



Published in final edited form as:

*Behav Res Methods*. 2010 August ; 42(3): 651–664. doi:10.3758/BRM.42.3.651.

## Stimulus Design is an Obstacle Course: 560 Matched Literal and Metaphorical Sentences for Testing Neural Hypotheses about Metaphor

Eileen R. Cardillo, Gwenda L. Schmidt, Alex Kranjec, and Anjan Chatterjee

Center for Cognitive Neuroscience, University of Pennsylvania

### Abstract

Despite the ubiquity and importance of metaphor in thought and communication, its neural mediation remains elusive. We suggest that this uncertainty reflects, in part, stimuli that have not been designed with recent conceptual frameworks in mind or have been hampered by inadvertent differences between metaphoric and literal conditions. In this paper we begin addressing these shortcomings by developing a large, flexible, extensively normed, and theoretically motivated set of metaphoric and literal sentences. Based on the results of three norming studies, we provide 280 pairs of closely matched metaphoric and literal sentences that are characterized along 10 dimensions: length, frequency, concreteness, familiarity, naturalness, imageability, figurativeness, interpretability, valence, and valence judgment reaction time. In addition to allowing control of these potentially confounding lexical and sentential factors, these stimuli are designed to address questions about the role of novelty, metaphor type, and sensory-motor grounding in determining the neural basis of metaphor comprehension.

### Keywords

figurative language comprehension; nominal and predicate metaphors; verbs of sound and motion; norming study

## 1. Introduction

Metaphor is not just for poets. Over the last twenty years, cognitive scientists have recognized the pervasiveness of metaphor in ordinary language. Metaphor's universality, and perhaps more importantly, the systematic ways it is used, appear to reflect something important about the structure of the mind (Lakoff & Johnson, 1980; Gentner, 2003; Boroditsky, 2000). In order to appreciate this claim about the importance of metaphor, first we must move beyond traditional literary definitions. At the most prosaic level, metaphor refers simply to the ways that we can conceptualize one domain in terms of another superficially dissimilar domain. Often this involves the figurative use of a noun, as in "The data are a headache." In other cases the basis for the metaphor might be a verb phrase, as in "The scientists got over their disappointment." Both kinds of metaphors occur frequently in language, and are indicative of the ubiquity and heterogeneity of metaphor. Despite this prevalence of use, and consensus in the cognitive sciences about the importance of metaphoric thought and expression, the neural substrates of metaphor are surprisingly murky. We believe this uncertainty reflects, in part, the influence of uncontrolled properties of the stimuli used in past research. In this paper, we attempt to address

these shortcomings by developing stimuli specifically tailored to cognitive neuroscience methods and neural hypotheses about metaphor.

Although now largely discounted, early accounts of metaphor comprehension assumed the existence of distinct processes required to understand literal and figurative language (Grice, 1975; Searle, 1979), with presumably different neural bases for each. Initial investigations with brain-injured patients supported this conjecture, suggesting a critical role for the right hemisphere in the comprehension of metaphor (Brownell, et al, 1984, 1990; Winner & Gardner, 1977). The accumulated evidence is less clear than this now familiar story suggests, though, with early studies suffering from several methodological weaknesses (e.g. small numbers of items, anatomical distinctions lacking precision, task confounds; for review see Schmidt et al, 2010). More recent work with patients has not been as vulnerable to these limitations and, coupled with the advent of neuroimaging and TMS, has challenged this straightforward assignment of hemispheric duties. Rather, the left as well as the right hemisphere have now been implicated in the processing of metaphors, and some research suggests that the right-hemisphere may not be uniquely sensitive to figurativeness, *per se*. However, at this point no single brain area or discrete network has emerged as uniquely responsive to metaphoric language (Schmidt et al, 2010).

These inconsistencies likely reflect, in part, a failure to properly consider factors of non-interest that nonetheless tend to vary between metaphoric and literal conditions. In the review that follows we discuss the confounding influence of several psycholinguistic variables (frequency, concreteness, length, imageability, naturalness, and interpretability) that likely make the metaphorical sentences used in previous studies more difficult to process than the literal sentences. Other differences between studies may be attributed to more conceptual factors that have not yet been adequately addressed (Schmidt et al, 2010). Specifically, we argue for the theoretical importance of metaphor novelty and type when investigating the neural substrates for metaphor. Following this discussion, we present data from three norming tasks used to develop a set of metaphoric and literal stimuli that we believe to be well suited to address these methodological and theoretical issues in future empirical work on the neural basis of metaphor.

### 1.1 Difficulty

Previous comparisons between literal and metaphorical processing have likely been complicated by differences in processing difficulty related to uncontrolled lexical and sentential characteristics. Behavioral studies at the single word level consistently demonstrate that frequency, concreteness, and length affect the ease (i.e., speed) with which words are accessed (Balota, Yap, & Cortese, 2006). Neuroimaging studies also demonstrate the importance of these three factors in driving neural activity in single-word tasks. Low frequency words are read more slowly and also more strongly engage language-sensitive areas of the brain, including the left prefrontal cortex, middle and superior temporal gyri, and the so-called “visual word-form area” of the fusiform gyrus (for review, see Hauk et al, 2008). Although concreteness can be captured along a continuum with abstractness, unlike low frequency words, abstract words do not simply take longer to read and more strongly engage the same brain areas recruited by concrete words. Rather, concrete and abstract words show both shared and distinct neural substrates even after controlling for reaction time differences (Binder et al, 2005b; see also Fiebach & Friederici, 2003; Scott, 2004). As for length, longer words generally elicit slower responses in behavioral tasks (Balota et al, 2006). Such processing time differences are not insignificant; the BOLD signal in fMRI is especially sensitive to the time taken to perform a task, and activity in many regions of both hemispheres are independently modulated by reaction time in language tasks (e.g. Binder et al, 2005a, 2005b).

The impact of the frequency, concreteness, and length of words on neural activity when they are embedded in sentences or narratives has not generally been addressed. However, the

modulatory effects of these factors at the single word level strongly suggest their continued influence at higher levels of language processing, an inference supported by several recent studies (Keller et al, 2001; Prat et al, 2007; Yarkoni et al, 2008). Similarly, other features of sentences such as imageability (Just et al, 2004), syntactic complexity (Constable et al, 2004; Friederici et al, 2006; Just et al, 1996), and semantic plausibility (Cardillo et al, 2004; Kuperberg et al, 2003; Menenti et al, 2008) also alter neural demands in areas both within and beyond the classic language areas of the brain. The more syntactically complex and the less semantically plausible or predictable a sentence, the more time it requires to be processed and the more strongly it activates the left-hemisphere language network and its right-hemisphere homologues. With respect to imageability, signal intensity is positively correlated in some areas (e.g. the intraparietal sulcus) and negatively in others (e.g. the middle and superior temporal gyri).

These patterns indicate that qualities that make sentences more challenging to understand often make them slower to read and increase neural activation. Accordingly, these findings suggest two other sentence level properties – naturalness and interpretability - that are important to control, as both factors are also likely to affect reading times. Naturalness refers to the normality of an utterance, or the likelihood that a speaker might spontaneously express an idea in a particular manner. Literal sentences are likely to seem more natural than metaphors, especially if the metaphors are not very familiar, but naturalness is rarely directly addressed in neuroimaging studies. Interpretability refers to the ease with which a clear meaning can be derived from an expression. It is not operationalized in a consistent way across studies and, more problematic, is likely to be lower for metaphoric stimuli than for literal items.

Taken together, it appears that the BOLD signal in fMRI studies is sensitive to a range of lexical and sentential characteristics not typically controlled for in neuroimaging studies of metaphor, all of which influence processing time. Although time on task is only an indirect indicator of cognitive effort, it is a critical dimension to address. The extent to which previous reports of increased neural activity in certain regions associated with metaphors reflects processing difficulty rather than figurativeness remains unknown since with few exceptions (Bottini et al, 1994; Chen, Widick, & Chatterjee, 2008; Lee & Dapretto, 2006; Mashal et al, 2009; Rapp et al, 2004; Stringaris et al, 2007), condition differences in reaction time have either not been measured, controlled, or reported.

## 1.2 Novelty

Another critical factor contributing to the empirical muddle is how the processing of metaphors changes over time. Novel metaphoric uses of words may gain in popularity, with those same figurative senses that were once so creative becoming familiar and fairly unremarkable with increased use. Bowdle and Gentner (2005) provide behavioral evidence that this shift from novel to conventional usage is accompanied by a shift in how metaphors are understood, a process they describe as the “career of metaphor.” Although the Career of Metaphor account is motivated by behavioral data, the implication of this proposed shift in cognitive processing with increased familiarity (i.e. diminished novelty) is that it is paralleled by a shift in neural recruitment. Similarly, the hypothesis that the right hemisphere may be sensitive to novelty rather than metaphoricity, per se, has been recently proposed to clarify the neural evidence. Many neuroimaging and all patient studies to date consider conventional or idiomatic expressions. Those studies that consider novel metaphors are more likely to implicate the right hemisphere (Arzouan, Goldstein, & Faust, 2007; Bottini et al, 1994; Mashal et al, 2005, 2007; Pobric et al, 2008; Sotillo et al, 2005). Less promising for the novelty hypothesis, nearly as many neuroimaging studies manipulating novelty/familiarity have *not* found right hemisphere engagement during processing of novel metaphors as have found it (Rapp et al, 2004, 2007; Kircher et al, 2007; Mashal et al, 2009; Shibata et al, 2007).

One reason these studies with novel metaphors have differed in their findings may be related to variability in how their stimuli were normed on the previously mentioned lexical and sentential characteristics. A further possible explanation for the discrepant findings is that the right hemisphere is sensitive to the *salience* or prominence of a metaphorical interpretation, rather than the novelty or figurativeness of the utterance (Giora, 1999). Giora and colleagues (2000) have proposed that a literal interpretation may be more prominent (i.e. more quickly generated) than the figurative meaning of unfamiliar metaphors, and that the right hemisphere may play an important role in generating the novel, nonsalient interpretation (which frequently happens to also be the figurative one).

Alternatively, the inconsistencies may reflect differences related to generating metaphorical meaning at the level of words versus sentences (Mashal et al, 2009). For the most part, the studies indicating right hemisphere sensitivity to novelty have compared metaphorical word pairs of high and low conventionality (“bright student” versus “pearl tears”), whereas those failing to find right hemisphere sensitivity to novelty have used sentence stimuli. Mashal and colleagues suggest that while the right hemisphere may be biased to extract novel meanings entailed by two uncommonly paired words, the syntactic and semantic integration demands imposed by novel sentences may necessitate the recruitment of the language-dominant left hemisphere. Although no patient studies have considered novel metaphor comprehension, work by Kempler and colleagues is consistent with this suggestion (Van Lancker & Kempler, 1987; Kempler et al, 1999). In two studies, right hemisphere injured patients performed worse than left hemisphere injured patients on a sentence-picture matching task involving familiar, idiomatic expressions (“He’s turning over a new leaf”) – but outperformed left hemisphere injured patients when tested on novel, literal sentences (“He’s sitting deep in the bubbles”). Although these studies cannot disentangle the effects of novelty from figurativeness, they are consistent with the notion that the left, not the right, hemisphere is required for handling novel meaning at the sentence level.

A recent fMRI study considering novel sentence-level metaphors taken from poetry is also suggestive of this proposed difference in how the brain processes novel lexical and sentential metaphors (Mashal, et al, 2009), but thus far no neuroimaging study has directly compared lexical and sentential metaphors, nor has any study yet contrasted conventional and novel metaphors at the sentence level, or considered neural changes with increased experience with a particular metaphoric sense of a word. Thus, determining hemispheric contributions to the processing of novel metaphors requires sentences that vary in familiarity, but for which a metaphorical interpretation is still more salient than a literal one. Additionally, novelty may have different consequences for the processing of different types of metaphor.

### 1.3 Metaphor Type

Metaphor is not a unitary construct (Chen et al, 2008). As any conversation quickly reveals, even the least poetic of speakers is likely to use a variety of metaphoric expressions. These different types of metaphor reflect differences in the class of word that is being used figuratively (the base term). While the majority of metaphor research has addressed the metaphorical extension of nouns, or **nominal** metaphors (“The stock is a rollercoaster”), other parts of speech are also frequently used metaphorically. For example, we extend verbs metaphorically in **predicate** metaphors (“The stock soared”), prepositions in **locative** metaphors (“The stock is down”), and adjectives in **attributive** metaphors (“the hot stock”).

For two reasons, these differences in the nature of metaphor base terms are likely to be paralleled by different neural substrates. First, the cognitive demands required for the comprehension of their figurative use is likely to differ. Nominal metaphors, for instance, appear to be understood via a process of comparison (Gentner et al, 2001), categorization (Glucksberg, 2003), or some combination of the two (Bowdle & Gentner, 2005). In contrast,

predicate metaphors may be understood by a process of *abstraction* whereby the concrete, sensory-motor features of a verb are stripped away, retaining only a few core conceptual attributes during its metaphoric use (Chen et al, 2008; see also Torreano, Cacciari, & Glucksberg, 2005). Second, both fMRI and patient studies have demonstrated differences in the brain areas important for literal processing of these word classes. Noun processing is typically associated with inferior occipito-temporal cortex, verbs with postero-lateral temporal, prefrontal and motor cortex, and prepositions with parietal cortices (for reviews see Martin, Ungerleider, & Haxby, 2000; Chatterjee, 2008; and Kemmerer, 2006, respectively).

Importantly, these different neural signatures for nouns, verbs, and prepositions are very similar to those areas involved in the perception of objects (Ishai et al, 1999; Ishai et al, 2000), actions (Kable et al, 2005, 2006; Tranel, Manzel, Asp & Kemmerer, 2008), and spatial relations (Amorapanth, Widick, & Chatterjee, in press; Kosslyn et al, 1995), respectively. Thus, there is growing evidence that linguistic representations are stored either in the same, overlapping, or adjacent cortex as is responsible for the perception of their concrete referents (Pullvermuller, 2006; Simmons & Barsalou, 2003; Thompson-Schill, 2003). Such sensory-motor grounding for the literal senses of words suggests that the neural mediation of their metaphorical extensions may similarly recruit relevant sensory or motor regions (Gibbs, 2006). Or, it may be that the imperfect overlap between activations elicited by perceptual and linguistic stimuli indicates a neural principle of organization by which more abstract representations are shifted relative to their literal representations and “perceptual points of entry” (Wu, Waller, & Chatterjee, 2007). That is, representation of word meaning at a neural level may not be *identical* to the neural processes involved in perception of its concrete referent, but may be mediated instead by adjacent cortical areas. Three recent fMRI studies provide support for such an organization, all implicating the left postero-lateral middle temporal cortex, just anterior to motion-sensitive area V5, in the comprehension of metaphorical extensions of motion verbs (Chen et al, 2008; Wallentin et al, 2005a, 2005b).

However, it remains to be seen whether these posterior temporal activations reflect a sensitivity to figurative senses of verbs in general, or verbs of motion in particular. The former possibility would indicate different neural resources for processing nominal versus predicate metaphors, as would be expected given their hypothesized differences in cognitive demands. The latter possibility would suggest that the perceptual qualities of the literal sense of a base term, rather than its grammatical class or the type of metaphor it is embedded within, determine the neural substrates for its metaphorical extensions. If supported, this latter possibility would also suggest that heterogeneity in the sensory-motor features associated with base terms may have contributed to observed inconsistencies in previous neuroimaging studies of metaphor. It is also of course possible that both the form of the metaphor and the sensory-motor qualities of the base term interact to determine neural engagement. Thus, to distinguish the relative importance of sensory-motor grounding and metaphor type for dictating the neural basis of metaphor it is necessary to disentangle sensory-motor qualities of base terms from the syntactic environments in which they occur (e.g., nominal versus predicate metaphors). It is also important to control or intentionally manipulate metaphor familiarity in such designs as readers may be more likely to draw upon sensory-motor knowledge when interpreting novel expressions (cf. Aziz-Zadeh & Damasio, 2008).

In sum, the neural basis of metaphor remains elusive but when coupled with rigorous stimulus control, recent conceptual frameworks motivate several potential ways to clarify the observed inconsistencies. At this point, it remains uncertain whether figurativeness, grammatical class, or sensory-motor properties are the most important determinants of neural processing of words – nor how the familiarity of an expression might alter such neural engagement. We suggest that interactions between these factors are possible, probable, and worth exploring. A necessary first step for testing these neural hypotheses, and for distinguishing brain areas critical for

metaphor comprehension from those more strongly recruited for any more difficult text, is the development of a large set of stimuli that is characterized along the dimensions reviewed above (frequency, concreteness, length, interpretability, familiarity, naturalness, imageability, figurativeness, processing time, novelty, metaphor type, sensory-motor qualities of base terms). If sufficiently large, such a set would enable selection of literal and metaphoric sentences such that nuisance variables can be controlled or covaried out while factors of theoretical interest (e.g. novelty, metaphor type, sensory-motor grounding) can be experimentally manipulated in a categorical or a parametric fashion. The goal of the present paper is to develop such a well-characterized set of literal and metaphoric stimuli and to suggest how it can be sampled to test relevant questions.

## 2. Methods

### 2.1 Construction of Stimuli

An initial pool of 628 sentences of two syntactic forms, predicate or nominal, was generated (for examples see Table 1). Predicate sentences ( $n = 316$ ) consisted of a noun phrase and an action verb followed by a prepositional phrase. Nominal sentences ( $n = 312$ ) consisted of two noun phrases linked by a copula (i.e., “An X is a Y”). Predicate sentences had zero or one adjective and nominal sentences had up to two. Half of each set expressed a literal meaning and half expressed a metaphorical meaning. In the predicate sentences the verb was the base term to be used figuratively in the metaphors; in the nominal sentences the second noun was the base term in the metaphors.

To generate predicate items, 79 verbs of visual motion and 79 verbs of sound were first selected as base terms. Next, for each verb both a literal and metaphorical sentence was created, resulting in 158 literal-metaphor sentence pairs. In this way, in each pair the identical verb implied a literal or a figurative interpretation depending on its context. To generate nominal items, 78 nouns with salient motion qualities and 78 nouns with salient sound qualities were selected as base terms. Next, for each noun both a literal and metaphorical sentence was created, resulting in 156 literal-metaphor sentence pairs. In this way, in each pair the identical noun implied a literal or a figurative interpretation depending on the noun phrase and adjectives with which it was paired. Critically, in both predicate and nominal sets, all metaphors involving auditory base terms were designed such that no sound was implied by the figurative interpretation of the sentence. Likewise, all metaphors involving motion base terms were designed such that no physical or fictive motion was implied by the figurative interpretation of the sentence.

To make nominal items maximally comparable to predicate items, base terms in nominal sentences were always nominalized verbs. In this way, although the syntax differed between predicate and nominal items, the word being used metaphorically was very similar across sentence types, ensuring the strictest test that predicate and nominal metaphors entail different cognitive processes. In order to provide a final stimulus set with the option of further maximizing similarity between predicate and nominal sentences, as many of the verbs used in the predicate sentences as possible were identical in form to the nominalized verbs used as base terms in the nominal sentences (for practical purposes, this amounted to roughly half ( $n=86$ ) of base terms). A small number of predicate metaphors ( $n = 13$ ) were modifications of those used in two previous studies (Chen et al, 2008; Torreano et al, 2005); the rest were created by the authors.

### 2.2 Overview of Norming Studies

This initial pool of sentences was normed both offline and online, and at both the word and sentence level, in order to characterize its psycholinguistic properties and to highlight any problematic items (Figure 1). Before norming, three measures of length (number of characters,

number of words, and number of content words) were calculated for each sentence, as well as an average frequency and concreteness score based on values for the content words of the sentence (i.e., nouns, verbs, and adjectives). In light of recent evidence that the most commonly used frequency values (Kucera & Francis, 1967) may be outdated at this point (Brysbaert & New, 2009), frequency values were calculated using both this traditional measure as well as values from SUBTLEXus, a more recent and larger corpus (Brysbaert & New, 2009). Concreteness values were taken from the MRC Psycholinguistic Database (Coltheart, 1981) and the University of South Florida Norms (Nelson et al, 1998). For those words for which concreteness ratings were not found in either of these databases, we collected our own (Norming Study 1). Participants in this task also judged the strength of auditory and visual imagery associated with all the base terms in order to ensure a valid manipulation of sensory modality. A different set of individuals normed the stimuli at the sentence level, interpreting them as well as rating them in terms of familiarity, naturalness, imageability, and figurativeness (Norming Study 2). Additionally, given the sensitivity of fMRI to reaction time, a valence judgment task was administered to a third group of individuals in order to generate an online measure of comprehension for each item (Norming Study 3)<sup>1</sup>. Together, the results of the norming studies indicated weak items to be discarded, ultimately dictating a final set of 280 metaphor-literal sentence pairs characterized on 12 lexical and sentential properties.

### 2.3 Norming Study 1: Words

**2.3.1 Participants**—Forty participants were recruited from the University of Pennsylvania community in compliance with the procedures established by the university's Institutional Review Board and were compensated \$15 or given course credit for their participation. All participants were proficient English speakers (all learned English from birth, except one that reported learning before age 5). Twenty participants made judgments on words from the predicate sentences (mean age = 28.6 years, SD = 9.2; males = 5; mean education = 17.7 years, SD = 2.4) and 20 participants made judgments on words from the nominal sentences (mean age = 29.6 years, SD = 10.5; males = 7; mean education = 18.4 years, SD = 3.4).

**2.3.2 Stimuli**—The initial pool of sentences contained 1860 content words, 339 of which lacked published values and thus required norming. 155 of these words to be normed came from predicate sentences and 184 came from nominal sentences (Appendix A). For auditory and visual imagery associated with base terms, all 156 verbs used in the predicate sentences were rated. Since many of these verbs were identical in form to the nominalized verbs used as base terms in the nominal sentences, only 70 of the base terms used in nominal sentences required ratings (Appendix B).

**2.3.3 Task**—For both predicate and nominal items, an Excel workbook was generated with separate worksheets corresponding to the three rating tasks (concreteness, auditory imagery, visual imagery), and one line per worksheet corresponding to each item. For concreteness, participants were instructed to rate words in terms of their accessibility to one or more of the senses using a scale from 1 (*very abstract*) to 7 (*very concrete*). For auditory imagery, participants were instructed to rate words in terms of the speed and “ease or difficulty with which they arouse a particular sound” using a scale from 1 (*no sound*) to 5 (*clear sound*). For visual imagery, participants were instructed to rate words in terms of the speed and “ease or difficulty with which they arouse a mental picture or visual image” using a scale from 1 (*no image*) to 5 (*clear image*). In all cases, instructions were coupled with several examples and

<sup>1</sup>This particular task was chosen not only because it has been used in previous metaphor research but because the other most commonly used task - a plausibility judgment - may not be well-suited to studies involving novel metaphors. The inclusion of implausible sentences risks biasing readers away from making metaphoric interpretations of unfamiliar, but interpretable, metaphors. Additionally, a valence judgment task enables controlling for differences in verbal processing associated with the affective content of stimuli (cf. Kutas, 2006).

explanations<sup>2</sup>. Participants worked at their own pace and made these judgments as part of a larger word norming task. The task required approximately 40 minutes to complete.

**2.3.4 Data Analysis**—For all words, ratings were averaged over the 20 participants for each of the three judgments. The 339 new concreteness values supplemented the previously published values for the other 1521 content words in the stimulus set. These individual concreteness ratings were then used to determine an average concreteness value for each of the 628 candidate sentences (i.e., the sum of the concreteness values associated with each content word in any particular sentence divided by the number of content words in that sentence). The imagery ratings of base terms indicated six problematic base terms: four base terms used in the auditory conditions (*blubber*, *serenade*, *splash*, and *yawn*) elicited stronger visual than auditory imagery and two used in the motion conditions (*wind* and *stomp*) elicited stronger auditory than visual imagery.

## 2.4 Norming Study 2: Sentences

**2.4.1 Participants**—Eighty participants were recruited from the University of Pennsylvania community in compliance with the procedures established by the university's Institutional Review Board and were compensated \$25 or given course credit for their participation. All participants were proficient English speakers (all learned English from birth, except again one person that reported learning before age 5), and none had participated in Norming Task 1. Forty participants made judgments on predicate sentences (mean age = 20.8 years, SD = 2.5; males = 12; mean education = 14.7 years; SD = 1.6) and 40 participants made judgments on nominal sentences (mean age = 20.4 years, SD = 2.9; males = 11; mean education = 14.6 years; SD = 1.8).

**2.4.2 Stimuli**—All 628 candidate sentences were assessed.

**2.4.3 Task**—The sets of predicate (n = 316) and nominal (n = 312) items were randomly divided in half, and for each of these subsets an Excel workbook was generated with separate worksheets corresponding to the five norming tasks (familiarity, naturalness, imageability, figurativeness, interpretation), and one line per worksheet corresponding to each item. In this way, participants in the predicate condition saw one of two possible lists of predicate stimuli (both literal and metaphorical) and participants in the nominal condition saw one of two possible lists of nominal stimuli (both literal and metaphorical), with ratings for each sentence based on the responses of 20 individuals.

For the Familiarity task, participants were instructed to rate their frequency of experience with the sentence and its meaning, using a scale from 1 (*very unfamiliar*) to 7 (*very familiar*). For the Naturalness task, participants were instructed to rate each sentence for how “natural and normal” it seemed, using a scale from 1 (*very unnatural*) to 7 (*very natural*). For the Imageability task, participants were instructed to rate “how quickly and easily each sentence brings a visual image to mind”, using a scale from 1 (*no image*) to 7 (*clear, immediate image*). For the Figurativeness task, participants were instructed to rate how literal an interpretation each sentence suggested using a scale from 1 (*very literal*) to 7 (*very figurative*). For the Interpretation task, participants were instructed to write the meaning of each sentence using their own words. Familiarity, Naturalness, Imageability, and Figurativeness ratings were collected for both literal and metaphorical sentences. Given the difficulty of restating a concrete, literal sentence in novel words and the absence of any

<sup>2</sup>Instructions were slightly modified directions from Paivio and colleagues (1968), and also very similar to those used in the two other major sources of concreteness and imageability norms in the MRC Psycholinguistic Database (i.e., Toglia & Battig (1978) and Gilhooly & Logie (1980)).



theoretical relevance for such descriptions, interpretations were only collected for the metaphors. The task required approximately 90 minutes to complete.

**2.4.4 Data Analysis**—To determine Interpretability, several steps were necessary for each item and for each subject. First, for each metaphor, three of the authors independently judged the number of interpretations that reflected a plausible, figurative construal of the sentence. For some sentences, all interpretations reflected a single meaning; for many others, responses indicated multiple or overlapping meanings. For instance, despite being rated as fairly familiar, the metaphor “The day's events were a whirl” received various responses. Some interpreted the metaphorical sense of *whirl* to mean the day was busy or passed quickly (e.g. “There were a lot of events that happened in the day and they went quickly”) while others interpreted the base term to mean the day's events were hard to remember (e.g. “ Looking back, what happened today is fuzzy and unclear”). Still others interpreted the sentence to mean both (e.g. “The day's events went by so quickly, they seemed blurred in retrospect”). This variability in response is unlikely to indicate that the item is difficult to understand; without context, all of these interpretations are reasonable. Both for this reason, and given the subjectivity in determining where one meaning ends and another begins, rather than tally the incidence of the most common interpretation, any plausible, figurative interpretation was taken to indicate that the metaphor had been understood. In contrast, blank, nonsensical, literal, or uninformative (e.g. “Just what it says”) interpretations were not taken to indicate metaphoric comprehension.

All three judges were in full agreement (all three agreed an interpretation was plausible or all three agreed an interpretation was not plausible) on 91.0% of the responses across the four lists (range = 90.2 – 92.0%). Similarly, at least two out of three judges agreed an interpretation was plausible for 91.4% of responses across the four lists (range = 89.0 – 94.2%).

Lastly, an interpretability score for each subject was calculated by dividing the number of their interpretations that were deemed plausible by at least two judges by the total number of items assessed by that subject (# plausible interpretations/all possible interpretations). This assessment revealed poor comprehension by four participants (> 30% of their interpretations were not considered plausible), so their data were excluded from subsequent analyses. To generate Familiarity, Naturalness, Imageability, and Figurativeness ratings for each item, averages were derived from the responses made by the remaining participants. To generate an interpretability score for each item, the number of interpretations deemed plausible by at least two judges was divided by the total number of interpretations for that item (# plausible interpretations/all possible interpretations). Results indicated 18 metaphors that failed to reach our minimum desired comprehensibility criteria of 70% plausible interpretations.

## 2.5 Norming Study 3: Online Comprehension

**2.5.1 Participants**—Forty participants were recruited from the University of Pennsylvania community in compliance with the procedures established by the university's Institutional Review Board and were compensated \$15 or given course credit for their participation. All participants were native English speakers. Twenty participants (mean age = 23.9 years, SD = 3.4; males = 7) made judgments on predicate sentences and 20 participants (mean age = 19.1 years, SD = 1.1; males = 13) made judgments on nominal sentences. None had participated in Norming Tasks 1 or 2.

**2.5.2 Stimuli**—In the predicate condition, all 158 candidate predicate metaphors and their literal counterparts were assessed. In the nominal condition, all 156 candidate nominal metaphors and their literal counterparts were assessed.

**2.5.3 Task**—Sentences were presented centrally in black 18-point font on a white background, using E-prime 1.1 software on a Dell Inspiron laptop. Sentences were displayed for 3000ms and separated by a 1000 ms ITI. Participants were instructed to read each sentence and then judge its emotional valence, using the ‘f’ key to indicate a positive valence and the ‘j’ key to indicate a neutral or negative valence. They were informed that there was no right or wrong answer, and were encouraged to respond as quickly as possible. Twelve practice trials preceded four blocks of experimental trials. Each subject received a different random order of items and saw each sentence only once. The task required approximately 20 minutes to complete.

**2.5.4 Data Analysis**—For every sentence, response times were averaged across participants and the proportion of positive valence judgments was calculated. Due to computer errors, these values were not available for 18 metaphor-literal pairs in the predicate condition.

### 3. Results

Results of the norming studies were used to eliminate problematic stimuli from the initial pool. Imagery ratings from Norming Study 1 and interpretability scores from Norming Study 2 indicated 24 metaphor-literal sentence pairs to be discarded, resulting in 73, 75, 71, and 70 remaining possible sentence pairs in the Predicate-Auditory (PA), Predicate-Motion (PM), Nominal-Auditory (NA), and Nominal-Motion (NM) conditions, respectively. In order to generate a final stimulus set with equal numbers of items in each condition (i.e., 70 metaphors and 70 matched literal sentences), items in the PA, PM, and NA conditions with the lowest interpretability values were also discarded. In this final set, the overlap in base terms between predicate and nominal sentences remained roughly half (67 out of 140). The lexical and sentential characteristics of the final set of 560 sentences are summarized in Table 2 and an example set of items can be found in Appendix C (see the supplemental materials for the full set of items and their normative values).

We intentionally did not calculate statistical differences between sentence types because our aim is not for the stimulus set to be used in its entirety, but rather, for it to be sampled in ways that control for condition differences or for the normative data to be used to covary out the influence of nuisance variables. Nonetheless, overall means indicate some areas where such control is likely to be necessary. Although literal and metaphorical sentences had comparable length, frequency, and concreteness values, unsurprisingly, literal sentences were judged to be less figurative and more familiar, natural, and imageable than metaphorical sentences in both predicate and nominal sets. Also, despite the inclusion of additional adjectives, nominal sentences tended to be shorter than predicate sentences on all length measures. Importantly, our online comprehension measure suggested no major differences in the time required to make semantic judgments about the different classes of metaphorical and literal sentences and interpretability was consistently very high across all metaphor types. The only difference to emerge at this group level regarded predicate metaphors with verbs of sound, which were overall judged to be more figurative and less familiar, natural, and imageable than the other three metaphor types. As intended, base terms consisting of verbs of motion or nominalized verbs of motion were rated as having more salient visual imagery and less salient auditory imagery than base terms consisting of verbs of sound or nominalized verbs of sound (and vice versa).

To further explore relationships between sentence-level factors of theoretical interest, the five normative values collected for each sentence (familiarity, naturalness, imageability, figurativeness, and interpretability) were correlated with each other separately for the predicate and nominal metaphors (Table 3). Results indicated several expected relationships. In both predicate and nominal metaphors, familiarity and naturalness were highly correlated, indicating that these constructs are either conceptually indistinguishable or at least practically

difficult to disentangle. We suggest then that there is little utility for future researchers to norm on both measures. Both metaphor sets also indicated that sentences rated higher in familiarity and naturalness tend to evoke greater visual imagery, are perceived as less figurative, and are more easily understood. Yet, these patterns are not sufficiently strong that they cannot be orthogonalized with careful item selection. Most important with respect to future use, it is possible with these stimuli to disentangle ease of comprehension from metaphoricity. Despite the relative novelty of the metaphors in this stimulus set, in neither predicate nor nominal metaphor sets were correlations between interpretability and figurativeness significant.

By and large, the metaphor sets showed very similar relationships between the sentence-level factors. The only clear difference to emerge concerned imageability. Separate correlational analyses by modality and metaphor type (Table 4) indicated that this difference between metaphor types related to the modality of their base terms. No relationship between figurativeness and imageability was observed in nominal-auditory, nominal-motion, or predicate-auditory items. In contrast, for predicate metaphors with motion base terms, the more figurative a metaphor was judged, the less it evoked a visual image.

#### 4. Discussion

We have developed a large set of metaphoric and literal sentences that, while certainly unable to address all questions of interest, have several advantages over existing sources of stimuli that might be used in cognitive neuroscience studies of metaphor. To our knowledge, this is one of the largest available stimulus sets (see also Katz et al, 1988; Torreano et al, 2005) and it is characterized on more dimensions than in most extant sets. In addition, we have included more than one type of metaphor, both in terms of syntactic form and sensory modality, and our metaphors are relatively novel. Our metaphors also have higher interpretability than is generally reported, especially for novel metaphors, and are matched on an item-by-item basis with a sentence using the figurative base term in a literal sense. With the exception of interpretability, these literal sentences have been normed on all the same dimensions as the metaphors, making them optimal control items in cognitive neuroscience studies of metaphor (as well as suitable sentences for studies of literal language processing).

Most importantly, to our knowledge, this is the only set of stimuli specifically designed to address hypotheses relevant to cognitive neuroscience. If widely adopted, their use could facilitate cross-study comparison and, critically, avoid potential lexical and sentential confounds that have made integrating results of previous studies difficult. Despite some of the correlations observed, the stimulus set allows disentangling typically correlated characteristics by selectively sampling it in either of two ways: 1) splitting by a dimension of interest while balancing on nuisance variables or using them as covariates, or 2) parametrically varying a dimension of interest while balancing on nuisance variables or using them as covariates. For instance, the first approach could be used in order to test proposed differences in the cognitive processing entailed by different metaphor types. To do so, one could consider only items of a single base term modality (motion or sound), divide that modality by metaphor type (nominal versus predicate), and then selectively remove items in such a way that nominal and predicate metaphors, and their literal counterparts, closely match on lexical and sentential properties. This sorting approach could also be used to test whether the sensory-motor features of the base term determine the neural substrates of its literal and metaphoric senses. In this case, one could consider only metaphors of a single type, divide them instead by modality, then selectively remove items in such a way that auditory and motion metaphors, and their literal counterparts, closely match on lexical and sentential properties. Careful sampling of the stimuli also enables the parametric manipulation of novelty within either of the previous examples of categorical sampling. Alternatively, novelty could be addressed by restricting initial selection of items to items of high or low familiarity, or by using familiarity ratings as a covariate of non-interest.

By combining sampling approaches in this way, interactions between novelty and metaphor type or modality can be considered.

These are possibilities we are currently pursuing in our lab, but we can imagine other ways in which these stimuli could be extended, either by the addition of items or by behavioral norming with different tasks. For example, the notion of novelty requires unpacking. The novelty of these stimuli is currently described in terms of their familiarity at the level of the whole sentence, but they could also be characterized by several other closely related concepts – conventionality, aptness, and salience. Although often used interchangeably with *familiarity*, more precisely, *conventionality* refers to how strongly a figurative meaning is associated with a specific base term (Gentner & Wolff, 1997). *Aptness* loosely refers to the “goodness” of a metaphor, or more specifically, the degree to which the base expresses important features of the target (Chiappe, Kennedy, & Chiappe, 2003). Recent behavioral studies demonstrate that aptness may be driving effects otherwise attributed to conventionality (Jones & Estes, 2006; Chiappe et al, 2003). *Salience* is a composite construct referring to an expression's most prominent meaning, as determined by familiarity, conventionality, context, and frequency (Giora, 1999). As familiarity, conventionality, aptness, and salience are often highly correlated dimensions, the degree to which they reflect cognitively and neurally relevant distinctions remains an open question. Although conventionality and familiarity are constructs easily applied to other classes of metaphor, as it is strictly defined, aptness is less easily extended to predicate metaphors. This stimulus set presents a challenge as well as an opportunity to develop the conceptualization of novelty and a precise model for how these factors conspire to affect comprehension.

More pressing, neuroimaging experiments cannot distinguish areas *necessary* for successful performance in a metaphor task from those areas that are simply *involved* in the task; to make such inferences fMRI and PET research is best complemented by studies involving brain-injured individuals. However, the patient literature with respect to metaphor is quite limited. In addition to the already highlighted methodological weaknesses of some studies (Schmidt et al, 2010), one obvious shortcoming is the narrowness of the sampled metaphor probes. Most studies have only considered the metaphoric extension of adjectives (Winner & Gardner, 1977; Brownell et al, 1984; Brownell et al, 1990; Tompkins, 1990; Mackenzie et al, 1999; Giora et al, 2000; Zaidel et al, 2002; Gangon et al, 2003), or the potentially quite different domain of idioms (Van Lancker & Kempler, 1987; Tompkins, Boada, & McGarry, 1992; Kempler et al, 1999; Papagno & Caporali, 2007; Papagno & Genoni, 2004; Papagno et al, 2004; Rinaldi, Marangolo, & Baldassarri, 2004; Cacciari et al, 2006). At this point, no patient study has considered nominal or predicate metaphors. Neither has any patient study considered comprehension of novel metaphors.

As these various suggested lines of research illustrate, these stimuli are designed to address current questions about the neural basis of metaphor comprehension. They are, of course, not without their limitations. For instance, despite the inclusion of additional adjectives in the nominal sentences, these items are generally still shorter than the predicate items. Similarly, although literal sentences were made as similar to metaphors as possible, they were still rated as more imageable and familiar. Circumventing these differences, however, would have created more problematic issues. For example, adding even more adjectives to nominals would equalize length with predicates – but it would also mean that nominals might require greater semantic processing in order to integrate the additional semantic information. Using more abstract words in literal sentences would likely equate imageability with metaphors – but at the expense of differences in their concreteness values. Making literal sentences less familiar would on first pass seem desirable, but doing so would require such unlikely combinations of words (e.g. “He is sitting deep in the bubbles”, Kempler et al, 1999) that we feared they might seem metaphorical when paired with our novel metaphors.

In addition, we constrained the stimulus set in certain ways. We have provided normative data for only one online task, a valence judgment. We believe this to be an appropriate index for ensuring comparability between conditions in terms of semantic processing time, but future researchers may want to add others. We have also avoided familiar metaphors, limiting the set to relatively novel expressions instead since they are scarce in the literature.

Careful stimulus selection and the use of covariates can address some of these limitations. For others, we hope to have provided a protocol for generating similar stimuli in other languages or to augment the current set as theories of metaphors evolve. Although most of the measures we have included consist of straightforward ratings or easily acquired psycholinguistic values, we have taken a laborious approach to assessing interpretability that has important benefits. Ratings of ‘comprehensibility’ using an ordinal scale are a frequently employed and easy to acquire measure (e.g. Rate how easy this statement is to understanding using a scale from 1 *not at all* to 7 *very easy*) but are coarse in what they can indicate. One cannot know whether the interpretation any given reader generated when ranking a stimulus was actually reasonable or even metaphorical, nor does one have any sense if the particular interpretation assigned to the item varied between individuals. These shortcomings are especially problematic when one is interested in how anything but the most conventional of metaphors is understood. In everyday language, metaphors are rarely encountered out of context. Foregrounding information likely strongly biases the salience of a figurative interpretation (Giora, 1999). However, as our interpretation task demonstrated, an isolated metaphor of even fairly high familiarity may evoke several different senses. For this reason, it seems sensible to operationalize the interpretability of a metaphor in terms of the plausibility of interpretations, regardless of consensus or lack thereof. Using more than one rater of plausibility is also preferable given the subjective nature of interpretation, especially for less familiar metaphors.

## 5. Conclusion

Given metaphor's likely standing as a hallmark of human intelligence, characterizing its neural basis is a goal with broad interest. Nonetheless, the accumulated evidence does not fall out coherently. We suggest that cognitive neuroscience research on metaphor has been hampered by a lack of adequately controlled stimuli, and by the fact that stimulus designs have not always kept pace with emergent theoretical frameworks, or have addressed them too narrowly. We aim to fill this gap with this large, extensively normed, flexible, and theoretically motivated set of stimuli.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This research was supported by National Institute of Health grants (RO1-HD-050199-01A2 and RO1-DC-008779-01) awarded to A. Chatterjee. We would like to thank Bianca Bromberger and Page Widick for their help collecting and organizing the data.

## Appendix A Words Normed for Concreteness

accountant, admonishment, alibi, amused, anthology, applause, archeological, aside, assignment, ATM, babble, bankruptcy, bashful, beckoning, billboard, blast, bleat, belch, blubber, blurt, bold, braking, bully, bungee, cackle, campaign, canter, canvas, carousel, cartwheel, cathartic, celebrity, celebrity, chant, chase, chat, chatter, cheer, chime, chirp, clamber, clatter, clatter, click, clomp, cluck, collapse, colorful, comeback, comment, commotion, competitive, complaint, complicated, composition, confident, constant,

controversial, coo, cooking, corporate, corrupt, couple, course, creep, curl, current, dash, declined, demo, designer, desperate, dice, dieter, disorder, disturbance, divorcee, dodge, drift, drive, drone, drummer, editorial, elitist, email, embarrassed, embarrassment, endless, environmentalist, eviction, ex, excursion, exhausting, exhibition, eyelashes, fart, feminist, fizzle, flip, flit, flop, flounder, forced, forgotten, furtive, gambler, gardener, gaze, gear, girlfriend, glance, gleeful, glide, goodbye, grasp, greeting, growl, grumble, grunt, guffaw, gurgle, haircut, handshake, hangover, headline, heartbroken, hipster, hiss, hobble, homework, hoot, hopeful, hormonal, hostess, housewife, huff, icy, ill-timed, immediate, injection, input, insistent, interjection, internet, interviewer, irrepressible, irresistible, irritating, karate, landing, Latin, legal, leisurely, license, literary, logging, loophole, lope, lost, lurch, massage, media, mediocre, memoirs, mere, model, monk, mosey, motif, mumble, murmur, negotiations, nervous, objection, obstacle, oink, optimistic, outburst, outraged, overhead, packaging, painful, paisley, pamphlet, paperwork, parking, partnership, password, patriotic, perfect, perspective, petition, photographer, plod, plummet, poignant, polluted, popular, posture, pounce, prance, preacher, pregnancy, press, pretentious, Prius, privileged, programmer, promised, purr, rant, reception, recession, recording, rejection, relationship, relay, reproach, request, rescuer, resume, retort, revealing, romantic, romp, roommate, rousing, runner, runway, sashay, scamper, screech, screech, script, scrutiny, scurry, seizure, serenade, sexist, shameless, shipwreck, shuffle, shy, sidle, single, sip, sizzle, skater, skulk, skydive, slam-dunk, sleepwalk, slink, slogan, slouch, slurp, smirk, snarl, sneak, snicker, sniff, snigger, snore, sputter, squawk, squeal, stammer, stampede, status, steady, stir, stomp, stoplight, straggler, strategy, stroll, struggle, stutter, successful, suitor, supportive, surfer, SUV, swagger, swimmer, symptom, tacky, tailspin, take-off, tango, tattoo, teen, teenager, textbook, theater, therapy, tiptoe, totter, traipse, trajectory, trek, trickle, triumphant, trudge, t-shirt, tsunami, tug, twitch, twitter, unexpected, unpaid, urgent, valley, vocalization, wade, waiting, Wallstreet, waterfall, whimper, whine, whinny, whir, winner, word, wrecking, wrestle, wrinkled, yip, yodel

## Appendix B Words Normed for Auditory and Visual Imagery

argue, babble, balloon, bang, bark, belch, blast, bleat, blubber, blurt, bounce, buzz, cackle, call, canter, cartwheel, chant, charge, chat, cheer, chime, chirp, chop, chuckle, clamber, clamor, clash, clatter, click, climb, clomp, cluck, coast, collapse, coo, cough, crackle, crawl, creep, cry, dance, dart, dash, dig, dive, dodge, drift, drive, drone, drop, drum, fall, fart, fizzle, flip, flit, flop, flounder, flow, flush, fly, gasp, gesture, giggle, glide, groan, growl, grumble, grunt, guffaw, gurgle, hiss, hobble, holler, hoot, hop, howl, huff, hug, hum, hush, inch, jingle, jog, jump, knock, laugh, launch, leap, lift, limp, lope, lumber, lurch, march, moan, mosey, move, mumble, murmur, oink, plod, plow, plummet, polka, pop, pounce, prance, press, puff, pull, punch, purr, push, quack, race, rain, rant, rattle, reel, retreat, ride, roar, roll, run, sail, sashay, scamper, scream, screech, scurry, serenade, shatter, shout, shriek, shuffle, sidle, sigh, sing, sizzle, skulk, skydive, slam-dunk, slap, sleepwalk, slide, slink, slither, slouch, slurp, smash, snake, snap, snarl, sneak, sneeze, snicker, sniff, snigger, snore, snort, sob, spill, spin, splash, spring, sprint, sputter, squawk, squeal, stab, stammer, stampede, stand, stir, stomp, stream, stretch, stroll, strut, stumble, stumble, stutter, surf, surge, swagger, swarm, sweep, swim, swing, tailspin, take-off, tango, thunder, tiptoe, toss, totter, traipse, trudge, tug, twitter, voice, wade, wail, walk, waltz, wander, wave, weep, whimper, whine, whinny, whir, whirl, whisper, whistle, whoop, wiggle, wind, worm, wrestle, yawn, yell, yelp, yip, yodel, yowl, yowl, zigzag

## Appendix C

## Appendix C

## Example Items from Final Stimulus Set

	Base	Aud	Vis	Sentence Pair		Pos	RT	Int	Fam	Nat	Img	Fig	L1	L2	L3	F1	F2	Con
PA	sigh	3.95	2.65	Metaphor	The letter was a lonely sigh.	0.00	1372	1.00	4.10	4.70	3.40	5.90	29	6	3	60	43	403
				Literal	Her only comment was a sigh.	0.06	1233	n/a	6.50	6.61	5.17	1.80	28	6	3	600	367	319
PM	stumble	1.80	3.47	Metaphor	The first date was a stumble.	0.00	1179	1.00	4.25	4.80	4.35	5.10	29	6	3	488	328	419
				Literal	The skater's shame was a stumble.	0.00	1547	n/a	4.15	4.20	5.40	2.85	33	6	3	7	15	423
NA	bark	4.80	3.00	Metaphor	The unpaid bills barked at the father.	0.05	1659	1.00	2.28	2.18	2.06	6.56	38	7	4	60	144	498
				Literal	The mean bulldog barked at the burglar.	0.15	1367	n/a	6.26	6.70	6.45	1.00	39	7	4	50	314	515
NM	hop	1.65	3.85	Metaphor	The insults hopped on her tongue.	0.00	1233	1.00	2.37	2.10	2.50	6.50	33	6	3	15	12	500
				Literal	The rabbits hopped in the yard.	0.45	1545	n/a	6.00	6.60	6.85	1.05	31	6	3	15	11	561

**Key:** PA = Predicate-Auditory; PM = Predicate-Motion; NA = Nominal-Auditory; NM = Nominal-Motion; Aud = Base Auditory Imagery; Vis = Base Visual Imagery; Pos = % positive valence; RT = Valence judgment reaction time; Int = Interpretability; Fam = Familiarity; Nat = Naturalness; Img = Imageability; Fig = Figurativeness; L1 = Length in characters; L2 = Length in words; L3 = Length in content words; F1 = Kucera-Francis frequency; F2 = Brysbaert & New frequency; Con = concreteness

## References

- Amorapant PX, Widick P, Chatterjee A. The Neural basis of spatial relations. *Journal of Cognitive Neuroscience*. in press.
- Arzouan Y, Goldstein A, Faust M. Dynamics of hemispheric activity during metaphor comprehension: Electrophysiological measures. *NeuroImage* 2007;36(1):222–231. [PubMed: 17428685]
- Aziz-Zadeh L, Damasio A. Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology, Paris* 2008;102:35–39.
- Balota, DA.; Yap, MJ.; Cortese, MJ. Visual word recognition: The journey from features to meaning (A travel update).. In: Traxler, M.; Gernsbacher, MA., editors. *Handbook of Psycholinguistics*. Elsevier; New York: 2006. p. 285-375.
- Binder JR, Westbury CF, McKiernan KA, Possing ET, Medler DA. Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience* 2005a;17(6):905–917. [PubMed: 16021798]
- Binder J, Medler D, Desai R, Conant L, Liebenthal E. Some neurophysiological constraints on models of word naming. *NeuroImage* 2005b;27(3):677–693. [PubMed: 15921937]
- Boroditsky L. Metaphoric structuring: understanding time through spatial metaphors. *Cognition* 2000;75(1):1–28. [PubMed: 10815775]
- Bottini G, Corcoran R, Sterzi R, Paulesu E, Schenone P, Scarpa P, et al. The role of the right hemisphere in the interpretation of figurative aspects of language A positron emission tomography activation study. *Brain* 1994;117(6):1241–1253. [PubMed: 7820563]
- Bowdle BF, Gentner D. The Career of Metaphor. *Psychological Review* 2005;112(1):193–216. [PubMed: 15631593]
- Brownell HH, Simpson TL, Bihle AM, Potter HH, Gardner H. Appreciation of metaphoric alternative word meanings by left and right brain-damaged patients. *Neuropsychologia* 1990;28(4):375–383. [PubMed: 1692980]
- Brownell HH, Potter HH, Michelow D, Gardner H. Sensitivity to lexical denotation and connotation in brain-damaged patients: A double dissociation. *Brain and Language* 1984;22:253–265. [PubMed: 6204711]
- Brysbaert M, New B. Moving beyond Kucera and Francis: A Critical Evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 2009;41:977–990. [PubMed: 19897807]
- Cacciari C, Reati F, Colombo MR, Padovani R, Rizzo S, Papagno C. The Comprehension of ambiguous idioms in aphasic patients. *Neuropsychologia* 2006;44(8):1305–1314. [PubMed: 16504224]

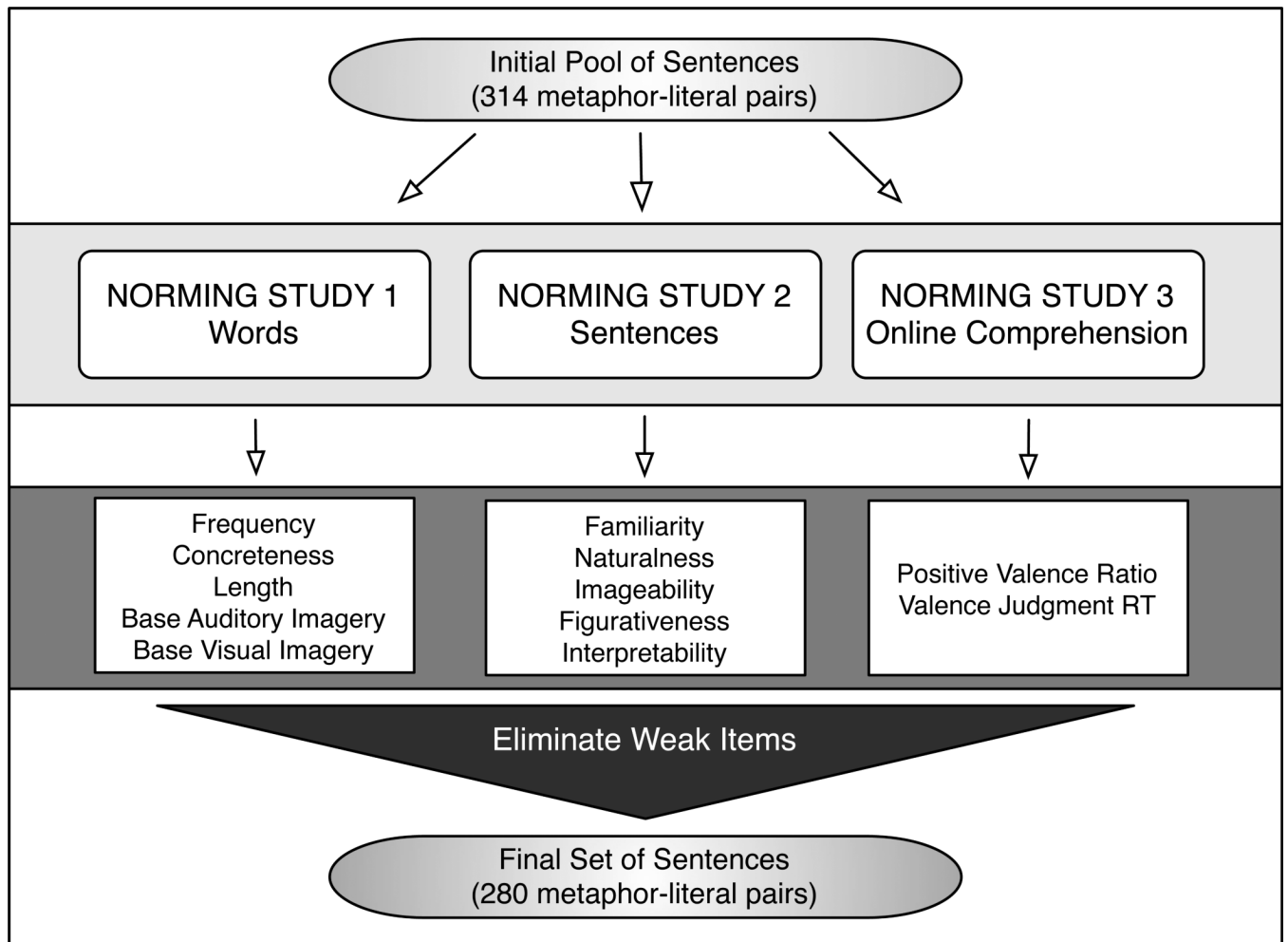
- Cardillo ER, Aydelott J, Matthews PM, Devlin JT. Left inferior prefrontal cortex activity reflects inhibitory rather than facilitatory priming. *Journal of Cognitive Neuroscience* 2004;16(9):1552–1561. [PubMed: 15601518]
- Chatterjee A. The Neural organization of spatial thought and language. *Seminars in Speech and Language* 2008;29(3):226–252. [PubMed: 18720319]
- Chen E, Widick P, Chatterjee A. Functional-anatomical organization of predicate metaphor processing. *Brain and Language* 2008;107(3):194–202. [PubMed: 18692890]
- Chiappe DL, Kennedy JM, Chiappe P. Aptness is more important than comprehensibility in preference for metaphors and similes. *Poetics* 2003;31(1):51–68.
- Coltheart M. The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology* 1981;33(4):497–505.
- Constable RT, Pugh KR, Berroya E, Mencl WE, Westerveld M, Ni W, Shankweiler D. Sentence complexity and input modality effects in sentence comprehension: an fMRI study. *NeuroImage* 2004;22(1):11–21. [PubMed: 15109993]
- Fiebach CJ, Friederici AD. Processing concrete words: fMRI evidence against a specific right-hemisphere involvement. *Neuropsychologia* 2003;42(1):62–70. [PubMed: 14615076]
- Friederici AD, Fiebach CJ, Schlesewsky M, Bornkessel ID, von Cramon DY. Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex* 2006;16(12):1709–1717. [PubMed: 16400163]
- Gagnon L, Goulet P, Giroux F, Joannette Y. Processing of metaphoric and non-metaphoric alternative meanings of words after right- and left-hemispheric lesion. *Brain and Language* 2003;87(2):217–226. [PubMed: 14585291]
- Gentner, D. Why we're so smart.. In: Gentner, Dedre; Goldin-Meadow, Susan, editors. *Language in Mind: Advances in the Study of Language and Thought*. MIT Press; Cambridge, MA: 2003. p. 195–235.
- Gentner, D.; Bowdle, BF.; Wolff, P.; Boronat, C.; Centner, D.; Bowdle, B., et al. Metaphor is like analogy.. In: Gentner, D.; Holyoak, KJ.; Kokinov, BN., editors. *The Analogical Mind: Perspectives from Cognitive Science*. MIT Press; Cambridge, MA: 2001. p. 199–253.
- Gentner D, Wolff P. Alignment in the processing of metaphor. *Journal of Memory and Language* 1997;37(3):331–355.
- Gibbs RW. Metaphor Interpretation as Embodied Simulation. *Mind and Language* 2006;21(3):434–458.
- Gilhooly KL, Logey RH. Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behavioural Research Methods and Instrumentation* 1980;12:395–427.
- Giora R. On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics* 1999;31(7):919–929.
- Giora R, Zaidel E, Soroker N, Batori G, Kasher A. Differential effects of right- and left-hemisphere damage on understanding sarcasm and metaphor. *Metaphor and Symbol* 2000;15(1):63.
- Glucksberg S. The Psycholinguistics of metaphor. *Trends in Cognitive Sciences* 2003;7(2):92–96. [PubMed: 12584028]
- Grice, HP. Logic and Conversation.. In: Cole, P.; Morgan, JL., editors. *Syntax and Semantics: Speech Acts*. Vol. 3. Academic Press; New York: 1975. p. 41–58.
- Hauk O, Davis MH, Pulvermüller F. Modulation of brain activity by multiple lexical and word form variables in visual word recognition: A parametric fMRI study. *NeuroImage* 2008;42(3):1185–1195. [PubMed: 18582580]
- Ishai A, Ungerleider LG, Martin A, Haxby JV. The Representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience* 2000;12(Supplement 2):35–51. [PubMed: 11506646]
- Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV. Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America* 1999;96(16):9379–9384. [PubMed: 10430951]
- Jones LL, Estes Z. Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language* 2006;55(1):18–32.
- Just MA, Carpenter PA, Keller TA, Eddy WF, Thulborn KR. Brain activation modulated by sentence comprehension. *Science* 1996;274(5284):114–116. [PubMed: 8810246]



- Just MA, Newman SD, Keller TA, McEleney A, Carpenter PA. Imagery in sentence comprehension: an fMRI study. *NeuroImage* 2004;21(1):112–124. [PubMed: 14741648]
- Kable JW, Chatterjee A. Specificity of Action Representations in the Lateral Occipitotemporal Cortex. *Journal of Cognitive Neuroscience* 2006;18(9):1498–1517. [PubMed: 16989551]
- Kable JW, Kan IP, Wilson A, Thompson-Schill SL, Chatterjee A. Conceptual representations of action in the lateral temporal cortex. *Journal of Cognitive Neuroscience* 2005;17(12):1855–1870. [PubMed: 16356324]
- Katz AN, Paivio A, Marschark M, Clark JM. Norms for 204 Literary and 260 Nonliterary Metaphors on 10 Psychological Dimensions. *Metaphor and Symbol* 1988;3(4):191.
- Keller TA, Carpenter PA, Just MA. The Neural bases of sentence comprehension: a fMRI examination of syntactic and lexical processing. *Cerebral Cortex* 2001;11(3):223–237. [PubMed: 11230094]
- Kemmerer D. The Semantics of space: Integrating linguistic typology and cognitive neuroscience. *Neuropsychologia* 2006;44:1607–1621. [PubMed: 16516934]
- Kempler D, Van Lancker D, Marchman V, Bates E. Idiom comprehension in children and adults with unilateral brain damage. *Developmental Neuropsychology* 1999;15(3):327.
- Kircher TT, Leube DT, Erb M, Grodd W, Rapp AM. Neural correlates of metaphor processing in schizophrenia. *NeuroImage* 2007;34(1):281–289. [PubMed: 17081771]
- Kosslyn SM, Maljkovic V, Hamilton SE, Horwitz G, Thompson WL. Two types of image generation: Evidence for left and right hemisphere processes. *Neuropsychologia* 1995;33(11):1485–1510. [PubMed: 8584182]
- Kucera, H.; Francis, WN. Computational analysis of present-day American-English. Brown University Press; Providence, RI: 1967.
- Kuperberg GR, Holcomb PJ, Sitnikova T, Greve D, Dale AM, Caplan D. Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience* 2003;15(2):272–293. [PubMed: 12676064]
- Kutas M. One lesson learned: Frame language processing—literal and figurative—as a human brain function. *Metaphor and Symbol* 2006;21(4):285.
- Lakoff, G.; Johnson, M. *Metaphors We Live By*. University of Chicago Press; Chicago: 1980.
- Lee SS, Dapretto M. Metaphorical vs. literal word meanings: fMRI evidence against a selective role of the right hemisphere. *NeuroImage* 2006;29(2):536–544. [PubMed: 16165371]
- Mackenzie C, Begg T, Lees KR, Brady M. The communication effects of right brain damage on the very old and the not so old. *Journal of Neurolinguistics* 1999;12(2):79–93.
- Martin, A.; Ungerleider, LG.; Haxby, JV. Category specificity and the brain: The Sensory/Motor model of semantic representations of objects.. In: Gazzaniga, MS., editor. *The New Cognitive Neurosciences*. 2nd Edition. MIT Press; Cambridge: 2000. p. 1023-1036.
- Mashal N, Faust M, Hendler T. The role of the right hemisphere in processing nonsalient metaphorical meanings: Application of principal components analysis to fMRI data. *Neuropsychologia* 2005;43(14):2084–2100. [PubMed: 16243053]
- Mashal N, Faust M, Hendler T, Jung-Beeman M. An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain and Language* 2007;100(2):115–126. [PubMed: 16290261]
- Mashal N, Faust M, Hendler T, Jung-Beeman M. An fMRI-study of processing novel metaphoric sentences. *Laterality: Asymmetries of Body, Brain and Cognition* 2009;14(1):30.
- Menenti L, Petersson KM, Scheeringa R, Hagoort P. When elephants fly: Differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. *Journal of Cognitive Neuroscience* 2008;21(12):2358–2368. [PubMed: 19016600]
- Nelson, DL.; McEvoy, CL.; Schreiber, TA. The University of South Florida word association, rhyme, and word fragment norms. 1998. Retrieved from <http://www.usf.edu/FreeAssociation/>
- Paivio A, Yuille JC, Madigan SA. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology* 1968;76(1)(Suppl):1–25. [PubMed: 5672258]
- Papagno C, Caporali A. Testing idiom comprehension in aphasic patients: The effects of task and idiom type. *Brain and Language* 2007;100(2):208–220. [PubMed: 16487581]

- Papagno C, Genoni A. The role of syntactic competence in idiom comprehension: a study on aphasic patients. *Journal of Neurolinguistics* 2004;17(5):371–382.
- Papagno C, Tabossi P, Colombo MR, Zampetti P. Idiom comprehension in aphasic patients. *Brain and Language* 2004;89(1):226–234. [PubMed: 15010254]
- Pobric G, Mashal N, Faust M, Lavidor M. The Role of the right cerebral hemisphere in processing novel metaphoric expressions: A transcranial Magnetic stimulation study. *Journal of Cognitive Neuroscience* 2008;20(1):170–181. [PubMed: 17919080]
- Prat CS, Keller TA, Just MA. Individual differences in sentence comprehension: A functional magnetic resonance imaging investigation of syntactic and lexical processing demands. *Journal of Cognitive Neuroscience* 2007;19(12):1950–1963. [PubMed: 17892384]
- Pullvermuller F. Brain mechanisms linking language and action. *Nature Reviews Neuroscience* 2006;6(7):576–582.
- Rapp AM, Leube DT, Erb M, Grodd W, Kircher TTJ. Neural correlates of metaphor processing. *Cognitive Brain Research* 2004;20(3):395–402. [PubMed: 15268917]
- Rapp AM, Leube DT, Erb M, Grodd W, Kircher TT. Laterality in metaphor processing: Lack of evidence from functional magnetic resonance imaging for the right hemisphere theory. *Brain and Language* 2007;100(2):142–149. [PubMed: 16677700]
- Rinaldi MC, Marangolo P, Baldassarri F. Metaphor comprehension in right brain-damaged patients with visuo-verbal and verbal material: A dissociation (re)considered. *Cortex* 2004;40(3):479–490. [PubMed: 15259328]
- Schmidt GL, Cardillo ER, Kranjec A, Chatterjee A. Beyond laterality: A critical assessment of research on the neural basis of metaphor. *International Journal of Neuropsychology* 2010;16:1–5.
- Scott SK. The neural representation of concrete nouns: What's right and what's left? *Trends in Cognitive Sciences* 2004;8(4):151–153. [PubMed: 15551484]
- Searle, J. *Expression and Meaning*. Cambridge University Press; Cambridge, England: 1979.
- Shibata M, Abe J, Terao A, Miyamoto T. Neural mechanisms involved in the comprehension of metaphoric and literal sentences: An fMRI study. *Brain Research* 2007;1166:92–102. [PubMed: 17662699]
- Simmons WK, Barsalou LW. The Similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology* 2003;20:451–486.
- Sotillo M, Carretié L, Hinojosa JA, Tapia M, Mercado F, López-Marín S, et al. Neural activity associated with metaphor comprehension: spatial analysis. *Neuroscience Letters* 2005;373(1):5–9. [PubMed: 15555767]
- Stringaris AK, Medford NC, Giampietro V, Brammer MJ, David AS. Deriving meaning: Distinct neural mechanisms for metaphoric, literal, and non-meaningful sentences. *Brain and Language* 2007;100(2):150–162. [PubMed: 16165201]
- Thompson-Schill SL. Neuroimaging studies of semantic memory: inferring “how” from “where”. *Neuropsychologia* 2003;41(3):280–292. [PubMed: 12457754]
- Toglia, MP.; Battig, WF. *Handbook of Semantic Word Norms*. Erlbaum; Hillsdale, NJ: 1978.
- Tompkins CA. Knowledge and strategies for processing lexical metaphor after right or left hemisphere brain damage. *Journal of Speech and Hearing Research* 1990;33(2):307–316. [PubMed: 2359271]
- Tompkins CA, Boada R, McGarry K. The Access and processing of familiar idioms by brain-damaged and normally aging adults. *Journal of Speech and Hearing Research* 1992;35(3):626–637. [PubMed: 1608254]
- Torreano LA, Cacciari C, Glucksberg S. When dogs can fly: Level of abstraction as a cue to metaphorical use of verbs. *Metaphor and Symbol* 2005;20(4):259.
- Tranel D, Manzel K, Asp E, Kemmerer D. Naming dynamic and static actions: Neuropsychological evidence. *Journal of Physiology - Paris* 2008;102(1):80–94.
- Van Lancker DR, Kempler D. Comprehension of familiar phrases by left-but not by right-hemisphere damaged patients. *Brain and Language* 1987;32(2):265–277. [PubMed: 2446699]
- Wallentin M, Lund TE, Ostergaard S, Ostergaard L, Roepstorff A. Motion verb sentences activate left posterior middle temporal cortex despite static context. *Neuroreport* 2005a;16(6):649–652. [PubMed: 15812326]

- Wallentin M, Østergaard S, Lund TE, Østergaard L, Roepstorff A. Concrete spatial language: See what I mean? *Brain and Language* 2005b;92(3):221–233. [PubMed: 15721955]
- Winner E, Gardner H. The comprehension of metaphor in brain-damaged patients. *Brain: A Journal of Neurology* 1977;100(4):717–729. [PubMed: 608117]
- Wu DH, Waller S, Chatterjee A. The Functional Neuroanatomy of Thematic Role and Locative Relational Knowledge. *Journal of Cognitive Neuroscience* 2007;19(9):1542–1555. [PubMed: 17714015]
- Yarkoni T, Speer NK, Balota DA, McAvoy MP, Zacks JM. Pictures of a thousand words: Investigating the neural mechanisms of reading with extremely rapid event-related fMRI. *NeuroImage* 2008;42(2):973–987. [PubMed: 18554928]
- Zaidel E, Kasher A, Soroker N, Batori G. Effects of right and left hemisphere damage on performance of the “Right Hemisphere Communication Battery”. *Brain and Language* 2002;80(3):510–535. [PubMed: 11896655]



**Figure 1.**  
Overview of Procedure for Stimulus Selection

**Table 1**

## Examples of Each Sentence Type

Sentence Type	Literal	Metaphorical
PM	<i>The rabbits hopped in the yard.</i>	<i>The insults hopped on her tongue.</i>
	<i>The heavy box pressed against his side.</i>	<i>The lawyer pressed for a new trial.</i>
PA	<i>His daughter chuckled at the big glasses.</i>	<i>His eyes chuckled at the cute note.</i>
	<i>The hard candy rattled in the box.</i>	<i>The violent image rattled in her head.</i>
NM	<i>The blow was a single punch.</i>	<i>The editorial was a brass-knuckle punch.</i>
	<i>The injury was a knife stab.</i>	<i>The declined invitation was a stab.</i>
NA	<i>The last sip was a noisy slurp.</i>	<i>The man's gaze was a shameless slurp.</i>
	<i>The sounds was a bitter sob.</i>	<i>Her marriage was a long sob.</i>

Key: PM = Predicate Motion; PA = Predicate Auditory; NM = Nominal Motion; NA = Nominal Auditory

Table 2

Summary of Final Stimuli Characteristics by Sentence Type

	Literal						Metaphorical											
	PA	SD	M	PM	SD	M	NA	SD	M	PM	SD	M	NA	SD	M	NM	SD	M
Base Auditory Imagery	3.96	0.5	1.46	0.4	4.03	0.5	1.69	0.6	3.96	0.5	1.46	0.4	4.03	0.5	1.69	0.6		
Base Visual Imagery	2.82	0.5	3.52	0.7	2.72	0.6	3.63	0.7	2.82	0.5	3.52	0.7	2.72	0.6	3.63	0.7		
Concreteness	510	48	502	50	434	65	440	55	492	155	477	57	434	65	422	61		
Frequency <sup>1</sup>	83	108	85	93	83	142	73	88	92	108	100	111	60	78	78	91		
Frequency <sup>2</sup>	103	118	82	123	92	140	86	173	70	93	113	169	72	137	78	94		
# Characters	36.5	5.7	37.4	5.2	30.8	4.1	31.1	3.9	37.3	5.2	38.8	4.8	32.1	4.4	31.9	3.8		
# Words	6.5	0.8	6.5	0.6	5.9	0.4	6.1	0.4	6.6	0.8	6.5	0.7	6.0	0.5	6.1	0.4		
# Content words	3.6	0.6	3.6	0.5	3.0	0.4	3.1	0.4	3.6	0.7	3.6	0.5	3.0	0.4	3.1	0.4		
Interpretability	-	-	-	-	-	-	-	-	0.95	0.06	0.96	0.06	0.93	0.08	0.92	0.09		
Familiarity	5.81	0.82	5.51	0.9	5.30	0.8	5.32	0.9	2.84	1.1	4.11	1.4	3.81	1.2	4.13	1.2		
Naturalness	5.88	0.84	5.66	0.9	5.78	0.8	5.76	0.8	3.00	1.0	4.15	1.3	4.09	1.2	4.56	1.2		
Imageability	6.07	0.72	6.15	0.7	5.12	0.8	5.84	0.8	3.14	0.8	3.45	1.3	3.90	0.9	4.06	0.8		
Figurativeness	1.50	0.56	1.69	0.7	2.27	0.7	2.16	0.8	6.15	0.6	5.33	1.0	5.31	0.7	5.42	0.8		
Valence RT (ms)	1455	220	1459	185	1388	189	1485	220	1528	185	1591	197	1491	256	1485	201		
Valence Positive Ratio	0.19	0.26	0.21	0.24	0.17	0.26	0.29	0.29	0.18	0.27	0.25	0.28	0.27	0.32	0.28	0.32		

Key: PA = Predicate Auditory; PM = Predicate Motion; NA = Nominal Auditory; NM = Nominal Motion

<sup>1</sup>Frequency = values from Kucera & Francis (1967)

<sup>2</sup>Frequency = SUBTLWF values from Brysbaert & New (2009).

**Table 3**  
Correlation Coefficients Between Sentence Scales, Collapsed Across Modality

	Predicate Metaphors				
	FAM	NAT	IMG	FIG	INT
Familiarity (FAM)	.96**	.57**	.61**	.30**	.30**
Naturalness (NAT)		.61**	.62**	.30**	.30**
Imageability (IMG)			.52**	.15~	.15~
Figurativeness (FIG)				.04	.04
Interpretability (INT)					.04

	Nominal Metaphors				
	FAM	NAT	IMG	FIG	INT
Familiarity (FAM)	.91**	.36**	.54**	.27**	.27**
Naturalness (NAT)		.42**	.46**	.31**	.31**
Imageability (IMG)			.03	.25**	.25**
Figurativeness (FIG)				.02	.02
Interpretability (INT)					.02

Key

\* =  $p < .05$   
 ~ =  $p < .10$   
 \*\* =  $p < .01$

**Table 4**  
Correlation Coefficients Between Sentence Scales, Separated by Modality and Metaphor Type

Predicate Auditory Metaphors					
	FAM	NAT	IMG	FIG	INT
Familiarity (FAM)	.93**	.33**	.55**	.36**	.36**
Naturalness (NAT)		.43**	.53**	.36**	.36**
Imageability (IMG)			.14	.21~	.21~
Figurativeness (FIG)				.01	.01
Interpretability (INT)					.01

Nominal Auditory Metaphors					
	FAM	NAT	IMG	FIG	INT
Familiarity (FAM)	.89**	.42**	.57**	.35**	.35**
Naturalness (NAT)		.40**	.56**	.40**	.40**
Imageability (IMG)			.11	.27*	.27*
Figurativeness (FIG)				.21~	.21~
Interpretability (INT)					.21~

Predicate Motion Metaphors					
	FAM	NAT	IMG	FIG	INT
Familiarity (FAM)	.97**	.69**	.50**	.24*	.24*
Naturalness (NAT)		.71**	.53**	.25*	.25*
Imageability (IMG)			.66**	.11	.11
Figurativeness (FIG)				.01	.01
Interpretability (INT)					.01

Nominal Motion Metaphors					
	FAM	NAT	IMG	FIG	INT
Familiarity (FAM)	.92**	.28*	.55**	.22~	.22~
Naturalness (NAT)		.41**	.42**	.26*	.26*



**Nominal Motion Metaphors**

	FAM	NAT	IMG	FIG	INT
Imageability (IMG)			.13	.24*	
Figurativity (FIG)				.13	
Interpretability (INT)					

Key

~ =  $p < .10$

\* =  $p < .05$

\*\* =  $p < .01$