Semantic properties, aptness, familiarity, conventionality, and interpretive diversity scores for 84 metaphors and similes

Carlos Roncero · Roberto G. de Almeida

Published online: 10 July 2014 © Psychonomic Society, Inc. 2014

Abstract For 84 unique topic-vehicle pairs (e.g., knowl*edge-power*), participants produced associated properties for the topics (e.g., knowledge), vehicles (e.g., power), metaphors (knowledge is power), and similes (knowledge is like power). For these properties, we also obtained frequency, saliency, and connotativeness scores (i.e., how much the properties deviated from the denotative or literal meaning). In addition, we examined whether expression type (metaphor vs. simile) impacted the interpretations produced. We found that metaphors activated more salient properties than did similes, but the connotativeness levels for metaphor and simile salient properties were similar. Also, the two types of expressions did not differ across a wide range of measures collected: aptness, conventionality, familiarity, and interpretive diversity scores. Combined with the property lists, these interpretation norms constitute a thorough collection of data about metaphors and similes, employing the same topic-vehicle words, which can be used in psycholinguistic and cognitive neuroscience studies to investigate how the two types of expressions are represented and processed. These norms should be especially useful for studies that examine the online processing and interpretation of metaphors and similes, as well as for studies examining how properties related to metaphors and similes affect the interpretations produced.

Electronic supplementary material The online version of this article (doi:10.3758/s13428-014-0502-y) contains supplementary material, which is available to authorized users.

C. Roncero · R. G. de Almeida

Department of Psychology, Concordia University, Montréal, Quebec, Canada

Present Address:

C. Roncero (🖂)

e-mail: tcroncero@gmail.com

Keywords Metaphor · Simile · Semantic norms · Aptness · Conventionality · Familiarity · Interpretive diversity (entropy) · Figurative language

As was first observed by Aristotle (in the Rhetoric), a metaphor might differ from a simile "only in the way it is put." Both metaphors and similes relate a topic (such as minds) to a vehicle (such as *computers*), but in contrast to metaphors, similes include a word such as *like* before the vehicle, to yield copular expressions such as Minds are (like) computers. Aristotle in fact claimed that, except for this difference, the two expressions "mean the same thing." However small, the differences between metaphors and similes-and how they might be computed by language comprehension mechanisms-have led to numerous studies aiming to understand the nature of literal and figurative language (see, e.g., the chapters in Gibbs, 2008, for reviews). Contemporary psycholinguistic research has focused not only on the potential differences and similarities in the ways that metaphors and similes are interpreted, but also on the multiple variables influencing how these expression are processed, and in particular, how a metaphor is comprehended, as well as when a metaphor is used in lieu of a simile (Black, 1955, 1979; Chiappe & Kennedy, 1999, 2000, 2001; Chiappe, Kennedy, & Smykowski, 2003; Gentner & Wolff, 1997; Harris, Friel, & Mickelson, 2006; Jones & Estes, 2006; Miller, 1979; Ortony, 1993; Pierce & Chiappe, 2008; Richards, 1936; Roncero, Kennedy, & Smyth, 2006; Shibata et al., 2012).

Thus far, however, most norming data collected for investigating the variables underlying metaphor and simile processing have been narrow in scope, serving for the production of particular experimental materials to test specific hypotheses (e.g., Bowdle & Gentner, 2005; Jones & Estes, 2006), and usually controlling for only a few of the numerous variables implicated in the interpretation of these expressions. Among

Lady Davis Institute, Jewish General Hospital, McGill University, 3755 Côte Ste-Catherine Road Ouest, Montréal, Quebec H4B 1R6, Canada

the more recent psycholinguistic studies examining copular metaphors and corresponding similes, only a few have actually published norming data with their materials (e.g., Chettih, Durgin, & Grodner, 2012; Chiappe, Kennedy, & Smykowski, 2003; Jones & Estes, 2005, 2006; Pierce & Chiappe, 2008). Most studies have employed rating tasks involving variables as different as metaphoricalness/figurativeness (i.e., how figurative or metaphorical an expression is; e.g., Coulson & van Petten, 2002; Tartter, Gomes, Dubrovsky, Molholm, & Stewart, 2002), meaningfulness (whether or not an expression makes sense or is difficult to interpret, including labels as diverse as interpretability, comprehensibility, and understandability; e.g., Campbell & Katz, 2006; Kazmerski, Blasko, & Dessalegn, 2003; Lai, Curran, & Menn, 2009; McGlone & Manfredi, 2001; Tartter et al., 2002), concreteness (Bowdle & Gentner, 2005), familiarity (Blasko & Connine, 1993; Kazmerski et al., 2003), aptness (e.g., Blasko & Connine, 1993; Gerrig & Healy, 1983; Glucksberg & Haught, 2006b; Jones & Estes, 2005, 2006; Pierce & Chiappe, 2008; Xu, 2010), and conventionality (e.g., Bowdle & Gentner, 2005; Chettih et al., 2012; Gentner & Wolff, 1997; Pierce & Chiappe, 2008; Xu, 2010). And although these norming data have served the purposes of individual studies in the investigation of particular variables affecting the interpretation of metaphors and similes, the availability of materials and norms has been rather limited, making it difficult to compare across these studies, and even to analyze results in terms of the effects of different variables on dependent measures. More importantly, the difficulty of access to materials and norms, combined with the lack of widely available standardized norms, has been an obstacle to replicability, which is the essence of any empirical science (see Asendorpf et al., 2013). In addition, the only two large normative studies on metaphor so far (Cardillo, Schmidt, Kranjec, & Chatterjee, 2010; Katz, Paivio, Marschark, & Clark, 1988) have lacked norms for similes-which many researchers consider the key contrast to metaphors.

In the present study, we first collected the properties (semantic features) of 84 metaphors and similes, as well as those of their topic and vehicle words in isolation. We then ranked the properties in terms of frequency and saliency, to determine which properties reflected the most salient interpretations. A separate group of participants provided connotative ratings for these salient properties, which allowed us to gather for each property and constituent expression a measure of deviation from denotative or literal meaning. These ratings thus also allowed us to examine the degree to which metaphors and similes produce similar interpretations, and whether one form produces a more connotative interpretation than does the other. We also obtained norms for the variables most commonly discussed in current psycholinguistic studies of metaphors-aptness, familiarity, and conventionality-and for a less often examined variable, interpretive diversity. Table 1

provides basic definitions for these variables, as well as an overview of the tasks that we employed and the norms that we obtained. These norms serve to inform the investigation of metaphors and similes with a set of stimuli and related attributes that have played an important role in distinguishing between theoretical predictions driving the empirical investigation of metaphor and simile processing. Furthermore, from the present materials and their associated norms, a variety of quantitative predictions can be made for testing how these expressions are processed online, interpreted by special populations (e.g., Alzheimer's Disease patients), and expressed in particular social environments (e.g., Internet postings), as well as how their processes are neurologically implemented.

The need for norms on metaphors and similes

To our knowledge, thus far only two major normative studies on metaphors have been published (Cardillo et al., 2010; Katz et al., 1988). In the first of these, Katz et al. presented a comprehensive set of norms, involving 464 expressions and ten rating scales divided into four categories: comprehensibility (ratings of comprehension as a sentence and ease of comprehension as a figurative expression), metaphoricity (the degree at which a sentence is literally or figuratively true, as well as the metaphor goodness-how good or apt a metaphor is), imagery (imageability of the whole sentence and of the topic and vehicle separately), and other (familiarity, semantic relatedness between topic and vehicle, and the number of alternative interpretations). Those norms included 204 literary expressions with complex topics and vehicles (e.g., The soul is a rope that binds heaven and earth, Memory is a heap of broken images where the sun beats and the dead tree gives no shelter) and 264 nonliterary metaphors (i.e., created by the authors or collected from other studies). It is to this second set that we can compare our norms. This set contains copular metaphors, as in the present study (e.g., Alcohol is a crutch), but also more complex topics and vehicles (e.g., The creative mind is a kettle on the stove; A white rabbit's fur in winter is a soldier's army-green uniform in a jungle). Of the 264 nonliterary metaphors that Katz et al. used, only 52 have complexity similar to the ones that we employed in the present study, with a Noun-is/are-(determiner)-Noun form (e.g., Alcoholism is a parasite). Of these 52, only 14 have the constituents Noun-is/ are-Noun (Money is penicillin), in contrast to all of our materials, and only six of the metaphors that they used also appear in our norms. The most critical difference between our norms and those of Katz et al., however, is in the kinds of variables investigated. Although both sets of norms collected ratings of familiarity and aptness, we also collected conventionality ratings and semantic properties related to the topicvehicle pairs, both when they were read as either a metaphor

Task, Variable	Variable Definition	Instruction Summary	Sample Item	Items (N)	Participants (N)	Obtained Norms
Properties	Semantic "features" associated with a particular word or expression. Interpretive diversity reflects the richness of interpretation of a given word based on the number of properties produced (see text)	List three properties associated with the vehicle word.	Alcohol is a crutch	84	20	Metaphor vehicle properties; rank-ordered properties
			Alcohol is like a crutch	84	20	Simile vehicle properties; rank-ordered properties
		List up to ten properties associated with each word.	alcohol, crutch	168	20	Topic and vehicle word properties; rank-ordered properties; interpretive diversity
Conventionality	Strength of association between a vehicle word and a figurative interpretation	How common (from 1 to 10) is it to use the word "drug" to mean "addiction" in statements such as	x is a drug	84	20	Metaphor vehicle conventionality rating
			x is like a drug	84	20	Simile vehicle conventionality rating
Connotativeness	Degree of deviation from a literal meaning	Rate (from 1 to 5) whether a property is denotative or connotative.	drug, addictive	68	100	Connotativeness ratings for topic and vehicle properties
Aptness	How well properties of the vehicle capture properties that can be predicated about the topic	Rate (from 1 to 10) how apt it is to say	Life is a journey	84	20	Aptness ratings for metaphors
			Life is like a journey	84	20	Aptness ratings for similes
Familiarity	How well known an expression is	Rate (from 1 to 10) how familiar you are with the statement (i.e., have you heard it before?).	Politics is a jungle	84	20	Familiarity ratings for metaphors
			Politics is like a jungle	84	20	Familiarity ratings for similes

Table 1 Tasks, sample materials, and norms obtained for metaphors, similes, and topic and vehicle words

or a simile and when the topic and vehicle were read in isolation. In addition, we calculated interpretive diversity scores for each topic–vehicle pair based on the collected semantic properties. As we will discuss below, the norms that we collected are associated with the most important variables driving the disputes between current theories of metaphor and simile processing (e.g., Bowdle & Gentner, 2005; Gentner & Bowdle, 2008; Gentner & Wolff, 1997; Glucksberg, 2003, 2008; Glucksberg & Haught, 2006a, 2006b).

The second major normative study, Cardillo et al. (2010), included a total of 280 metaphorical expressions and 280 literal controls. Of the 280 metaphors, more directly related to our set of norms were their 140 nominal (copular) metaphors. Their norms include imagery, familiarity, naturalness, imageability, and figurativeness (obtained with Likert scales), word frequencies (obtained from databases such as the MRC Psycholinguistic database; Coltheart 1981), and sentence norms, including positive–negative valence (using a timed response task). Like Katz et al.'s (1988) norms, however, those of Cardillo et al. (2010) differ from ours in several respects. First, they differ in the grammatical properties of their expressions: of the 140 nominal metaphors, only seven were similar to ours in verb tense (i.e., present singular, *is*), with all others being past singular or plural (*was/were*).¹ In addition, in Cardillo et al.'s (2010) norms, with the exception of only 11 items, the metaphors contained complex topics (e.g., *The secretary's promotion was a leap*), complex vehicles (e.g., *The reception was an icy swim*), or both (e.g., *The business card was an optimistic chirp*). Crucially, we provide norms for parallel metaphors and similes maintaining the same topic and vehicle terms, thus allowing for direct comparison between the two types of expressions, without the need to control for extraneous factors such as the word frequency and imageability of different topics and vehicles, which might affect how these expressions are processed.

¹ We found only one study that has investigated the processing of copular metaphors employing expressions in the past tense (Cardillo, Watson, Schmidt, Kranjec, & Chatterjee, 2012); the reason might be the implications that verb tense might have for a categorization (Glucksberg, 2008) or a comparison statement (Gentner & Bowdle, 2008). For example, if someone says *My lawyer was a shark*, it implies that (a)the lawyer is no longer a shark, although he or she was for an extended period of time; (b) the lawyer was briefly in a given state (*shark*); or (c)there is no lawyer at the present moment. None of these possible implicatures involve categorizations or comparisons that endure, and thus arguably they weaken the potential impact of a figurative statement. By contrast, *My lawyer is a shark* conveys only that the lawyer *is* either a member of an (ad hoc) category or that a lawyer can be compared with sharks on some metric.

Therefore, our norms differ substantially from the only two other published normative studies, in method, the norms obtained, and types of expressions. In addition, the previous normative studies lacked norms for similes, and instead employed different forms of metaphorical expressions rather than the productive copular forms. With regard to topics and vehicles, a key difference from the previous norms is that we did not employ complex phrases, thus allowing for the use of the present norms in experiments requiring precise measurements of activity at the word level-as in priming, eyetracking, and fMRI paradigms, among others. For these reasons, the set of norms collected in the present study is perhaps better suited than the other norms for studies investigating the key variables known to affect the interpretation of metaphors and similes. In particular, the norms should be of greater value for studies investigating the types of properties evoked by metaphors and similes during comprehension. This is so because, in contrast to the previous metaphor norms, which did not include property lists and paraphrases for their expressions, our study obtained the properties evoked by metaphors, similes, and their respective topic and vehicle words. The semantic features that we collected, together with feature saliency, connotativeness ratings, and interpretive diversity, allow for a better understanding of how metaphors and similes-as well as their constituent topic and vehicle words-might actually be interpreted.

Metaphor versus simile interpretation

In order to illustrate the role that particular variables play in metaphor and simile interpretation, consider how a given vehicle contributes to the interpretation of a simile or metaphor. The "career-of-metaphor" theory (Bowdle & Gentner, 2005; Gentner & Bowdle, 2008) proposes that metaphor vehicles gain their figurative meaning from the use of the vehicle word in simile expressions. As a vehicle becomes conventionally used to express a given meaning, its use in metaphor form would be preferred over its use in simile form. Because this theory assumes that metaphors get their meaning from originally being interpreted as similes, the prediction is that metaphors and similes with the same vehicle should elicit the same properties, because conventional vehicles consistently activate the same salient properties from memory. For example, gold mine used as a vehicle is conventionally associated with the idea of being the "source of something valuable" (Gentner & Bowdle, 2008, p. 116), regardless of which topic is used in the expression. Categorization theory (e.g., Glucksberg & Haught, 2006b) makes a different prediction of how vehicles contribute meaning to a metaphor or simile. According to this theory, vehicles have a dual reference: activating literal properties in similes, but both figurative and literal properties in metaphors (Glucksberg & Haught, 2006a, 2006b; Hasson, Estes, & Glucksberg, 2001). Hasson et al. found support for this prediction when they asked participants to write paraphrases of metaphor and simile statements. The paraphrases for similes such as *Ideas are like diamonds* elicited more literal properties (e.g., *shiny*) that were related to the vehicle. Possible evidence supporting dual reference in the present study would be a larger number of properties for metaphors than for similes, assuming that the metaphor would activate both figurative and literal properties. Likewise, if similes activate more literal properties than do metaphors, then higher ratings of connotativeness would be expected for the properties written for metaphor versions of the topic–vehicle pairs, because connotativeness reflects the degree to which properties are figurative or literal.

Method

Participants

A total of 280 Concordia University undergraduate students participated in this study. Sixty of them participated in properties list task, 40 in the aptness rating task, 40 in the familiarity rating task, 40 in the conventionality rating task, and 100 in the properties connotativeness rating task. All participants were native speakers of English and received credit toward a psychology course or monetary compensation for participation in the study.

Materials

The main stimuli consisted of 84 topic-vehicle pairs, written as metaphor and simile sentences and as topic and vehicle words in isolation. We chose to use items that had simple topic-vehicle constituents (i.e., copular metaphors with single-word topics and vehicles were preferred) and that had been used previously in research on metaphors and similes. Sixty-six of these topic-vehicle pairs had been used in previous research (Chiappe, Kennedy, & Chiappe, 2003; Roncero et al., 2006). To increase this cohort size, we employed Roncero et al.'s (2006) method, entering phrases such as "common metaphor," "common simile," "an example of a metaphor is," and "an example of a simile is" as queries into the Google search engine. Possible topicvehicle pairs were then accepted if they produced a minimum of three metaphor and simile statements combined. Eighteen topic-vehicle pairs were added as a result of this procedure, to create the total cohort of 84 topic-vehicle pairs. Booklets for the collection of the different norms are described in more detail below.

Collection of properties

The main semantic norms collected in the present study were based on the properties (features) provided by participants to the set of metaphors, similes, and their topics and vehicles presented in isolation. Our strategy was first to obtain the set of properties elicited by each of these expressions and words, to gather a basic metric of their interpretation. We wanted to use these semantic properties to refine our understanding of metaphors and similes by obtaining five other characteristics of these expressions. First, we calculated salience, on the basis of the properties that people wrote first in the list-by hypothesis, the properties that most easily came to mind when people read the metaphor and simile expressions (following Giora, 1997). Second, we calculated property frequency, on the basis of the most-often-produced properties for each word or expression. Third, the more frequent properties were used to determine the *conventionality* of the metaphors and similes. The conventionality rating task is described separately below. Fourth, we used the most salient properties in a separate connotativeness rating task, devised to determine whether the properties were deemed to be more connotative or denotative, comparing across words and expression types. This task is also described in more detail below. Finally, we also used the properties elicited to calculate interpretive diversity, following Utsumi's (2005) procedure. Diversity reflects the "richness of interpretation" (Utsumi, 2005, p. 153) of a given expression and is derived from the concept of entropy (Shannon, 1948). Details about the interpretive diversity norms are also presented separately below.

For the properties task, 60 participants were given one of three booklets containing either the metaphors, the similes, or the topic and vehicle words in isolation, with 20 participants assigned to each booklet type. For each expression listed in the metaphors and in the similes booklets, participants were asked to list three properties that they felt the vehicle word was expressing about the topic. For example, participants could have written the properties soothing, healing, and enjoyable for the metaphor Music is medicine. Three separate lines were written below each expression, and participants were asked to list each property on a separate line. This manipulation allowed us to measure whether a property had been written first (on the top line), second (on the middle line), or third (on the final line). No time limit was imposed, and participants were encouraged to write three properties for every expression, but told that it was fine to move onto the next expression if they had difficulties coming up with three properties.

The third booklet contained a list of 168 words, constituted by the 84 topics and 84 vehicles in isolation (e.g., *time* and *money* from the metaphor *Time is money*). Participants were asked to list properties related to the each word. For example, for the word *money*, participants could have written the properties *valuable* and *green*. However, because these words presented in isolation may be interpreted more literally than when they were presented in metaphors and similes, we were concerned that participants might only write literal properties (e.g., *green* for *money*) if they were restricted to listing only three properties. To obtain a wide range of properties for each topic and vehicle term in isolation, we encouraged participants to list up to ten properties for each word. As was the case with the metaphors and similes, no time limit was imposed, and participants could proceed to another expression when they felt that they had written all of the properties that came to mind while reading a particular topic or vehicle.

For each booklet type (i.e., metaphors, similes, and words), we tabulated how often a property was listed for the full expression and for each of the topic and vehicle words in isolation. Collapsing the ratings for semantically similar words can reduce the amount of variation inherent in the set of properties elicited. For example, although strong and powerful might reflect similar concepts, there are differences between these words. Therefore, we counted distinct properties even when they reflected potentially similar concepts, and restricted collapsing responses to when words shared the same morphological root (e.g., sleep and sleepiness). By tabulating how often the different properties were listed, we were able to determine the most commonly listed properties for each metaphor and simile, as well as for each topic and vehicle read in isolation. An associated saliency rating was also determined for each property by checking whether the property was listed first, second, or third by the participant. A property was given a score of 1 if written first, 2 if written second, and 3 if written third. Therefore, properties that were more often written first would have an associated saliency rating average closer to 1. These associated saliency ratings were used to calculate the interpretive diversity score of each expression and which properties would be used for the connotativeness and conventionality ratings. In particular, we considered properties to be salient when they were reported by at least 25 % of the participants sampled.

Aptness ratings

Apt metaphors are taken to be those whose vehicle term by hypothesis activates many properties from memory that might also be true of the topic (Chiappe & Kennedy, 1999). Activating properties that are salient for both the topic and the vehicle is also predicted to aid with establishing a categorical relationship and makes comprehending metaphors easier. More specifically, categorization is easier when the created classification is seen as relevant and informative, as determined by the number of salient properties being attributed to the topic by the vehicle (Glucksberg & Keysar, 1990, 1993; Jones & Estes, 2006). Supporting this prediction, multiple studies have found that more apt metaphors are processed faster and are more easily understood (Blasko & Connine, 1993; Chiappe & Kennedy, 1999; Chiappe, Kennedy, & Chiappe, 2003; Chiappe, Kennedy, & Smykowski, 2003; Glucksberg & McGlone, 1999; Jones & Estes, 2005, 2006).

In order to collect aptness ratings, two booklets were created: one that presented topic-vehicle pairs as metaphors (Life is a journey), and one presenting topic-vehicle pairs as similes (Life is like a journey). A scale ranging from 1 (not at all apt) to 10 (very apt) was presented beneath each statement. Twenty participants received booklets that listed metaphors, whereas another 20 participants received booklets with similes. Participants were told that they would read statements that involved a relationship between two terms, a topic and a vehicle, and that their task was to rate how apt they found each statement by circling the number that reflected their judgment. Participants were told that apt expressions were those in which the second term, the vehicle, captured salient properties of the first term, the topic. Politics is a jungle was given as an example of an apt statement, whereas Politics is a beach was given as an example of a less apt statement.

Familiarity ratings

Familiarity reflects how well known a given expression is. This measure is not the same as conventionality, because the latter reflects figurative properties of the vehicle only, nor is it the same as aptness, which reflects properties of the vehicle attributable to the topic. Thus, although it is unfamiliar, a metaphor such as Knowledge is light could be considered apt, because one can take *light* to predicate something of knowledge. Familiarity with a given metaphor would presumably make processing easier, reflecting the general effect that practice with any particular expression typically leads to faster processing. Consistent with this argument, studies have found that more familiar metaphors are read faster than less familiar metaphors (e.g., Blasko & Briihl, 1997; Blasko & Connine, 1993). This difference suggests that novel metaphors would require more effortful processing than familiar metaphors. Gentner and Bowdle (2008) and Glucksberg (2008), however, argued that conventionality and aptness are more important variables than familiarity, because both conventionality and aptness make metaphor comprehension permissible, even when the expressions are novel.

A procedure similar to the one employed in the aptness task was used for collecting familiarity ratings. However, rather than presenting scales for aptness, scales were created for familiarity from 1 (*not at all familiar*) to 10 (*very familiar*). Forty participants received booklets that contained either metaphors or similes, and they were asked to circle the number that reflected how familiar they found each expression. Specifically, they were asked to report the extent to which they had heard or read the statement in the past.

Conventionality ratings

Consistent with other studies (e.g., Bowdle & Gentner, 2005; Gentner & Wolff, 1997; Jones & Estes, 2006), we defined conventionality as the strength of association between a word (the vehicle) and a specific figurative meaning (the most frequently listed property for that vehicle). For this rating task, then, we took the property listed most frequently for each expression—as obtained in the properties task described above—and used it for determining vehicle conventionality. When several properties were equally frequent, we chose the one with the lowest associated saliency rating (i.e., the one whose mean was closer to 1).

A group of 20 participants received a booklet containing vehicles inserted in metaphor frames (e.g., x is a drug). For each metaphor vehicle, participants were presented with the most frequently generated property for that vehicle, which had been obtained when the vehicle was presented in a metaphor context in the earlier properties task. Participants were asked to rate the extent to which that vehicle was used to express that particular property. For example, addiction was found to be the most salient property for *Love is a drug* in the properties task, and thus in the conventionality booklet participants were asked to what extent the word *drug* is used to convey addiction in statements such as x is a drug. Together with each property and statement, participants were presented with a scale ranging from 1 (not at all conventional) to 10 (very conventional). Another group of 20 participants received the same list of vehicle words, but in simile frames (e.g., x is like a drug), together with the properties generated most frequently for those vehicles stemming from the similes booklet in the properties task.

Connotativeness ratings

Connotative properties are seen as reflecting more emergent properties, whereas denotative properties are those that reflect more literal properties (Danesi, 1998). For example, when describing a cat, *furry* would be a more denotative property than *sneaky*. For this task, we created a booklet containing a list of the associated salient properties produced for the metaphor and simile vehicles, to examine whether metaphors or similes elicit more literal properties. Eight metaphors (9.5 %) and 13 similes (15.4 %) had no salient properties (i.e., those stated by at least 25 % of the participants); therefore, connotative ratings were collected for properties related to only 68 metaphor and simile pairs. Also, because connotativeness can be a difficult variable to rate, we used a sample size of 100 participants to collect these ratings. The difference between a denotative and a connotative property was explained to the participants by means of a short quiz to ensure that they understood the distinction. In the quiz, they were asked whether rectangular and artistic were connotative or denotative properties of *painting*, and whether *fur* and *playful* were connotative or denotative properties of cat. Only participants who correctly distinguished between these properties were allowed to proceed with the ratings. If a participant did not answer these quiz questions correctly, the concepts of connotation and denotation were explained again, and they were given a similar quiz. If an incorrect answer was again given, the participant was excused from the study. From the initially recruited group of 104, only four participants failed the final quiz and were excused from the study. The participants permitted to provide connotativeness ratings were asked to rate the extent to which each property reflected either a connotative or a denotative property of that vehicle, on the following scale: 1 (strictly denotative), 2 (more denotative than connotative), 3 (both denotative and connotative), 4 (more connotative than denotative), and 5 (strictly connotative). This scale was presented for each salient property in relation to each one of the metaphors and similes.

The norms

The present norms were obtained by first collecting lists of properties for metaphors, similes, and their topics and vehicles separately. We also obtained independent ratings for metaphor and simile aptness as well as metaphor and simile familiarity. On the basis of the property lists, we calculated the most frequently produced property for each expression, topic, and vehicle. The most frequently produced property for each vehicle were then used in the conventionality task. We also calculated the most salient properties on the basis of the order in which properties were produced in the properties task. We then obtained connotativeness ratings for these salient properties, to estimate the degrees of connotativeness for metaphors and similes. Finally, the salient properties were also used to calculate the degree of interpretive diversity or "richness of interpretation" of each of our expressions. We present below the data analyses for the main norms that we produced: properties, aptness, familiarity, conventionality, connotativeness, and interpretive diversity. In addition, we present analyses on the correlations between these variables, on the reliability of our data, and on the differences and similarities between the metaphor and simile interpretations gathered from our norms. See the supplemental materials for the full set of norms.

Properties

Participants produced a large variety of properties for each metaphor and simile, as well as for each topic and vehicle word in isolation (see Tables S1 and S2 in the supplemental materials). The mean number of different properties listed for

metaphors per topic-vehicle was 8.15 (SD = 2.20), whereas for similes this value was 7.93 (SD = 2.32). The difference between metaphors and similes was not significant (t = 37.11, p = .27). We next examined whether the properties activated by the metaphors and similes were more related to the properties related to the topic or the vehicle. The three most listed properties for each topic and vehicle term in isolation were compared to the three most listed properties for the metaphors and similes. In cases of a tie for the third most frequent property, the property with the lower associated saliency rating was selected. Comparing the properties listed for each metaphor and simile to those listed for each topic and vehicle term in isolation, a match score out of 3 was given for how many of the properties were common between the term (topic or vehicle) and the metaphor or simile. Therefore, scores closer to 3 represent greater overlap, whereas scores closer to 0 reflect minimal match. We found greater overlap between those properties written for metaphor expressions and for vehicles in isolation (M = .65, SD = .63), relative to the overlap between metaphor expressions and topics in isolation (M = .40, SD = .56) [t(84) = 271, p < .05, r = -.29]. A similar pattern emerged for similes: The amount of overlap was greater for vehicles (M = .60, SD = .68) than for topics (M =.39, SD = .54) [t(84) = 240.50, p < .05, r = .25]. Therefore, both metaphors and similes appear to activate properties more related to the vehicle term than to the topic.

The low overlap scores (<1) between expressions and topic-vehicle terms in isolation further suggests that metaphors and similes activate properties not elicited by the topic and vehicle terms in isolation (i.e., emergent properties). To further examine this result, we computed overlap scores between the three most frequent properties for each topic-vehicle expressed as a metaphor and the three most frequent properties for expressions written as similes. This comparison produced an averaged overlap score of 1.46 (SD = 0.84) for metaphors to similes, and an average overlap score for similes to metaphors of 1.38 (SD = 0.83), which were significantly larger overlap scores than those found for vehicle properties [metaphors, t(84) = 200, p < .01, r = -.63; similes, t(84) = 112, p < .01, r = -.62]. Therefore, both metaphors and similes produced emergent properties-that is, properties not elicited when the topic and vehicle terms were presented in isolation.

Aptness

The metaphors had aptness ratings ranging from 1.65 to 9.22, with a mean of 5.95 (SD = 1.80). The aptness ratings for similes ranged from 2.04 to 9.52, with a mean of 5.69 (SD = 1.66). The aptness difference between the two expressions was significant [t(83) = 2.27, p < .05, r = .24], although the correlation between these ratings was also significant and large [r(84) = .82, p < .001]. This result suggests that the aptness level for a given expression depends more on the topic

and vehicle in that expression, rather than on whether the expression was a metaphor or a simile.

Familiarity

The familiarity scores for metaphors ranged from 1.1 to 10, with a mean score of 4.2 (SD = 2.09), whereas the familiarity scores for similes ranged from 1.12 to 9.68, with a mean score of 4.58 (SD = 2.23). The correlation between metaphor and simile familiarity ratings was large and significant [$r_s(84) = .85, p < .001$], although participants were found to give higher familiarity ratings for simile statements than for metaphor statements (t = 1,154, p < .01, r = ..31).

Conventionality

Metaphor conventionality scores ranged from 1.70 to 9.85 (M = 7.20, SD = 1.87), whereas simile conventionality scores ranged from 2.12 to 9.94 (M = 7.40, SD = 1.86). Conventionality ratings for metaphors and similes were not found to be significantly different (t = 1,444, p = .13, r = -.17), and a significant correlation was also found between the ratings [$r_s(84) = .34$, p < .001].

Interpretive diversity

First used by Utsumi (2005), interpretive diversity is calculated using Shannon's (1948) formula for estimating the entropy of a given source—that is, the amount of information (and thus, uncertainty) that a source generates:

$$H(X) = -\sum p(x_i)\log_2 p(x_i),$$

where H(X) is the entropy value at source X, and $p(x_i)$ is a range of possibilities $(x_1, x_2, x_3,...)$, not all of which have the same probability of occurrence. The formula then, serves to calculate the overall probability of a given event (in fact, a message), when there are several possible outcomes, each with its own probability of occurrence. Although discussion of the specifics of the equation goes beyond the scope of the present article, the equation provides us with the means for calculating the diversity produced in the properties of a given expression. Values of interpretive diversity are maximal when many properties are equally salient, but the minimum value is 0 if a word or expression has only one property (i.e., only one meaning). When examining the properties activated by metaphors and similes, greater values of interpretive diversity would reflect a larger amount of equally salient properties being activated. Therefore, if a topic-vehicle pair evokes three salient properties, but only one is activated for another topicvehicle pair, the first pair would have a larger interpretive diversity score. If two expressions evoke equal numbers of properties, however, then the expression whose salient properties are more similarly salient would be the one greater in interpretive diversity. These scores can be informative regarding the number of properties that an expression activates and the saliency differences among these properties.

To calculate interpretive diversity, we counted the number of times that a property was listed for a vehicle to create a list of properties (x_i), with saliency values corresponding to the relative saliencies of those properties (p). Consistent with Utsumi (2007), we excluded those properties listed only once. All values were then plugged into Shannon's entropy equation to determine an interpretive diversity value for each metaphor and simile. Metaphors were found to have interpretive diversity scores ranging from 3.10 to 5.52 (M = 4.49, SD = .48), whereas similes were found to have interpretive diversity scores ranging from 2.89 to 5.52 (M = 4.55, SD = .47); this difference was not significant (t = 1,559, p = .31, r = ..11), and the values for metaphors and similes were significantly correlated [$r_s(84) = .61$, p < .01].

Correlations between the expressions' variables

In order to examine possible correlations between our different variables, we collapsed the ratings across sentence types (i.e., metaphors and similes of the same topic-vehicle pair). This decision was motivated by the significant correlations (>.8) found for measures of aptness and familiarity, and by the nonsignificant differences found between metaphor and simile expressions on conventionality ratings and interpretive diversity scores. The correlation between conventionality ratings and aptness ratings was not significant $[r_s(84) = .19, p = .09]$, but a significant correlation was found between conventionality and familiarity $[r_s(84) = .30, p < .05]$. The correlation between interpretive diversity scores and conventionality ratings was also significant, but negative $[r_s(84) = -.36, p < ...]$.001]. Therefore, the higher the conventionality rating is for a vehicle, the lower the level of interpretive diversity associated with that vehicle. Aptness ratings were not found to correlate with interpretive diversity ratings $[r_s(84) = .09, p =$.41], but a positive correlation was found with averaged familiarity ratings $[r_s(84) = .73, p < .001]$. Thus, more familiar expressions were also considered more apt. Finally, we found no significant correlation between interpretive diversity ratings and familiarity ratings ($r_s = .19, p = .09$). The different ratings for metaphors and similes are presented in Table S3 of the supplemental materials.

Reliability

We also conducted reliability analyses for those measures for which participants were asked to provide subjective ratings: aptness, familiarity, conventionality, and connotativeness. For the first three measures, we compared the averaged scores from the first ten participants to those from the latter ten participants, but compared the first 50 to the latter 50 participants for connotativeness ratings. For aptness ratings, Cronbach's alpha was .89 for metaphors and .84 for similes. For familiarity ratings, Cronbach's alpha was .88 for metaphors and .80 for similes. For conventionality ratings, Cronbach's alpha was .92 for metaphors and .91 for similes. Comparing the first 50 participants' ratings to those of the latter 50, Cronbach's alpha was .98 for connotativeness ratings. In summary, all reliability measures were greater than .79, suggesting that our measures have good levels of reliability.

Metaphor versus simile interpretation

We restricted our comparison to salient properties, to ensure that our comparison would only compare properties that were frequently produced by participants. Thus, only the salient properties for 68 metaphor and simile expressions were compared, to see whether metaphors or similes activated more salient properties. Metaphors elicited a mean of 1.95 salient properties (SD = .94), which was significantly more than the mean of the simile-elicited properties (M = 1.62, SD = 0.84), t = 176, p < .001, r = .39. A significant correlation was also found, however, between the numbers of salient properties for metaphors and similes $[r_s(84) = .47, p < .001]$. Therefore, whenever a simile was associated with many salient properties, a larger number was also found for the metaphor. In order to determine the extent to which metaphors and similes had similar salient property lists, we gave metaphors and similes with the same topic-vehicle pair an overlap score, as had been done previously, by comparing the properties written for the expressions versus for the topic and vehicle presented in isolation. If a simile had three salient properties, whereas the metaphor had four salient properties, three of which were those that were determined to be salient for the simile expression, then the comparison was given a match score of 1. In contrast, if a simile had two salient properties, whereas the metaphor had three, and only one of the former properties was also salient for the metaphor, the comparison would receive a match score of .5, because only half of the salient properties in one were also present in the other. The same procedure was done for comparisons in which the simile had more salient properties. Thus, if a metaphor had two salient properties, whereas the simile had three, but two of those three properties were the ones found to be salient for the metaphor, then the match score was 1. In this manner, a mean match score close to 0 would suggest that the metaphors and similes elicited completely different salient properties, whereas a mean match score closer to 1 would suggest that the metaphors and similes activated matching sets. This procedure was similar to that used by Roberts and Kruez (1994) and Graeasser (1981) for the creation of overlap scores to measure the proportion of responses that were common to any two answer distributions—ranging from 0 (*no overlap*) to 1.0 (*perfect overlap*).

Comparing the salient properties for metaphors and similes in the present study, the mean match score was found to be quite high, .83 (SD = .35), indicating that the salient properties activated by our metaphors and similes were very similar. Indeed, high match scores were found regardless of whether the number of salient properties was low or high. For example, only one salient property was found for Love is a drug and Love is like a drug, but in both cases the property was addiction. Similarly, three salient properties were found for both Clouds are cotton and Clouds are like cotton, and for both expressions these properties were round, soft, and fluffy. We next examined whether the salient properties elicited by similes were more denotative or connotative than those written for metaphors. The salient properties for metaphors had a mean connotative rating of 3.05 (SD = 0.82), whereas the salient properties for similes had a mean connotative rating of 3.01 (SD = 0.88); that is, the sets of properties elicited for both metaphors and similes were rated as being exactly in the middle of the scale-both denotative and connotative. The difference between the metaphor and simile expressions was found to be not significant [t(70) = 0.64, p = .53]; therefore, the properties activated by similes were not found to be more denotative than those determined as being salient for metaphors.

Discussion

Our primary goal was to collect a set of interpretation norms that could be used as variables for future experimental studies investigating metaphor and simile interpretations. We collected property lists for 84 unique topic-vehicle pairs when they were read as metaphors or similes, as well as when the topic or vehicle term was read in isolation. From these property lists, we also collected frequency and saliency ranks. We then collected connotativeness ratings on these salient properties, to measure the extent to which the expressions produced a list of more connotative properties-which, by hypothesis, would represent expressions that produced a more figurative interpretation. We also examined the amount of overlap between the properties written for the expressions versus for the topic and vehicle in isolation, to measure the extent to which the expressions produced emergent properties. Finally, we collected ratings of conventionality, aptness, and familiarity for the topic and vehicle terms presented as metaphors and as similes, while also determining the interpretive diversity of these expressions.

We found that the properties listed for metaphors and similes were different from those listed when the topic or vehicle was read in isolation, which reflects the potential of metaphors and similes for activating emergent properties during comprehension. We also found, however, that these expression properties were more similar to those listed for the vehicle in isolation, which suggests that the vehicles provide the bulk of the semantic material for the interpretation of metaphors and similes. The conventionality, aptness, and familiarity ratings for metaphors and similes, as well as the interpretive diversity scores for both expressions, were also fairly equivalent. Metaphors and similes did not produce significantly different conventionality ratings, either. Also, although similes were rated as being significantly less apt but more familiar than metaphors, for both variables-aptness and familiarity-the correlations between metaphors and similes were greater than .8, which suggests that the combination between topic and vehicle might play a greater role in interpretation than the type of expression itself. In terms of interpretive diversity, there was also no difference between metaphors and similes.

Among our expression ratings, three significant correlations were found between the norms when the ratings were collapsed: a positive correlation between conventionality and familiarity, a positive correlation between aptness and familiarity, and a negative correlation between conventionality and interpretive diversity. The positive correlation between conventionality and familiarity ratings might be consistent with the career-of-metaphor theory's proposal for the way that vehicles change their properties over time. According to this proposal, a figurative meaning attributed to a particular vehicle word becomes stronger, the more often a vehicle word is used to convey that meaning. Thus, although these are taken to be fundamentally different variables, the familiarity of a given *expression* goes hand in hand with how conventional it is to use a given vehicle to convey a certain idea. The significant correlation between aptness and familiarity replicates Thibodeau and Durgin's (2011) findings. These authors suggested that participants' ratings of aptness and familiarity might not be fully independent. More specifically, because familiar expressions are those that are presumably activated faster, participants could view certain expressions as being apt due to how easily properties related to that expression come to mind. However, we would argue that certain expressions may also become familiar because they are considered "good" metaphors (Chiappe & Kennedy, 2000). People would more often produce the metaphors that they found apt, and in this manner, aptness would breed familiarity. Consequently, it can be generally expected that apt metaphors will also be familiar, and the positive correlations found in the present norms as well as in previous experimental studies (e.g., Chaippe, Kennedy, & Chaippe, 2003) support this view. For these reasons, it can at times be difficult to determine the extent to which ratings such as aptness and familiarity reflect different variables. Finally, because interpretive diversity reflects the extent to which a metaphor or simile potentially conveys many equally salient properties, the negative correlation between conventionality and interpretive diversity has implications for how vehicle conventionality could develop. More specifically, a vehicle word would be less likely to build a strong association with a specific property when other salient properties are competing to be associated with that vehicle. For example, the property lists show examples of conventional vehicles in which one property is dominant over all others: for instance, absorbent for Memory is a sponge, addictive for Love is a drug, slow for Time is a snail, or protecting for Trees are umbrellas. In several expressions, however, multiple properties are salient. Examples of such properties include soft, white, and fluffy for Clouds are like cotton; colorful and beautiful for Love is a rainbow; and flows and long for Memory is a river. Therefore, although it is possible for one salient meaning to become associated with a vehicle term, some vehicles can bring many salient properties to mind.

Metaphor versus simile interpretations

Examining whether metaphors and similes produce similar interpretations, we found that the property lists produced for metaphors and similes were equivalent. The numbers of properties listed for both expression types were not significantly different, and the properties listed for metaphors and similes containing the same topic-vehicle pair were also equivalent. One difference that did emerge was a greater number of salient properties for metaphors, but the levels of connotativeness between metaphor and simile properties were also indistinguishable. The finding that metaphors activate more salient properties than do similes could in part be accounted for by categorization theory's proposal of a "dual reference" for metaphor vehicles (Glucksberg, 2003). According to this proposal, the vehicle word is predicted to access the literal meaning when embedded in a simile expression, but the same vehicle accesses both the literal and figurative categories in metaphors-bringing both sets of related properties (literal and figurative) to mind. The salient properties for metaphors and similes, however, were often the same properties, which suggests that the two expressions might not differ from each other with respect to the properties activated. Similarly, the result found for connotativeness ratings suggests that the vehicle in both expressions brought to mind properties that were equally connotative and denotative, which is inconsistent with the prediction that similes would elicit more literal properties than metaphors do (Glucksberg & Haught, 2006b). We did note specific examples, however, in which the simile does seem to activate more literal properties, as predicted by Glucksberg and Haught (2006b). For example, processors was the most frequent property for Minds are like computers,

but complicated was the most frequent property for Minds are computers; and although red was a salient property for Love is *like a heart*, this property was less salient (<5) than for *Love is* a heart. We also observed cases in which properties that were less frequent for the simile expression showed more saliency and frequency for the metaphor expression. For example, powerful was not very frequent for God is like fire, but it was the most frequent property listed for God is fire. Thus, although there is great feature overlap between metaphors and similes, consistent with the career-of-metaphor theory (Bowdle & Gentner, 2005), something extra does appear to occur for the metaphors: Metaphors seem to increase the level of activation for properties that have a lower saliency value for similes. In other words, reading a metaphor rather than a simile seems to make certain properties more salient. Ultimately, these differences and commonalities in the interpretations of metaphors and similes can be determined by experimental studies employing a set of materials based on meticulous semantic norms.

Potential applications

Insults are daggers because they hurt, but a job is a jail because it is *boring*. The present norms allow one to marvel at the wide range of different properties that metaphors and similes bring to mind, but perhaps more importantly, they also allow researchers to investigate with greater confidence how various metaphor and simile expressions are interpreted. The lack of interpretation norms has compelled studies of metaphor and simile interpretation to collect their own norming data for the preparation of experimental materials (e.g., Amanzio, Geminiani, Leotta, & Cappa, 2008; Chiappe, Kennedy, & Smykowski, 2003). However, larger-scale norms have many advantages over smaller-scale norming data, including greater standardization of materials, allowing for replication and for comparisons between studies and across techniques. In addition, the present norms facilitate the selection of stimuli relying on several variables of interest. More specifically, the properties that we collected and their derived measures (saliency, connotativeness, and conventionality) are required for the manipulation of a wide set of variables and experimental protocols at the forefront of the current research on copular metaphors and similes. Thus, the present norms provide other scientists with a shortcut to the preparation of materials designed to investigate numerous variables bearing on the nature of the interpretation of metaphors and similes.

As an example of how the present norms can serve contemporary theoretical and empirical research, consider studies showing that metaphors such as *Lawyers are sharks* are less reversible than corresponding similes such as *Lawyers are like sharks* (Chiappe, Kennedy, & Smykowski, 2003). This distinction primarily reflects the grammatical structure of the expressions, with metaphors being understood as categorizations or predications, and similes as comparisons. This distinction is also dependent, however, on the salient property being attributed to the expression and on whether this salient property is shifted when reversals occur. For example, people generally prefer Cherries are like olives to Olives are like cherries, because the salient property has pits is more salient for olives than for cherries. The collected norms allow us to discriminate between those expressions whose topics and vehicles bring similar properties to mind and those whose salient properties are different. In principle, the more similar are the property lists, the more one should be able to predict those expressions that would be equally comprehensible when reversed, while expressing the same general idea. In contrast, when reversals bring to mind a property salient for the vehicle that is an anomalous predicate of the topic, comprehension should be greatly affected. Thus, the present norms can contribute to solving long-standing debates on differences between metaphor and similes with regard to their reversibility properties (e.g., Ortony, 1979; Tversky, 1977).

Regarding the structure of the items themselves, we see many advantages in the use of the simpler copular forms, in comparison to those used in other metaphor norming studies (Cardillo et al., 2010; Katz et al., 1988). For current controlled psycholinguistic studies of metaphor, there is an increasing need for expressions for which the variable being examined is isolated to a single item, such as the vehicle. In fMRI studies of metaphors, for instance, the BOLD signal is usually obtained relative to a single word or to a short constituent (e.g., Bambini, Gentili, Ricciardi, Bertinetto, & Pietrini, 2011; Prat, Mason, & Just, 2012), whereas other paradigms are typically focused on when a given word is read or heard-such as in techniques involving priming (e.g., Blasko & Connine, 1993), event-related potentials (Pynte, Besson, Robichon, & Poli, 1996), or evetracking and self-paced reading (e.g., Roncero & de Almeida, 2014). In the present investigation, for example, it was important for us to keep the topic and vehicle constituents as simple as possible, to allow us to collect properties specifically related to these constituents. This also allows for the use of the present materials in the investigation of metaphor comprehension in special populations, such as Alzheimer's patients. Longer and more complex sentencesliteral or figurative-can be difficult for patients to comprehend and can confound whether patients retain the ability to comprehend figurative language. In addition, shorter constituents also allow researchers to embed metaphors or similes in larger sentential contexts or to investigate how these expressions are produced in different contexts (see, e.g., Roncero et al., 2006). Thus, the simple topic-vehicle constituents used in the present study keep constant the grammatical properties of the expressions, while also being ideal for a wide range of experimental methods. In addition, all of our norms include similes, which can be used experimentally as controls for

metaphors to investigate a variety of hypotheses on the semantic similarities and differences between two expressions. Creating corresponding literal statements for our metaphors would also be relatively simple, because the vehicles often make reference to a particular category (e.g., *shark*, *snakes*, *drug*, or *candy*) that can be used to generate literal associates. Consequently, little work would be needed to develop studies that compare a subset of our items to a corresponding set of literal expressions, while controlling for widely available lexical norms (such as corpus frequency, semantic associates, imagery, and other variables; see, e.g., Balota et al. 2007; Coltheart, 1981; Davies, 2010).

The types of expressions within the scope of the present study certainly do not exhaust all forms of metaphor use. Our goal was to develop materials related to the form that is most commonly used in current psycholinguistic studies. Therefore, we focused exclusively on nominal or copular metaphors, rather than, for instance, on what have been called cognitive or conceptual metaphors (e.g., love being a journey, which, by hypothesis, gives rise to expressions such as Our love hit a dead-end street; Lakoff, 1993). Also, the present set of metaphors can be distinguished from the verbal and literary types of metaphors collected by Katz et al. (1988) and Cardillo et al. (2010). The numerous measures that we obtained allow for various contrasts crucial for understanding fine-grained aspects of metaphor and simile interpretation: familiar versus unfamiliar; apt versus inept; metaphors that bring numerous properties to mind (high diversity) versus those that bring only a single salient property (low diversity); and expressions whose interpretations are relatively figurative (high connotativeness) versus those drawn on more literal properties (low connotativeness). Also, because these variables are combined with the associated property lists, the study of how these variables affect metaphor and simile interpretations can be examined with numerous online and offline experimental paradigms.

Conclusion

The present norms represent a unique resource for scientists interested in the interpretation of metaphors and similes, with a focus on copular structures. Crucially, the materials are designed for the investigation of key variables known to affect metaphor and simile processing, with particular attention to the theoretical predictions made by contemporary theories of copular metaphor interpretation.

Author note This study was supported by graduate fellowships from the Social Sciences and Humanities Research Council of Canada (SSHR C) and by the Alzheimer's Society of Canada to C.R., and by a grant from SSHRC to R.G.d.A. We thank Effie Andreadakis for organizing and

tabulating the associated property lists, and are grateful to Gregory Francis and two anonymous reviewers for suggestions on an earlier version of this article.

References

- Amanzio, M., Geminiani, G., Leotta, D., & Cappa, S. (2008). Metaphor comprehension in Alzheimer's Disease: Novelty matters. *Brain and Language*, 107, 1–10.
- Asendorpf, J. B., Conner, M., De Fruyt, F., Houwer, J. D., Denissen, J. J. A., Fiedler, K., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27, 108–119.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., & Treiman, R. (2007). The english lexicon project. *Behavior Research Methods*, 39, 445–459. doi:10.3758/ BF03193014
- Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M., & Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin*, 86, 203–216. doi:10.1016/j.brainresbull.2011.07.015
- Black, M. (1955). Metaphor. Proceedings of the Aristotelian Society, 55, 273–294.
- Black, M. (1979). More about metaphor. In A. Ortony (Ed.), *Metaphor* and thought (pp. 19–43). Cambridge, UK: Cambridge University Press.
- Blasko, D. G., & Briihl, D. S. (1997). Reading and recall of metaphorical sentences: Effects of familiarity and context. *Metaphor and Symbol*, 12, 261–285.
- Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experiment Psychology: Learning, Memory, and Cognition, 19,* 295–308. doi: 10.1037/0278-7393.19.2.295
- Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, 112, 193–216. doi:10.1037/0033-295X.112. 1.193
- Campbell, J. D., & Katz, A. N. (2006). On reversing the topics and vehicles of metaphor. *Metaphor and Symbol*, 21, 1–22.
- 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, 651–664. doi:10.3758/BRM. 42.3.651
- Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *NeuroImage*, 59, 3212–3221.
- Chettih, S., Durgin, F. H., & Grodner, D. (2012). Mixing metaphors in the cerebral hemispheres: What happens when careers collide? *Journal* of Experimental Psychology: Learning, Memory, and Cognition, 38, 295–311.
- Chiappe, D. L., & Kennedy, J. M. (1999). Aptness predicts preference for metaphors or similes, as well as recall bias. *Psychonomic Bulletin & Review*, 6, 668–676.
- Chiappe, D. L., & Kennedy, J. M. (2000). Are metaphors elliptical similes? *Journal of Psycholinguistic Research*, 29, 371–398.
- Chiappe, D. L., & Kennedy, J. M. (2001). Literal bases for metaphor and simile. *Metaphor and Symbol*, 16, 249–276.
- Chiappe, D. L., Kennedy, J. M., & Chiappe, P. (2003a). Aptness is more important than comprehensibility in preference for metaphors and similes. *Poetics*, 31, 51–68.
- Chiappe, D. L., Kennedy, J. M., & Smykowski, T. (2003b). Reversibility, aptness, and the conventionality of metaphors and similes. *Metaphor and Symbol*, 18, 85–105.

- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33A, 497–505. doi:10.1080/14640748108400805
- Coulson, S., & van Petten, C. (2002). Conceptual integration and metaphor: An event-related brain potential study. *Memory & Cognition*, 30, 958–968.
- Danesi, M. (1998). Sign, thought, and culture. Toronto, ON, Canada: Canadian Scholars' Press.
- Davies, M. (2010). The corpus of contemporary American rnglish as the first reliable monitor corpus of english. *Literary & Linguistic Computing*, 25, 447–464. doi:10.1093/llc/fqq018
- Gentner, D., & Bowdle, B. (2008). Metaphor as structure-mapping. In R. W. Gibbs Jr. (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 109–128). New York, NY: Cambridge University Press.
- Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. Journal of Memory and Language, 37, 331–355.
- Gerrig, R. J., & Healy, A. F. (1983). Dual processes in metaphor understanding: Comprehension and appreciation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 667–675. doi:10.1037/0278-7393.9.4.667
- Gibbs, R. W., Jr. (Ed.). (2008). *The Cambridge handbook of metaphor and thought*. Cambridge, UK: Cambridge University Press.
- Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, 7, 183–206.
- Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, 7, 92–96.
- Glucksberg, S. (2008). How metaphors create categories—quickly. In R. W. Gibbs Jr. (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 109–128). New York, NY: Cambridge University Press.
- Glucksberg, S., & Haught, C. (2006a). Can Florida become *like* the new Florida? When metaphoric comparisons fail. *Psychological Science*, 17, 935–938.
- Glucksberg, S., & Haught, C. (2006b). On the relation between metaphor and simile: When comparison fails. *Mind and Language*, 21, 360– 378.
- Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, 97, 3–18. doi: 10.1037/0033-295X.97.1.3
- Glucksberg, S., & Keysar, B. (1993). How metaphors work. In A. Ortony (Ed.), *Metaphor and thought* (2nd ed., pp. 401–424). Cambridge, UK: Cambridge University Press.
- Glucksberg, S., & McGlone, M. S. (1999). When love is not a journey: What metaphors mean. *Journal of Pragmatics*, *31*, 1541–1558.
- Graesser, A. C. (1981). Prose comprehension beyond the word. New York, NY: Springer.
- Harris, R. J., Friel, B. M., & Mickelson, N. R. (2006). Attribution of discourse goals for using concrete- and abstract-tenor metaphors and similes with or without discourse context. *Journal of Pragmatics*, 38, 863–879.
- Hasson, U., Estes, Z., & Glucksberg, S. (2001). Metaphors communicate more effectively than do similes. *Abstracts of the Psychonomic Society, 42nd Annual Meeting* (p. 103). Austin, TX: Psychonomic Society.
- Jones, L. L., & Estes, Z. (2005). Metaphor comprehension as attributive categorization. *Journal of Memory and Language*, 53, 110–124.
- Jones, L. L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal* of Memory and Language, 55, 18–23.
- Katz, A. N., Paivio, A., Marschark, M., & Clark, J. M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor and Symbolic Activity*, 3, 191–214.

- Kazmerski, V., Blasko, D., & Dessalegn, B. (2003). ERP and behavioral evidence of individual differences in metaphor comprehension. *Memory & Cognition*, 31, 673–689.
- Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (2nd ed., pp. 202–251). Cambridge, UK: Cambridge University Press.
- Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, 1284, 145– 155.
- McGlone, M. S., & Manfredi, D. (2001). Topic–vehicle interaction in metaphor comprehension. *Memory & Cognition*, 29, 1209–1219.
- Miller, G. A. (1979). Images and models, similes and metaphors. In A. Ortony (Ed.), *Metaphor and thought* (pp. 202–250). New York, NY: Cambridge University Press.
- Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, 86, 161–180. doi:10.1037/0033-295X.86.3.161
- Ortony, A. (1993). The role of similarity in similes and metaphors. In A. Ortony (Ed.), *Metaphor and thought* (2nd ed., pp. 342–356). New York, NY: Cambridge University Press.
- Pierce, R. S., & Chiappe, D. L. (2008). The roles of aptness, conventionality, and working memory in the production of metaphors and similes. *Journal of Metaphor and Symbol*, 24, 1–19.
- Prat, C. S., Mason, R. A., & Just, M. A. (2012). An fMRI investigation of analogical mapping in metaphor comprehension: The influence of context and individual cognitive capacities on processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 282–294.
- Pynte, J., Besson, M., Robichon, F., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain* and Language, 55, 293–316.
- Richards, I. A. (1936). *The philosophy of rhetoric*. Oxford, UK: Oxford University Press.
- Roberts, R. M., & Kreuz, R. J. (1994). Why do people use figurative language? *Psychological Science*, 5, 159–163.
- Roncero, C., & de Almeida, R.G. (2014, July). Alike or different: Comparing the online processing of metaphors and similes. Paper to be presented at the 24th Annual Meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science (CSBBCS), Toronto, ON, Canada
- Roncero, C., Kennedy, J. M., & Smyth, R. (2006). Similes on the Internet have explanations. *Psychonomic Bulletin & Review*, 13, 74–77. doi: 10.3758/BF03193815
- Shannon, C. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379–423.
- Shibata, M., Toyomura, A., Motoyama, H., Itoh, H., Kawabata, Y., & Abe, J. (2012). Does simile comprehension differ from metaphor comprehension? A functional MRI study. *Brain and Language*, 121, 254–260.
- Tartter, V. C., Gomes, H., Dubrovsky, B., Molholm, S., & Stewart, R. V. (2002). Novel metaphors appear anomalous at least momentarily: Evidence from N400. *Brain and Language*, 80, 488–509.
- Thibodeau, P. H., & Durgin, F. H. (2011). Metaphor aptness and conventionality: A processing fluency account. *Metaphor and Symbol*, 26, 206–226.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352. doi:10.1037/0033-295X.84.4.327
- Utsumi, A. (2005). The role of feature emergence in metaphor appreciation. *Metaphor and Symbol*, 20, 151–172.
- Utsumi, A. (2007). Interpretive diversity explains metaphor-simile distinction. *Metaphor and Symbol*, *22*, 291–312.
- Xu, X. (2010). The relationship between content and the form of metaphorical statements. *Journal of Psycholinguistic Research*, 39, 165–178.