Dimensions of Metaphorical Meaning

Andrew Gargett[*], **Josef Ruppenhofer**[†] and **John Barnden**[*]
[*]University of Birmingham
United Kingdom
{a.d.gargett|j.a.barnden}@cs.bham.ac.uk
[†]Hildesheim University
Germany
ruppenho@uni-hildesheim.de

Abstract

Recent work suggests that *concreteness* and *imageability* play an important role in the meanings of figurative expressions. We investigate this idea in several ways. First, we try to define more precisely the context within which a figurative *expression* may occur, by parsing a corpus annotated for metaphor. Next, we add both concreteness and imageability as "features" to the parsed metaphor corpus, by marking up words in this corpus using a psycholinguistic database of scores for concreteness and imageability. Finally, we carry out detailed statistical analyses of the augmented version of the original metaphor corpus, cross-matching the features of concreteness and imageability with others in the corpus such as parts of speech and dependency relations, in order to investigate in detail the use of such features in predicting whether a given expression is metaphorical or not.

1 Introduction

Figurative language plays an important role in "grounding" our communication in the world around us. Being able to talk metaphorically about "the journey of life", "getting into a relationship", whether there are "strings attached" to a contract, or even just "surfing the internet", are important and useful aspects of everyday discourse. Recent work on such phenomena has pursued this kind of grounding in interesting directions, in particular, treating it as a way of injecting meanings that are somehow more "concrete" into daily discourse (Neuman et al., 2013; Turney et al., 2011; Tsvetkov et al., 2013), or else as a way of expressing abstract ideas in terms of concepts that are more "imageable", where imageability can be defined as *how easily a word can evoke mental imagery*, (Cacciari and Glucksberg, 1995; Gibbs, 2006; Urena and Faber, 2010). It should be noted that while it is generally accepted that imageability and concreteness are highly correlated, recent work has shown they are contrastive, in particular, in their interaction with additional cognitive dimensions such as affective states, so that they "can no longer be considered interchangeable constructs" (Dellantonio et al., 2014).

When someone describes love as a journey, or life as a test, one possible way of thinking about what they are doing is that they are trying to cast a fairly abstract idea or concept, such as love or life, in terms of more concrete or imageable experiences or concepts, such as a journey or a test. More formally, metaphor can be characterized as the mapping of properties from a "source" domain concept (typically more concrete) on to a "target" domain concept (typically more abstract). However, despite the ease with which people understand both established metaphors such as these, or even more novel ones[1], and despite well-established findings about the ubiquity of metaphor in everyday discourse (Lakoff and Johnson, 1980), explicit and testable proposals for the mechanisms underlying such forms of expression remain elusive.

When looking for such mechanisms, it seems natural to start with the patterns of language that so effectively convey metaphorical meanings. Along these lines, Deignan (2006) argues that:

---

[1]Consider how readily one can make sense of a novel, yet metaphorical utterance, such as "life is a box of chocolates" (from a recent film), despite never having heard it before.

> [M]etaphorical uses of words show differences in their grammatical behavior, or even their word class, when compared to their literal use. In addition, it shows that metaphorical uses of a word commonly appear in distinctive and relatively fixed syntactic patterns.

Focusing on word class of figurative expressions, so-called content words, such as nouns, adjectives and verbs, have long been considered to more strongly convey figurative meanings than so-called function words, such as prepositions (Neuman et al., 2013; Tsvetkov et al., 2013). Yet, Steen et al. (2010) find prepositions within figurative expressions to be as prevalent as content words such as nouns and verbs, and indeed, for particular genres (such as academic texts) prepositions are the most frequently attested part of speech for figurative expressions.

Further, there has been work on the interaction between metaphorical expressions and syntactically defined contexts (e.g. phrase, clause, sentence). For example, Neuman et al. (2013) investigate how metaphorical expressions apparently pattern by syntactically definable types, specifically: Type I, where "a subject noun is associated with an object noun via a form of the copula verb *to be*" (e.g. "God is a king"), Type II having the verb as "the focus of the metaphorical use representing the act of a subject noun on an object noun" (e.g. "The war absorbed his energy"), and Type III "involve an adjective-noun phrase" (e.g. "sweet girl"). While such work yields a useful typology of figurative expressions, such investigations into the syntactic patterns of figurative forms of expression is far from exhaustive. It would be useful to take this further somewhat, with a more rigorous, syntactically precise definition of the context of occurrence of figurative language.

Motivated by the above considerations, we have begun investigating the interaction of *concreteness* and *imageability* with figurative meanings in several ways. This paper reports the initial stages of this ongoing work into the dimensions of meaning of figurative language such as metaphor. As part of this work, we have attempted to define more precisely the context within which a figurative *expression* may occur, by parsing a corpus annotated for metaphor, the Vrije University Amsterdam Metaphor Corpus (VUAMC) (Steen et al., 2010), using an off the shelf dependency parser, the Mate parser (Bohnet, 2010). In addition, we add both concreteness and imageability as "features" to the dependency parsed metaphor corpus, by marking up words in this corpus using a psycholinguistic database of scores for concreteness and imageability, the MRC Psycholinguistic Database (Wilson, 1988). In this paper, we report detailed statistical analyses we have carried out of the resulting data set, cross-matching the features of concreteness and imageability with others in the corpus such as parts of speech (PsOS) and dependency relations, in order to investigate in detail the use of such features in determining whether a given expression is metaphorical or not.

2 Method

2.1 Data

Our data comes from the Vrije University Amsterdam Metaphor Corpus (VUAMC), consisting of approximately 188,000 words selected from the British National Corpus-Baby (BNC-Baby), and annotated for metaphor using the Metaphor Identification Procedure (MIP) (Steen et al., 2010). The corpus has four registers, of between 44,000 and 50,000 words each: academic texts, news texts, fiction, and conversations. We have chosen this corpus because of its broad coverage and its rich metaphorical annotation.

2.2 Procedure

PRE-PROCESSING. We have enriched the VUAMC in several ways. First, we have parsed the corpus using the graph-based version of the Mate tools dependency parser (Bohnet, 2010), adding rich syntactic information.[2] Second, we have incorporated the MRC Psycholinguistic Database[3] (Wilson, 1988), a dictionary of 150,837 words, with different subsets of these words having been rated by human subjects in psycholinguistic experiments. Of special note, the database includes 4,295 words rated with degrees of abstractness, these ratings ranging from 158 (meaning *highly abstract*) to 670 (meaning *highly concrete*),

---

[2]https://code.google.com/p/mate-tools/
[3]http://ota.oucs.ox.ac.uk/headers/1054.xml

and also 9,240 words rated for degrees of imageability, which can be defined as *how easily a word can evoke mental imagery*, these ratings also ranging between 100 and 700 (a higher score indicating greater imageability). It should be noted that it has long been known that the concreteness and imageability scores are highly correlated (Paivio et al., 1968), however, there are interesting differences between these sets of scores (Dellantonio et al., 2014), and we are currently investigating these differences in further studies (see Section (4) below). These scores have been used extensively for work that is similar to ours, e.g. (Neuman et al., 2013; Turney et al., 2011; Tsvetkov et al., 2013), and while our work is also largely computational in approach, a significant component of our research is devoted to investigating in some detail the cognitive aspects of figurative meanings.

EXPERIMENTAL DESIGN. We carried out five studies, all beginning with pre-processing tasks to prepare the data (additional to those listed immediately above, undertaken to prepare the entire corpus for these studies). We list the aims, details of pre-processing, and hypotheses below.

Study 1. This study initiated the investigation, and guided the setting up of the computational framework for our broader research activities. The VUAMC was extended with dependency information from the Mate dependency parser, enabling extraction of both dependency information and metaphorical annotation for each VUAMC word.[4] Hypotheses: H_1 = nouns are more prevalent in metaphorical expressions than verbs, verbs more than adjectives, adjectives more than prepositions; H_2 = metaphorical expressions are more likely to occur in sentences in which other metaphorical expressions occur.

Study 2. This study aimed to evaluate claims about syntactically-defined metaphor types (Neuman et al., 2013), and search for other types. The structure of a sentence revealed by a dependency parse is based on the relation between a word, known as a *head*, and its *dependents*. This extended VUAMC data provided variables for metaphor types I, II and III, respectively, *Noun-BE-Noun*, *Noun-ActiveVerb-Noun*, and *Adjective-Noun*, as well as the discovery of additional metaphor types.

Study 3. Going further than Studies 1 and 2, this study extended the VUAMC data with MRC concreteness and imageability scores, plus further processing of the VUAMC corpus, assigning MRC scores to each item in this corpus. Note here that the VUAMC data was examined word-by-word (rather than sentence-by-sentence, as for Study 2). However, the VUAMC data set is much larger than the MRC data set, so that many VUAMC words have no MRC scores. To smooth this discrepancy, for this initial stage of our investigations, we have implemented the fairly rudimentary approach of calculating global MRC scores by: first, from VUAMC words with MRC scores, a global average MRC score for each part of speech of the VUAMC data was calculated, and second, those VUAMC words without MRC scores (i.e. missing from the MRC database) were assigned a global score based on their part of speech. Of course, a range of possible smoothing strategies are available, and while at this stage we are employing a rather crude averaging of the score, this is an area we intend to investigate further in follow-up studies, inspired by the more sophisticated methods that have been implemented by others, e.g. (Feng et al., 2011; Tsvetkov et al., 2013).[5] For this study, we sought to answer the following two questions: Do concreteness and imageability scores correlate with metaphoricity of expressions? Do concreteness and imageability scores correlate with parts of speech of metaphorical expressions?

Study 4. This study replicated Study 3, but also considered the data sentence-by-sentence (cf. Study 2), to integrate syntactic information and MRC score. Examining MRC scores across syntactically fine-grained contexts, enabled collecting information about heads, their dependent/s, as well as the dependency relation/s, and this information could then be examined to see if it helped to distinguish between literal and nonliteral items. This approach enables us to investigate in detail the contexts in which *concreteness* and *imageability* with figurative meanings, a key aim of our work, as pointed out in Section (1). Hypotheses: H_3 = metaphorical expressions are more likely to occur in sentences where the head is more

---

[4]For more details on the VUAMC categories, see: `http://www.natcorp.ox.ac.uk/docs`.

[5]This work is part of a larger project, `http://www.cs.bham.ac.uk/~gargetad/genmeta-about.html`, which aims to annotate larger web-based corpora of discourse on illness and political conflict.
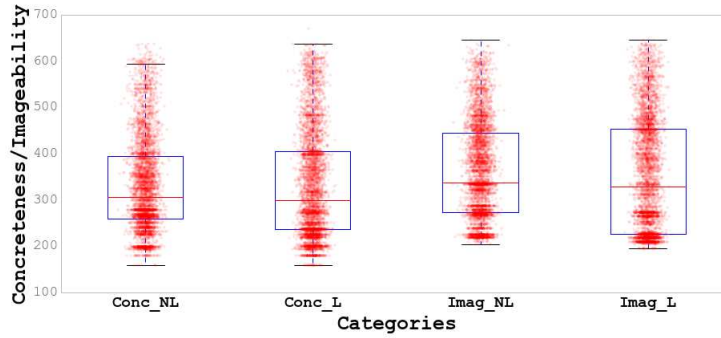
Figure 1: Plots of concreteness vs. imageability scores for literal vs. nonliteral words in the VUAMC (Conc=concreteness, Imag=imageability, NL=nonliteral, L=literal)

concrete than the dependent/s; H_4 = metaphorical expressions are more likely to occur in sentences where the head is more imageable than the dependent/s.

Study 5.   Finally, this study finished by examining the relative importance of the variables identified so far, for predicting literal vs. nonliteral expressions, another key aim of our work (as mentioned in Section (1)). We implemented this study through building and evaluating a series of logistic regression models.

3   Results

3.1   Study 1

The first hypothesis listed for this study above has not been refuted, with the percentage of all nonliteral sentences in our collection having only one nonliteral item being 27%, while the percentage of all nonliteral sentences having more than one nonliteral item is 73%: so after finding one nonliteral item in a sentence, we can expect to find more. Regarding the second hypothesis, our data set had the following proportions of occurrence of nonliteral items according to parts of speech: Adjectives=10.8%, Prepositions=28%, Nouns=22.5%, Verbs=27%, Adverbs=5%, Pronouns=0.2%, Conjunctions=0.5%, Other=6%. Consistent with Steen et al. (2010), that function words can occur more frequently than content words in metaphorical expressions, we found prepositions to be far more prevalent than adjectives in such expressions, and occur about as frequently as verbs.

3.2   Study 2

We found the following percentages of metaphor types (across all metaphors): Type I = 3.06%, Type II = 33.53%, Type III = 7.56% (note the reversal for Type II vs. Type III, contrary to (Neuman et al., 2013)). Such differences may be due to differences in data sets, as well as different syntactic models.[6] Additionally, we found a pattern of expression we have dubbed "Type IV" metaphors, consisting of preposition as head, together with noun phrase dependents (e.g. "at the end of the decade", "after the break-up"): these account for 35.53% of the total occurrence of metaphors.

3.3   Study 3

The boxplots in Figure (1) compare concreteness and imageability scores for nonliteral vs. literal items, suggesting nonliteral and literal items are indistinguishable from one another with respect to their concreteness and imageability scores. Next, we further categorise our data according to parts of speech, the boxplots in Figure (2) showing results for concreteness, and the boxplots Figure (3) presenting results for imageability – these figures suggest literal and nonliteral items can be better distinguished, with respect to their concreteness and imageability scores, by increasing the granularity of annotation of the context (e.g. by including parts of speech). Note that imageability scores for prepositions seem to show the

---

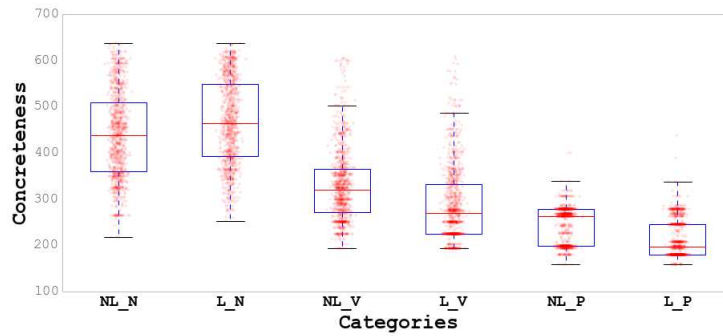[6]Neuman et al. (2013) used the Stanford Dependency Parser (De Marneffe and Manning, 2008).

Figure 2: Plots of concreteness scores for literal vs. nonliteral/metaphorical words in the VUAMC, grouped by parts of speech (L=literal, NL=nonliteral, N=noun, V=verb, P=preposition)
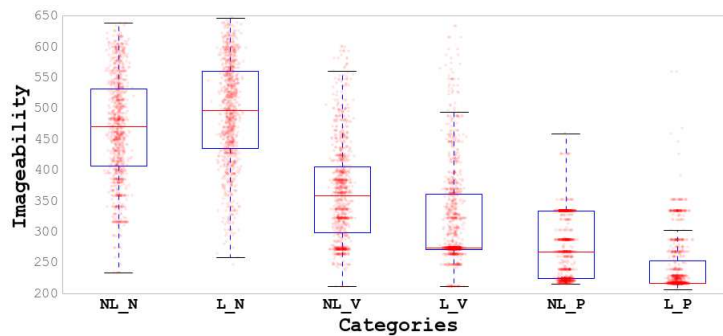

Figure 3: Plots of imageability scores for literal vs. nonliteral/metaphorical words in the VUAMC, grouped by parts of speech (L=literal, NL=nonliteral, N=noun, V=verb, P=preposition)

clearest distinction between literal vs. nonliteral items. But can we do better? What further categories in the data should we focus on in order to achieve even clearer distinctions between literal vs. nonliteral items?

3.4 Study 4

Figures (4) and (5) show the variation that can be achieved by making a more fine-grained distinction within our data set between heads and their dependents, plus MRC scores of each. Figure (4) shows that concreteness scores enable distinguishing between literal and nonliteral items for some parts of speech, such as nouns, where nonliteral heads have higher MRC scores than their dependents, distinct from literal head nouns (verbs appear to make no such a distinction). While literal and nonliteral head prepositions both seem indistinguishable from their dependents in terms of concreteness scores, nonliteral head prepositions seem to have imageability scores quite distinct from their dependents.

3.5 Study 5

Based on our previous studies, we here examine the following 5 independent variables: **POS** = part of speech of the head, **C_Head** = concreteness score of the head, **I_Head** = imageability score of the head, **C_Dep** = average concreteness score of the dependents, **I_Dep** = average imageability score of the dependents. Table (1) sets out the results for 7 logistic regression models we tested, and formulas representing these models **M1** to **M7** are as follows (**Nonliteral** of course being the dependent variable, its values being either "yes, this is nonliteral" or "no, this is not nonliteral"):
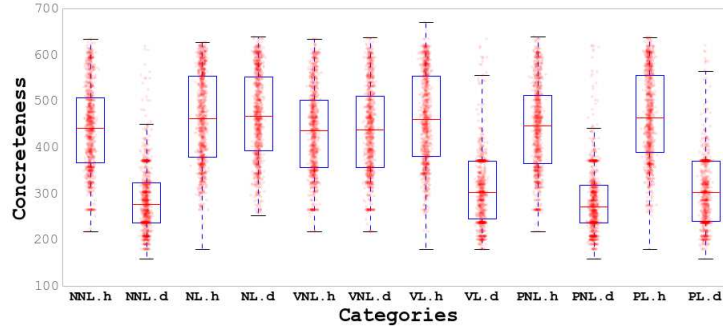
Figure 4: Plots of concreteness scores for literal vs. nonliteral/metaphorical heads vs. their dependents, in the VUAMC, grouped by parts of speech (L=literal, NL=nonliteral, N=noun, V=verb, P=preposition, h=head, d=dependents)
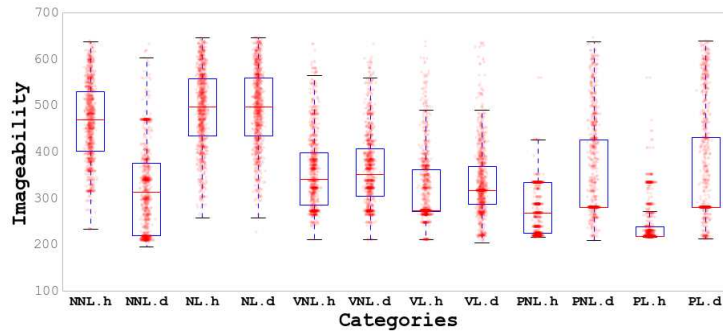

Figure 5: Plots of imageability scores for literal vs. nonliteral/metaphorical heads vs. their dependents, in the VUAMC, grouped by parts of speech (L=literal, NL=nonliteral, N=noun, V=verb, P=preposition, h=head, d=dependents)

M1: $Nonliteral \sim POS + C\_Head + I\_Head + C\_Dep + I\_Dep$
M2: $Nonliteral \sim C\_Head + I\_Head$
M3: $Nonliteral \sim POS + C\_Head + I\_Head$
M4: $Nonliteral \sim POS + C\_Head + C\_Dep + I\_Dep$
M5: $Nonliteral \sim POS + I\_Head + C\_Dep + I\_Dep$
M6: $Nonliteral \sim POS + C\_Head + C\_Dep$
M7: $Nonliteral \sim POS + I\_Head + I\_Dep$

In Table (1), p-values have three categories, $p < .0001$, $p < .001$, or $p < .01$: this value represents a test of the null hypothesis that the coefficient of the variable being considered is zero, i.e., the variable has no effect on the model (a lower p-value is stronger evidence for rejecting the null hypothesis). Where variables have significantly low p-values, Table (1) in effect presents optimal combinations of variables for specific models, with low p-values indicating variables likely to have a greater effect on the model and so more directly reflecting changes in the independent variable. For example, Table (1) shows that models selecting MRC scores for heads (e.g. **C_Head**) with the same kinds of scores for their dependents (e.g.**C_Dep**) seem most successful, which is perhaps to be expected, in light of studies 3 and 4.

It should be noted that no single variable models are reported here, since (1) while models such as $Nonliteral \sim I\_Head$ and $Nonliteral \sim C\_Head$ indeed achieve significant p-values, others such as $Nonliteral \sim I\_Dep$ and $Nonliteral \sim C\_Dep$ do not, (2) single variable models do not explain Figure (1), nor indeed the variation for multiple variable contexts as exhibited by Figures (4) and (5). We are currently comparing single vs. multiple variables, and early machine learning results suggest multiple variable models are superior compared to single variable models as predictive tools.

Variables	M1	M2	M3	M4	M5	M6	M7
Intercept	-7.534***	-2.609*	-9.088***	-7.836***	-7.522***	-7.816***	-7.614***
POS	9.265***		8.884***	9.330***	9.163***	9.316***	9.082***
C_Head	1.555	0.288	1.382	4.844***		4.876***	
I_Head	0.459	-1.312	0.513		4.611***		4.660***
C_Dep	-1.964			-1.982	-1.919	-3.799***	
I_Dep	0.682			0.699	0.660		-3.325**

Table 1: Results (t scores) of logistic regression model for predicting non/literal items from the VUAMC, n=1855 (nb. p-values are shown by asterisks, ***=p<.0001, **=p<.001, *=p<.01)

4   Discussion

This paper reports results from ongoing work we are carrying out toward building a tool for identifying metaphorical expressions in everyday discourse, through fine-grained analysis of the dimensions of meaning of such expressions. We have presented evidence that detecting metaphor can usefully be pursued as the problem of modeling how conceptual meanings such as concreteness and imageability, interact with syntactically definable linguistic contexts. We increase the granularity of our analyses by incorporating detailed syntactic information about the context in which metaphorical expressions occur. By increasing the granularity of context, we were able to distinguish between metaphorical expressions according to different parts of speech, and further, according to heads and their dependents.

We were able to show that for the purpose of determining whether a specific linguistic expression is metaphorical or not, the most successful approach seems to be to combine information about parts of speech with either concreteness scores for both heads and their dependents, or else with imageability scores for both heads and their dependents. Note that this result is in part a direct consequence of the high correlation between concreteness and imageability, whereby their combination will typically not result in an optimal regression model. Such high correlation between concreteness and imageability has been understood for some time (Paivio et al., 1968), yet, of course, there is good reason to think that concreteness and imageability do not in fact pattern identically, and that they are at some level distinct phenomena. Indeed, concreteness and imageability are likely related to distinct cognitive systems, and we are currently undertaking further investigations in this direction.

Finally, we should note that while our results are likely to be language-specific, it is reasonable to assume the general approach could be replicated across languages. We are currently planning such cross-linguistic research for future work.

Acknowledgements

References

Bernd Bohnet. 2010. Very high accuracy and fast dependency parsing is not a contradiction. In *The 23rd International Conference on Computational Linguistics (COLING 2010), Beijing, China*.

Christina Cacciari and Sam Glucksberg. 1995. Imaging idiomatic expressions: literal or figurative meanings. In Martin Everaert, Erik-Jan van der Linden, Andr Schenk, and Rober Schreuder, editors, *Idioms: Structural and psychological perspectives*, pages 43–56. Lawrence Erlbaum.

Marie-Catherine De Marneffe and Christopher D Manning. 2008. The stanford typed dependencies representation. In *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, pages 1–8. Association for Computational Linguistics.

Alice Deignan. 2006. The grammar of linguistic metaphors. In Anatol Stefanowitsch and Stefan Gries, editors, *Corpus-based approaches to metaphor and metonymy*, pages 106–122. Walter de Gruyter.

Sara Dellantonio, Claudio Mulatti, Luigi Pastore, and Remo Job. 2014. Measuring inconsistencies can lead you forward: The case of imageability and concreteness ratings. *Frontiers in Psychology*, 5(708).

Shi Feng, Zhiqiang Cai, Scott A Crossley, and Danielle S McNamara. 2011. Simulating human ratings on word concreteness. In *FLAIRS Conference*.

Raymond W Gibbs. 2006. Metaphor interpretation as embodied simulation. *Mind & Language*, 21(3):434–458.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago.

Yair Neuman, Dan Assaf, Yohai Cohen, Mark Last, Shlomo Argamon, Newton Howard, and Ophir Frieder. 2013. Metaphor identification in large texts corpora. *PloS one*, 8(4):e62343.

Allan Paivio, John C Yuille, and Stephen A Madigan. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology*, 76(1, pt.2):1–25.

G.J. Steen, A.G. Dorst, J.B. Herrmann, A.A. Kaal, and T. Krennmayr. 2010. *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. Converging Evidence in Language and Communication Research. John Benjamins Publishing Company.

Yulia Tsvetkov, Elena Mukomel, and Anatole Gershman. 2013. Cross-lingual metaphor detection using common semantic features. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 45–51, Atlanta, Georgia, June. Association for Computational Linguistics.

Peter D Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 Conference on the Empirical Methods in Natural Language Processing*, pages 680–690.

Jose Manuel Urena and Pamela Faber. 2010. Reviewing imagery in resemblance and non-resemblance metaphors. *Cognitive Linguistics*, 21(1):123–149.

Michael Wilson. 1988. MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, 20(1):6–10.