

# Hunting Elusive Metaphors Using Lexical Resources

**Saisuresh Krishnakumaran\***

Computer Sciences Department  
University of Wisconsin-Madison  
Madison, WI 53706  
ksai@cs.wisc.edu

**Xiaojin Zhu**

Computer Sciences Department  
University of Wisconsin-Madison  
Madison, WI 53706  
jerryzhu@cs.wisc.edu

## Abstract

In this paper we propose algorithms to automatically classify sentences into metaphoric or normal usages. Our algorithms only need the WordNet and bigram counts, and does not require training. We present empirical results on a test set derived from the Master Metaphor List. We also discuss issues that make classification of metaphors a tough problem in general.

## 1 Introduction

Metaphor is an interesting figure of speech which expresses an analogy between two seemingly unrelated concepts. Metaphoric usages enhance the attributes of the source concept by comparing it with the attributes of the target concept. Abstractions and enormously complex situations are routinely understood via metaphors (Lakoff and Johnson, 1980). Metaphors begin their lives as Novel Poetic Creations with marked rhetoric effects whose comprehension requires special imaginative leap. As time goes by, they become part of general use and their comprehension becomes automatic and idiomatic and rhetoric effect is dulled (Nunberg, 1987). We term such metaphors whose idiomatic effects are dulled because of common usage as *dead metaphors* while metaphors with novel usages as *live metaphors*. In this paper we are interested only in identifying live metaphors.

Metaphors have interesting applications in many NLP problems like machine translation, text summarization, information retrieval and question answering. Consider the task of summarizing a parable which is a metaphoric story with a moral. The best summary of a parable is the moral. Paraphrasing a metaphoric passage like a parable is difficult without understanding the metaphoric uses. The performance of the conventional summarizing systems will be ineffective because they cannot identify such metaphoric usages. Also it is easy to create novel and interesting uses of metaphors as long as one concept is explained in terms of another concept. The performance of machine translation systems will be affected in such cases especially if they have not encountered such metaphoric uses before.

Metaphor identification in text documents is, however, complicated by issues including context sensitiveness, emergence of novel metaphoric forms, and the need for semantic knowledge about the sentences. Metaphoric appeal differs across language or people's prior exposure to such usages. In addition, as (Gibbs, 1984) points out, literal and figurative expressions are end points of a single continuum along which metaphoricity and idiomaticity are situated, thereby making clear demarcation of metaphoric and normal usages fuzzy.

We discuss many such issues that make the task of classifying sentences into metaphoric or non-metaphoric difficult. We then focus on a subset of metaphoric usages involving the nouns in a sentence. In particular, we identify the subject-object, verb-noun and adjective-noun relationships in sentences and classify them as metaphoric or

---

\* The first author is currently affiliated with Google Inc, Mountain View, CA.

non-metaphoric. Extensions to other metaphoric types will be part of future work. Our algorithms use the hyponym relationship in WordNet (Fellbaum, 1998), and word bigram counts, to predict the metaphors. In doing so we circumvent two issues: the absence of labeled training data, and the lack of clear features that are indicative of metaphors.

The paper is organized as follows. Section 2 presents interesting observations that were made during the initial survey, and presents examples that makes metaphor identification hard. Section 3 discusses our main techniques for identifying metaphors in text documents. Section 4 analyzes the effect of the techniques. Section 5 discusses relevant prior work in the area of metaphor processing and identification. Finally we conclude in Section 6.

## 2 Challenges in Metaphor Identification

In this section we present some issues that make metaphor identification hard.

### 2.1 Context Sensitivity

Some metaphoric usages are sensitive to the context in which they occur. For example, the following sentence can act as a normal sentence as well as a metaphoric sentence.

*Men are animals.*

It is a normal sentence in a biology lecture because all human beings fall under the animal kingdom. However this is a metaphoric sentence in a social conversation when it refers to animal qualities. Also the word ‘Men’ has two different senses in WordNet and hence it is necessary to disambiguate the senses based on the context. Sense disambiguation is beyond the scope of this paper.

### 2.2 Pronoun Resolution

Consider the following sentence,

*This homework is a breeze. The previous one was on calculus. It was a tornado.*

The techniques we discuss in this paper can classify the reference to ‘breeze’ as metaphoric. In order to correctly classify the reference to ‘tornado’ as metaphoric, however, the system needs to resolve

the reference to the pronoun ‘It’. Strictly speaking, this example might be solved without resolution because any of the potential antecedents render the sentence metaphoric, but in general resolution is necessary.

### 2.3 Word Usages

Consider the following two sentences,

*He is a Gandhi. vs. He is Gandhi.*

The first sentence is a metaphor which attributes the qualities of Gandhi to the actor, while the second sentence is a normal one. Here the article ‘a’ distinguishes the first sentence from the second. Similarly, in the following example, the phrase ‘among men’ helps in making the second usage metaphoric.

*He is a king. vs. He is a king among men.*

A comprehensive list of such uses are not known and incorporating all such grammatical features would make the system quite complex.

### 2.4 Parser Issues

The techniques that we propose work on the parsed sentences. Hence the accuracy of our technique is highly dependent on the accuracy of the parser.

### 2.5 Metaphoric Usages in WordNet

Some metaphoric senses of nouns are already part of the WordNet.

*He is a wolf.*

The metaphoric sense of ‘wolf’ is directly mentioned in the WordNet. We call such usages as ‘dead metaphors’ because they are so common and are already part of the lexicon. In this paper we are interested in identifying only novel usages of metaphors.

## 3 Noun-Form Metaphors

We restrict ourselves to metaphoric usages involving nouns. In particular, we study the effect of verbs and adjectives on the nouns in a sentence. We categorize the verb-noun relationship in sentences as Type I and Type II based on the verb. We call the adjective-noun relationship as Type III, see Table 1.

For Type I, the verb is one of the ‘be’ form verbs like ‘is’, ‘are’, ‘am’, ‘was’, etc. An example of Type I form metaphor is

Table 1: Terminology

Sentence Type	Relationship
Type I	Subject IS-A Object
Type II	Verb acting on Noun (verb not 'be')
Type III	Adjective acting on Noun

*He is a brave lion.*

An example of Type II form metaphor is

*He planted good ideas in their minds.*

An example for Type III form metaphor is

*He has a fertile imagination.*

We use two different approaches for Type I vs. Types II, III. In Type I form we are interested in the relationship between the subject and the object. We use a hyponym heuristic. In Types II and III, we are interested in the subject-verb, verb-object, or adjective-noun relations. We use hyponym together with word co-occurrence information, in this case bigrams from the Web 1T corpus (Brants and Franz, 2006). Sections 3.1 and 3.2 discuss the two algorithms, respectively. We use a parser (Klein and Manning, 2003) to obtain the relationships between nouns, verbs and adjectives in a sentence.

### 3.1 Identifying Type I metaphors

We identify the WordNet hyponym relationship (or the lack thereof) between the subject and the object in a Type I sentence. We classify the sentence as metaphoric, if the subject and object does not have a hyponym relation. A hyponym relation exists between a pair of words if and only if one word is a subclass of another word. We motivate this idea using some examples. Let us consider a normal sentence with a subject-object relationship governed by a 'be' form verb, 'is'.

*A lion is a wild animal.*

The subject-verb-object relationship of this normal sentence is shown in Figure 1.

The subject and the object in the above example is governed by 'IS-A' relationship. Thus, Lion 'IS-A' type of animal. The 'IS-A' relationship is captured

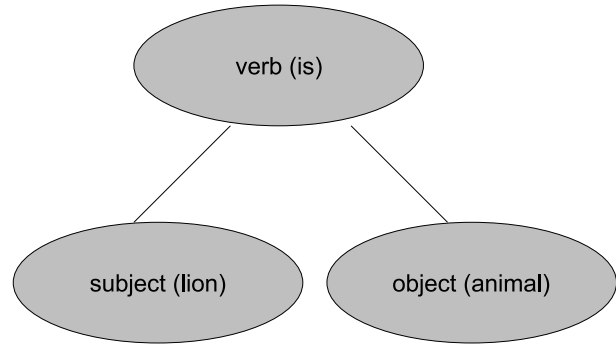


Figure 1: The Subject-Verb-Object relationship for 'A lion is a wild animal.'

as the 'hyponym' relationship in WordNet, where 'Lion' is the hyponym of 'animal'. Consider another example,

*He is a scientist.*

Here the object 'scientist' is the occupation of the subject 'He', which we change to 'person'. 'Scientist' is a hyponym of 'person' in WordNet. The above two examples show that we expect a subject-object hyponym relation for normal Type I relations. On the other hand, consider a metaphoric example in Type I form,

*All the world's a stage.*  
- William Shakespeare

The subject-verb-object relationship is represented by Figure 2.

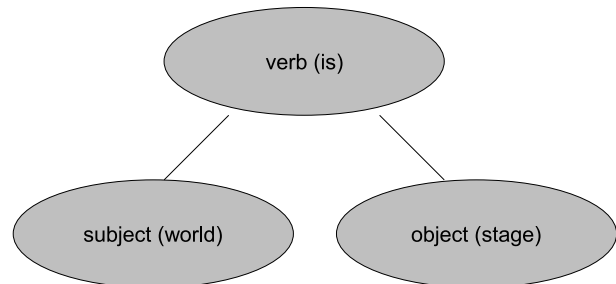


Figure 2: The Subject-Verb-Object relationship for 'All the world's a stage.'

There is a subject-object relation between 'World' and 'Stage', but they do not hold a hyponym relation in WordNet. This is an important observation which we use in classifying relationships of this

form. Consider another example with complex sentences,

*Men are April when they woo, December  
when they wed. Maids are May when  
they are maids, but the sky changes  
when they are wives.  
-Shakespeare's 'As You Like It'.*

In this case, there are two explicit subject-object relations, namely Men-April and Maids-May. The WordNet hyponym relation does not exist between either pair.

From the examples considered above, it seems that when a hyponym relation exists between the subject and the object, the relationship is normal, and metaphoric otherwise. The effectiveness of this approach is analyzed in detail in Section 4. The pseudo code for classifying Type I relations is given below:

1. Parse the sentences and get all  $R \leftarrow \{\text{subject, be, object}\}$  relations in those sentences.
2. for each relation  $R_{sub,obj}$ 
  - if  $\text{Hyponym}(\text{sub,obj}) = \text{true}$
  - then  $R_{sub,obj}$  is normal usage
  - else  $R_{sub,obj}$  is a metaphoric relation
3. All sentences with at least one metaphoric relation is classified as metaphoric.

### 3.2 Identifying Type II and Type III metaphors

We use a two dimensional V/A-N co-occurrence matrix, in addition to WordNet, for detecting Type II and Type III metaphors. V/A-N matrix stands for Verb/Adjective-Noun matrix, which is a two dimensional matrix with verbs or adjectives along one dimension, and nouns along the other. The entries are co-occurrence frequency of the word pair, from which we may estimate the conditional probability  $p(w_n|w)$  for a noun  $w_n$  and a verb or adjective  $w$ . Ideally the matrix should be constructed from a parsed corpus, so that we can identify V/A-N pairs from their syntactic roles. However parsing a large corpus would be prohibitively expensive. As a practical approximation, we use bigram counts from the Web 1T corpus (Brants and Franz, 2006). Web 1T corpus consists of English word n-grams

(up to 5-grams) generated from approximately 1 trillion word tokens of text from public Web pages. In this paper we use the bigram data in which a noun follows either a verb or an adjective. We note that this approximation thus misses, for example, the pair (plant, idea) in phrases like 'plant an idea'. Nonetheless, the hope is that the corpus makes it up by sheer size.

#### 3.2.1 Type II metaphors

We discuss the metaphoric relationship between a verb-noun pair  $(w_v, w_n)$ . The idea is that if neither  $w_n$  nor its hyponyms or hypernyms co-occur frequently with  $w_v$ , then the pair is a novel usage, and we classify the pair as metaphoric. To this end, we estimate the conditional probability  $p(w_h|w_v) = \text{count}(w_v, w_h) / \text{count}(w_v)$  from the V/A-N matrix, where  $w_h$  is  $w_n$  itself, or one of its hyponyms / hypernyms. If at least one of these  $w_h$  has high enough conditional probability as determined by a threshold, we classify it as normal usage, and metaphoric otherwise. Consider the following example

*He planted good ideas in their minds.*

The verb 'planted' acts on the noun 'ideas' and makes the sentence metaphoric. In our corpus the objects that occur more frequently with the verb 'planted' are 'trees', 'bomb' and 'wheat', etc. Neither the noun 'ideas' nor its hyponyms / hypernyms occurs frequently enough with 'planted'. Hence we predict this verb-object relationship as metaphoric. The pseudo code for classifying Type II metaphors is given below:

1. Parse the sentences and obtain all  $R \leftarrow \{\text{verb, noun}\}$  relations in those sentences.
2. for each relation  $R_{verb,noun}$ 
  - Sort all nouns  $w$  in the vocabulary by decreasing  $p(w|verb)$ . Take the smallest set of top  $k$  nouns whose conditional probability sum  $\geq$  threshold  $T$ .
  - if  $\exists w_h$  such that  $w_h$  is related to  $noun$  by the hyponym relation in WordNet, and  $w_h \in$  top  $k$  words above,
  - then  $R_{verb,noun}$  is normal usage
  - else  $R_{verb,noun}$  is a Type II metaphoric relation

- All sentences with at least one metaphoric relationship is classified as a metaphor.

### 3.2.2 Type III metaphors

The technique for detecting the Type III metaphors is the same as the technique for detecting the Type II metaphors except that it operates on different relationship. Here we compare the Adjective-Noun relationship instead of the Verb-Noun relationship. For example,

*He has a fertile imagination.*

Here the adjective ‘fertile’ acts on the noun ‘imagination’ to make it metaphoric. The nouns that occur frequently with the ‘fertile’ in our corpus are ‘soil’, ‘land’, ‘territory’, and ‘plains’, etc. Comparison of the WordNet hierarchies of the noun ‘imagination’ with each of these nouns will show that there does not exist any hyponym relation between ‘imagination’ and any of these nouns. Hence we classify them as metaphors. As another example,

*TV is an idiot box.*

The adjective ‘idiot’ qualifies nouns related to people such as ‘boy’, ‘man’, etc. that are unrelated to the noun ‘box’. Thus we classify it as a Type III metaphor.

## 4 Experimental Results

We experimented with the Berkeley Master Metaphor List (Lakoff and Johnson, 1980) to compute the performance of our techniques. The Berkeley Master Metaphor List is a collection of nearly 1728 unique sentences and phrases. We corrected some typos and spelling errors in the Master list and expanded phrases to complete sentences. The list has many metaphoric uses which has become very common usages in today’s standards, and thus no longer have any rhetoric effects. Therefore, we manually label the sentences in the Master List into 789 ‘live metaphors’ and the remaining ones ‘dead metaphors’ as the ground truth<sup>1</sup>.

Table 2 shows the initial performance of the Type I algorithm. There are 129 sentences in the

<sup>1</sup>Our processed and labeled dataset is available at [http://www.cs.wisc.edu/~ksai/publications/2007/HLT\\_NAACL\\_metaphors/metaphors.html](http://www.cs.wisc.edu/~ksai/publications/2007/HLT_NAACL_metaphors/metaphors.html)

Master List that contain subject-be-object form. Our algorithm has a precision of 70% and a recall of 61% with respect to the live/dead labels. Note that although the accuracy is 58%, the algorithm is better than a random classification in terms of precision and recall. One thing to note is that our negative examples are (subjectively labeled) dead metaphors. We thus expect the task to be harder than with random non-metaphoric sentences. Another point to note here is that the live/dead labels are on sentences and not on particular phrases with type I relations. A sentence can contain more than one phrases with various types. Therefore this result does not give a complete picture of our algorithm.

Table 2: Type I Performance

	Predicted as Metaphoric	Predicted as Normal
Annotated as live	50	32
Annotated as dead	22	25

A few interesting metaphors detected by our algorithm are as follows:

*Lawyers are real sharks.*

Smog pollution is an environmental malaise.

Some false negatives are due to phrases qualifying the object of the sentence as in the following example,

*He is a budding artist.*

There is a Type I relation in this sentence because the subject ‘He’ and the object ‘artist’ are related by the ‘be’ form verb ‘is’. In this case, the Type I algorithm compares the hyponyms relation between ‘person’ and ‘artist’ and declares it as a normal sentence. However the adjective ‘budding’ adds Type III figurative meaning to this sentence. Therefore although the Type I relation is normal, there are other features in the sentences that make it metaphoric. We observed that most of false negatives that are wrongly classified because of the above reason have pronoun subject like ‘he’, ‘she’ etc.

Another major source of issue is the occurrences of pronoun ‘it’ which is hard to resolve. We replaced it by ‘entity’, which is the root of WordNet, when

comparing the hyponyms. ‘Entity’ matches the hyponym relation with any other noun and hence all these sentences with ‘it’ as the subject are classified as normal sentences.

Table 3: Type I Performance for sentences with non-pronoun subject

	Predicted as Metaphoric	Predicted as Normal
Annotated as live	40	1
Annotated as dead	19	4

Table 3 shows the performance of our Type I algorithm for sentences with non-pronoun subjects. It clearly shows that the performance in Table 2 is affected by sentences with pronoun subjects as explained in the earlier paragraphs.

In some cases, prepositional phrases affects the performance of our algorithm. Consider the following example,

*He is the child of evil.*

Here the phrase ‘child of evil’ is metaphoric. But the parser identifies a subject-be-object relationship between ‘He’ and ‘child’ and our algorithm compares the hyponym relation between ‘person’ and ‘child’ and declares it as a normal sentence.

Our current algorithm does not deal with cases like the following example

*The customer is a scientist. vs. The customer is king.*

Since there is no direct hyponym relation between scientist/king with customer we declare both these sentences as metaphors although only the latter is.

Unlike the algorithm for Type I, there is a threshold  $T$  to be set for Type II and III algorithm. By changing  $T$ , we are able to plot a precision recall curve. Figure 3 and figure 4 show the precision recall graph for Type II and Type III relations respectively. Figure 5 shows the overall precision recall graph for all three types put together.

False positives in Type II and Type III were due to very general verbs and adjectives. These verbs and adjectives can occur with a large number of nouns, and tend to produce low conditional probabilities even for normal nouns. Thereby they are

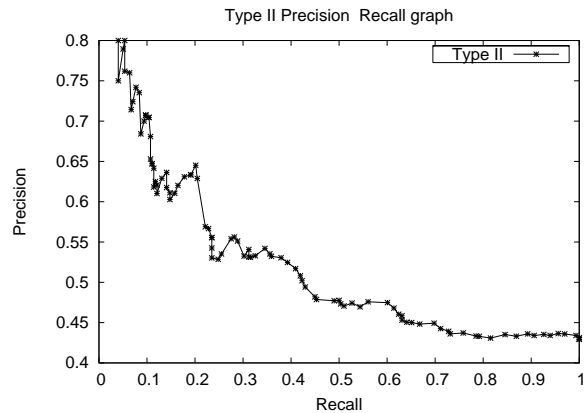


Figure 3: Precision Recall curve for Type II relations.

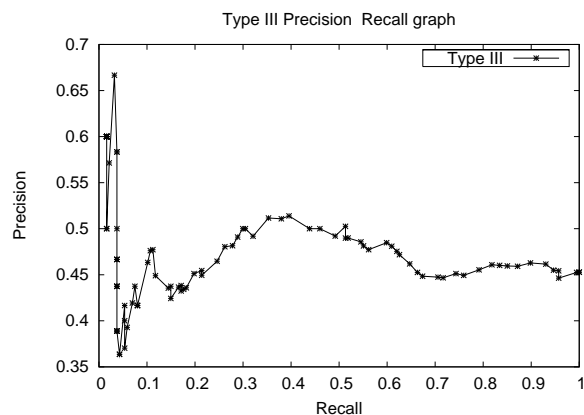


Figure 4: Precision Recall curve for Type III relations.

often mistakenly classified as metaphoric relations. We expect the performance to improve if these general verbs and adjectives are handled properly. Some general verbs include ‘gave’, ‘made’, ‘has’, etc., and similarity general adjectives include ‘new’, ‘good’, ‘many’, ‘more’, etc. The plot for Type III is more random.

Most errors can be attributed to some of the following reasons:

- As mentioned in the challenges section, the parser is not very accurate. For example,

*They battled each other over the chess board every week.*

Here the parser identifies the verb-object relation as (battled, week), which is not correct.

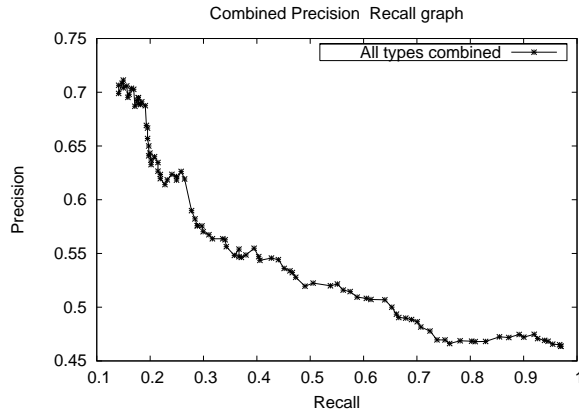


Figure 5: Overall Precision Recall curve for all three types combined.

- Pronoun resolution: As discussed earlier, the pronoun ‘it’ is not resolved and hence they introduce additional source of errors.
- Manual annotations could be wrong. In our experiment we have used only two annotators, but having more would have increased the confidence in the labels.
- Many of the verb-noun forms are most naturally captured by trigrams instead of bigram. For example, (developed , attachment) most likely occurs in a corpus as ‘developed an attachment’ or ‘developed the attachment’. Our bigram approach can fail here.
- Sense disambiguation: We don’t disambiguate senses while comparing the WordNet relations. This increases our false negatives.
- Also as mentioned earlier, the labels are on sentences and not on the typed relationships. Therefore even though a sentence has one or more of the noun form types, those may be normal relationships while the whole sentence may be metaphoric because of other types. Note, however, that some of these mismatches are corrected for the ‘All types combined’ result.

## 5 Related Work

There has been a long history of research in metaphors. We briefly review some of them here.

One thing that sets our work apart is that most previous literatures in this area tend to give little empirical evaluation of their approaches. In contrast, in this study we provide detailed analysis of the effectiveness of our approaches.

(Fass and Wilks, 1983) proposes the use of preference semantics for metaphor recognition. Techniques for automatically detecting selections preferences have been discussed in (McCarthy and Carroll, 2003) and (Resnik, 1997). Type II and Type III approaches discussed in this paper uses both these ideas for detecting live metaphors. Fass (Fass, 1991) uses selectional preference violation technique to detect metaphors. However they rely on hand-coded declarative knowledge bases. Our technique depends only on WordNet and we use selection preference violation based on the knowledge learned from the bigram frequencies on the Web.

Markert and Nissim (Markert and Nissim, 2002) presents a supervised classification algorithm for resolving metonymy. Metonymy is a closely related figure of speech to metaphors where a word is substituted by another with which it is associated. Example,

*A pen is mightier than a sword.*

Here sword is a metonymy for war and pen is a metonymy for articles. They use collocation, co-occurrence and grammatical features in their classification algorithm.

MetaBank (Martin, 1994) is a large knowledge base of metaphors empirically collected. The detection technique compares new sentences with this knowledge base. The accuracy is dependent on the correctness of the knowledge base and we expect that some of these metaphors would be dead in the present context. The techniques we discuss in this work will drastically reduce the need for manually constructing such a large collection.

Goatly (Goatly, 1997) proposes using analogy markers such as ‘like’, ‘such as’, ‘illustrated by’ and lexical markers like ‘literally’, ‘illustrating’, ‘metaphorically’ etc. These would be useful for identifying simile and explicit metaphoric relations but not metaphors where the relation between the target concept and the source concept is not explicit.

The CorMet system (Mason, 2004) dynamically mines domain specific corpora to find less frequent

usages and identifies conceptual metaphors. However the system is limited to extracting only selectional preferences of verbs. Verbal selectional preference is the verb's preference for the type of argument it takes.

Dolan (Dolan, 1995) uses the path and path length between words in the knowledge base derived from lexical resources for interpreting the interrelationship between the component parts of a metaphor. The effectiveness of this technique relies on whether the metaphoric sense is encoded in the dictionaries. This approach however will not be effective for novel metaphoric usages that are not encoded in dictionaries.

## 6 Conclusion

In this paper we show that we can use the hyponym relation in WordNet and word co-occurrence information for detecting metaphoric uses in subject-object, verb-noun and adjective-noun relationships. According to (Cameron and Deignan, 2006), non literal expressions with relatively fixed forms and highly specific semantics are over-represented in the metaphor literature in comparison to corpora occurrences. Therefore as part of future work we would be studying the effect of our algorithms for naturally occurring text. We are also interested in increasing the confidence of the labels using more and diverse annotators and see how the techniques perform. The study can then be extended to incorporate the role of prepositions in metaphoric uses.

## 7 Acknowledgment

We would like to thank our anonymous reviewers for their constructive suggestions that helped improve this paper. We would also like to thank Mr. Krishna Kumaran Damodaran for annotating the Master Metaphor List.

## References

- Thorsten Brants and Alex Franz. 2006. *Web 1T 5-gram Version 1*. Linguistic Data Consortium, Philadelphia.
- Lynne Cameron and Alice Deignan. 2006. The emergence of metaphor in discourse. *Applied Linguistics*, 27(4):671–690.
- William B. Dolan. 1995. Metaphor as an emergent property of machine-readable dictionaries. *AAAI 1995 Spring Symposium*, 95(1):27–32.
- Dan Fass and Yorick Wilks. 1983. Preference semantics, ill-formedness, and metaphor. *American Journal of Computational Linguistics*, 9(3):178–187.
- Dan Fass. 1991. Met: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.
- Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.
- Raymond Gibbs. 1984. Literal meaning and psychological theory. *Cognitive Science*, 8:275–304.
- Andrew Goatly. 1997. *The Language of Metaphors*. Routledge, London.
- Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics*.
- George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago, Illinois.
- Katja Markert and Malvina Nissim. 2002. Metonymy resolution as a classification task. In *Proceedings of ACL-02 conference on Empirical Methods in Natural Language Processing*, pages 204–213.
- James H. Martin. 1994. Metabank: a knowledge-base of metaphoric language conventions. *Computational Intelligence*, 10(2):134–149.
- Zachary J. Mason. 2004. Cormet: A computational, corpus-based conventional metaphor extraction system. *Computational Linguistics*, 30(1):23–44.
- Diana McCarthy and John Carrol. 2003. Disambiguating nouns, verbs and adjectives using automatically acquired selectional preferences. *Computational Linguistics*, 29(4):639–654.
- Geoffrey Nunberg. 1987. Poetic and prosaic metaphors. In *Proceedings of the 1987 workshop on Theoretical issues in natural language processing*, pages 198–201.
- Philip Resnik. 1997. Selectional preferences and word sense disambiguation. In *Proceedings of ACL Siglex Workshop on Tagging Text with Lexical Semantics, Why, What and How?*, Washington, D.C., pages 52–57.