whether the literal meaning of conventional metaphors could be recovered immediately afterward. If so, this would indicate that the literal interpretation was available during metaphor processing. To that end, I employed a maze task in which participants read sentences containing metaphors and were then presented with a word selection task. The maze juncture included the correct answer paired with either a distractor that was related to the literal meaning of the metaphor or an unrelated one. As predicted, results from Chapter 3 indicated that trials that included the related cue were slower and less accurate than those with an unrelated cue. In the maze task, participants must read each word of the sentence and combine it with the preceding context to be able to perform an accurate word selection. In this sense, the maze task forces semantic composition, wherein the meaning of each word is integrated into the ongoing proposition. At the moment of the word selection, the metaphorical expression has already been processed according to its conventional interpretation. Nonetheless, trials with the related cue were slower and less accurate. This effect can be interpreted as a conflict between the literal meaning and the conventional content of the metaphor. This suggests that the literal interpretation of the conventional metaphor was being prompted—and (re)accessed—by the related cue. The sentence processing scheme elicited during the maze task can be divided into two levels of representation and processing after lexical input is recognized: linguistic and conceptual processing. The awakening effect is examined in the latter.

Lexical input. In the maze task, sentences were presented word by word. Participants were required to read and comprehend each word before moving to the next one in order to make an appropriate choice at the maze juncture. Thus, lexical input refers to the incoming tokens of the constituent parts of the sentence. Each sentence was composed of a subject, a verb, and a determinant (*John is an*), a two-word metaphor combination (*early bird*), continuing words (*so* *he can*), the word selection in one of two conditions (*attend*/*fly* or *attend*/*cry*), and a short ending (*morning classes*). It is important to mention that in the original maze, there were 1 – 3 continuing words (*so he can*), which increased to 6 – 8 words (*so he can eat a healthy breakfast and*) in the medium maze, and to 10 – 13 words in the large maze (*so he can go to the gym and exercise and then*). The added lexical material between the metaphor and the word selection was created to be as neutral as possible rather than to bias the related distractor (see Table 8A in Appendix C for the complete set of materials). Yet, it remains the possibility that adding semantic content provides information about the event, which can cause the word selection to be easier or harder for specific items across mazes. However, in all analyses, items were included as random effects as we did not control for the change in the degree of difficulty in each item from maze to maze. This accounts for the variability coming from each unique item within each maze, but not across mazes.

Linguistic processing. This refers to the first level of representation during sentence comprehension. Once a word is recognized, its conceptual representation is accessed. Then, these concepts are incrementally composed into propositions faithful to the literal meaning of the words. Every time a concept is accessed, it is integrated into the ongoing proposition by applying syntactic-semantic operations (e.g., simple predication). For instance, upon hearing *My lawyer is a shark*, one is able to determine that *sharkness* is predicated on *lawyer*.

Conceptual processing. This refers to the second level of representation during processing. Here, the literal proposition is combined with pragmatic information to derive the intended meaning. For instance, upon building the proposition of *My lawyer is a shark* and applying one's knowledge of the world, one is able to determine that this sentence means—metaphorically—that my lawyer is an aggressive advocate. Thus, a proposition according to the metaphorical content of the full sentence is formed. At this time, the related cue triggers the associated concepts accessed during literal interpretation, which conflicts with the metaphorical interpretation, leading to slower response times and lower accuracy rates at the word selection. I argue that both propositions, literal and metaphorical, are built in parallel.

The awakening effect further downstream

In Chapter 4, I examined whether the literal meaning of conventional metaphors could be recovered further during sentence processing. If so, that would inform about the time course—at two later time points—of the literal meaning during metaphor comprehension. To that end, I adapted the materials employed in the original maze so that there would be 6 – 8 words between the metaphor and the word selection in the medium maze and 10 – 13 words in the large maze. In the medium maze, participants arrived at the word selection between 1500 ms and 2000 ms after having read the metaphorical expression, while this time was set to approximately 2500 ms and 3250 ms in the large maze. Even so, the pattern of results remained constant. This indicates that the literal meaning remains available beyond the 1 – 3 words (250 ms to 750 ms) set in the original maze, and it can be recovered despite the intervening processing of words and phrases. However, the availability of the literal meaning waned significantly. On the one hand, response time for trials with the related distractor decreased significantly from the original to the medium maze and from the original to the large maze, but remained unchanged from the medium to the large maze. On the other hand, the accuracy rate for trials with the related distractor increased significantly from the original to the large maze but remained unchanged from the original to the medium maze and from the medium to the large maze. That is, as the distance between the metaphor (*early bird*) and the maze juncture (*attend* versus *fly*) increased, participants became better—faster and more accurate—at selecting the appropriate word (*attend*). These results provide further evidence for the awakening effect during conceptual processing.

Conceptual processing further downstream. As incoming token items are being integrated into the ongoing proposition to derive the intended meaning, they trigger associated concepts. The results from the medium and large mazes suggest that these associated concepts linger for about 3250 ms but begin to fade after the initial 750 ms. After this window, the related cue elicits a weaker awakening effect. This may be so because the concepts initially accessed during literal interpretation are no longer needed after the conventional content has been attained and, thus, they start to wane promptly. I constrain my previous argument that a proposition related to the literal interpretation is available together with one built with pragmatic information by specifying that the latter takes priority over the former rapidly.

The awakening effect modulated by familiarity and aptness

In Appendices A and B, I examined whether familiarity and aptness modulated the awakening effect. To that end, I regressed response time and accuracy rate to the word selection in the original maze (Chapter 3) and in the medium and large mazes (Chapter 4) on familiarity and aptness ratings (Chapter 2). Immediately after metaphor comprehension, delayed response time for trials with the related distractor remained constant when modeled as a function of familiarity and aptness. However, the accuracy rate increased as familiarity and aptness increased in trials with the related distractor. So, although the awakening effect is strong immediately after the metaphor has been processed, this effect can be attenuated—only in terms of accuracy—if the preceding metaphor is highly familiar or highly apt. In contrast, long after the metaphor has been processed, both response time and accuracy rate are affected by familiarity and aptness. In both the medium and large mazes, response time decreased and accuracy rate increased as familiarity and aptness increased in trials with the related distractor. These overall effects were not found in trials with the unrelated distractor, which was not significantly affected by familiarity or aptness. Similarly, although the awakening effect was present long after the metaphor had been processed, this effect could be further attenuated—in terms of response time and accuracy—if the preceding metaphor was highly familiar or highly apt. As argued in Appendices A and B, this may be the case because the conventional content associated with highly familiar and highly apt metaphors is more salient due to their frequent use—which, in turn, makes the literal meaning less accessible—and, as a result, more difficult to recover by the subsequent cue.

Nevertheless, a few of the effects obtained are not clear-cut. In the related condition, these are (a) the decrease in accuracy for aptness in isolation (see Table 6B and 7D) and (b) the increase in RT for familiarity in context (see Table 7C). In the unrelated condition, these are (c) the decrease in RT for aptness in context (see Table 7C) and (d) the increase in accuracy for familiarity in isolation (see Table 7D). On the one hand, in the related condition, accuracy analyses yielded a significant effect of aptness in isolation in the original maze, whereby for every increase of one unit in aptness in isolation, the accuracy rate decreased by 0.71 %. This effect was replicated in the large maze, wherein the accuracy rate decreased by 0.15 %. Further, in the large maze, analyses of response time yielded a significant effect of familiarity in context, whereby for every increase of one unit in familiarity in context, RT increased by 240.41 ms. These effects, although minimal, are in the opposite direction than the prediction that, as aptness and familiarity increase, the accuracy rate will increase and RT will decrease proportionately. On the other hand, in the unrelated condition, RT analyses yielded a significant effect of aptness in context in the large maze, whereby for every increase of one unit in aptness in context, response time decreased by 178.73 ms. Moreover, accuracy analyses yielded a significant effect of familiarity in isolation, whereby for every increase of one unit in familiarity in isolation, the accuracy rate increased by 16.37 %. These results are against the predictions that in the unrelated condition, familiarity and aptness will not affect either response time or accuracy given that there was no relationship between the metaphor and the *unrelated* cue. Overall, most of the results are compatible with the prediction that familiarity and aptness affect the awakening effect obtained in the related condition. Thus, I speculate that the incompatible results were so because the analyses were based on increases of one unit in each factor, however, the variability within the ratings of the 24 items was minimal (i.e., familiarity in context: $M = 5.59$, $SD = 0.38$; familiarity in isolation: $M = 5.47$, $SD = 0.43$; aptness in context: $M = 5.23$, $SD = 0.43$; aptness in isolation: $M = 4.93$, $SD = 0.41$), which may not have allowed the algorithm to calculate increases of one unit with high precision.

Several authors have argued that the way in which metaphors are processed is contingent on their degree of familiarity and aptness (Blasko & Conine, 1993; Bowdle & Gentner, 2005; Chiappe et al., 2003; Keysar et al., 2000; Jones & Estes, 2006; Roncero & de Almeida, 2014). I agree with this statement with a few caveats. To wit, Blasko and Conine (1993) argued that highly familiar metaphors are processed faster and more easily than low familiar ones due to their frequent use, and even if a metaphor is less familiar, its metaphorical content can be easily attained if the metaphor is highly apt. Thus, the authors advocate for a version of the three-stage model of metaphor comprehension in which the first stage can be truncated in the case of highly familiar or highly apt metaphors. Instead, based on the analyses reported in Appendices A and B, I contend that even in highly familiar and highly apt metaphors, the first stage—the derivation of the literal meaning—holds, but their conventional content is attained rapidly and takes priority over the literal interpretation faster than in low familiar and low apt metaphors. Furthermore, the career-of-metaphor theory (Bowdle & Gentner, 2005) states that whether metaphors are processed directly (i.e., by directly retrieving the conventional content) or indirectly (i.e., by first deriving the literal interpretation) depends on their degree of familiarity. Thus, when a metaphor is first encountered, it is interpreted by comparing the topic and the vehicle, which yields a novel category. Over time, this category becomes more salient; therefore, conventional metaphors are interpreted via categorization. Rather, I argue that both novel and conventional metaphors are processed indirectly—via literal interpretation—however, in the case of conventional metaphors, the conventional content is readily available and can be retrieved faster than in the case of novel metaphors. Thus, the literal interpretation fades promptly in favour of the conventional content.

The direct-indirect access debate

The three-stage model (Clark & Lucy, 1975; Grice, 1975; Searle, 1979) and the direct access view (Gibbs, 1994; Gibbs & Colston, 2012; Glucksberg, 2003; Keysar et al., 2000) make different predictions about the path of the literal meaning during metaphor comprehension.

The three-stage model states that the listener goes through different stages upon hearing an expression such as *Amy is a bright student*. First, the listener derives the literal meaning of the sentence, that Amy radiates light. Second, the listener determines that the literal interpretation is not appropriate—if it was, the speaker would be violating the cooperative principle (i.e., the maxim of quality that poses that one should be truthful in conversation). Therefore, the listener rejects the literal interpretation, that Amy radiates light. Third, the listener uses pragmatic information to infer that the most plausible interpretation is the metaphorical one, that Amy is a smart student. In line with the three-stage model, results from Chapters 3 and 4 showed that trials with the related cue were slower and less accurate than those with the unrelated cue. This suggests that the related cue triggered the literal meaning of the conventional metaphor thus creating a conflict between them, which in turn resulted in a delay in response time and a decay in accuracy rate. Crucially, this effect was not only found immediately after the metaphorical expression had been processed, but also further away from the metaphorical expression. Thus, against the three-stage model, these results indicate that the literal meaning is not rejected as soon as the conventional content has been attained, but rather remains accessible—presumably less salient—until it fades further downstream.

In contrast, the direct access view argues that upon hearing *Amy is a bright student*, the listener derives the metaphorical meaning directly. For instance, according to the categorization model (Glucksberg, 2008), the listener understands the vehicle *bright* as a dual-reference category of, on the one hand, 'things that radiate light' and, on the other, 'things or people that are smart'. Then, after intermediary pragmatic processes, the listener includes *Amy* in the second category. This approach distinguishes conventional and novel metaphors, wherein the former are associated with preexisting categories, while the latter create said categories in real time. Contrary to this view, results from Chapters 3 and 4 showed that trials in which the cue is related to the literal meaning of the preceding metaphor are slower and less accurate than those with an unrelated cue. Thus, if the literal meaning of conventional metaphors was not accessed before the metaphorical content was attained, then trials with the related cue would have yielded comparable response times and accuracy rates to trials with the unrelated one. Thus, even in the case of highly conventional metaphors, I argue that the literal meaning is first accessed.

A minimalist proposal of metaphor comprehension

The awakening effect offers evidence in favour of a minimalist model of metaphor comprehension. This model accounts for the results herein presented and can be considered as a refinement of the three-stage model of metaphor comprehension. This minimalist model also holds that metaphor comprehension occurs in stages with the first stage being the literal interpretation, however, this interpretation is not rejected but rather, it remains less salient until it fades gradually.

First stage. As words are recognized, their linguistic information is processed and their associated concepts are accessed. This first stage is purely linguistic and insulated from pragmatic knowledge**.** Thus, a proposition faithful to the lexical denotation of the token sentence is formed. However, the methodology employed here does not inform about the nature of this proposition, whether these are full propositions or partial ones.

Second stage. After the literal interpretation is derived, it is combined with postulates about the intended meaning to derive a proposition consistent with its conventional content. Contrary to the three-stage model, I argue that both propositions linger, with the metaphorical interpretation prevailing over the less salient literal interpretation, which remains dormant until it fades over time. Before then, a related cue can awaken the literal interpretation built during the first stage.

The case of two-word metaphor combinations

One of the contributions of the present thesis is the investigation of metaphor comprehension in two-word metaphor combinations rather than employing the vastly studied copular metaphor. First, the effects obtained here provide further support to the claim that metaphors are processed via literal meaning. To date, most studies in support of the literal-first hypothesis have employed copular metaphors (Ashby et al., 2018; Brisard et al., 2001; Patalas & de Almeida, 2019; Weiland et al., 2014). The current thesis provides further support by demonstrating that this effect is also true of two-word metaphor combinations. Second, I argued that this type of metaphor is particularly interesting due to the relationship between the head and the modifier. Two-word metaphor combinations differ in their syntactic type, the position of the vehicle, and whether they are meaningful if interpreted literally. In the present series of studies, I have explored the effects of vehicle position, whether the metaphorical content was carried in the first (*warm welcome*), second (*early bird*), or both words (*red flag*) of the expression. To that effect, I included the position of the vehicle as a predictor in all response time analyses across mazes. In the original and the medium mazes, all metaphors were affected by the related cue regardless of vehicle position. However, in the large maze, metaphors in which both words were metaphorical (*red flag*) were not affected by the related cue. That is, whether the metaphorical content was in the first, second, or both words did not affect the awakening effect early on, but only after approximately 2500 ms of having processed the metaphor. After this time, expressions with both words being metaphorical were not prone to the awakening effect. I hypothesize that the metaphorical content is attained faster when both words are metaphorical, which in turn causes the literal meaning to fade earlier when compared to metaphors with one word being metaphorical and the other one being literal.

Individual differences

The ability to comprehend metaphor involves several factors including creativity (Craig & Baron-Cohen, 1999; Gold et al., 2012; Kasirer & Mashal, 2014; Kenett et al., 2014), analogical reasoning (Holyoak et al., 1984; Gentner & Toupin, 1986; Gentner, 1988; Rattermann & Gentner, 1998), and executive functions such as working memory and flexibility (Mashal & Kasirer, 2011; Russell, 1997). However, these factors were not controlled for in the experiments conducted here. The present thesis focuses on the processes that are common in the human architecture, rather than on the processes that deviate from it. I aimed at proposing a general model of metaphor comprehension that assumes a certain degree of homogeneity among individuals. Thus, including individual information regarding cognitive abilities (e.g., reading abilities, working memory capacity, attention span) or cultural and socio-economic variables (e.g., general knowledge, socio-economic status, level of education) would not serve the purpose of the present thesis. This is so because I do not have a theory on how these factors are relevant or how they may affect metaphor comprehension. Some studies have examined, for instance, whether general knowledge affects hemispheric processing during metaphor comprehension (Briner et al., 2008). Their approach, however, aimed to find what differs between individuals rather than what is common to them. The current thesis is devoted to the latter. The goal of the present series of studies is to inform about metaphor comprehension as a human linguistic and cognitive capacity, and how the properties of the expressions themselves influence the effects obtained.

Future directions

In the present thesis, I have investigated the awakening effect by employing a word selection task. In this task, the response time and the accuracy rate inform about the processing cost to select the appropriate word as a function of the type of distractor. If the distractor is related to the literal meaning of the metaphor, one assumes that the conflict between the two leads to longer response times and lower accuracy rates. Yet, it remains uncertain whether this delay in time is, for instance, due to a reinterpretation of the metaphor or to higher cognitive effort at the maze juncture to inhibit the literal meaning prompted by the related cue. To further examine these aspects, I am currently running an eye-tracking study, where the literal meaning will be prompted later in the sentence without the need to interrupt it with a word selection task. This technique also allows one to track eye movements at the precise moment in which the literal cue is presented to observe whether it yields regressions to the metaphor or, instead, longer fixations at the cue. Further, I aim to conduct an ERP study to examine whether the awakening effect can also be evaluated in terms of cognitive effort—as measured by the N400 amplitude—in addition to response time and accuracy rate.

Conclusion

I argue that, during metaphor comprehension, the literal interpretation occurs first even in the case of highly conventional metaphors. The initial input that the listener uses to calculate the communicative intentions of the speaker is the linguistic information derived during the first stage, which is insulated from pragmatic information. Because this processing stage happens early on—before the conventional content has been attained—studies that do not tap into the early stages of metaphor processing have failed to trace it. In the present thesis, I have used a novel approach to recover the literal meaning of conventional metaphors by using a subsequent cue related to their literal interpretation. The outcome induced by this cue has given rise to the metaphor awakening effect.

References

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

Allott, N., & Textor, M. (2022). Literal and metaphorical meaning: in search of a lost distinction. *Inquiry*. https://doi.org/10.1080/0020174X.2022.2128867

Antal, C., & de Almeida, R. G. (2021). Indeterminate and Enriched Propositions in Context Linger: Evidence From an Eye-Tracking False Memory Paradigm. *Frontiers in Psychology*, 4376.

Aristotle. (1997). *Poetics*. (S. H. Butcher, Trans.). Mineola, NY: Dover Publications.

Arzouan, Y., Goldstein, A., and Faust, M. (2007a). Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Res.* 60, 69–81. doi: 10.1016/j.brainres.2007.05.034

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

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

Blank, G. D. (1988). Metaphors in the lexicon. *Metaphor & Symbolic Activity*, 3(1), 21–36.

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.

Bowdle, B. F. & Gentner, D. (2005). The career of metaphor. *Psychological Review*, 112(1), 193–216.

Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze Made Easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language, 111*, 104082.

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414.

Briner, S. W., Schutzenhofer, M. C., & Virtue, S. M. (2018). Hemispheric processing in conventional metaphor comprehension: The role of general knowledge. *Neuropsychologia*, 114, 101-109.

Brisard, F., Frisson, S., & Sandra, D. (2001). Processing unfamiliar metaphors in a self-paced Reading task. *Metaphor and Symbol*, *16*(1-2), 87–108 .

Campbell, S. J., & Raney, G. E. (2015). A 25-year replication of Katz et al.'s (1988) metaphor norms. *Behavior research methods*, *48*, 330-340.

Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior research methods*, *42*(3), 651-664.

Cardillo, E. R., Watson, C., & Chatterjee, A. (2017). Stimulus needs are a moving target: 240 additional matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior research methods*, *49*, 471-483.

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

Chiappe, D. L., Kennedy, J. M., & Chiappe, P. (2003). Aptness is more important than comprehensibility in preference for metaphors and similes. *Poetics*, *31*(1), 51-68.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42(4), 368–407

Clark, H. H., & Lucy, P. (1975). Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, 14(1), 56-72.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology* Section A, 33(4), 497-505.

Craig, J., & Baron-Cohen, S. (1999). Creativity and imagination in autism and asperger syndrome. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1023/A:1022163403479

Davies, M. (2008) *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.

de Almeida, R. G. (2018). Composing meaning and thinking. In G. Preyer (Ed.), *Beyond semantics and pragmatics: New essays* (pp. 201-229). Oxford: Oxford University Press.

De Grauwe, S., Swain, A., Holcomb, P. J., Ditman, T., & Kuperberg, G. R. (2010). Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia*, 48(7), 1965-1984.

Ferreira, F. & Nye, J. (2018). The modularity of sentence processing reconsidered. In R. G. de Almeida & L. R. Gleitman, *On Concepts, modules, and language* (pp. 63-86). Oxford University Press.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard university press.

Fodor, J. A. (2001). Language, thought and compositionality. *Royal Institute of Philosophy Supplements*, *48*, 227-242.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford University Press on Demand.

Forgács, B., Bardolph, M. D., Amsel, B. D., DeLong, K. A., & Kutas, M. (2015). Metaphors are physical and abstract: ERPs to metaphorically modified nouns resemble ERPs to abstract language. *Frontiers in Human Neuroscience*, *9*, 28.

Forster, K. (1970). Visual perception of rapidly presented word sequences of varying complexities. *Perception & Psychophysics*, *8*, 215-221.

Forster, K. (2010). Using a maze task to track lexical and sentence processing. *The Mental Lexicon*, 5, 347-357.

Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, *41*(1), 163-171.

Frege, G. (1884). *Die Grundlagen der Arithmertik* translated as Foundations of Arithmetic by J. L. Austin, NY: Bantam Books.

Gagné, C. L. (2002). Metaphoric interpretations of comparison-based combinations. *Metaphor and Symbol*, *17*(3), 161-178.

Gallant, J., & Libben, G. (2020). Can the maze task be even more amazing? Adapting the maze task to advance psycholinguistic experimentation. *The Mental Lexicon,* 15(2), 366–38.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2), 155-170.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, 59, 47-59.

Gentner, D., & Bowdle, B. (2008). Metaphor as structure-mapping. *The Cambridge Handbook of Metaphor and Thought*, *109*, 128.

Gentner, D. & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*(3), 277-300.

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

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

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

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in cognitive sciences*, *7*(2), 92-96.

Glucksberg, S., Gildea, P., & B. Bookin, H. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior, 21*(1), 85-98.

Gold, R., Faust, M., and Ben-Artzi, E. (2012). Metaphors and verbal creativity: the role of the right hemisphere. *Laterality* 17, 602–614. doi: 10.1080/1357650X. 2011.599936

Goldstein, A., Arzouan, Y., & Faust, M. (2012). Killing a novel metaphor and reviving a dead one: ERP correlates of metaphor conventionalization. *Brain and Language*, *123*(2), 137-142.

Grandy, R. E. (1990). Understanding and the principle of compositionality. *Philosophical Perspectives*, *4*, 557-572.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics, vol. 3: Speech Acts* (pp. 41-58). New York: Academic Press.

Grootswagers, T. (2020). [In Press] A primer on running human behavioural experiments online. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01395-3

Harris, R. J. (1976). Comprehension of metaphors: A test of the two-stage processing model. *Bulletin of the Psychonomic Society*, *8*(4), 312-314.

Holyoak, K. J. (1984). Mental models in problem solving. In *Tutorials in Learning and Memory: Essays in Honor of Gordon Bower*.

Inhoff, A. W., Lima, S. D., & Carroll, P. J. (1984). Contextual effects on metaphor comprehension in Reading. Memory & Cognition, 12, 558–567.

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

Jones, L. L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language*, *36*, 50–67.

Kasirer, A., & Mashal, N. (2014). Verbal creativity in autism: comprehension and generation of metaphoric language in high-functioning autism spectrum disorder and typical development. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2014.00615

Katz, A., Paivio, A., Marschark, M., & Clark, J. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor and Symbolic Activity, 3,* 191–214.

Kenett, Y. N., Anaki, D., & Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2014.00407

Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of memory and language*, *28*(4), 375-385.

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.

Kintsch, W. (1974). *The representation of meaning in memory.* Lawrence Erlbaum.

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain research*, 1284, 145-155.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought*, 2nd Ed. (pp. 202–251). Cambridge University Press.

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

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

Mashal, N., & Kasirer, A. (2011). Thinking maps enhance metaphoric competence in children with autism and learning disabilities. Research in Developmental Disabilities. https://doi.org/10.1016/j.ridd.2011.08.012

McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, *6*(3), 486.

Morgan, J. L. (1979). Observations on the pragmatics of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 136-147). Cambridge University Press.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Lawrence Erlbaum Associates, Inc.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Yarkoni, T. S. C. I. E. N. T. I. F. I. C. S. T. A. N. D. A. R. D. S. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422-1425.

Olkoniemi, H., Ranta, H., & Kaakinen, J. K. (2016). Individual differences in the processing of written sarcasm and metaphor: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 433.

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, *17*(4), 465-477.Partee, B. (1995). Lexical semantics and compositionality. *An invitation to cognitive science: Language*, *1*, 311-360.

Patalas, I., & de Almeida, R. (2019). Interpreting metaphors in real-time: Cross-modal evidence for exhaustive access. In *CogSci* (pp. 2530-2536).

Peirce, J. W., & MacAskill, M. R. (2018). *Building Experiments in PsychoPy.* Sage.

Peirce, J., Hirst, R., & MacAskill, M. (2022). *Building experiments in PsychoPy*. Sage.

Pelletier, F. J. (2017). Compositionality and concepts—a perspective from formal semantics and philosophy of language. In J. A. Hampton & Y. Winter (eds.), *Compositionality and concepts in linguistics and psychology* (pp. 31-94). Cham, Switzerland: Springer.

Pissani, L., & de Almeida, R. G. (2021). Can you mend a broken heart? Awakening conventional metaphors in the maze. *Psychonomic Bulletin & Review*, 1-9.

Pissani, L., & de Almeida, R. G. (2023) Early birds can fly: Awakening the literal meaning of conventional metaphors further downstream. *Metaphor and Symbol*

Pollio, H. R., Fabrizi, M. S., Sills, A., & Smith, M. K. (1984). Need metaphoric comprehension take longer than literal comprehension? *Journal of Psycholinguistic Research*, *13*, 195-214.

Potter, M. C. (2018). The immediacy of conceptual processing. In R. G. de Almeida & L. R. Gleitman (Eds.), *On Concepts, Modules, and Language: Cognitive Science at its Core* (pp. 239-249). Oxford: Oxford University Press.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Radford, A., Wu, J., Amodei, D., Amodei, D., Clark, J., Brundage, M., & Sutskever, I. (2019a). Better language models and their implications. *OpenAI blog*, *1*, 2.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019b). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.

Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*, *13*(4), 453-478.Recanati, F. (2001). Literal/Nonliteral. *Midwest Studies in Philosophy*, *25*(1), 264–274. https://doi.org/10.1111/1475-4975.00048

Reyna, V. F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition*, *5*(1), 1-9.

Riven, L., & de Almeida, R. G. (2021). Context breeds false memories for indeterminate sentences. *Frontiers in Psychology*, *12*, 616065.

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

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

Roncero, C., de Almeida, R. G., Martin, D., & de Caro, M. (2016). Aptness predicts metaphor preference in the lab and on the Internet. *Metaphor & Symbol, 31,* 31-46

Rubio Fernández, P. (2007). Suppression in metaphor interpretation: differences between meaning selection and meaning construction. *Journal of Semantics*, 24, 345-371.

Russell, James (ed.) (1997). *Autism as an Executive Disorder*. Oxford University Press UK.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*(9), 437-442.

Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory & Cognition, 2*(1-A), 95–100. https://doi.org/10.3758/BF03197498

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

Senaldi, M. S., Titone, D. A., & Johns, B. T. (2022). Determining the importance of frequency and contextual diversity in the lexical organization of multiword expressions. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *76*(2), 87.

Shibata, M., Toyomura, A., Motoyama, H., Kawabata, Y., Abe, J. (2012). Does simile comprehension differ from metaphor comprehension? A functional MRI study. Brain and Language, 121 (3), 254-260.

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

Thibodeau, P. H., Sikos, L., & Durgin, F. H. (2018). Are subjective ratings of metaphors a red herring? The big two dimensions of metaphoric sentences. *Behavior Research Methods*, *50*, 759-772.

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.

Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP. *Frontiers in Human Neuroscience*, *8*, 583.

Wilson, D., & Sperber, D. (2012). A deflationary account of metaphors. In *Meaning and Relevance* (pp. 97–122). Cambridge University Press. https://doi.org/10.1017/CBO9781139028370.007

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge.

Wisniewski, E. J., & Murphy, G. L. (2005). Frequency of relation type as a determinant of conceptual combination: a reanalysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 169.

Witzel, J., & Forster, K. (2014). Lexical co-occurrence and ambiguity resolution. *Language, Cognition and Neuroscience*, 29(2), 158-185.

Appendix A

The effect of familiarity and aptness on the awakening effect

In the following sections, I examined whether familiarity and aptness modulated the metaphor awakening effect in the original maze. To that effect, the dependent measures—response time and accuracy rate—were regressed on familiarity and aptness ratings, both in context and in isolation. I predicted that higher familiarity and aptness ratings would speed up response time and increase the accuracy rate at the maze juncture only for trials in which the correct answer was paired with the related distractor. I reasoned that the literal meaning of highly conventional and highly apt metaphors would be less salient and, therefore, more difficult to recover. Thus, given that trials in which the correct answer was paired with the unrelated distractor were not meant to recover the literal meaning of metaphorical expressions, the higher familiarity and aptness ratings of the metaphors would not affect response time or accuracy rate at the maze juncture.

Experiment 1: The original maze

RT as a function of familiarity and aptness ratings

I obtained familiarity and aptness ratings, both in context and in isolation, from the norming study presented in Chapter 2[9]. I then performed a linear mixed-effects (LME) analysis of Response time (RT) to word selection, using the same data-analytic procedures employed in Chapters 3 and 4. As in the main analyses, the full model was a significantly better fit to the data than the null model, $\chi^2(5) = 18.08$, p = .003. The type of distractor affected the response time,

[9] The norming study in Chapter 2 did not include ratings for *sharp tongue.* Thus, to perform a fair comparison, I re-ran all analyses with 23 items (rather than 24 items) before including familiarity and aptness as factors.

increasing it by 116 ms, 95% CI [60.29, 172.17][10]. Further, to assess whether familiarity and aptness affect response time in the related condition, I examined the response times for the related and unrelated conditions separately. I included familiarity and aptness ratings, both in context and in isolation, as predictors in both models. There was no main effect of familiarity or aptness on RTs in either related or unrelated conditions (see Table 6A).

Table 6A: Model of RT as a function of familiarity and aptness

	Related			Unrelated		
Predictor	Estimates	95% CI	p-value	Estimates	95% CI	p-value
Constant	2294.28	[887.55, 3701.01]		1490.83	[188.31, 2793.36]	
FAM-C	-125.93	[-520.51, 268.65]	p = .531	-22.78	[-376.80, 331.23]	p = .899
FAM-I	117.74	[-237.58, 473.06]	p = .515	29.12	[-283.90, 342.14]	p = .855
APT-C	-134.03	[-332.20, 64.13]	p = .184	-32.60	[-213.19, 147.98]	p = .723
APT-I	-28.71	[-237.04, 179.62]	p = .787	-15.41	[-205.08, 174.26]	p = .873

Units are in milliseconds. Predictors are familiarity in context (FAM-C), familiarity in isolation (FAM-I), aptness in context (APT-C), and aptness in isolation (APT-I).

Accuracy rate as a function of familiarity and aptness ratings

For the accuracy analyses, I employed the same data-analytic procedures from Chapters 3 and 4. I performed logistic regression by modeling the probability of observing a correct word selection as a function of the type of distractor. Results, again, indicated a reliable effect of type of distractor (logit difference: +1.81, $SE = 0.29$, $z = 6.21$, $p < .001$). I then calculated the probabilities of the log odds, which predicted that the probability of observing a correct word

[10] As a reminder, when including all 24 items, the increase in RT in the related trials was 116 ms, 95% CI [61.44, 171.77], and after excluding *sharp tongue*, it remained consistent at 116 ms, 95% CI [60.29, 172.17].

selection was 82% when the correct answer (*attend*) was paired with a related distractor (*fly*), while the probability of observing a correct word selection increased to 97% when the correct answer (*attend*) was paired with an unrelated distractor (*cry*)[11]. Similar to the previous section, I also examined accuracy rates for the related and unrelated conditions separately. For the related condition, there was a main effect of familiarity in context (logit difference: -1.61, $SE = 0.7$, $z = -2.31$, $p = .021$), familiarity in isolation (logit difference: +2.07, $SE = 0.63$, $z = 3.29$, $p < .001$), and aptness in isolation (logit difference: -1.25, $SE = 0.38$, $z = -3.27$, $p = .001$), but no main effect of aptness in context. For the unrelated condition, there was only a main effect of aptness in context (logit difference: -1.59, $SE = 0.81$, $z = -1.98$, $p = .048$), but no main effect of familiarity in context, familiarity in isolation, or aptness in isolation. Table 6B shows the probability (in percentage) of making a correct response in the related and unrelated conditions when each variable is set to 0 or 1 unit.

Table 6B: Model of accuracy probabilities as a function of familiarity and aptness

Related			Unrelated		
FAM-C = 0	FAM-C = 1		FAM-C = 0	FAM-C = 1	
48.85 %	55.92 %	$p = .021$	96.54 %	96.55 %	$p = .428$
FAM-I = 0	FAM-I = 1		FAM-I = 0	FAM-I = 1	
17.27 %	26.97 %	$p < .001$	68.64 %	77.82 %	$p = .194$
APT-C = 0	APT-C = 1		APT-C = 0	APT-C = 1	
79.81 %	80.31 %	$p = .156$	99.99 %	99.99 %	$p = .048$
APT-I = 0	APT-I = 1		APT-I = 0	APT-I = 1	
99.24 %	98.53 %	$p < .001$	99.79 %	99.63 %	$p = .689$

Probabilities are expressed as percentage. Predictors are familiarity in context (FAM-C), familiarity in isolation (FAM-I), aptness in context (APT-C), and aptness in isolation (APT-I).

[11] As a reminder, for the full set of 24, the probability of making a correct response was 81% (related) and 97% (unrelated) and after excluding *sharp tongue*, it remained consistent at 82% (related) and 97% (unrelated).

Discussion

In Appendix A, I examined whether familiarity and aptness ratings influenced the awakening effect by regressing response time and accuracy rate to the word choice in the original maze (Chapter 3) on familiarity and aptness ratings (Chapter 2). There was no main effect or interaction of any form of familiarity or aptness in RT in either the related or unrelated conditions. That is, the degree of familiarity or aptness did not affect the response time to the word choice. However, accuracy was affected by both measures of familiarity and aptness in isolation in the related condition and by aptness in context in the unrelated condition. As predicted, the higher the familiarity the higher the accuracy rate in the related condition, but not in the unrelated one. That is, even though participants were slower in the related condition, they were more accurate when the preceding metaphor was highly familiar in contrast to a low familiar one. These findings are in line with most research showing that familiarity is an integral factor to metaphor processing (e.g., Blasko & Conine, 1993). These results can further constrain the model proposed for the awakening effect by differentiating between familiar and unfamiliar metaphors, where the stages may be expedited for the former. The effect of aptness is not clear from these initial results, but were discussed together with the overall results in the final discussion.

Appendix B

The effect of familiarity and aptness on the awakening effect further downstream

In the present Appendix, I replicated the reanalyses (reported in Appendix A) on the data from Chapter 4 taking into account the norms reported in Chapter 2. I examined whether familiarity and aptness modulated the metaphor awakening effect in the medium and large mazes. Again, the dependent measures, response time and accuracy rate, were regressed on familiarity and aptness ratings, in context and in isolation. As in the original maze, I predicted that the higher the familiarity and aptness ratings were, the lower the response time and accuracy rate would be at the maze juncture. I reasoned that the literal meaning of highly conventional and highly apt metaphors would be less salient and, therefore, more difficult to recover.

Experiment 1: The medium maze

RT as a function of familiarity and aptness ratings

I performed a linear mixed-effects analysis of RTs to word selection, using the same data-analytic procedures employed in Chapters 3 and 4[12]. As in the main analyses, the full model was a significantly better fit to the data than the null model, $\chi^2(5) = 17.73$, $p = .003$. The type of distractor affected the response time, increasing it by 60 ms, 95% CI [11.43, 109.54] in the related condition. Also, I examined the response times for the related and unrelated conditions separately. I included familiarity and aptness ratings, both in context and in isolation, as predictors in both models. For the related condition, there was a main effect of familiarity in isolation, whereby for each increase in familiarity by one unit, response time decreased by 255

[12] As a reminder, when including all 24 items, the increase in RT in the related trials was 51 ms, 95% CI [3.27, 98.35], and after excluding *sharp tongue*, it remained consistent at 60 ms, 95% CI [11.43, 109.54].

ms, 95% CI [11.01, 499.40]. There was also a main effect of aptness in context, whereby for each increase in aptness by one unit, response time decreased by 137 ms, 95% CI [2.31, 270.98]. For the unrelated condition, none of the predictors was significant (see Table 7A). Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality.

Table 7A: Model of RT as a function of familiarity and aptness

	Related			Unrelated		
Predictor	Estimates	95% CI	p-value	Estimates	95% CI	p-value
Constant	2219.22	[1255.99, 3182.44]		2177.98	[1200.9, 3155.1]	
FAM-C	112.01	[-155.33, 379.35]	$p = .411$	50.03	[-212.6, 312.64]	$p = .708$
FAM-I	-255.21	[-499.40, -11.01]	$p = .041$	-211.17	[-448.50, 26.15]	$p = .081$
APT-C	-136.64	[-270.98, -2.31]	$p = .046$	-111.16	[-246.90, 24.58]	$p = .108$
APT-I	112.42	[-31.38, 256.22]	$p = .125$	102.09	[-42.15, 246.33]	$p = .165$

Units are in milliseconds. Predictors are familiarity in context (FAM-C), familiarity in isolation (FAM-I), aptness in context (APT-C), and aptness in isolation (APT-I).

Accuracy rate as a function of familiarity and aptness ratings

I performed logistic regression by modeling the probability of observing a correct word selection as a function of the type of distractor. Results, again, indicated a reliable effect of type of distractor (logit difference: +1.85, $SE = 0.32$, $z = 5.77$, $p < .001$). I also calculated the probabilities of the log odds, which predicted that the probability of observing a correct word selection was 85% when the correct answer (*attend*) was paired with a related distractor (*fly*), while the probability of observing a correct word selection increased to 97% when the correct answer (*attend*) was paired with an unrelated distractor (*cry*)[13]. I also examined accuracy rates for the related and unrelated conditions separately. For the related condition, there was a main effect of familiarity in context (logit difference: -1.61, $SE = 0.7$, $z = -2.31$, $p = .021$), familiarity in isolation (logit difference: +2.07, $SE = 0.63$, $z = 3.29$, $p < .001$), aptness in context (logit difference: +0.48, $SE = 0.34$, $z = 1.42$, $p = .156$), and aptness in isolation (logit difference: -1.25, $SE = 0.38$, $z = -3.27$, $p = .001$). For the unrelated condition, however, none of these four predictors was significant. Table 7B shows the probability of making a correct response in the related and unrelated conditions when each variable is set to 0 or 1 unit.

Table 7B: Model of accuracy probabilities as a function of familiarity and aptness

Related			Unrelated		
FAM-C = 0	FAM-C= 1		FAM-C= 0	FAM-C= 1	
18.19 %	28.6 %	$p < .001$	0.97 %	4.22 %	$p = .199$
FAM-I= 0	FAM-I= 1		FAM-I= 0	FAM-I= 1	
1.25 %	3.82 %	$p < .001$	8.04 %	21.31 %	$p = .878$
APT-C = 0	APT-C= 1		APT-C= 0	APT-C= 1	
10.12 %	19.43 %	$p = .001$	67.23 %	78.19 %	$p = .370$
APT-I= 0	APT-I = 1		APT-I = 0	APT-I = 1	
64.6 %	69.82 %	$p = .05$	99.06 %	98.84 %	$p = .306$

Probabilities are expressed as percentages. Predictors are familiarity in context (FAM-C), familiarity in isolation (FAM-I), aptness in context (APT-C), and aptness in isolation (APT-I).

[13] As a reminder, for the full set of 24, the probability of making a correct response was 85% (related) and 98% (unrelated) and after excluding *sharp tongue*, it remained consistent at 85% (related) and 97% (unrelated).

Experiment 2: The large maze

RT as a function of familiarity and aptness ratings

For these reanalyses, I also performed a linear mixed-effects analysis of RTs to word selection, and used the same data-analytic procedures from Chapters 3 and 4[14]. As in the main analyses, the full model was a significantly better fit to the data than the null model, $\chi^2(5) = 26.64$, $p < .001$. The type of distractor affected the response time, increasing it by 71 ms, 95% CI [32.12, 109.77] in the related condition. Again, I examined the response times for the related and unrelated conditions separately. I included familiarity and aptness ratings, both in context and in isolation, as predictors in both models. For the related condition, there was a main effect of familiarity in context, whereby for each increase in familiarity by one unit, response time increased by 240 ms, 95% CI [30.16, 450.67]. There was also a main effect of familiarity in isolation, whereby for each increase in familiarity by one unit, response time decreased by 380 ms, 95% CI [183.83, 576.11]. There was also a main effect of aptness in context, whereby for each increase in aptness by one unit, response time decreased by 127 ms, 95% CI [16.11, 237.00]. For the unrelated condition, there was a main effect of aptness in context, whereby for each increase in aptness by one unit, response time decreased by 179 ms, 95% CI [18.20, 339.26] (see Table 7C).

[14] As a reminder, when including all 24 items, the increase in RT in the related trials was 55 ms, 95% CI [17.38, 93.22], and after excluding *sharp tongue*, it remained consistent at 71 ms, 95% CI [32.12, 109.77].

Table 7C: Model of RT as a function of familiarity and aptness

	Related			Unrelated		
Predictor	Estimates	95% CI	p-value	Estimates	95% CI	p-value
Constant	2233.69	[1468.8, 2998.6]		2943.03	[1793.5, 4092.6]	
FAM-C	240.41	[30.2, 450.67]	$p = .025$	9.42	[-301.1, 320.01]	$p = .953$
FAM-I	-379.97	[-576.1, -183.83]	$p < .001$	-167.20	[-448.3, 113.92]	$p = .243$
APT-C	-126.56	[-237, -16.11]	$p = .025$	-178.73	[-339.26, -18.20]	$p = .029$
APT-I	86.20	[-29.8, 202.22]	$p = .145$	10.97	[-160.1, 182.08]	$p = .900$

Units are in milliseconds. Predictors are familiarity in context (FAM-C), familiarity in isolation (FAM-I), aptness in context (APT-C), and aptness in isolation (APT-I).

Accuracy rate as a function of familiarity and aptness ratings

I replicated the data-analytic procedures in the previous section. Results indicated a reliable effect of the type of distractor (logit difference: +1.29, $SE = 0.30$, $z = 4.25$, $p < .001$). Then, the calculation of the log odds predicted that the probability of observing a correct word selection was 89% when the correct answer (*attend*) was paired with a related distractor (*fly*), while the probability of observing a correct word selection increased to 97% when the correct answer (*attend*) was paired with an unrelated distractor (*cry*)[15]. The analyses of accuracy rates in the related and unrelated conditions indicated that, in the former, there was a main effect of familiarity in context (logit difference: -5.41, $SE = 1.06$, $z = -5.08$, $p < .001$), familiarity in isolation (logit difference: +6.22, $SE = 1.02$, $z = 6.11$, $p < .001$), aptness in context (logit difference: +2.71, $SE = 0.48$, $z = 5.62$, $p < .001$), and aptness in isolation (logit difference: -1.65,

[15] As a reminder, for the full set of 24, the probability of making a correct response was 89% (related) and 97% (unrelated) and after excluding *sharp tongue*, it remained consistent at 89% (related) and 97% (unrelated).

SE = 0.48, *z* = -3.46, *p* < .001). For the unrelated condition, however, there was only a main effect of familiarity in isolation (logit difference: +3.22, *SE* = 1.34, *z* = 2.40, *p* = .016). Table 7D shows the probability of making a correct response in the related and unrelated conditions when each variable is set to 0 or 1 unit.

Table 7D: Model of accuracy probabilities as a function of familiarity and aptness

Related			Unrelated		
FAM-C = 0	FAM-C= 1		FAM-C= 0	FAM-C= 1	
19.83 %	31.72 %	*p* < .001	60.38 %	72.23 %	*p* = .064
FAM-I= 0	FAM-I = 1		FAM-I = 0	FAM-I = 1	
0.79 %	2.83 %	*p* < .001	14.92 %	31.29 %	*p* = .016
APT-C = 0	APT-C = 1		APT-C = 0	APT-C = 1	
0.19 %	0.95 %	*p* < .001	29.03 %	48.36 %	*p* = .128
APT-I= 0	APT-I = 1		APT-I = 0	APT-I = 1	
89.89 %	89.74 %	*p* < .001	76.56 %	83.68 %	*p* = .664

Probabilities are expressed as percentages. Predictors are familiarity in context (FAM-C), familiarity in isolation (FAM-I), aptness in context (APT-C), and aptness in isolation (APT-I).

Discussion

In Appendix B, I examined whether familiarity and aptness ratings influenced the awakening effect by regressing response time and accuracy rate to the word choice in the medium and large mazes (Chapter 4) on familiarity and aptness ratings (Chapter 2). In the medium maze, results indicated an effect of familiarity in isolation and aptness in context on response time, whereby the higher the ratings were, the faster the response times were at the maze juncture in the related condition. Conversely, none of these factors influenced response time in the unrelated condition. Further, accuracy was affected by all measures of familiarity and aptness in the related condition, while none of these had an effect in the unrelated condition. As predicted, as familiarity and aptness increased, response time decreased and accuracy increased in the related condition, but not in the unrelated one. Altogether these results suggest that the awakening effect—when prompted further downstream (6 to 8 words after the metaphorical expression)—is modulated by familiarity and aptness.

In the large maze, results indicated an effect of familiarity in isolation and aptness in context on response time, whereby the higher these ratings were, the faster the responses time at the maze juncture were in the related condition. In contrast, only aptness in context affected response times in the unrelated condition. Moreover, accuracy was affected by all measures of familiarity and aptness in the related condition, while only familiarity in context had an effect in the unrelated condition. However, aptness in isolation decreased the accuracy rate in the related condition, which is at odds with the predictions, which was discussed in the final discussion in conjunction with other conflicting results. For the most part, these results are compatible with the idea that the awakening effect—when prompted further downstream (10 to 13 words after the metaphorical expression)—is partially modulated by familiarity and aptness. As argued earlier, this may be so because highly familiar and highly apt metaphors have their conventional content more salient as a result of their frequent use—which, in turn, makes the literal meaning less accessible—and, as a consequence, more difficult to recover.

Overall, these analyses suggest that, although our original results are confirmed when examined under the norms for familiarity and aptness, these factors seem to play an important role in the comprehension of metaphorical meaning.

Appendix C

Experimental materials employed in the original (Chapter 3) and medium and large mazes (Chapter 4)

Conventional metaphors are in boldface. The appropriate word choice at the maze juncture is in italics. Sentences are grouped in triplets: (1) the first sentence belongs to the original maze, (2) the second to the medium maze, and (3) the third one to the large maze. The appropriate word choice was presented with one of the two distractors (related, unrelated) in the right column. These remained constant within triplets.

Table 8A: Experimental materials employed in all mazes

Maze	Sentence	Related	Unrelated
1	His family thought that it was a **bitter pill** to *swallow* but still necessary.		
2	His family thought that it was a **bitter pill** to take after learning the facts and *swallow* but still necessary.	sweeten	unscrew
3	His family thought that it was a **bitter pill** to take after learning the facts of the complicated situation and *swallow* but still necessary.		
1	People agree that a **black sheep** can become a *good* person.		
2	People agree that a **black sheep** can become a very creative and *good* person.	white	early

3	People agree that a **black sheep** can become someone who has a lot of creative potential while remaining a *good* person.		
1	They say that a **blind date** can *be* the best for you.		
2	They say that a **blind date** can allow you to have fun which may *be* the best for you.	see	war
3	They say that a **blind date** can allow you to have new experiences and some fun which may *be* the best for you.		
1	Moms always say that a **bright student** can *boost* a whole class.		
2	Moms always say that a **bright student** can create a positive learning environment that will *boost* a whole class.	illuminate	masquerade
3	Moms always say that a **bright student** can create a positive learning environment which is able to help *boost* a whole class.		
1	Most people agree that a **broken heart** can be *difficult* to overcome.		
2	Most people agree that a **broken heart** can be a learning experience that is *difficult* to overcome.	stitched	unfolded
3	Most people agree that a **broken heart** can be a learning experience and also a pain that is *difficult* to overcome.		
1	Melissa has a **bubbly personality** that is *pleasing* all her friends.		
2	Melissa has a **bubbly personality** that brings joy to the room and is *pleasing* all her friends.	sparkling	buttoning
3	Melissa has a **bubbly personality** that brings joy to everybody she meets or talks to and is *pleasing* all her friends.		

1	The old neighbour has a **bucket list** so he can *write* his wishes in.		
2	The old neighbor has a **bucket list** where he can go ahead and *write* his wishes in.	shovel	gallop
3	The old neighbor has a **bucket list** full of activities where he can go ahead and happily choose to *write* his wishes in.		
1	Sara has **butter fingers** and she will *serve* very important guests.		
2	Sara has **butter fingers** which makes her drop anything and she will *serve* very important guests tomorrow.	melt	tame
3	Sara has **butter fingers** which makes her drop absolutely anything she carries and touches and she will *serve* very important guests tomorrow.		
1	If a groom gets **cold feet,** he should *reflect* and reconsider.		
2	If a groom gets **cold feet**, and feels unsure about his decision, he should *reflect* and reconsider.	warm	camp
3	If a groom gets **cold feet**, and feels great uncertainty when faced with such a decision, he should *reflect* and reconsider.		
1	John is an **early bird** so he can *attend* morning classes.		
2	John is an **early bird** so he can eat a healthy breakfast and *attend* morning classes.	fly	cry
3	John is an **early bird** so he can go to the gym and exercise and then *attend* morning classes.		
1	People know that a **foggy memory** can *affect* younger and older people.	fade	ache
2	People know that a **foggy memory** can be a problem that may *affect* younger and older people.		

3	People know that a **foggy memory** can be a problem when you are writing an exam as it can *affect* younger and older people alike.		
1	Everyone agrees that a **hungry mind** can *learn* to read faster.		
2	Everyone agrees that a **hungry mind** can be quick and can also *learn* to read faster.	devour	harass
3	Everyone agrees that a **hungry mind** can be very clever, flexible, and quick and can certainly also *learn* to read faster.		
1	Professionals agree that an **ice breaker** can *entertain* shy people.		
2	Professionals agree that an **ice breaker** can start a conversation and help to *entertain* shy people.	freeze	inject
3	Professionals agree that an **ice breaker** can be a great way to start a conversation and help to *entertain* shy people.		
1	The housewife has an **iron fist** that is really *intimidating* to everybody.		
2	The housewife has an **iron fist** that nobody wants to face and is really *intimidating* to everybody.	heavy	aware
3	The housewife has an **iron fist** that not a single person wants to face and that is really *intimidating* to everybody.		
1	People believe that a **melting pot** can *create* a balanced society.		
2	People believe that a **melting pot** can be a good thing that can *create* a balanced society.	heat	ship
3	People believe that a **melting pot** can be a good thing that can bring people together to *create* a balanced society.		
1	Mike is a **night owl** and he *hates* you if you don't go to bars.		

2	Mike is a **night owl** who would dance all night long and he *hates* you if you don't go to bars.	hunts	coins
3	Mike is a **night owl** who refuses to go to bed before the sun rises and he *hates* you if you don't go to bars.		
1	Most guys think that dating an **old flame** can *wreck* your life.		
2	Most guys think that dating an **old flame** can be bad for you and *wreck* your life.	burn	swim
3	Most guys think that dating an **old flame** can be bad for you and even really do damage and *wreck* your life.		
1	Some people advise that when a **red flag** occurs, one should *fix* it.		
2	Some people advise that when a **red flag** occurs, one should attempt to try to *fix* it.	wave	milk
3	Some people advise that when a **red flag** occurs, one should examine the situation closely in order to help *fix* it.		
1	Some people say that a **sharp tongue** can *ruin* a friendship.		
2	Some people say that a **sharp tongue** can be negative and it could even *ruin* a friendship if you are not careful.	cut	age
3	Some people say that a **sharp tongue** can be negative for you and your loved ones and it could even *ruin* a friendship if you are not careful.		
1	Erik's father has a **short fuse** that he should *manage* soon.	enlarge	testify
2	Erik's father has a **short fuse** that surprises others and that he should *manage* soon.		

3	Erik's father has a **short fuse** that can sometimes surprise the family dog and that he should *manage* soon.		
1	Everyone thinks that a **silver lining** should be *encouraging* for victims.		
2	Everyone thinks that a **silver lining** should be both promising as well as *encouraging* for everyone.	gold	pale
3	Everyone thinks that a **silver lining** should be something that will be both promising as well as *encouraging* for everyone.		
1	Mr. Harrison has a **smoking gun** that he will *show* to the attorney.		
2	Mr. Harrison has a **smoking gun** that he will most certainly go on to *show* to the attorney.	shoot	paint
3	Mr. Harrison has a **smoking gun** that he will most certainly take advantage of and go on to *show* to the attorney.		
1	Daniel got himself a **trophy wife** so he could *brag* about her.		
2	Daniel got himself a **trophy wife** so he could show off and *brag* about her.	display	stomach
3	Daniel got himself a **trophy wife** so he could introduce her to all of his friends and *brag* about her.		
1	Jane said that the **warm welcome** better be *fun* if it is for her boyfriend.		
2	Jane said that the **warm welcome** better be a large event that is extra *fun* if it is for her boyfriend.	hot	due
3	Jane said that the **warm welcome** better be a large event that is extra special and also tons of *fun* if it is for her boyfriend.		

Appendix D
Instructions for familiarity

This experiment consists of two blocks. In one block, you will be presented with sentences containing expressions in UPPERCASE; whereas in another block, you will be presented with expressions in UPPERCASE without any context; the order of the blocks will be randomly assigned. Your task is to rate how familiar you find each expression in uppercase. That is, you should report the extent to which you have heard or read that expression in the past. You are not supposed to rate the sentence—which is provided just for context—but the expression in uppercase only.

After you read the full sentence (or expression, depending on the block), a rating scale will appear on the screen. You should then record your response by clicking on a number between 1 (not familiar at all) and 7 (very familiar).

For example:
After moving abroad, Mary has been FEELING BLUE for a while.

If FEELING BLUE is a well-known expression to you, you might give it a high rating (perhaps 6 or 7). However, CRYING WOLF might not be as well-known to you, so you may want to give a lower rating (perhaps 2 or 3).

Try to use the full scale, with not so well-known expressions rated more towards the middle (2, 3, 4, 5, 6), reserving 1 for truly not familiar expressions and 7 for very familiar ones.

Appendix E

Instructions for aptness

This experiment consists of three blocks. In each block, you will be presented with metaphorical expressions in UPPERCASE (which may or may not be embedded in a sentence). Your task is to rate how apt you find each expression in UPPERCASE. You are not supposed to rate the sentence—which is provided just for context—but the expression in uppercase only.

After you read the full sentence (or expression, depending on the block), a rating scale will appear on the screen. You should then record your response by clicking on a number between 1 (not apt at all) and 7 (very apt).

Try to use the full scale, with not so apt expressions rated more towards the middle (2, 3, 4, 5, 6), reserving 1 for truly not apt expressions and 7 for very apt ones.

We will explain what aptness means next. Please, press the SPACE BAR to continue.

How to rate for aptness in the FIRST BLOCK[16]:

The expressions you will see are composed of two words, such as "SILKY HAIR". These expressions can be considered as containing two parts: the topic and the vehicle, where the topic is the subject of the metaphor and the vehicle is the word that modifies or describes this topic.

[16] Some of the experimental trials that participants saw during the first block were *piercing facts, pink noise, plum job, poker face, pyramid scheme, and rocky end.*

In the first block, the vehicle is the first word. For instance, in “SILKY HAIR”, “SILKY” is the vehicle, and it describes “HAIR”, the topic. Note that vehicles can also be used to describe other topics, such as “SILKY SUNSET” or “SILKY LAKE”.

In summary,

Expression: SILKY HAIR

Topic: HAIR

Vehicle: SILKY

Your task is to rate each expression for the extent to which the vehicle captures important features of the topic on a scale from 1 to 7, with 1 meaning it captures no features, and 7 meaning that it captures many features.

For example, hair can be shiny and smooth. Consequently, “SILKY HAIR” can be considered an apt expression, because “SILKY” captures important features of the “HAIR” (namely, shine and smoothness). A less apt statement would be “SILKY SUNSET” since it may be less common for a sunset to be both shiny and smooth. Hence “SILKY HAIR” would receive a higher rating (perhaps 6 or 7), whereas “SILKY SUNSET” would receive a lower rating (perhaps 3 or 4).

How to rate for aptness in the SECOND BLOCK[17]:

In the next block, the vehicle is the second word. For instance, in "BRIDAL SHOWER", "SHOWER" is the vehicle, and it is modified by "BRIDAL". Remember that vehicles can also be used with other topics, such as "WEDDING SHOWER" or "DIAPER SHOWER".

In summary,

Expression: BRIDAL SHOWER

Topic: BRIDAL (PARTY)

Vehicle: SHOWER

Your task is to rate each expression for the extent to which the vehicle captures important features of the topic on a scale from 1 to 7. 1 meaning it captures no features, and 7 meaning it captures many features.

For example, at a bridal party, the bride is 'showered' with gifts. Consequently, "BRIDAL SHOWER" can be considered an apt expression because "shower" captures important features of a bridal party (namely, receiving many gifts resembling a rain shower). A less apt statement would be "ANNIVERSARY SHOWER" since it may be less common to celebrate an anniversary by receiving numerous gifts. Hence "BRIDAL SHOWER" would receive a higher rating (perhaps 6 or 7), whereas "ANNIVERSARY SHOWER" would receive a lower rating (perhaps 3 or 4).

[17] Some of the experimental trials that participants saw during the second block were, *moral compass, movie star, night owl, office monkey, family tree,* and *old flame.*

How to rate for aptness in the THIRD BLOCK[18]:

In the next block, the vehicle comprises both words. For instance, in “ROUGH DIAMOND”, both words together refer to someone with potential, but who lacks certain skills, such as education or social skills.

In summary,

Expression: ROUGH DIAMOND

Implicit topic: someone with potential

Vehicle: ROUGH DIAMOND

Your task is to rate each expression for the extent to which the vehicle captures important features of the implicit topic on a scale from 1 to 7. 1 meaning it captures no features, and 7 meaning it captures many features.

For example, a diamond in the rough is the unpolished state of the diamond. Consequently, “ROUGH DIAMOND” can be considered an apt expression, because it captures important features of someone with potential but not fully developed (namely, being ‘unpolished’). A less apt statement would be “CHIN MUSIC”, which refers to idle chatter, because there are no apparent features being transferred. Hence “ROUGH DIAMOND” would receive a higher rating (perhaps 6 or 7), whereas “CHIN MUSIC” would receive a lower rating (perhaps 1 or 2).

[18] Some of the experimental trials that participants saw during the third block were *bat wings, smoking gun, silver lining, sock puppet, stalking horse,* and *wild west.*

Appendix F

Comprehensive list of materials employed in the norming study (Chapter 2)

Metaphorical expressions are in alphabetical order. Types are noun-noun (NN), adjective-noun (AN), and adjective-adjective (AA). COCA reflects the frequency scores for the co-occurrence of both constituents of the expression obtained from the Corpus of Contemporary American English. Lastly, the average ratings (from a 1-to-7 Likert scale) are provided for familiarity in context (FAM-C), familiarity in isolation (FAM-I), aptness in context (APT-C), and aptness in isolation (APT-I).

Table 11A: List of 309 normed metaphor combinations

Expression	Type	Carrier Sentence	COCA	FAM-C	FAM-I	APT-C	APT-I
ACID TEST	AN	Watkins' patients were put to the ACID TEST.	215	3.80	4.80	4.36	4.12
ALMOND EYES	NN	The model was extremely sought after for her beautiful, dark ALMOND EYES.	88	4.76	4.32	5.28	3.96
ANGRY SEA	AN	They screamed as if tossed by an ANGRY SEA.	23	4.40	4.28	5.12	4.68
ARM CANDY	NN	She was his ARM CANDY at so many cocktail parties.	68	5.12	4.92	3.52	4.52
BABY BLUES	NN	Mood swings, feeling sad, and being unable to concentrate are all signs of having the BABY BLUES.	159	4.60	4.80	4.04	4.40

BABY FEVER	NN	When one of her coworkers brought her baby to work, Breanna got a case of the BABY FEVER.	45	5.00	4.88	4.60	4.68
BABY SHOWER	NN	Most mothers have a lot of fun at a BABY SHOWER and will have at least one in their lifetime.	533	5.64	6.00	5.88	6.40
BABY STEPS	NN	Learning a new language requires BABY STEPS at first.	753	5.72	5.36	5.76	5.24
BAD APPLE	AN	The poor performance of one BAD APPLE can have serious consequences.	196	5.64	5.76	5.36	4.00
BAD BLOOD	AN	There was BAD BLOOD between the Eagles and the Cowboys.	618	5.12	5.48	5.60	4.28
BAD EGG	AN	He's a bit of a handful, a real BAD EGG.	63	5.28	5.48	4.76	4.12
BAD SEED	AN	That guy is a BAD SEED because he makes life very difficult for all of us.	165	5.52	5.16	4.72	3.64
BALLPARK FIGURE	AN	The BALLPARK FIGURE for fatalities is at least 20.	78	4.72	4.64	4.72	4.44
BARE BONES	AN	Alexander got the details of the book down to its BARE BONES.	293	4.96	4.64	4.56	5.20
BAT WINGS	NN	Mary's trainer suggested lifting weights to get rid of her BAT WINGS.	53	4.04	4.52	4.68	3.96
BEAR HUG	NN	The partners celebrated with a BEAR HUG after winning the contest.	384	5.80	5.20	5.20	5.08
BEER GOGGLES	NN	Emily later realized it was the BEER GOGGLES that made the man seem attractive at the bar.	35	5.00	4.92	4.36	4.16
BIG BROTHER	AN	Everyone is careful in the office because BIG BROTHER is always watching.	3766	5.80	5.40	4.88	4.60
BIG CHEESE	AN	He was a BIG CHEESE in the business world.	146	4.80	5.16	4.40	4.40

BIG FISH	AN	The prosecutor made a deal with the lawyer to catch the BIG FISH.	1277	5.44	5.72	4.68	5.04
BITTER END	AN	They stayed there and fought until the BITTER END.	465	5.60	5.24	5.72	4.40
BITTER PILL	AN	The disappointment had been a BITTER PILL for her parents.	234	5.12	4.92	5.24	5.24
BITTER TRUTH	AN	After the fight, she learned the BITTER TRUTH.	82	5.48	5.88	5.36	5.04
BLACK BOX	AN	Psychology research attempts to understand how the mind's mysterious BLACK BOX works.	1323	4.80	5.04	4.28	4.48
BLACK HUMOUR	AN	The BLACK HUMOUR wore out the audiences over time.	18	4.16	4.68	5.16	4.52
BLACK MARKET	AN	Those stolen items are easy to sell on the BLACK MARKET.	2082	5.56	5.96	5.04	4.72
BLACK SHEEP	AN	She soon became the BLACK SHEEP of the family.	535	5.64	5.76	5.40	5.76
BLACK SWAN	AN	Among natural disasters, 1992’s Hurricane Andrew was a BLACK SWAN.	695	3.68	4.76	4.00	4.76
BLACK WIDOW	AN	They suggest Andrea might be a BLACK WIDOW, not a grieving one.	475	4.72	5.72	4.60	4.56
BLANK SLEEP	AN	After drinking too much, the boy passed into a BLANK SLEEP.	2	3.64	3.04	3.64	3.00
BLIND DATE	AN	A friend in Miami set her up on a BLIND DATE with Jason.	725	5.84	5.52	5.32	4.52
BLIND EYE	AN	They can point out the other team's flaws but have a BLIND EYE to their own.	1370	5.56	5.52	5.56	4.48
BLUE BLOOD	AN	Future club members don't need to have BLUE BLOOD to be accepted.	157	4.92	3.80	4.32	4.24
BLUE MOON	AN	Karen only drinks every BLUE MOON, so this must be a special occasion.	671	5.32	4.96	5.00	4.48

BOILING MAD	AA	The professor gets BOILING MAD when students didn't read the syllabus.	1	4.80	4.36	4.76	4.96
BOTTOMLESS PIT	AN	He was a BOTTOMLESS PIT of ideas which made him a great writer.	284	5.32	5.16	5.28	4.72
BRAIDED FREEWAY	AN	Jim had such a long drive on the BRAIDED FREEWAY yesterday.	0	3.28	3.20	3.56	3.72
BRAIN FOG	NN	Chronic fatigue syndrome causes tiredness, poor sleep, BRAIN FOG, and/or muscle pain.	67	4.68	4.92	4.84	4.84
BRIGHT SIDE	AN	Since the general public abhors violence, the BRIGHT SIDE is that they will be able to see the truth.	1658	5.96	5.76	5.56	4.84
BRIGHT STUDENT	AN	She was a BRIGHT STUDENT who learned quickly.	43	5.92	5.44	5.88	5.28
BRILLIANT IDEA	AN	Jessie had the BRILLIANT IDEA to try sledding down the mountain in summer.	625	5.88	5.72	5.88	5.52
BROKEN FACE	AN	She couldn't get a date because she was a BROKEN FACE.	23	3.36	4.16	3.60	4.60
BROKEN HEART	AN	She was left with a BROKEN HEART after the split with her partner.	996	5.92	5.96	5.88	5.20
BUBBLY PERSONALITY	AN	Abby's BUBBLY PERSONALITY blended well at the party.	42	5.28	5.00	5.32	4.88
BUCKET LIST	NN	Here are 25 surprising new destinations to put on your BUCKET LIST.	557	5.92	5.52	5.04	4.20
BULLET POINT	AN	The outline was organized in BULLET POINTS.	118	5.84	5.64	5.60	4.64
BULLETPROOF ARGUMENT	AN	Ms. Allens used a BULLETPROOF ARGUMENT to make her case.	1	5.28	4.64	5.28	4.64
BUMPY LIFE	AN	Anna had overcome many difficulties in her BUMPY LIFE.	1	4.96	4.36	4.84	4.28
BURNING AMBITION	AN	He had a BURNING AMBITION to climb to the upper reaches of management.	19	5.16	4.96	4.72	4.92

BURNING MOMENT	AN	That interview was a BURNING MOMENT in his life.	1	4.28	4.56	4.36	4.32
BURNING QUESTION	AN	The woman had a BURNING QUESTION she couldn't wait to ask.	232	5.40	5.20	4.40	4.28
BUSY BEE	AN	My supervisor is going to be a BUSY BEE for a couple of months.	110	5.28	5.24	5.36	5.52
BUTTER FINGERS	NN	Steve is constantly dropping things because of his BUTTER FINGERS.	9	5.28	5.16	4.40	4.72
CABIN FEVER	NN	The kids will develop CABIN FEVER after too many days in isolation.	239	5.92	5.08	4.92	4.56
CAKE FACE	NN	To avoid having a CAKE FACE, it is advisable to apply fewer layers of foundation.	4	4.32	3.44	4.36	3.08
CASH COW	NN	He bought a bankrupt shop and turned it into a CASH COW.	497	4.88	5.48	4.00	4.92
CEILING EFFECT	NN	The researchers agreed that a CEILING EFFECT was the best explanation available.	98	4.00	3.76	4.36	3.80
CHAIN REACTION	NN	The mayor’s ideas caused a CHAIN REACTION with grave political consequences.	1185	5.72	5.80	5.68	5.12
CHEESE WRITING	NN	The first draft always turns out to be CHEESE WRITING.	0	3.60	2.64	2.88	3.00
CHICKEN FEED	NN	The store owner pays his employees CHICKEN FEED.	164	4.88	4.88	4.80	4.48
CHICKEN LEGS	NN	The runway model is so skinny that she has CHICKEN LEGS.	150	5.40	5.16	5.64	4.52
CHOPPING BLOCK	AN	The associate is just the next one to line himself up for the CHOPPING BLOCK.	486	5.00	5.36	5.04	5.24
CLEAN SLATE	AN	The previous negotiations did not work, so they will start with a CLEAN SLATE.	614	5.48	5.68	5.08	5.20

CLEAR INTENTION	AN	Most candidates have a CLEAR INTENTION of what they want to accomplish.	64	5.72	5.40	6.08	5.28
CLOSED MIND	AN	A lot of people had a CLOSED MIND about the case until the evidence was revealed.	74	5.44	5.60	5.80	5.16
CLOUD NINE	NN	Martin was on CLOUD NINE after hearing the news.	180	5.72	4.88	4.48	4.44
COLD FEET	AN	Bonnie gets COLD FEET when contemplating her future with Adam.	687	5.60	6.00	5.00	4.80
COLD FISH	AN	Scott Berg knew that Wilson was no COLD FISH.	112	4.40	4.56	4.32	4.40
COLD IRON	AN	Her boyfriend's heart is a COLD IRON.	44	3.76	3.56	4.08	3.72
COLD TURKEY	AN	He quit smoking COLD TURKEY after his doctor told him to.	538	5.60	5.40	4.56	4.56
COOL CAT	AN	Their guitar player is one COOL CAT.	66	5.72	5.72	5.32	4.80
COOL HEAD	AN	It is important to keep a COOL HEAD and not show anger when dealing with people in debate class.	142	5.84	5.24	5.32	4.40
COUCH POTATO	NN	The cold weather makes it easier for people to become a COUCH POTATO.	280	5.40	5.40	4.40	4.40
CROCODILE TEARS	NN	Last night, Jasmine tried to convince the jury with CROCODILE TEARS.	186	5.32	5.00	4.72	4.72
CROOKED MATH	AN	He used CROOKED MATH to become a billionaire.	0	3.64	3.60	3.96	3.68
CROSS FIRE	NN	She was a simple woman caught in the CROSS FIRE of other people's drama.	148	5.76	5.56	5.24	4.68
CROWD SURFING	NN	There is always a little bit of CROWD SURFING during pop concerts.	14	5.44	5.52	4.64	4.60
CRYSTAL CLEAR	AA	The new president made it CRYSTAL CLEAR that all debts will be honored.	1322	6.04	5.80	5.84	5.44
CURTAIN BANGS	NN	CURTAIN BANGS used to be a super-popular '70s hairstyle.	0	3.68	2.76	4.64	3.48

DARK LIFE	AN	It was the only good moment in his otherwise DARK LIFE.	19	4.80	4.28	5.32	4.56
DARK PERSONALITY	AN	People describe him as having a DARK PERSONALITY.	11	4.80	4.68	5.44	4.72
DARK THOUGHTS	AN	A few people saw Paul struggling with DARK THOUGHTS.	157	5.52	5.44	5.52	5.12
DEAD ACT	AN	Most of the audience left halfway through the movie because it was such a DEAD ACT.	0	4.16	3.92	4.48	3.68
DEAD END	AN	I found myself embarrassed when I hit a DEAD END in my life.	2020	5.48	5.72	5.68	4.80
DEAD FACE	AN	He had a DEAD FACE look after being asked the same question for the 5th time.	57	4.56	3.64	4.40	3.52
DEAD MATCH	AN	Putting those two together to work on that case was a DEAD MATCH.	8	3.76	3.72	3.64	3.76
DEAD METAPHOR	AN	The word "pedigree" is a DEAD METAPHOR that no one remembers.	6	4.08	3.28	4.56	3.48
DEAD SILENCE	AN	There was DEAD SILENCE in the waiting room.	235	5.76	5.56	5.64	5.28
DEATH BURGER	NN	The couple bought a DEATH BURGER from their usual takeaway place.	1	3.04	2.76	3.16	3.16
DEEP LOVE	AN	Everyone saw the DEEP LOVE between them.	291	5.60	5.32	5.96	5.68
DEEP WATER	AN	Tom is in some DEEP WATER with the boss after all the mistakes he made this week.	1063	5.32	5.40	5.32	4.92
DELICATE FLOWER	AN	She’s dating a DELICATE FLOWER who likes to talk about his feelings.	74	5.68	4.96	5.44	5.68
DIRTY DANCE	AN	That was quite a DIRTY DANCE at the night club.	14	5.24	5.08	5.32	4.48
DIRTY JOB	AN	It's a DIRTY JOB, but somebody has to do it.	130	5.60	5.04	6.04	5.16

DIRTY LAUNDRY	AN	I didn't want to risk airing our family's DIRTY LAUNDRY for millions to see.	665	5.52	5.32	5.52	4.64
DIRTY MONEY	AN	The wife believed his husband was giving her DIRTY MONEY for her to hide it.	219	5.00	5.60	5.48	4.84
DIRTY MOUTH	AN	That kid has a very DIRTY MOUTH and should be punished.	51	5.44	5.56	5.36	5.40
DRAGON TEACHER	NN	The students are scared of their DRAGON TEACHER.	0	3.36	2.48	3.92	3.48
DRUG MULE	NN	The girl agreed to be a DRUG MULE to help her family financially.	57	4.72	5.16	4.80	4.28
DRY BUZZ	AN	The DRY BUZZ of the computers annoyed the secretary.	0	3.44	3.08	3.92	3.64
DRY LAUGH	AN	He let out a DRY LAUGH without a shred of cheerfulness.	29	4.40	3.80	4.56	4.52
EAGLE EYES	NN	He was considered a sharp shooter in the army, but his EAGLE EYES have weakened with time.	41	5.52	5.52	5.84	5.48
EARLY BIRD	AN	The EARLY BIRD always gets the best seat at the movie theatre.	405	5.44	5.80	5.36	5.40
ELEPHANT EARS	NN	ELEPHANT EARS are a great choice to make any garden look like a tropical forest.	49	4.48	4.32	4.72	4.60
ELEPHANT MOM	NN	Amy feels protected by her ELEPHANT MOM who is always there to protect her daughter.	0	3.20	3.52	3.40	3.24
EMOTIONAL ROLLERCOASTER	AN	Ever since Itamar received the devastating news, his life has been an EMOTIONAL ROLLERCOASTER.	48	5.36	5.56	5.32	5.44
EMPTY NEST	AN	Once the kids moved out, John was enjoying the EMPTY NEST with his wife	208	5.44	5.08	5.52	4.60

EYE BAGS	NN	Greensations makes a variety of anti-aging products that target EYE BAGS and dark circles.	12	5.08	5.16	4.72	4.32
EYE CANDY	NN	Derek married the last EYE CANDY in town and took her off the market.	444	5.16	5.60	4.36	4.44
FAMILY TREE	NN	The orphan was tracing her FAMILY TREE to see where she came from.	1346	6.12	6.08	5.40	5.20
FAN BRUSH	NN	A FAN BRUSH is ideal to apply powder products in small regions of the face.	20	3.64	3.12	4.52	3.48
FIREFLY EYES	NN	The guy was looking at his girlfriend with FIREFLY EYES.	0	3.48	3.52	3.52	3.40
FIRM DEADLINE	AN	The department imposed a FIRM DEADLINE for the submission of forms.	21	5.16	5.00	5.56	5.36
FIRM HAND	AN	Supervisors need to have a FIRM HAND with their volunteers.	210	5.00	5.44	4.88	5.44
FISH FINGERS	NN	Children love to eat FISH FINGERS for lunch.	31	4.92	4.16	4.16	3.64
FLAKY IDEA	AN	Since he always has FLAKY IDEAS, it's best not to listen to him.	1	4.48	4.28	4.40	4.48
FLASH SALE	NN	The FLASH SALE will include designers like Vera Wang, Alberta Ferretti, and Steuben.	25	5.12	4.92	5.24	4.96
FLAT ART	AN	She wasn't hired at the museum because of her FLAT ART.	2	3.04	3.44	3.68	3.92
FLOATING SHELF	AN	A FLOATING SHELF is ideal for items that you need easy access to.	3	4.12	4.40	4.80	4.32
FLOOR EFFECT	NN	These procedures were used to reduce the chance of observing a FLOOR EFFECT in the economy.	20	3.20	3.72	4.36	3.88

FOGGY MEMORY	AN	Due to his FOGGY MEMORY, he doesn't remember seeing the doctor again.	14	5.32	5.16	5.76	5.20
FOOT SOLDIERS	NN	Programmers are the FOOT SOLDIERS of the computer revolution.	622	5.64	4.92	4.24	4.40
FORBIDDEN FRUIT	AN	Mary was a FORBIDDEN FRUIT for John because she was married.	246	5.84	5.40	5.20	4.80
FRAGRANT SHADOW	AN	He was surrounded by the FRAGRANT SHADOW of the church and felt at peace.	0	3.56	3.04	3.60	3.52
FRESH COURAGE	AN	He transmitted FRESH COURAGE and hope to the oppressed.	4	3.68	3.92	4.60	3.96
GENTLE ART	AN	He completed his work with GENTLE ART, always giving it the attention it needs.	0	3.96	3.44	4.80	3.88
GHOST TOWN	NN	During the pandemic, Montreal became a GHOST TOWN.	887	5.88	5.68	5.24	4.92
GLASS CEILING	NN	In any company, a GLASS CEILING is broken when a woman can advance to a higher position.	658	4.76	5.32	4.48	4.48
GLASS RIVER	NN	The family saw the beautiful GLASS RIVER on their road trip.	0	3.40	3.00	4.24	3.68
GOLD DIGGER	NN	I thought she was a GOLD DIGGER and had other motives to marry him.	285	5.80	5.96	5.12	5.44
GOLD NUGGET	NN	The motivational speaker gave his audience another GOLD NUGGET in his recently uploaded video.	67	5.24	4.36	4.92	4.68
GOLD STANDARD	NN	This is a must have album, the one that set the new GOLD STANDARD for Helloween albums.	2337	5.08	5.64	5.48	5.04
GOLDEN EGG	AN	The American taxpayer has become a reliable GOLDEN EGG for greedy politicians.	123	5.12	4.88	4.72	4.52

GOLDEN GOOSE	AN	College athletics, especially football, have always been a GOLDEN GOOSE for universities.	211	5.00	5.20	4.52	4.92
GOLDEN HEART	AN	The boy has a GOLDEN HEART, he's always helping his family.	50	4.96	4.76	4.76	4.36
GOLDEN HOUR	AN	The patient suffered major blood loss, fortunately they got him to the hospital within the GOLDEN HOUR.	73	4.88	5.08	4.16	4.72
GOLDEN RIVER	AN	Her long hair was a flowing GOLDEN RIVER.	18	4.44	3.40	4.44	3.56
GREEN LIGHT	AN	Quickly after his recovery, Jim was given the GREEN LIGHT to walk a mile or two.	2311	5.92	5.64	5.32	5.20
GREEN THUMB	AN	I lack the GREEN THUMB required to nurture a garden.	195	5.44	5.28	4.60	5.04
GREY FACE	AN	Her GREY FACE suggested that something bad had happened.	6	3.92	3.60	4.48	3.96
HAIR DONUT	NN	A HAIR DONUT can give the appearance of a great hair bun for people with thin hair.	0	3.52	3.12	3.68	3.44
HAPPY DAWN	AN	The new year sees a HAPPY DAWN and a fresh start.	2	4.04	3.16	4.32	4.00
HAPPY HOUR	AN	Undergrads go to bars only during HAPPY HOUR to save money.	1221	6.04	5.92	5.36	4.84
HELICOPTER PARENT	NN	HELICOPTER PARENTS never stop obsessing over their kids.	42	4.92	4.60	4.56	4.72
HEN PARTY	NN	Her bridesmaids had already thrown her a HEN PARTY.	19	3.72	3.56	3.68	3.76
HIDDEN GEM	AN	The travel blogger needed to find a new HIDDEN GEM to write about.	113	5.68	5.60	5.08	4.96
HOLY GRAIL	AN	Finding an HIV vaccine is the HOLY GRAIL of medical research.	1492	6.00	5.56	4.72	4.96

HOME BIRD	NN	As a HOME BIRD, Paul simply does not leave his house.	1	3.72	3.84	3.56	3.92
HOT HEAD	AN	The notorious Hollywood HOT HEAD Eric Bolling showed everyone his rage again.	68	5.56	5.68	5.24	5.00
HOT MESS	AN	Her life turned into a HOT MESS after the child's birth.	242	5.64	5.44	4.36	4.68
HOT POTATO	AN	Abortion is a political HOT POTATO.	364	5.04	5.64	4.48	5.08
HOT SEAT	AN	After being involved in the settlement talks, he was in the HOT SEAT with management.	695	5.36	5.44	5.04	5.44
HOT SPOT	AN	Influencers can get into every HOT SPOT in town.	1191	5.68	5.80	5.28	5.20
HOT WATER	AN	He landed himself in HOT WATER after his mom realized he lied to her.	5124	6.00	5.72	4.96	5.24
HOURGLASS FIGURE	NN	Wrap tops and dresses accentuate the waist and enhance an HOURGLASS FIGURE.	76	5.44	4.96	5.76	4.84
HUNGRY MIND	AN	Undergraduates with a HUNGRY MIND make better graduate students.	14	4.40	4.16	5.28	4.56
ICE BREAKER	NN	The teacher started with a quick ICE BREAKER on the first day of school.	88	5.96	5.96	4.88	5.24
ICY CLEAN	AA	The surgery room was left ICY CLEAN and ready for the next procedure.	0	3.80	3.60	3.68	3.28
ICY LOOK	AN	The angry lady gave him an ICY LOOK.	11	5.16	4.40	4.88	4.28
IMPATIENT MACHINERY	AN	The IMPATIENT MACHINERY kept on running.	0	3.40	3.12	3.60	3.36
INNOCENT LAMB	AN	Jones might be an INNOCENT LAMB but investors are nervy.	8	3.80	4.60	4.64	5.20
IRON FIST	NN	Big law firms are managed with an IRON FIST.	537	5.16	5.36	5.04	4.72

JUICY GOSSIP	AN	He brought home the JUICY GOSSIP about Drucilla Hawkins.	48	5.24	5.24	5.32	5.00
KILLER PRESENTATION	AN	A KILLER PRESENTATION always captures people's attention.	0	5.20	4.40	4.60	4.12
KITTEN HEEL	NN	A KITTEN HEEL is better for people with smaller calves as it is easier on the feet.	9	3.40	4.04	3.24	3.56
LAME DUCK	AN	That politician is a LAME DUCK but he's still so powerful.	820	4.76	4.36	4.56	4.28
LAST STRAW	AN	The departure of Amy and Jack was the LAST STRAW for Millie.	737	5.48	5.44	4.60	4.68
LATE BLOOMER	AN	Raphael was a LATE BLOOMER when he finally arrived at the department.	278	5.28	5.32	4.88	4.76
LAWNMOWER PARENTS	NN	That kid never deals with his problems because he has LAWNMOWER PARENTS.	1	3.04	2.60	3.36	3.12
LAZY EYE	AN	As a child, her doctors told her she had a LAZY EYE.	99	5.04	5.88	4.96	4.76
LIGHTNING REFLEXES	NN	Andrew is good at ping-pong because he has LIGHTNING REFLEXES.	19	5.48	5.04	5.52	4.96
LIP SERVICE	NN	The only thing administrators give is LIP SERVICE to their employees.	1141	5.04	5.00	4.48	4.12
LONE WOLF	AN	Chris considers himself a LONE WOLF because he is very independent.	475	5.44	5.72	5.36	4.96
LONELY OVAL	AN	The cloud was a LONELY OVAL in the clear blue sky.	0	3.48	2.76	3.60	3.04
LONG FUSE	AN	I've got a LONG FUSE, but after weeks of tolerating issues, I get annoyed very quickly.	21	4.84	4.28	4.68	4.60
LOVE BUG	NN	Nick got bit by the LOVE BUG on Valentine's Day.	122	5.40	5.04	4.96	4.40

LOVE HANDLES	NN	Vanessa goes to the gym to get rid of unwanted LOVE HANDLES and a bit of a belly.	173	5.04	5.56	4.44	4.32
LOVE TRIANGLE	NN	There was a LOVE TRIANGLE of tension building in the office.	563	5.36	5.76	4.64	4.72
LOW KEY	AN	Although he was guaranteed a promotion, Paul kept it LOW KEY.	355	5.44	4.96	4.12	3.84
LUKEWARM RECEPTION	AN	Alex's best idea got a LUKEWARM RECEPTION.	22	4.92	4.68	4.96	4.92
MAD SLAP	AN	The music Andrew listened to yesterday was a MAD SLAP.	0	3.88	3.60	3.56	3.24
MAGIC BULLET	AN	People believe that the adaptive textbook is the MAGIC BULLET of educational technology.	451	5.32	4.80	4.00	4.52
MAGNETIC ATTRACTION	AN	Holly felt an undeniable MAGNETIC ATTRACTION towards a boy she barely knew.	48	5.48	5.28	5.68	5.40
MEAT PUPPET	NN	George is Pat's MEAT PUPPET, he does whatever she says.	13	3.76	3.64	3.44	4.00
MEDIA STORM	NN	The resignation of the duke and duchess caused a MEDIA STORM.	75	5.08	4.68	5.12	4.68
MELTING POT	AN	There are different heritages mixed into the American MELTING POT.	1050	5.48	5.28	5.44	4.40
MEMORY LANE	NN	Taking a trip down MEMORY LANE helps when you're having a bad day.	3	5.28	5.44	4.52	4.84
MERMAID DRESS	NN	A MERMAID DRESS is the best option for those who want to show off their curves.	9	4.68	3.48	5.00	3.96
MILK MOUSTACHE	NN	The toddler with the MILK MOUSTACHE was eating by himself.	9	5.52	4.88	4.56	4.36
MOLTEN SOUND	AN	That band has such a MOLTEN SOUND.	0	3.88	3.24	3.76	3.44

MONEY LAUNDERING	NN	The politician was accused of MONEY LAUNDERING.	1526	5.76	5.52	5.44	4.80
MONKEY BUSINESS	NN	I bet Paula and Frank are up to some MONKEY BUSINESS upstairs.	189	5.48	5.12	4.36	4.08
MONKEY DANCE	NN	Everyone is doing the MONKEY DANCE at night clubs these days.	14	3.44	4.28	4.56	4.08
MORAL COMPASS	NN	Most villains lack a MORAL COMPASS.	774	5.40	5.88	4.80	5.04
MOTHER CELL	NN	If the MOTHER CELL is introduced into the DNA of a mutated animal, a cure could be possible synthesized.	15	3.64	3.36	5.08	3.72
MOTHER LODE	NN	He believes he has hit the MOTHER LODE.	372	5.16	4.64	4.76	4.16
MOVIE STAR	NN	Meryl Streep is without a doubt a MOVIE STAR.	2687	5.80	5.68	5.48	4.96
MUFFIN TOP	NN	Cindy has noticed a little MUFFIN TOP when she puts on her best pair of jeans.	77	4.88	5.32	4.40	4.56
NAKED EYE	AN	Although not visible to the NAKED EYE, the professionals understood what makes a successful image composition.	1144	6.04	5.52	5.20	4.64
NIGHT OWL	NN	You have to be a very early riser or a lively NIGHT OWL to enjoy that festival.	210	5.92	5.84	5.56	5.12
NINJA WRITING	NN	The NINJA WRITING course will show you the techniques and strategy you need to write effectively.	0	3.52	3.04	3.64	3.52
NODDING LEAVES	AN	The gentle NODDING LEAVES in fall are my favorite.	0	3.60	2.76	3.40	3.16
NUCLEAR ANGER	AN	His NUCLEAR ANGER made him a frightening person to be around.	0	3.72	3.16	4.80	4.36
NUN GIRLFRIEND	NN	Nathan and his NUN GIRLFRIEND never party because she goes to church every Sunday morning.	0	3.28	2.92	3.96	3.36

OFFICE MONKEY	NN	Daniel is Microsoft's OFFICE MONKEY, constantly doing other people's errands.	2	4.20	3.24	3.40	3.24
OLD DINOSAUR	AN	That professor is an OLD DINOSAUR who doesn't know how to use a computer.	24	5.00	4.52	5.16	4.80
OLD FLAME	AN	Charles reconnected with an OLD FLAME named Camilla.	255	5.68	5.52	4.56	4.12
OLD HAT	AN	My favorite joke might be a bit of an OLD HAT by now.	358	4.28	5.00	3.96	3.88
ONLINE SCRUB	AN	The publicist had to do an ONLINE SCRUB to delete any evidence from the internet.	0	4.60	3.76	4.44	3.68
OPEN MIND	AN	He has an OPEN MIND about what people have to say.	1763	5.96	5.84	5.96	5.40
PADDLE BRUSH	NN	The PADDLE BRUSH is known to work wonders on thicker, longer hair.	24	3.88	4.12	4.84	3.72
PANDA EYE	NN	A way of preventing the dreaded PANDA EYE is to go for waterproof mascara.	2	4.00	3.72	4.40	3.52
PARTY POOPER	NN	Everyone in the room wondered who invited the PARTY POOPER.	118	5.76	5.12	4.68	4.68
PEAK HOURS	NN	Increasing traffic in the area has made travel difficult during PEAK HOURS.	201	5.92	5.76	5.84	4.80
PENGUIN HUDDLE	NN	The students are doing a PENGUIN HUDDLE to stay warm.	1	4.12	3.20	4.36	3.68
PIE CHART	NN	The website shows you what your budget looks like in a PIE CHART divided into the various expense categories.	188	5.92	5.92	5.52	5.08
PINK NOISE	AN	Researchers found that steady PINK NOISE increases stable sleep.	18	3.52	3.40	2.96	3.08

PLASTER SHACKLE	NN	The cast on his broken leg was a PLASTER SHACKLE.	0	4.04	3.16	4.16	3.16
PLUM JOB	NN	She later landed a PLUM JOB at the prestigious Library of Congress.	49	4.04	3.80	3.56	3.76
POKER FACE	NN	Melissa discovered quite early that she had a good POKER FACE.	316	5.72	5.60	5.20	5.04
POOL NOODLE	NN	Everyone brings a POOL NOODLE to the town swimming spots.	8	4.76	4.96	4.36	4.12
PUNCHING BALL	AN	Whenever her husband was frustrated, Jane became his PUNCHING BALL.	1	3.76	4.52	3.96	4.36
PYRAMID SCHEME	NN	New York's elite have been running a decades-long PYRAMID SCHEME.	247	5.96	5.40	5.40	4.36
RABBIT HOLE	NN	It is very difficult to read papers without falling into a RABBIT HOLE.	696	5.56	5.80	5.00	4.36
RAGING BULL	AN	Surfing in the Pacific Ocean is a dangerous sport because the RAGING BULL can easily knock surfers over.	323	4.44	4.76	4.76	5.04
RARE BIRD	AN	Having odd hobbies means you'll be picked on as a RARE BIRD by school bullies.	105	3.88	4.40	3.88	4.56
REBOUND EFFECT	NN	The patient experienced a REBOUND EFFECT after having liposuction done.	101	4.48	3.84	4.72	4.96
RED FLAG	AN	Loss of appetite is a RED FLAG for malnutrition.	1865	5.92	5.76	5.08	5.12
RESTAURANT CHAIN	NN	McDonald's is the most famous RESTAURANT CHAIN in the world.	319	5.80	5.80	5.60	4.60
RISING STAR	AN	The new member of parliament is a RISING STAR in politics.	987	5.96	5.92	5.12	4.92
ROAD DOG	NN	Bill was visiting all the good places with his ROAD DOG.	17	3.84	4.12	3.48	4.12
ROAD HOG	NN	If it were not for that ROAD HOG, he would have gotten to work on time.	15	4.36	4.48	4.20	3.96

ROCKY END	AN	The couple had a ROCKY END to their relationship.	2	5.36	3.44	5.20	4.08
ROTTING EDUCATION	AN	The last government collapsed because of the ROTTING EDUCATION system.	0	4.68	3.20	4.48	3.56
ROUGH DAY	AN	Monday is always a ROUGH DAY at school.	497	5.60	5.52	6.04	5.00
RUG RATS	NN	Family restaurants are always noisy because RUG RATS are constantly running around.	50	4.68	4.88	3.80	4.76
RUNNING HEAD	AN	Research reports normally require a RUNNING HEAD.	23	3.60	3.60	3.16	3.20
SALTY DOG	AN	A SALTY DOG knows everything about the seas.	33	3.72	4.00	3.32	3.72
SEAMLESS RECOVERY	AN	The patient went back home and had a SEAMLESS RECOVERY.	0	4.76	5.36	4.84	4.48
SHARK LAWYER	NN	The SHARK LAWYER will always defends her actions as an overzealous advocate.	0	4.12	3.68	4.44	3.56
SHARP MIND	AN	She had a SHARP MIND and was knowledgeable on a broad range of topics.	101	6.20	5.68	5.76	5.60
SHARP SCENT	AN	He disliked the SHARP SCENT of cleaner on her clothes.	26	4.28	3.28	5.16	4.72
SHIVERING LIFE	AN	What draws the reader to the novel is the hope of warming his SHIVERING LIFE.	0	3.52	3.40	3.68	3.64
SHORT FUSE	AN	He has a SHORT FUSE and when pushed, may become violent.	180	5.68	5.56	5.36	4.68
SILENT WOUND	AN	The mother had a SILENT WOUND because of her son’s death.	2	3.72	3.88	4.12	3.84
SILVER LINING	NN	It is important to see the SILVER LINING in every tragedy.	1299	5.84	5.64	4.92	4.64

SITTING DUCK	AN	Not wanting to be a SITTING DUCK for the storm, he left the worn down farmhouse.	201	5.40	5.76	5.00	4.96
SIXTH SENSE	AN	He always had a SIXTH SENSE for when police were around, so he never got caught.	771	5.96	6.40	5.64	4.84
SKELETON CREW	NN	The company went from a SKELETON CREW of 25 employees to 150.	136	4.88	4.32	4.92	4.60
SLAM DUNK	AN	After the prosecutor threw away a SLAM DUNK conviction, he decided to quit.	768	5.16	5.32	4.80	4.76
SLY FOX	AN	There is a SLY FOX in every office who always gets others to do his work for him.	36	4.80	4.84	5.28	4.96
SMALL POTATOES	AN	The lawyer was told to stop fighting for SMALL POTATOES.	23	4.88	4.52	4.64	4.28
SMALL TALK	AN	The advisor doesn't enjoy SMALL TALK, she goes straight to the point.	1924	6.12	5.48	5.24	4.88
SMART COOKIE	AN	She is one SMART COOKIE and a great role model for the rest of us.	81	5.28	5.80	4.12	4.44
SMART PANTS	AN	Diana was wearing her SMART PANTS today, nit-picking what everyone else is saying.	102	4.44	4.68	3.64	3.68
SMOKING GUN	AN	That last piece of evidence provided investigators with their SMOKING GUN.	1157	5.56	5.00	5.00	5.08
SMOOTH SAILING	AN	After researchers learned the new software, everything was SMOOTH SAILING.	377	5.52	5.88	5.32	5.16
SOCIAL BUBBLE	AN	Her SOCIAL BUBBLE consisted mostly of childless couples with little responsibilities.	4	5.36	4.52	4.40	4.52

SOCIAL BUTTERFLY	AN	SOCIAL BUTTERFLIES always make the most friends at parties.	99	5.52	5.20	4.96	5.16
SOCIAL LADDER	AN	They now had a chance to earn college degrees and to climb the SOCIAL LADDER in their new country.	165	5.52	5.16	4.92	5.24
SOCK PUPPET	NN	When someone votes as a SOCK PUPPET, we never learn their true identity.	190	3.92	4.60	4.36	3.84
SOFT LANDING	AN	After her terrible dismissal, Alexa had a SOFT LANDING at her new job.	274	4.44	5.32	4.92	4.40
SOFT SKILLS	AN	SOFT SKILLS are great to know to advance in your career.	219	4.48	4.64	4.72	4.60
SOFT SPOT	AN	Daniel has a SOFT SPOT for street dogs.	869	5.72	5.88	5.08	4.64
SPIDER PLANK	NN	A SPIDER PLANK can give you a full body workout by toning your abs and building flexibility.	0	4.12	3.24	3.64	3.24
SPORTS CLOWN	NN	Everyone was looking at the SPORTS CLOWN who was wearing his Chargers jersey at church.	0	3.52	3.28	3.64	3.32
SPRING ROLL	NN	He sometimes eats a SPRING ROLL at the Chinese takeaway.	65	5.92	4.88	5.12	4.52
SQUARE MEAL	NN	In the school, children receive three SQUARE MEALS a day.	36	5.36	4.52	4.44	4.60
STALKING HORSE	AN	The client was used as a STALKING HORSE to get an even bigger client.	96	3.76	3.00	3.20	3.08
STAR JUMP	NN	A STAR JUMP can be substituted by a squat and a side leg rise.	4	3.84	3.08	4.28	3.44
STARFISH CRUNCH	NN	To minimize injuries, do not perform the STARFISH CRUNCH on a full stomach.	0	3.20	3.08	3.76	2.88

STIFF NOISE	AN	They hate distractions such as funny sounds caused by STIFF NOISE reduction.	0	3.16	3.08	3.52	3.32
STRAIGHT ARROW	AN	He has always been such a STRAIGHT ARROW in his career.	153	5.68	5.60	5.08	5.92
STRAWBERRY LEGS	NN	Implementing the right products into your skincare routine can help treat STRAWBERRY LEGS.	0	3.56	3.00	3.72	3.00
STUMBLING BLOCK	AN	A lack of innovation is a STUMBLING BLOCK for small organizations.	680	4.96	4.28	4.64	4.48
SUGAR BABY	NN	Most people did not understand her work as a SUGAR BABY.	62	4.48	5.40	3.84	4.20
SUGAR DADDY	NN	All of a sudden, he became the big SUGAR DADDY for anybody who needed their rent paid.	317	5.92	6.00	4.68	4.00
SUNNY DISPOSITION	AN	Amanda's SUNNY DISPOSITION was appreciated by everyone.	99	5.52	4.64	5.20	4.84
SWEET PEA	AN	Britney is such a lucky girl, she got the SWEAT PEA of the class to ask her to prom.	392	4.28	5.16	3.52	4.32
SWEET SPOT	AN	There must be a SWEET SPOT that maximizes the efficacy of the drugs and minimizes the side effects.	1121	5.96	5.96	5.08	4.80
SWEET TOOTH	AN	Ryan has a SWEET TOOTH and keeps a stash of candy in his top drawer.	537	5.96	5.92	4.88	4.76
TENDER SKY	AN	The TENDER SKY made for a romantic evening.	0	3.92	3.32	3.48	3.96
TENT DRESS	NN	A TENT DRESS is perfect for women who don’t like to wear anything remotely form-fitting.	11	4.00	3.92	4.44	4.16
TEST BED	AN	The artists used the fourth-floor space as a photography TEST BED.	126	3.84	3.76	3.44	3.72

THIN ICE	AN	John is treading on THIN ICE after all the money he borrowed.	507	5.72	5.28	5.16	4.76
THIRD EYE	AN	Everyone thinks that she has a THIRD EYE for anticipating drama.	390	5.28	4.68	4.20	4.36
TIGER PARENTS	NN	Martha is pushed by her TIGER PARENTS to get the best grades possible.	3	4.00	3.72	3.64	3.36
TOUGH COOKIE	AN	The girl next door is a TOUGH COOKIE.	152	5.68	5.52	4.92	4.68
TOXIC PERSON	AN	A TOXIC PERSON deserves to be unfriended.	13	5.68	5.44	5.96	4.96
TROPHY WIFE	NN	Mark introduced his TROPHY WIFE to the members as his guest.	165	5.60	5.60	4.68	5.36
TUBE TOP	NN	Allegra was wearing a TUBE TOP with sequins to the club.	129	5.08	5.00	5.20	4.36
TUNNEL VISION	NN	TUNNEL VISION can also be considered as a symptom of glaucoma.	407	4.72	5.52	5.20	5.40
UGLY DUCKLING	AN	Hathaway blossomed from an UGLY DUCKLING into a princess in her breakthrough role.	183	5.16	4.96	5.16	4.72
UGLY TRUTH	AN	People are never ready to hear the UGLY TRUTH about themselves.	191	5.72	5.36	5.84	5.24
UMBRELLA TERM	NN	The committee examined which UMBRELLA TERM best encompasses the problems they face.	152	4.40	3.28	4.44	4.16
UNKEMPT AFTERNOON	AN	The UNKEMPT AFTERNOON resulted in a horrible evening.	0	3.48	3.20	3.40	3.20
VELVET SKIN	AN	After she applied the foundation she was left with VELVET SKIN.	7	3.96	4.52	5.40	4.76
VIOLIN GRAPH	NN	A VIOLIN GRAPH is very informative and visually intuitive.	0	2.96	3.08	3.60	3.24
VIRGIN DRINK	NN	The couple had to pay $20 for a VIRGIN DRINK at the bar.	2	4.64	4.64	4.60	4.44

WALKING ENCYCLOPEDIA	AN	Jennifer, the WALKING ENCYCLOPEDIA of the group, is always contributing to the conversation.	48	5.36	4.96	4.68	4.44
WARM WELCOME	AN	The newcomer thanked everyone for the WARM WELCOME during her visit.	500	6.12	5.88	6.12	5.16
WEAK ARGUMENT	AN	When someone has a WEAK ARGUMENT, they resort to personal attacks.	75	5.16	5.12	5.88	4.72
WET BLANKET	AN	Joseph is such a WET BLANKET who takes the fun out of every party.	157	5.00	4.96	4.44	4.52
WHITE BLANKET	AN	During winter, the city of Montreal is covered by a WHITE BLANKET.	60	5.16	3.76	5.32	4.08
WHITE NOISE	AN	WHITE NOISE is currently available in an endless number of apps.	722	4.84	5.32	5.00	4.52
WIGGLE ROOM	AN	He had some WIGGLE ROOM in his schedule.	699	5.60	5.16	4.64	4.64
WILD DOG	AN	The managing partner becomes a WILD DOG when it comes to his firm.	158	4.20	4.64	4.72	4.48
WILD WEST	AN	Now a WILD WEST, the class had no teacher and could do whatever they wanted.	1535	5.40	5.12	4.64	4.64
WINE LEGS	NN	WINE LEGS indicate either high alcohol content and/or high sugar content in wine.	0	3.56	2.96	3.12	3.88