Metaphors: Where the neighborhood in which one resides interacts with (interpretive) diversity

J. Nick Reid[1] (jreid256@uwo.ca)
Hamad Al-Azary[2] (alazary@ualberta.ca)
Albert N. Katz[1] (katz@uwo.ca)

[1]Department of Psychology, Western University, 1151 Richmond St., London, ON N6A 3K7
[2]Department of Psychology, University of Alberta, 116 St. & 85 Ave., Edmonton, AB T6G 2R3

Abstract

The effect of semantic neighborhood density (SND) on metaphor interpretation was examined by asking participants to list features for both metaphors (e.g., "*music is medicine*") and the individual words in the metaphors (e.g., the metaphor topic, "music" and the metaphor vehicle, "medicine"). Topic word SND was associated with greater interpretive diversity for individual words, but lesser interpretive diversity for metaphors. High-SND metaphor topics decreased in interpretive diversity when the topic was framed in a metaphor (compared to when presented alone) whereas for Low-SND topics, there was an increase in interpretive diversity when framed in a metaphor. We propose that the function of the vehicle differs depending on the nature of the metaphor topic. Vehicles convey or emphasize a select few relevant features of topic meaning when the topic has many close semantic associates, whereas they highlight many potentially relevant low-salience features for topic concepts with few close associates.

Keywords: metaphor, feature-listing, interpretive diversity, semantic neighborhood density, conceptual representation.

Introduction

A metaphor involves the conjunction of two seemingly unrelated concepts to make some communicative point. For instance, in the classic Shakespearian metaphor, "*All the world's a stage,*" (*As You Like It, Act II Scene VII*) the metaphor topic, the "world," is framed in terms of the metaphor vehicle, a "stage." This framing highlights certain aspects of the topic: much like on a stage, in the world people play certain roles and need to act in certain ways depending on the social context.

In most psycholinguistic models of metaphor processing, the topic and vehicle play different roles. Although models differ in how specific relevant properties of the vehicle subsequently modify our understanding of the topic, the general consensus is that such properties are nonetheless conveyed to the topic, with the semantic representations of the topic and vehicle interacting to produce metaphoric meaning.

Metaphors vary in terms of "interpretive diversity" (Utsumi, 2005), that is, the amount and distribution of different possible interpretations. With some metaphors, the vehicle may highlight one salient feature of the topic, for instance, for the metaphor "*a zebra is a piano*," the interpretation that a zebra, like a piano, is "black and white" may dominate. In contrast, for a metaphor such as "*time is money*," there may be a variety of interpretations, such as time is valuable, limited, needs to be carefully managed, etc., and no single interpretation may dominate over the rest. When a metaphor generates a small number of interpretations, and a single interpretation is far more frequent than all other interpretations, it is considered low in interpretive diversity. In contrast, when a metaphor generates many interpretations and they are all about equally as frequent, it is considered high in interpretive diversity.

Although interpretive diversity has been investigated in terms of how it influences metaphor aptness and appreciation (Utsumi, 2005), the difference between similes and metaphors (Utsumi, 2007), and how it affects metaphor comprehension (Utsumi, 2011), there has been little research on what factors influence interpretive diversity itself. The purpose of this study was to investigate a word-level variable, *semantic neighborhood density,* to see whether it influences the diversity of metaphor interpretations.

Semantic neighborhood density (SND) is a measure of how closely associated in meaning a word is to its neighbors (i.e., other related words) in semantic space (Buchanan, Westbury & Burgess, 2001). Words with many close associates are considered to reside in a high-density semantic space, whereas words with more distant associates are considered to reside in a low-density space. Semantic neighborhood density can be estimated using a vector space model of word meaning (Reid & Katz, 2018), for instance, by calculating the average cosine of a target word to its *n* nearest neighbors.

Semantic distance (or dissimilarity of topic and vehicle) influences both metaphor comprehension and production. Trick and Katz (1986) found that metaphors were best understood and appreciated when the topic and vehicle concepts were from dissimilar higher-order categories (high between-domain distance), but shared similar distinguishing features within those categories (low within-domain distance). Semantic distance between topic and vehicle has also been used in computer simulations to predict whether categorical statements are considered metaphorical or literal with high accuracy (85% correct in both English and Chinese; Su, Huang, & Chen, 2017). Additionally, Katz (1989) found in a metaphor completion task that participants preferred to select metaphor vehicles that were moderately semantically distant from the topic, not too close and not too far.

Semantic neighborhood density has received less attention in the metaphor literature. Katz and Al-Azary (2017) found

that metaphors were more bidirectional (i.e., decreased less in comprehensibility when the topic and vehicle terms switched positions; "A is B" to "B is A") when both the topic and vehicle were from high-density space. Recently, Al-Azary and Buchanan (2017) found that metaphors with Low-SND topic and vehicle words were more comprehensible than metaphors with High-SND topic and vehicle words (see also Al-Azary, McAuley, Buchanan & Katz, 2019). To interpret their effect, Al-Azary and Buchanan reasoned that High-SND words are too semantically rich to take on new meanings in metaphors. Low-SND words, on the other hand, have more room for making new associations.

To date, SND has not been examined in terms of how it might affect the features evoked by metaphors and the diversity of interpretations. High-SND words have many close associates, so when the topic and vehicle terms in a metaphor are High-SND, it may evoke more diversity of features and multiple interpretations. On the other hand, High-SND words may be very similar to their neighbors, leading to a more constrained overall meaning, and thus, when employed in a metaphor, may elicit less diversity of interpretations.

In terms of the relative contributions of the topic and vehicle to a metaphor's interpretive diversity, the semantic space of the topic may be particularly important. Glucksberg, Manfredi, and McGlone (1997) argue that the topic provides relevant dimensions for constraining the semantic properties of the vehicle. Therefore, the topic and its associates may play an important role in constraining which features of the vehicle, and how many, are attributed to the topic.

The Current Study

The effects of SND on interpretive diversity were examined in a feature-listing task in which each participant was instructed to list three features per metaphor or word. We obtained features for 88 metaphors, and also for the 146 words used in those metaphors (some metaphors had overlapping words, e.g., "*history is a mirror*" and "*history is a sponge*"). This way, we can compare the interpretive diversity for the words both when they are presented alone, and when they are presented in a metaphor.

Methods

Participants

One hundred and fifty-five (90 female) undergraduate psychology students completed the study in partial fulfilment of course requirements. The reported ages ranged from 17 to 77 (M = 18.91, SD = 5.09). Participants were recruited through the department of psychology's Sona system website, a cloud-based service for connecting researchers with potential participants.

Materials

The metaphors consisted of 88 nonliterary metaphors taken from the Katz, Paivio, Marschark, and Clark (1988) norms. Only metaphors that contained a single word (or hyphenated word) for both the topic and the vehicle were selected as experimental stimuli. For instance, "*Education is a lantern*" was selected as the topic was the single word "education" and the vehicle was the single word "lantern." However, "*Thought is a boiling kettle*" was not selected as the vehicle, "boiling kettle," consisted of two words. The word stimuli were simply the 146 unique topic and vehicle words contained in the 88 selected metaphors.

The stimuli were combined into four groups, two each with 44 metaphors, and two each with 73 words. The stimuli were initially randomly sorted into groups, but after randomization, we then sorted similar metaphors or words (e.g., "*history is a mirror*" and "*history is a sponge*") into separate groups as much as possible to reduce carryover effects. Each participant was randomly assigned to one of the four groups of stimuli. The stimuli within each group were presented in random order to participants. Unlike some previous studies on metaphor features (e.g., Becker, 1997; Utsumi, 2005), each participant saw only words or metaphors, not both (see Roncero & de Almeida, 2015). We did this to obtain purer features of the words; we did not want to encourage the participant to think metaphorically about the words.

Procedure

The entire task took place online over the Qualtrics survey platform. The first screen was a letter of information explaining the study and the next screen asked demographic questions on age and gender. Following this, a screen explaining the basic task was presented. For the metaphor groups, participants were instructed to list three features or characteristics of the topic that are being described by the vehicle (see Utsumi, 2005). An example with the metaphor "*music is medicine*" was given in which the three listed features were "soothing," "healing," and "enjoyable." For the word groups, participants were instructed to list three features or characteristics of the word. The word "*music*" was given as an example, and the features listed were "artists," "beautiful," and "creative" (examples taken from Roncero & de Almeida, 2015). Following this screen, the metaphors or words were presented one at a time with a textbox underneath. Once a participant completed their response, they were not allowed to return to previously answered items.

Results

Pre-processing

The feature data were processed in Python using tools from the nltk package (Loper & Bird, 2002). Aside from saving time, automated analysis of feature data also increases consistency and transparency across different labs and studies (Buchanan, De Deyne, & Montefinese, 2019). There were four basic steps to data processing. First, the raw response from the textbox was split into responses using the re.split() command in Python; whenever a paragraph indent or a comma occurred, a new response started (e.g., "soothing,

healing, enjoyable" would be split into three responses, "soothing," "healing," and "enjoyable," based on the commas). The second step converted all responses into lower case letters using the .lower() command in Python. This was so that identical words would be counted together even when the capitalization differed, for instance, "Healing" and "healing." Third, stop words (words that hold little semantic content such as "the," "a," "has," and "it") were removed, which allows similar responses to be grouped together more accurately (Buchanan, De Deyne, & Montefinese, 2019). For instance, for the word *butterfly*, one participant may list "wings" whereas another may list "has wings," which is essentially the same feature. If the stop word "has" is removed, both responses are now simply "wings," and the program will count both responses together. Finally, word suffixes were removed using Porter's (1980) stemming algorithm using the PorterStemmer() command from the nltk package. This allowed features to be counted together when they shared the same morphological root (see Roncero & de Almeida, 2015). For instance, words such as *excite, excited, exciting,* and *excitement* would all be reduced to "*excit*," which means all four responses would be counted as the same feature (note that unlike lemmatization, stemming algorithms sometime result in non-words). Similar to other metaphor feature studies, features were only retained in the final analysis if at least two participants listed it (Becker, 1997; Utsumi 2005, 2007). An example of the processed features for the metaphor "*A fisherman is a spider*" is displayed in Table 1.

Table 1: Processed feature list for the metaphor "*A fisherman is a spider*"

Features	Count
patient	5
catch	4
hunter	4
resourc	3
sneaki	3
consum	2
net	2
prey	2
quiet	2
work	2

Semantic Neighborhood Density

For our SND calculations, we used the Global Vectors (GloVe) model of word representation, a vector-based model that predicts word similarity based on word occurrences in large text corpora (Pennington, Socher, & Manning, 2014). Vector-space models assume that words that occur in similar contexts are more similar but vary in terms of what is defined as a "context." Some models focus on occurrences within a small context window, such as a span of 10 words, whereas others focus on occurrences within a large context, such as an entire document (see Reid & Katz, 2018, for a review). GloVe combines a mixture of both methods and has been demonstrated to outperform singular value decomposition models (e.g., Latent Semantic Analysis; Landauer & Dumais, 1997) and word2vec's models on analogy and word similarity tasks (Pennington et al., 2014).

The GloVe model we employed was pre-trained on a dump of Wikipedia from February 2017 (available for download from vectors.nlpl.eu/repository, model ID = 8). Recall that SND can be calculated by taking the average cosine of a target word to its n nearest neighbors. We selected a neighborhood size of 500, as is commonly used in computational simulations of metaphor comprehension (Kintsch, 2000; Reid & Katz, 2018). For a given word, the 500 nearest neighbors were the 500 words that had the highest cosine similarity to the word out of all the words in the model (the model included just over 300,000 words). Cosine is typically preferred over distance in computing similarity between word vectors because some words occur far more frequently than others (Clark, 2015; Erk, 2012). Distance underestimates similarity when two words often occur in similar contexts, but one word is far more frequent than the other. After the 500 nearest neighbors were found, SND was calculated as the average cosine similarity of these 500 words to the target word.

Three metaphors ("*memory is a trash-masher*," "*a storm is a coffeepot*" and "*wounds are fiords*") contained words that were not in the pre-trained GloVe model. Therefore, these metaphors were removed from all subsequent analyses, yielding a total of 85 metaphors and 140 unique words.

Interpretive Diversity

Interpretive diversity was calculated using Utsumi's (2005) method, which is based on Shannon's (1948) measure of entropy. The equation is as follows:

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

$p(x)$ in the above equation refers to the probability of a feature being listed. This can be calculated by dividing the number of times a feature was listed by the total number of features listed. For example, imagine for the metaphor "*music is medicine*" the features "soothing," "healing," and "enjoyable" are listed by 5, 3, and 2 participants respectively. $p(x)$ for the feature "soothing" would be .5 (5/10). The entire calculation for the interpretive diversity in this example would be: –(5/10) log(5/10) –(3/10) log(3/10) –(2/10) log(2/10) = 1.49. Interpretive diversity scores increase as the probability distribution becomes more uniform across the features and decrease as one feature becomes more probable than all other features.

Correlation analyses

Words The first correlation analysis involved the interpretive diversity scores and SND values for the 140 unique words in the study. There was a marginally significant positive correlation between interpretive diversity and SND, $r(138) = .16$, $p = .054$, wherein words from dense semantic space

tended to be more interpretively diverse. We next examined the words that serve as topics (n=66) and vehicles (n=78) separately, finding that the correlation between interpretive diversity and SND was only significant for the topics, $r(64) = .24$, $p = .048$, but not for the vehicles, $r(76) = .02$, $p = .837$.

An independent t-test was conducted to compare the mean SND values for topic vs. vehicle words. Four words served as both topics and vehicles and were removed from analysis. The t-test revealed that topics had significantly higher SND on average (.37) than vehicles (.30), $t(106.99) = 6.50$, $p < .001$ (degrees of freedom were adjusted from 134 to 106.99 as Levene's test indicated unequal variances, $F = 7.44$, $p = .007$). Furthermore, the vehicles had a more restricted range of SND (.23–.42, range = .19) compared to the topics (.25–.53, range = .28). The lower SND and restricted range of the vehicles may have diminished the power to detect a significant correlation with interpretive diversity.

Metaphors Separate correlations were conducted between interpretive diversity for the metaphors and both topic SND and vehicle SND. The correlation between interpretive diversity and vehicle SND was non-significant, $r(83) = -.03$, $p = .756$. However, unlike the *positive* effect with diversity when topic words were presented alone, there was a significant *negative* correlation between interpretive diversity and topic SND, $r(83) = -.29$, $p = .007$, when the topic was presented in a metaphor. Thus, greater topic SND was associated with less diverse metaphor interpretations.

As suggested by a reviewer, we explored whether there were interactive effects of topic and vehicle SND on interpretive diversity. This was examined using a linear regression model with topic SND, vehicle SND, and their interaction entered as predictors. The overall model was significant, $F = 3.36$, $p = .023$, $R^2 = .11$; however, none of the predictors contributed significantly to the prediction on their own, t's < 1.6, p's > .1. An ANOVA comparing this interactive model to a model with only topic SND indicated that the interactive model did not account for significantly more variance than the topic only model, $F = 1.16$, $p = .320$. Therefore, there was no evidence of a significant interaction between topic and vehicle SND, and because the interactive model did not account for significantly more variance, the more parsimonious topic only model is preferred.

To further unpack the opposing correlations between topic SND and interpretive diversity for words and metaphors, we examined the change in interpretive diversity between topic words when presented alone (e.g., "*music*") to when presented in a metaphor (e.g., "*music is medicine*"). Recall that the instructions for the metaphor feature-listing task were to list three characteristics of the topic being highlighted by the vehicle. Therefore, essentially the metaphor features were topic features made salient by the vehicle. We calculated the change in interpretive diversity by subtracting the interpretive diversity score for the topic word presented alone from the interpretive diversity score for the metaphor. Therefore, higher values correspond to a greater increase in diversity for the topic when presented in a metaphor vs. the topic alone. If the value was negative, it means that the topic concept actually exhibited less diversity of interpretations when presented in a metaphor than when presented alone. A median split in terms of topic SND revealed that topics from High-SND space were associated with metaphors that decreased interpretive diversity (mean = -0.28) whereas topics from Low-SND space were associated with metaphors that increased interpretive diversity (mean = .16). This difference was reliable, $t(82) = 3.99$, $p < .001$.

We also examined whether the change in interpretive diversity between topics presented alone and topics embedded in metaphors was predicted by an interaction of topic and vehicle SND. A linear regression model with topic SND, vehicle SND, and their interaction yielded a significant overall prediction, $F = 7.34$, $p < .001$, $R^2 = .21$, but none of the three predictors accounted for a significant unique portion of variance, t's < 1.8, p's > .05. A topic only model also yielded a significant prediction, $F = 18.65$, $p < .001$, $R^2 = .18$, and the interactive model did not account for significantly more variance than the topic only model, $F = 1.56$, $p = .216$. Therefore, again the more parsimonious topic only model is preferred.

Discussion

We find that SND differentially influenced interpretive diversity for words, depending on whether they were interpreted individually (e.g., "*music*") or in metaphors (e.g., "*music is medicine*"). For topic words alone, SND was associated with an increased diversity of features listed, but for metaphors, the association was opposite as higher topic SND was associated with less diversity of features. Further analyses revealed that High-SND topics were associated with a decrease in interpretive diversity when the topic was framed in a metaphor vs. presented alone. In contrast, Low-SND topics were associated with an increase in diversity when presented in a metaphor vs. alone.

Our results indicate that, in metaphor, the vehicle may serve a different purpose depending on the nature of the topic concept being framed. In everyday metaphor usage, when the topic concept has many close associates (High-SND), the vehicle employed may serve to constrain topic meaning, emphasizing only certain characteristics of the semantically rich concept. For example, in "*a mosquito is a vampire*," the vehicle vampire emphasizes the blood meaning of mosquito (a High-SND topic). In contrast, when the topic concept has few close associates, the vehicle may serve to highlight other, non-obvious or low-salience features of the concept (see Ortony, 1979), bringing about more diversity of interpretations. For example, in "*a library is a sanctuary,*" the vehicle "sanctuary" may bring numerous less-salient features of "library" (a Low-SND topic) to mind, such as "peaceful," "contemplative," "safe," etc. According to the interaction theory of metaphor (Black, 1962, 1979; Gineste, Indurkhya, & Scart, 2000), the vehicle functions as a "filter" that both highlights and hides aspects of the topic. However, our data suggests that the topic shapes that filter, widening the filter when the topic concept is semantically poor (i.e., Low-SND),

allowing for more possible interpretations, and constraining the filter for semantically rich (i.e., High-SND) topics. This is also consistent with Glucksberg et al. (1997) who argue that the metaphor topic provides relevant dimensions upon which specific values from the vehicle on those dimensions are conveyed to the topic. As these authors argue, metaphor topics can vary in terms of the number of relevant dimensions. Our data suggests that SND is a strong predictor of the number of relevant dimensions a topic will provide in a metaphor. Furthermore, the lack of an effect for vehicle SND is also consistent with this model, as the vehicle does not provide the dimensions themselves, but only values on those dimensions.

It was also found that SND only affected interpretive diversity for the topic terms, but not the vehicle terms, even when these terms were presented alone. This was not due to the task or the metaphors as the participants who listed features for words never saw metaphors. Therefore, it seems there is something different about the nature of the words themselves. Recall we chose the metaphors first and elicited features for these words and it remains likely that words that serve often as metaphor vehicles differ systematically from those that rarely do so. Here we find that the concepts used as vehicles had lower SND on average and a more restricted range of SND values than the concepts used as topics. Restricted range is a factor that is known to hinder the ability to find significant correlations. There is evidence that in general Low-SND words make for good metaphor vehicles. For instance, Al-Azary (2018) found that participants preferred to employ Low-SND vehicles when creating metaphors, which he suggests may be to reduce the overall semantic richness of the metaphor. Our findings were consistent with this as it was found that High-SND topics decreased in interpretive diversity when framed in terms of a vehicle, suggesting that the vehicle helped to reduce the semantic richness of the topic. Furthermore, as mentioned earlier, Glucksberg et al. (1997) argue that topics provide dimensions that are relevant to the vehicle, which in turn creates metaphor meaning through class attribution. Arguably, when the vehicle is from a low-density neighborhood, the relevant value is easier to find, leading to efficient metaphor meaning resolution.

Metaphor topics and vehicles may differ in other ways as well, and these differences may also contribute to why SND level only affects interpretive diversity for topics but not vehicles. Lakoff and Johnson (1980) posit that the function of metaphor is to comprehend abstract concepts in terms of more concrete, directly experienced concepts. Therefore, metaphor vehicles are often more concrete and imageable than metaphor topics. It is possible that because these vehicles have such semantically rich representations already, there is little room for word associations (as measured by SND) to have an impact. In contrast, word associations have been proposed as a major component for how abstract concepts are represented (Barsalou & Wiemer-Hastings, 2005). To this point, Danguecan and Buchanan (2016) found that SND affects response times associated with processing abstract words but not concrete words. A full analysis of concreteness and imageability is beyond the scope of this paper, but could be an avenue for future research on SND and interpretive diversity, as SND and concreteness tend to interact in various semantic processing tasks (Al-Azary & Buchanan, 2017; Danguecan & Buchanan, 2016).

Lastly, it should be noted that we only employed non-literary "A is B" metaphors, which are somewhat artificial, and it is unclear whether the findings apply to more creative, natural, and literary uses of metaphor. We are currently working on a study with a set of literary metaphors to explore this question.

Conclusion

In this study, we found that words employed as metaphor topics differed in interpretive diversity when these words were presented by themselves or in the context of a metaphor, depending on SND level. Greater SND was associated with increased diversity for words' interpretations when presented alone but, in a reversal, was associated with decreased diversity when presented in the context of a metaphor. We propose that the function of the metaphor vehicle may differ depending on the semantic richness of the metaphor topic. When topics are semantically rich, the vehicle may function to constrain meaning by limiting features to consider. In contrast, for semantically poorer topics, the vehicle may function to highlight aspects of the topic, leading to increased diversity of interpretations.

References

Al-Azary, H. (2018). *Semantic processing of nominal metaphor: Figurative abstraction and embodied simulation* (Doctoral dissertation). Retrieved from https://ir.lib.uwo.ca/etd/5845

Al-Azary, H., & Buchanan, L. (2017). Novel metaphor comprehension: Semantic neighbourhood density interacts with concreteness. *Memory & Cognition*, *45*(2), 296-307.

Al-Azary, H., McAuley, T., Buchanan, L., & Katz, A. N. (2019). Semantic processing of metaphor: A case-study of deep dyslexia. *Journal of Neurolinguistics, 51,* 297-308.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. A. Zwaan (Eds.), *Grounding Cognition: The Role of Perception and Action in Memory, Language and Thinking* (pp. 129-163). Cambridge, England: Cambridge University Press.

Becker, A. H. (1997). Emergent and common features influence metaphor interpretation. *Metaphor and Symbol, 12*(4), 243-259.

Black, M. (1962). Metaphor. In M. Black (Ed.), *Models and Metaphors* (pp. 25-47). Ithaca, NY: Cornell University Press.

Black, M. (1979). More about metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 19-45). Cambridge, England: Cambridge University Press.

Buchanan, E. M., De Deyne, S., & Montefinese, M. (2019). A practical primer on processing semantic property norm data. *Cognitive Processing,* 1-13.

Buchanan, L., Westbury, C., Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8*(3), 531-544.

Clark, S. (2015). Vector space models of lexical meaning. In S. Lappin & C. Fox (Eds.), *The Handbook of Contemporary Semantic Theory, 2nd ed.* (pp. 493-522). Malden, MA: Wiley-Blackwell.

Danguecan, A N., & Buchanan, L. (2016). Semantic neighborhood effects for abstract versus concrete words. *Frontiers in Psychology, 7*, 1034.

Erk, K. (2012). Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass, 6*(10), 635-653.

Gineste, M-D, Indurkhya, B., & Scart, V. (2000). Emergence of features in metaphor comprehension. *Metaphor and Symbol, 15*(3), 117-135.

Glucksberg, S., Manfredi, D. A., & McGlone, M. S. (1997). Metaphor comprehension: How metaphors create new categories. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative Thought: An Investigation of Conceptual Structures and Processes* (pp. 327-350). Washington, DC: American Psychological Association.

Katz, A. N. (1989). On choosing the vehicles of metaphors: Referential concreteness, semantic distances, and individual differences. *Journal of Memory and Language, 28*(4), 486-499.

Katz, A. N., & Al-Azary, H. (2017). Principles that promote bidirectionality in verbal metaphor. *Poetics Today*, *38*(1), 35-59.

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*(4), 191-214.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, *7*(2), 257-266.

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

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211-240.

Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. In *Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics* (pp. 63-70). Philadelphia, PA: Association for Computational Linguistics. Retrieved from https://arxiv.org/abs/cs/0205028

Ortony, A. (1979). Beyond literal similarity. *Psychological Review, 86*(3), 161-180.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532-1543). Doha, Qatar: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/D14-1162/

Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14*(3), 130-137.

Reid, J. N., & Katz, A. N. (2018). Vector space applications in metaphor comprehension. *Metaphor and Symbol, 33*(4), 280-294.

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

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379-423.

Su, C., Huang, S., & Chen, Y. (2017). Automatic detection and interpretation of nominal metaphor based on the theory of meaning. *Neurocomputing*, *219*(5), 300-311.

Trick, L., & Katz, A. N. (1986). The domain interaction approach to metaphor processing: Relating individual differences and metaphor characteristics. *Metaphor and Symbolic Activity, 1*(3), 185-213.

Utsumi, A. (2005). The role of feature emergence in metaphor appreciation. *Metaphor and Symbol, 20*(3), 151-172.

Utsumi, A. (2007). Interpretive diversity explains metaphor—simile distinction. *Metaphor and Symbol, 22*(4), 291-312.

Utsumi, A. (2011). Computational exploration of metaphor comprehension processes using a semantic space model. *Cognitive Science*, *35*(2), 251-296.