Detecting Metaphor by Contextual Analogy

Eirini Florou
Dept of Linguistics, Faculty of Philosophy
University of Athens, Greece
eirini.florou@gmail.com

Abstract

As one of the most challenging issues in NLP, metaphor identification and its interpretation have seen many models and methods proposed. This paper presents a study on metaphor identification based on the semantic similarity between literal and non literal meanings of words that can appear at the same context.

1 Introduction

A *metaphor* is a literary figure of speech that describes a subject by asserting that it is, on some point of comparison, the same as another otherwise unrelated object. Metaphor is a type of analogy and is closely related to other rhetorical figures of speech that achieve their effects via association, comparison or resemblance including allegory, hyperbole, and simile. Rhetorical theorists and other scholars of language have discussed numerous dimensions of metaphors, though these nomenclatures are by no means universal nor necessarily mutually exclusive.

A very challenging task in linguistics is the metaphor identification and the its interpretation. *Metaphor identification procedure (MIP)* is a method for identifying metaphorically used words in discourse. It can be used to recognize metaphors in spoken and written language. The procedure aims to determine the relationship of a particular lexical unit in the discourse and recognize its use in a particular context as possibly metaphorical. Since many words can be considered metaphorical in different contexts, MIP requires a clear distinction between words that convey metaphorical meaning and those that do not, despite the fact that language generally differs in the degrees of metaphoricity.

In this paper we propose a method for identifying metaphorical usage in verbs. Our method is looking for semantic analogies in the context of a verb by comparing it against prior known instances of literal and non-literal usage of the same verb in different contexts. After discussing the metaphor identication literature (Section 2), we proceed to present our research proposal (Section 3) and to present and discuss our first experiments based on WordNet similarity measures (Section 4). Experiment results help us to draw conclusions and insights about analogical reasoning and memory-based learning for this task and to outline promising research paths (Section 5).

2 Background

According to Lakoff and Johnson (1980), metaphor is a productive phenomenon that operates at the level of mental processes. Metaphor is thus not merely a property of language, but rather a property of thought. This view was subsequently acquired and extended by a multitude of approaches (Grady, 1997; Narayanan, 1997; Fauconnier and Tuner, 2002; Feldman, 2006; Pinker, 2007) and the term *conceptual metaphor* was adopted to describe it.

In cognitive linguistics, conceptual metaphor, or cognitive metaphor, refers to the understanding of an idea, or conceptual domain, in terms of another, for example, understanding quantity in terms of directionality as in, for example, 'prices are rising'. A conceptual metaphor uses an idea and links it to another in order to better understand something. It is generaly accepted that the conceptual metaphor of viewing communication as a conduit is a large theory explained with a metaphor. These metaphors are prevalent in communication and everyone actually perceives and acts in accordance with the metaphors.

2.1 Metaphor Identification

Automatic processing of metaphor can be clearly divided into two subtasks: metaphor identifica-

Proceedings of the ACL Student Research Workshop, pages 23–30,
Sofia, Bulgaria, August 4-9 2013.

tion (distinguishing between literal and metaphorical language in text) and metaphor interpretation (identifying the intended literal meaning of a metaphorical expression). Both of them have been repeatedly attempted in NLP.

The most influential account of metaphor identification is that of Wilks (1978). According to Wilks, metaphors represent a violation of selectional restrictions in a given context. Consider an example such as *My car drinks gasoline*; the verb *drink* normally takes an *animate* subject and a *liquid* object.

This approach was automated by Fass (1991) in his MET* system. However, Fass himself indicated a problem with the method: it detects any kind of non-literalness or anomaly in language (metaphors, metonymies and others), i.e., it overgenerates with respect to metaphor. The techniques MET* uses to differentiate between those are mainly based on hand-coded knowledge, which implies a number of limitations. First, literalness is distinguished from non-literalness using selectional preference violation as an indicator. In the case that non-literalness is detected, the respective phrase is tested for being a metonymic relation using hand-coded patterns. If the system fails to recognize metonymy, it proceeds to search the knowledge base for a relevant analogy in order to discriminate metaphorical relations from anomalous ones.

Berber Sardinha (2002) describes a collocation-based method for spotting metaphors in corpora. His procedure is based on the notion that two words sharing collocations in a corpus may have been used metaphorically. The first step was to pick out a reasonable number of words that had an initial likelihood of being part of metaphorical expressions. First, words with marked frequency (in relation to a large general corpus of Portuguese) were selected. Then, their collocations were scored for closeness in meaning using a program called 'distance' (Padwardhan et al., 2003), under the assumption that words involved in metaphorical expressions tend to be denotationally unrelated. This program accesses WordNet in order to set the scores for each word pair. The scores had to be adapted in order for them to be useful for metaphor analysis. Finally, those words that had an acceptable semantic distance score were evaluated for their metaphoric potential. The results indicated that the procedure did pick up some major metaphors in the corpus, but it also captured metonyms.

Another approach to finding metaphor in corpora is CorMet, presented by Mason (2004). It works by searching corpora of different domains for verbs that are used in similar patterns. When the system spots different verbs with similar selectional preferences (i.e., with similar words in subject, object and complement positions), it considers them potential metaphors.

CorMet requires specific domain corpora and a list of verbs for each domain. The specific domain corpora are compiled by searching the web for domain-specific words. These words are selected by the author, based on his previous knowledge of subject areas and are stemmed. The most typical verbs for each specific corpus are identified through frequency markedness, by comparing the frequencies of word stems in the domain corpus with those of the BNC. The resulting words have a frequency that is statistically higher in the domain corpus than in the reference corpus. These stems are then classified according to part of speech by consulting WordNet.

Alternative approaches search for metaphors of a specific domain defined a priori in a specific type of discourse. The method by Gedigian et al. (2006) discriminates between literal and metaphorical use. They trained a maximum entropy classifier for this purpose. They obtained their data by extracting the lexical items whose frames are related to MOTION and CURE from FrameNet (Fillmore et al., 2003). Then, they searched the PropBank Wall Street Journal corpus (Kingsbury and Palmer, 2002) for sentences containing such lexical items and annotated them with respect to metaphoricity.

Birke and Sarkar (2006) present a sentence clustering approach for non-literal language recognition implemented in the TroFi system (Trope Finder). This idea originates from a similarity-based word sense disambiguation method developed by Karov and Edelman (1998). The method employs a set of seed sentences, where the senses are annotated, computes similarity between the sentence containing the word to be disambiguated and all of the seed sentences and selects the sense corresponding to the annotation in the most similar seed sentences. Birke and Sarkar (2006) adapt this algorithm to perform a two-way classification: literal vs. non-literal, and they do not clearly define the kinds of tropes they aim to discover. They attain a performance of 53.8% in terms of f-score.

Both Birke and Sarkar (2006) and Gedigian et al. (2006) focus only on metaphors expressed by a verb. As opposed to that the approach of Krishnakumaran and Zhu (2007) deals with verbs, nouns and adjectives as parts of speech. They use hyponymy relation in WordNet and word bigram counts to predict metaphors at the sentence level. Given an IS-A metaphor (e.g. The world is a stage) they verify if the two nouns involved are in hyponymy relation in WordNet, and if this is not the case then this sentence is tagged as containing a metaphor. Along with this they consider expressions containing a verb or an adjective used metaphorically. Hereby they calculate bigram probabilities of verb-noun and adjective-noun pairs (including the hyponyms/hypernyms of the noun in question). If the combination is not observed in the data with sufficient frequency, the system tags the sentence containing it as metaphorical. This idea is a modification of the selectional preference view of Wilks (1978).

Berber Sardinha (2010) presents a computer program for identifying metaphor candidates, which is intended as a tool that can help researchers find words that are more likely to be metaphor vehicles in a corpus. As such, it may be used as a device for signalling those words that the researcher might want to focus on first, because these have a higher probability of being metaphors in their corpus, or conversely, it may indicate those words that are worth looking at because of their apparent low probability of being metaphors. The program is restricted to finding one component of linguistic metaphors and has been trained on business texts in Portuguese, and so it is restricted to that kind of text.

Shutova et al. (2012) present an approach to automatic metaphor identification in unrestricted text. Starting from a small seed set of manually annotated metaphorical expressions, the system is capable of harvesting a large number of metaphors of similar syntactic structure from a corpus. Their method captures metaphoricity by means of verb and noun clustering. Their system starts from a seed set of metaphorical expressions exemplifying a range of source-target domain mappings; performs unsupervised noun clustering in order to harvest various target concepts associated with the same source domain; by means of unsupervised verb clustering creates a source domain verb lexicon; searches the BNC for metaphorical expressions describing the target domain concepts using the verbs from the source domain lexicon. According to Shutova et al. (2012), abstract concepts that are associated with the same source domain are often related to each other on an intuitive and rather structural level, but their meanings, however, are not necessarily synonymous or even semantically close. The consensus is that the lexical items exposing similar behavior in a large body of text most likely have the same meaning. They tested their system starting with a collection of metaphorical expressions representing verb-subject and verb-object constructions, where the verb is used metaphorically. They evaluated the precision of metaphor identification with the help of human judges. Shutova's system employing unsupervised methods for metaphor identification operates with precision of 0.79.

For verb and noun clustering, they used the *subcategorization frame acquisition* system by Preiss et al. (2007) and *spectral clustering* for both verbs and nouns. They acquired selectional preference distributions for Verb-Subject and Verb-Object relations from the BNC parsed by RASP; adopted Resnik's selectional preference measure; and applied to a number of tasks in NLP including word sense disambiguation (Resnik, 1997).

3 Detecting the metaphor use of a word by contextual analogy

The first task for metaphor processing is its identification in a text. We have seen above how previous approaches either utilize hand-coded knowledge (Fass, 1991), (Krishnakumaran and Zhu, 2007) or reduce the task to searching for metaphors of a specific domain defined a priori in a specific type of discourse (Gedigian et al., 2006).

By contrast, our research proposal is a method that relies on distributional similarity; the assumption is that the lexical items showing similar behaviour in a large body of text most likely have related meanings. Noun clustering, specifically, is central to our approach. It is traditionally assumed that noun clusters produced using distributional clustering contain concepts that are similar to each other.

3.1 Word Sense Disambiguation and Metaphor

One of the major developments in metaphor research in the last several years has been the focus on identifying and explicating metaphoric language in real discourse. Most research in Word Sense Disambiguation has concentrated on using contextual features, typically neighboring words, to help infer the correct sense of a target word. In contrast, we are going to discover the predominant sense of a word from raw text because the first sense heuristic is so powerful and because manually sense-tagged data is not always available.

In word sense disambiguation, the first or predominant sense heuristic is used when information from the context is not sufficient to make a more informed choice. We will need to use parsed data to find distributionally similar words (nearest neighbors) to the target word which will reflect the different senses of the word and have associated distributional similarity scores which could be used for ranking the senses according to prevalence.

The predominant sense for a target word is determined from a prevalence ranking of the possible senses for that word. The senses will come from a predefined inventory which might be a dictionary or WordNet-like resource. The ranking will be derived using a distributional thesaurus automatically produced from a large corpus, and a semantic similarity measure will be defined over the sense inventory. The distributional thesaurus will contain a set of words that will be 'nearest neighbors' Lin (1998) to the target word with respect to similarity of the way in which they will be distributed. The thesaurus will assign a distributional similarity score to each neighbor word, indicating its closeness to the target word.

We assume that the number and distributional similarity scores of neighbors pertaining to a given sense of a target word will reflect the prevalence of that sense in the corpus from which the thesaurus was derived. This is because the more prevalent senses of the word will appear more frequently and in more contexts than other, less prevalent senses. The neighbors of the target word relate to its senses, but are themselves word forms rather than senses. The senses of the target word have to be predefined in a sense inventory and we will need to use a semantic similarity score which will be defined over the sense inventory to relate the neighbors to the various senses of the target word.

The measure for ranking the senses will use the sum total of the distributional similarity scores of the k nearest neighbors. This total will be divided between the senses of the target word by apportioning the distributional similarity of each neighbor to the senses. The contribution of each neighbor will be measured in terms of its distributional similarity score so that 'nearer' neighbors count for more. The distributional similarity score of each neighbor will be divided between the various senses rather than attributing the neighbor to only one sense. This is done because neighbors can relate to more than one sense due to relationships such as systematic polysemy. To sum up, we will rank the senses of the target word by apportioning the distributional similarity scores of the top k neighbors between the senses. Each distributional similarity score (dss) will be weighted by a normalized semantic similarity score (sss) between the sense and the neighbor.

We chose to use the distributional similarity score described by Lin (1998) because it is an unparameterized measure which uses pointwise mutual information to weight features and which has been shown Weeds et al. (2004) to be highly competitive in making predictions of semantic similarity. This measure is based on Lin's information-theoretic similarity theorem (Lin, 1997) : The similarity between A and B is measured by the ratio between the amount of information needed to state the commonality of A and B and the information needed to fully describe what A and B are.

3.2 Similarity-based metaphorical usage estimation

After the noun clustering and finding the predominant sense of an ambiguous word, as the local context of this word can give important clues to which of its senses was intended, the metaphor identification system will start from a small set of seed metaphors (the seed metaphors are a model extracted from metaphor-annotated and dependency-parsed sentences), to point out if a word is used literaly or non literaly at the certain context. For the purposes of this work as context should be considered the verb of the seed metaphors. We are going to take as seed metaphors the examples of Lakoff's Master Metaphor List (Lakoff et al., 1991).

Then, as we will have already find the k nearest neighbors for each noun and we will have created clusters for nouns which can appear at the same context, we will be able to calculate their semantic similarity. We then will use the WordNet similarity package Padwardhan et al. (2003) in order to measure the semantic similarity between each member of the cluster and the noun of the annotated metaphor. The WordNet similarity package supports a range of WordNet similarity scores. We will experiment using a lot of these in order to find those which perform the best. Each time, we want to estimate if the similarity between the target noun and the seed metaphor will be higher than the similarity between the target noun and another literal word which could appear at the certain context. Calculating the target word's semantic similarity with the seed words (literal or non literal) we will be able to find out if the certain word has a literal or metaphorical meaning at the concrete context.

By this way, starting from an already known metaphor, we will be able to identify other non literal uses of words which may appear at the same context, estimating the similarity measure of the target word between the seed metaphor and another literal meaning of a word at the same context. If the semantic similarity's rate of the target word (for instance the word 'assistance' at the context of the verb 'give') and the annotated metaphor (like 'quidance' at the certaincontext) is higher that the rate of the target word and the seed word with the literal meaning (for example the word 'apple' at the same context) , then we will be able to assume that the tartget word is used metaphorically, at the concrete context.

4 First Experiments and Results

In order to evaluate our method we search for common English verbs which can take either literal or non literal predicates. As the most common verbs (be, have and do) can function as verbs and auxiliary verbs, we didn't use them for our experiments. As a consequence, we chose common function verbs which can take a direct object as predicate. More specifically, at our experiments we concentrated on literal and non literal predicates of the verbs: *break, catch, cut, draw, drop, find, get, hate, hear, hold, keep, kill, leave, listen, lose, love, make, pay, put, save, see, take, want*.

We used the *VU Amsterdam Metaphor Corpus*[1] in order to extract data for our experiments. We used shallow heuristics to match verbs and direct objects, with manually checking and correcting the result. We have also used the *British National Corpus (BNC)*, in order to take more samples, mostly literal. In the case of he BNC, we were able to extract the direct object from the depency parses, but had manually controlled metaphorical vs. literal usage. In all, we collected 124 instances of literal usage and 275 instances of non-literal usage involving 311 unique nouns.

With this body of literal and non-literal contexts, we tried every possible combination of one literal and one non-literal object for each verb as seed, and tested with the remaining words. The mean results are collected in Table 1, where we see how the LCS-based measures by Resnik (1997) and Wu and Palmer (1994) performed the best.

One observation is that the differences between the different measures although significant, they are not as dramatic as to effect reversals in the decision. This is apparent in the *simple voting* results (right-most column in Table 1) where all measures yield identical results. Only when differences in the similarities accumulate before the comparison between literal and non-literal context is made (three left-most columns in Table 1), does the choice of similarity measure make a difference.

Another observation pertains to relaxing the dependency on WordNet so that method can be based on similarities defined over more widely available lexical resources. In this respect, the low F-score by the adapted Lesk measure is not very encouraging, as variations of the Lesk measure could be defined over the glosses in digital dictionaries without explicit WordNet-style relations. Combined with the high valuation of methods using the LCS, this leads us to conclude that the relative taxonomic position is a very important factor.

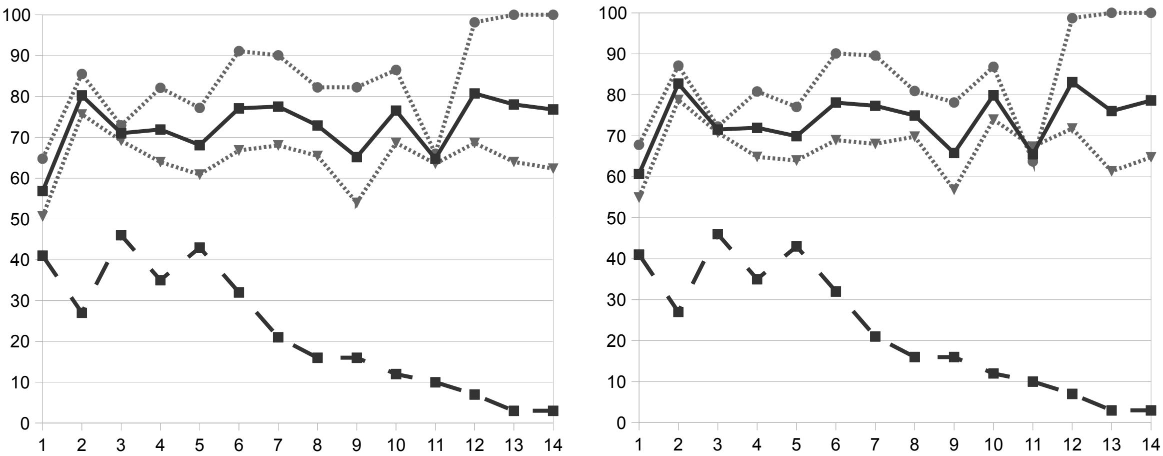
Finally, and happily counter to our prior intuition, we would like to note the robustness of the method to the number of different senses test words have: plotting the F-score against the number of senses did not result in consistently deteriorating results as the senses multiply (Figure 1).[2] If this had happened, we would have con-

[1] Please see `http://www.metaphorlab.vu.nl/en/research/funded_research/VU-Amsterdam-Metaphor-Corpus`

[2] Although some of the nouns in our collection have as many as 33 senses, we have only plotted the data for up to 15 senses; the data is too sparse to be reasonably usuable beyond that point.

Table 1: Fβ=1 scores for all combinations of seven different similarity measures and five ways of deriving a single judgement on literal usage by testing all senses of a word against all senses of the seed words.

Measure	Maximum		Average		Sum		Simple Voting	
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
Adapted Lesk	63.87	6.96	63.39	9.41	64.77	6.47	68.64	10.69
Jiang *et al.* (1997)	70.92	9.19	64.31	8.41	65.14	6.45	68.64	10.69
Lin (1998)	71.35	10.70	70.39	10.02	70.07	9.47	68.64	10.69
Path length	67.63	9.83	**72.60**	8.83	65.33	6.91	68.64	10.69
Resnik (1993)	66.14	9.13	**72.92**	9.08	70.54	8.24	68.64	10.69
Wu and Palmer (1994)	70.84	9.38	**72.97**	9.05	66.02	6.82	68.64	10.69



Resnik (1997) Wu and Palmer (1994)

Figure 1: Plot of precision (dotted line, circles), recall (dotted line, triangles), and $F_{\beta=1}$ score (solid line) versus the number of different senses for a word. Also includes the frequency of each sense count (dashed line, squares). For both measures, final judgement is made on average similarity of all senses.

fronted a Catch-22 situation where disambiguation is needed in order to carry out metaphora identification, a disambiguation task itself. The way things stand, our method can be successfully applied to shallow NLP tasks or as a pre-processing and optimization step for WSD and parsing.

5 Conclusions

In this paper, we presented a mildly supervised method for identifying metaphorical verb usage by taking the local context into account. This procedure is different from the majority of the previous works in that it does not rely on any metaphor-specic hand-coded knowledge, but rather on previous observed unambiguous usages of the verb. The method can operates on open domain texts and the memory needed for the seeds can be relatively easily collected by mining unannotated corpora. Furthermore, our method differs as compares the meaning of nouns which appear at the same context without associating them with concepts and then comparing the concepts. We selected this procedure as words of the same abstract concept maybe not appear at the same context while words from different concepts could appear at the same context, especially when the certain context is metaphorical. Although the system has been tested only on verb-direct object metaphors, the described identi- cation method should be immediately applicable to a wider range of word classes, which is one of the future research directions we will pursue. Another promising research direction relates to our observation regarding the importance of measuring similarities by considering the relative taxonomic position of the two concepts; more specifically, we will experiment with clustering methods over unannotated corpora as a way of producing the taxonomy over which we will dene some Resnik-esque similarity measure.

References

Tony Berber Sardinha. 2002. Metaphor in early applied linguistics writing: A corpus-based analysis of lexis in dissertations. In *I Conference on Metaphor in Language and Thought*.

Tony Berber Sardinha. 2010. Creating and using the Corpus do Portugues and the frequency dictionary of portuguese. *Working with Portuguese Corpora*.

Julia Birke and Anoop Sarkar. 2006. A clustering approach for the nearly unsupervised recognition of nonliteral language. In *Proceedings of EACL-06*, pages 329–336. Trento, Italy.

Dan Fass. 1991. met*: a method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1).

Gilles Fauconnier and Mark Tuner. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.

Jerome Feldman. 2006. *From Molecule to Metaphor: A Neutral Theory of Language*. The MIT Press.

Charles Fillmore, Christopher Johnson and Miriam Petruck. 2003. Background to framenet. *International Journal of Lexicography*, 16(3):235–250.

Matt Gedigian, John Bryant, Srini Narayanan, and Branimir Ciric. 2006. Catching metaphors. In *Proceedings of the 3rd Workshop on Scalable Natural Language Understanding*, pages 41–48. New York.

Joseph Edward Grady. 1997. *Foundations of meaning: primary metaphors and primary scenes*. University Microfilms International.

Yael Karov and Shimon Edelman. 1998. Similarity-based word sense disambiguation. *Computational Linguistics*, 24(1):41–59.

Paul Kingsbury and Martha Palmer. 2002. From treebank to propbank. In *Proceedings of LREC-2002*, pages 1989–1993. Gran Canaria, Canary Islands, Spain.

Saisuresh Krishnakumaran and Xiaojin Zhu. 2007. Hunting elusive metaphors using lexical resources. In *Proceedings of the Workshop on Computational Approaches to Figurative Language, April, 2007, Rochester, New York*, pages 13–20. Association for Computational Linguistics, Rochester, New York. URL `http://www.aclweb.org/anthology/W/W07/W07-0103`.

George Lakoff, Jane Espenson, and Alan Schwartz. 1991. *The master metaphor list*. University of California at Berkeley.

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

Dekang Lin. 1997. Using syntactic dependency as local context to resolve word sense ambiguity. In *Proceedings of ACL-97*, pages 64–71. Madrid, Spain.

Dekang Lin. 1998. An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning (ICML-98). Madison, WI, USA, July 1998.*, page 296304.

Zachary Mason. 2004. Cormet: a computational, corpus-based conventional metaphor extraction system. *Computational Linguistics*, 30(1):23–44.

Srini Narayanan. 1997. *Knowledge-based Action Representations for Metaphor and Aspect (KARMA)*. University of Californial.

Siddharth Padwardhan, Satanjeev Banerjee, and Ted Pedersen. 2003. Using measures of semantic relatedness for word sense disambiguation. In *Proceedings of the 4th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-03), Mexico City*, pages 241–257.

Stephen Pinker. 2007. *The Stuff of Thought: Language as a Window into Human Nature*. Viking Adult.

Judita Preiss, Ted Briscoe, and Anna Korhonen. 2007. A system for large-scale acquisition of verbal, nominal and adjectival subcategorization frames from corpora. In *Proceedings of ACL-07*, volume 45, page 912.

Philip Resnik. 1997. Selectional preference and sense disambiguation. In *ACL SIGLEX Workshop on Tagging Text with Lexical Semantics*. Washington, D.C.

Ekaterina Shutova, Simone Teufel and Anna Korhonen. 2012. Statistical metaphor processing. *Computational Linguistics*, 39(2).

Julie Weeds, David Weir, and Diana McCarthy. 2004. Characterising measures of lexical distributional similarity. In *Proceedings of Coling 2004*, pages 1015–1021. COLING, Geneva, Switzerland.

Yorick Wilks. 1978. Making preferences more active. *Artificial Intelligence*, 11(3).

Zhibiao Wu and Martha Palmer. 1994. Verb semantics and lexical selection. In *Proceedings of the 32nd Annual Meeting of the ACL (ACL-04)*, pages 133–138.