

# Metaphor Identification Using Verb and Noun Clustering

Ekaterina Shutova, Lin Sun and Anna Korhonen

Computer Laboratory, University of Cambridge

es407, ls418, alk23@cam.ac.uk

## Abstract

We present a novel 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. Our method is distinguished from previous work in that it does not employ any hand-crafted knowledge, other than the initial seed set, but, in contrast, captures metaphoricity by means of verb and noun clustering. Being the first to employ unsupervised methods for metaphor identification, our system operates with the precision of 0.79.

## 1 Introduction

Besides enriching our thought and communication with novel imagery, the phenomenon of metaphor also plays a crucial structural role in our use of language. Metaphors arise when one concept is viewed in terms of the properties of the other. Below are some examples of metaphor.

- (1) How can I *kill* a process? (Martin, 1988)
- (2) Inflation has *eaten up* all my savings. (Lakoff and Johnson, 1980)
- (3) He *shot down* all of my arguments. (Lakoff and Johnson, 1980)
- (4) And then my heart with pleasure *fills*,  
And *dances* with the daffodils.<sup>1</sup>

In metaphorical expressions seemingly unrelated features of one concept are associated with another concept. In the computer science metaphor

in (1) the *computational process* is viewed as something *alive* and, therefore, its forced termination is associated with the act of killing. Lakoff and Johnson (1980) explain metaphor as a systematic association, or a *mapping*, between two concepts or conceptual domains: the *source* and the *target*. The metaphor in (3) exemplifies a mapping of a concept of *argument* to that of *war*. The *argument*, which is the target concept, is viewed in terms of a *battle* (or a *war*), the source concept. The existence of such a link allows us to talk about arguments using the war terminology, thus giving rise to a number of metaphors.

Characteristic to all areas of human activity (from poetic to ordinary to scientific) and, thus, to all types of discourse, metaphor becomes an important problem for natural language processing (NLP). In order to estimate the frequency of the phenomenon, Shutova and Teufel (2010) conducted a corpus study on a subset of the British National Corpus (BNC) (Burnard, 2007) representing various genres. They manually annotated metaphorical expressions in this data and found that 241 out of 761 sentences contained a metaphor, whereby in 164 phrases metaphoricity was introduced by a verb. Due to such a high frequency of their use, a system capable of recognizing and interpreting metaphorical expressions in unrestricted text would become an invaluable component of any semantics-oriented NLP application.

Automatic processing of metaphor can be clearly divided into two subtasks: *metaphor identification* (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.

To date the most influential account of metaphor identification is that of Wilks (1978).

<sup>1</sup>“I wandered lonely as a cloud”, William Wordsworth, 1804.

According to Wilks, metaphors represent a violation of selectional restrictions in a given context. Consider the following example.

(5) My car *drinks* gasoline. (Wilks, 1978)

The verb *drink* normally takes an *animate* subject and a *liquid* object. Therefore, *drink* taking a *car* as a subject is an anomaly, which may as well indicate metaphorical use of *drink*.

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-literality 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. In a similar manner manually created knowledge in the form of WordNet (Fellbaum, 1998) is employed by the system of Krishnakumaran and Zhu (2007), which essentially differentiates between highly lexicalized metaphors included in WordNet, and novel metaphorical senses.

Alternative approaches (Gedigan et al., 2006) search for metaphors of a specific domain defined a priori (e.g. MOTION metaphors) in a specific type of discourse (e.g. Wall Street Journal). In contrast, the scope of our experiments is the whole of the British National Corpus (BNC) (Burnard, 2007) and the domain of the expressions we identify is unrestricted. However, our technique is also distinguished from the systems of Fass (1991) and Krishnakumaran and Zhu (2007) in that it does not rely on any hand-crafted knowledge, but rather captures metaphoricity in an unsupervised way by means of verb and noun clustering.

The motivation behind the use of clustering methods for metaphor identification task lies in the nature of metaphorical reasoning based on association. Compare, for example, the target concepts of *marriage* and *political regime*. Having quite distinct meanings, both of them are cognitively mapped to the source domain of *mechanism*, which shows itself in the following examples:

(6) Our relationship is not really *working*.

(7) Diana and Charles did not succeed in *mending* their marriage.

(8) The *wheels* of Stalin's regime were *well oiled* and already *turning*.

We expect that such relatedness of distinct target concepts should manifest itself in the examples of language use, i.e. target concepts that are associated with the same source concept should appear in similar lexico-syntactic environments. Thus, clustering concepts using grammatical relations (GRs) and lexical features would allow us to capture their relatedness **by association** and harvest a large number of metaphorical expressions beyond our seed set. For example, the sentence in (6) being part of the seed set should enable the system to identify metaphors in both (7) and (8).

In summary, our system (1) starts from a seed set of metaphorical expressions exemplifying a range of source–target domain mappings; (2) performs unsupervised noun clustering in order to harvest various target concepts associated with the same source domain; (3) by means of unsupervised verb clustering creates a source domain verb lexicon; (4) searches the BNC for metaphorical expressions describing the target domain concepts using the verbs from the source domain lexicon.

We tested our system starting with a collection of metaphorical expressions representing verb-subject and verb-object constructions, where the verb is used metaphorically. We evaluated the precision of metaphor identification with the help of human judges. In addition to this we compared our system to a baseline built upon WordNet, whereby we demonstrated that our method goes far beyond synonymy and captures metaphors not directly related to any of those seen in the seed set.

## 2 Experimental Data

### 2.1 Seed Phrases

We used the dataset of Shutova (2010) as a seed set. Shutova (2010) annotated metaphorical expressions in a subset of the BNC sampling various genres: literature, newspaper/journal articles, essays on politics, international relations and history, radio broadcast (transcribed speech). The dataset consists of 62 phrases that are single-word

metaphors representing verb-subject and verb-object relations, where a verb is used metaphorically. The seed phrases include e.g. *stir excitement*, *reflect enthusiasm*, *accelerate change*, *grasp theory*, *cast doubt*, *suppress memory*, *throw remark* (verb - direct object constructions) and *campaign surged*, *factor shaped [..]*, *tension mounted*, *ideology embraces*, *changes operated*, *approach focuses*, *example illustrates* (subject - verb constructions).

## 2.2 Corpus

The search space for metaphor identification was the British National Corpus (BNC) that was parsed using the RASP parser of Briscoe et al. (2006). We used the grammatical relations output of RASP for BNC created by Andersen et al. (2008). The system searched the corpus for the source and target domain vocabulary within a particular grammatical relation (verb-object or verb-subject).

## 3 Method

Starting from a small seed set of metaphorical expressions, the system implicitly captures the associations that underly their production and comprehension. It generalizes over these associations by means of unsupervised verb and noun clustering. The obtained clusters then represent potential source and target concepts between which metaphorical associations hold. The knowledge of such associations is then used to annotate metaphoricity in a large corpus.

### 3.1 Clustering Motivation

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 results of previous research on corpus-based lexical semantics suggest that the linguistic environment in which a lexical item occurs can shed light on its meaning. A number of works have shown that it is possible to automatically induce semantic word classes from corpus data via clustering of contextual cues (Pereira et al., 1993; Lin, 1998; Schulte im Walde, 2006). The consensus is that

the lexical items exposing similar behavior in a large body of text most likely have the same meaning. However, the concepts of *marriage* and *political regime*, that are also observed in similar lexico-syntactic environments, albeit having quite distinct meanings are likewise assigned by such methods to the same cluster. In contrast to concrete concepts, such as *tea*, *water*, *coffee*, *beer*, *drink*, *liquid*, that are clustered together due to **meaning similarity**, abstract concepts tend to be clustered together **by association** with the same source domain. It is the presence of this association that explains the fact that they share common contexts. We exploit this idea for identification of new target domains associated with the same source domain. We then use unsupervised verb clustering to collect source domain vocabulary, which in turn allows us to harvest a large number of new metaphorical expressions.

### 3.2 Verb and Noun Clustering

Since Levin (1993) published her classification, there have been a number of attempts to automatically classify verbs into semantic classes using supervised and unsupervised approaches (Lin, 1998; Brew and Schulte im Walde, 2002; Korhonen et al., 2003; Schulte im Walde, 2006; Joanis et al., 2008; Sun and Korhonen, 2009). Similar methods were also applied to acquisition of noun classes from corpus data (Rooth et al., 1999; Pantel and Lin, 2002; Bergsma et al., 2008).

We adopt a recent verb clustering approach of Sun and Korhonen (2009), who used rich syntactic and semantic features extracted using a shallow parser and a clustering method suitable for the resulting high dimensional feature space. When Sun and Korhonen evaluated their approach on 204 verbs from 17 Levin classes, they obtained 80.4 F-measure (which is high in particular for an unsupervised approach). We apply this approach to a much larger set of 1610 verbs: all the verb forms appearing in VerbNet (Kipper et al., 2006) with the exception of highly infrequent ones. In addition, we adapt the approach to noun clustering.

#### 3.2.1 Feature Extraction

Our verb dataset is a subset of VerbNet compiled as follows. For all the verbs in VerbNet we

extracted their occurrences (up to 10,000) from the raw corpus data collected originally by Korhonen et al. (2006) for construction of VALEX lexicon. Only the verbs found in this data more than 150 times were included in the experiment.

For verb clustering, we adopted the best performing features of Sun and Korhonen (2009): automatically acquired verb subcategorization frames (SCFs) parameterized by their selectional preferences (SPs). We obtained these features using the SCF acquisition system of Preiss et al. (2007). The system tags and parses corpus data using the RASP parser and extracts SCFs from the resulting GRs using a rule-based classifier which identifies 168 SCF types for English verbs. It produces a lexical entry for each verb and SCF combination occurring in corpus data. We obtained SPs by clustering argument heads appearing in the subject and object slots of verbs in the resulting lexicon.

Our noun dataset consists of 2000 most frequent nouns in the BNC. Following previous works on semantic noun classification (Pantel and Lin, 2002; Bergsma et al., 2008), we used GRs as features for noun clustering. We employed all the argument heads and verb lemmas appearing in the subject, direct object and indirect object relations in the RASP-parsed BNC.

The feature vectors were first constructed from the corpus counts, and subsequently normalized by the sum of the feature values before applying clustering.

### 3.2.2 Clustering Algorithm

We use spectral clustering (SPEC) for both verbs and nouns. This technique has proved to be effective in previous verb clustering works (Brew and Schulte im Walde, 2002; Sun and Korhonen, 2009) and in related NLP tasks involving high dimensional data (Chen et al., 2006). We use the MNCut algorithm for SPEC which has a wide applicability and a clear probabilistic interpretation (Meila and Shi, 2001).

The task is to group a given set of words  $W = \{w_n\}_{n=1}^N$  into a disjoint partition of  $K$  classes. SPEC takes a similarity matrix as input. We construct it using the Jensen-Shannon divergence (JSD) as a measure. The JSD between two feature

vectors  $w$  and  $w'$  is  $d_{j_{sd}}(w, w') = \frac{1}{2}D(w||m) + \frac{1}{2}D(w'||m)$  where  $D$  is the Kullback-Leibler divergence, and  $m$  is the average of the  $w$  and  $w'$ .

The similarity matrix  $S$  is constructed where  $S_{ij} = \exp(-d_{j_{sd}}(w, w'))$ . In SPEC, the similarities  $S_{ij}$  are viewed as weights on the edges  $ij$  of a graph  $G$  over  $W$ . The similarity matrix  $S$  is thus the adjacency matrix for  $G$ . The degree of a vertex  $i$  is  $d_i = \sum_{j=1}^N S_{ij}$ . A cut between two partitions  $A$  and  $A'$  is defined to be  $\text{Cut}(A, A') = \sum_{m \in A, n \in A'} S_{mn}$ .

The similarity matrix  $S$  is then transformed into a stochastic matrix  $P$ .

$$P = D^{-1}S \quad (1)$$

The degree matrix  $D$  is a diagonal matrix where  $D_{ii} = d_i$ .

It was shown by Meila and Shi (2001) that if  $P$  has the  $K$  leading eigenvectors that are piecewise constants<sup>2</sup> with respect to a partition  $I^*$  and their eigenvalues are not zero, then  $I^*$  minimizes the multiway normalized cut (MNCut):

$$\text{MNCut}(I) = K - \sum_{k=1}^K \frac{\text{Cut}(I_k, I_k)}{\text{Cut}(I_k, I)}$$

$P_{mn}$  can be interpreted as the transition probability between the vertexes  $m, n$ . The criterion can thus be expressed as  $\text{MNCut}(I) = \sum_{k=1}^K (1 - P(I_k \rightarrow I_k | I_k))$  (Meila, 2001), which is the sum of transition probabilities across different clusters. This criterion finds the partition where random walks are most likely to happen within the same cluster. In practice, the leading eigenvectors of  $P$  are not piecewise constants. However, we can extract the partition by finding the approximately equal elements in the eigenvectors using a clustering algorithm, such as K-Means.

Since SPEC has elements of randomness, we ran the algorithm multiple times and the partition that minimizes the distortion (the distances to cluster centroid) is reported. Some of the clusters obtained as a result of applying the algorithm to our noun and verb datasets are demonstrated in Figures 1 and 2 respectively. The noun clusters represent target concepts that we expect to be associated with the same source concept (some suggested source concepts are given in Figure 1, although the system only captures those implicitly).

<sup>2</sup>An eigenvector  $v$  is piecewise constant with respect to  $I$  if  $v(i) = v(j) \forall i, j \in I_k$  and  $k \in 1, 2 \dots K$

Source: MECHANISM  
 Target Cluster: consensus relation tradition partnership  
 resistance foundation alliance friendship contact reserve  
 unity link peace bond myth identity hierarchy relation-  
 ship connection balance marriage democracy defense  
 faith empire distinction coalition regime division  
 Source: STORY; JOURNEY  
 Target Cluster: politics practice trading reading occupa-  
 tion profession sport pursuit affair career thinking life  
 Source: LOCATION; CONTAINER  
 Target Cluster: lifetime quarter period century succes-  
 sion stage generation decade phase interval future  
 Source: LIVING BEING; END  
 Target Cluster: defeat fall death tragedy loss collapse de-  
 cline disaster destruction fate

Figure 1: Clustered target concepts

Source Cluster: sparkle glow widen flash flare gleam  
 darken narrow flicker shine blaze bulge  
 Source Cluster: gulp drain stir empty pour sip spill swal-  
 low drink pollute seep flow drip purify ooze pump bub-  
 ble splash ripple simmer boil tread  
 Source Cluster: polish clean scrape scrub soak  
 Source Cluster: kick hurl push fling throw pull drag haul  
 Source Cluster: rise fall shrink drop double fluctuate  
 dwindle decline plunge decrease soar tumble surge spiral  
 boom

Figure 2: Clustered verbs (source domains)

The verb clusters contain coherent lists of source domain vocabulary.

### 3.3 Selectional Preference Strength Filter

Following Wilks (1978), we take metaphor to represent a violation of selectional restrictions. However, not all verbs have an equally strong capacity to constrain their arguments, e.g. *remember*, *accept*, *choose* etc. are weak in that respect. We suggest that for this reason not all the verbs would be equally prone to metaphoricity, but only the ones exhibiting strong selectional preferences. We test this hypothesis experimentally and expect that placing this criterion would enable us to filter out a number of candidate expressions, that are less likely to be used metaphorically.

We automatically acquired selectional preference distributions for Verb-Subject and Verb-Object relations from the BNC parsed by RASP. We first clustered 2000 most frequent nouns in the BNC into 200 clusters using SPEC as described in the previous section. The obtained clusters formed our selectional preference classes. We adopted the selectional preference

measure proposed by Resnik (1993) and successfully applied to a number of tasks in NLP including word sense disambiguation (Resnik, 1997). Resnik models selectional preference of a verb in probabilistic terms as the difference between the posterior distribution of noun classes in a particular relation with the verb and their prior distribution in that syntactic position regardless of the identity of the predicate. He quantifies this difference using the relative entropy (or Kullback-Leibler distance), defining the *selectional preference strength* (SPS) as follows.

$$S_R(v) = D(P(c|v)||P(c)) = \sum_c P(c|v) \log \frac{P(c|v)}{P(c)}, \quad (2)$$

where  $P(c)$  is the prior probability of the noun class,  $P(c|v)$  is the posterior probability of the noun class given the verb and  $R$  is the grammatical relation in question. SPS measures how strongly the predicate constrains its arguments.

We use this measure to filter out the verbs with weak selectional preferences. The optimal SPS threshold was set experimentally on a small held-out dataset and approximates to 1.32. We excluded expressions containing the verbs with preference strength below this threshold from the set of candidate metaphors.

## 4 Evaluation and Discussion

In order to prove that our metaphor identification method generalizes well over the seed set and goes far beyond synonymy, we compared its output to that of a baseline taking WordNet synsets to represent source and target domains. We evaluated the quality of metaphor tagging in terms of precision with the help of human judges.

### 4.1 Comparison against WordNet Baseline

The baseline system was implemented using synonymy information from WordNet to expand on the seed set. Assuming all the synonyms of the verbs and nouns in seed expressions to represent the source and target vocabularies respectively, the system searches for phrases composed of lexical items belonging to those vocabularies. For example, given a seed expression *stir excitement*, the baseline finds phrases such as *arouse fervour*,

*stimulate agitation, stir turmoil* etc. However, it is not able to generalize over the concepts to broad semantic classes, e.g. it does not find other *feelings* such as *rage, fear, anger, pleasure* etc., which is necessary to fully characterize the target domain. The same deficiency of the baseline system manifests itself in the source domain vocabulary: the system has only the knowledge of direct synonyms of *stir*, as opposed to other verbs characteristic to the domain of *liquids*, e.g. *pour, flow, boil* etc., successfully identified by means of clustering.

To compare the **coverage** achieved by unsupervised clustering to that of the baseline in quantitative terms, we estimated the number of WordNet synsets, i.d. different word senses, in the metaphorical expressions captured by the two systems. We found that the baseline system covers only 13% of the data identified using clustering and does not go beyond the concepts present in the seed set. In contrast, most metaphors tagged by our method are novel and represent a considerably wider range of meanings, e.g. given the seed metaphors *stir excitement, throw remark, cast doubt* the system identifies previously unseen expressions *swallow anger, hurl comment, spark enthusiasm* etc. as metaphorical.

#### 4.2 Comparison with Human Judgements

In order to access the **quality** of metaphor identification by both systems we used the help of human annotators. The annotators were presented with a set of randomly sampled sentences containing metaphorical expressions as annotated by the system and by the baseline. They were asked to mark the tagged expressions that were metaphorical in their judgement as correct.

The annotators were encouraged to rely on their own intuition of metaphor. However, we also provided some guidance in the form of the following definition of metaphor<sup>3</sup>:

1. For each verb establish its meaning in context and try to imagine a more basic meaning of this verb on other contexts. Basic meanings normally are: (1) more concrete; (2) re-

<sup>3</sup>taken from the annotation procedure of Shutova and Teufel (2010) that is in turn partly based on the work of Pragglejaz Group (2007).

<p>CKM 391 Time and time again he would stare at the ground, hand on hip, if he thought he had received a bad call, and then <b>swallow his anger</b> and play tennis.</p> <p>AD9 3205 He tried to <b>disguise the anxiety</b> he felt when he found the comms system down, but Tammuz was nearly hysterical by this stage.</p> <p>AMA 349 We will <b>halt the reduction</b> in NHS services for long-term care and community health services which support elderly and disabled patients at home.</p> <p>ADK 634 <b>Catch their interest</b> and <b>spark their enthusiasm</b> so that they begin to see the product's potential.</p> <p>K2W 1771 The committee heard today that gangs regularly <b>hurled</b> abusive <b>comments</b> at local people, making an unacceptable level of noise and leaving litter behind them.</p>
--

Figure 3: Sentences tagged by the system (metaphors in bold)

lated to bodily action; (3) more precise (as opposed to vague); (4) historically older.

2. If you can establish the basic meaning that is distinct from the meaning of the verb in this context, the verb is likely to be used metaphorically.

We had 5 volunteer annotators who were all native speakers of English and had no or sparse linguistic knowledge. Their agreement on the task was 0.63 in terms of  $\kappa$  (Siegel and Castellan, 1988), whereby the main source of disagreement was the presence of highly lexicalized metaphors, e.g. verbs such as *adopt, convey, decline* etc. We then evaluated the system performance against their judgements in terms of *precision*. Precision measures the proportion of metaphorical expressions that were tagged correctly among the ones that were tagged. We considered the expressions tagged as metaphorical by at least three annotators to be correct. As a result our system identifies metaphor with the precision of 0.79, whereas the baseline only attains 0.44. Some examples of sentences annotated by the system are shown in Figure 3.

Such a striking discrepancy between the performance levels of the clustering approach and the baseline can be explained by the fact that a large number of metaphorical senses are included in WordNet. This means that in WordNet synsets source domain verbs are mixed with more abstract terms. For example, the metaphorical sense of *shape* in *shape opinion* is part of the synset (*de-*

*termine, shape, mold, influence, regulate*). This results in the baseline system tagging literal expressions as metaphorical, erroneously assuming that the verbs from the synset belong to the source domain.

The main source of confusion in the output of our clustering method was the conventionality of some metaphorical expressions, e.g. *hold views, adopt traditions, tackle a problem*. The system is capable of tracing metaphorical etymology of conventional phrases, but their senses are highly lexicalized. This lexicalization is reflected in the data and affects clustering in that conventional metaphors are sometimes clustered together with literally used terms, e.g. *tackle a problem* and *resolve a problem*, which may suggest that the latter are metaphorical. It should be noted, however, that such errors are rare.

Since there is no large metaphor-annotated corpus available, it was impossible for us to reliably evaluate the *recall* of the system. However, the system identified a total number of 4456 metaphorical expressions in the BNC starting with a seed set of only 62, which is a promising result.

## 5 Related Work

One of the first attempts to identify and interpret metaphorical expressions in text automatically is the approach of Fass (1991). Fass developed a system called *met\**, capable of discriminating between literalness, metonymy, metaphor and anomaly. It does this in three stages. 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 (such as CONTAINER-for-CONTENT). 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. E.g., the sentence in (5) would be represented in this framework as (*car,drink,gasoline*), which does not satisfy the preference (*animal,drink,liquid*), as *car* is not a hyponym of *animal*. *met\** then searches its knowledge base for a triple containing a hypernym of both the actual ar-

gument and the desired argument and finds (*thing,use,energy\_source*), which represents the metaphorical interpretation.

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.

The method of Gedigan 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. They used PropBank annotation (arguments and their semantic types) as features to train the classifier and report an accuracy of 95.12%. This result is, however, only a little higher than the performance of the naive baseline assigning majority class to all instances (92.90%). These numbers can be explained by the fact that 92.00% of the verbs of MOTION and CURE in the Wall Street Journal corpus are used metaphorically, thus making the dataset unbalanced with respect to the target categories and the task notably easier.

Both Birke and Sarkar (2006) and Gedigan 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*<sup>4</sup>) 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 (e.g. *He planted good ideas in their minds* or *He has a fertile imagination*). 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. However, by using bigram counts over verb-noun pairs as opposed to verb-object relations extracted from parsed text Krishnakumaran and Zhu (2007) lose a great deal of information. The authors evaluated their system on a set of example sentences compiled from the Master Metaphor List (Lakoff et al., 1991), whereby highly conventionalized metaphors (they call them *dead metaphors*) are taken to be negative examples. Thus, they do not deal with literal examples as such: essentially, the distinction they are making is between the senses included in WordNet, even if they are conventional metaphors, and those not included in WordNet.

## 6 Conclusions and Future Directions

We presented a novel approach to metaphor identification in unrestricted text using unsupervised methods. Starting from a limited set of metaphorical seeds, the system is capable of capturing the regularities behind their production and annotating a much greater number and wider range of previously unseen metaphors in the BNC.

Our system is the first of its kind and it is capable of identifying metaphorical expressions with a high precision (0.79). By comparing its coverage to that of a WordNet baseline, we proved that our method goes far beyond synonymy and generalizes well over the source and target domains. Although at this stage we tested our system on verb-subject and verb-object metaphors only, we are

<sup>4</sup>William Shakespeare

convinced that the described identification techniques can be similarly applied to a wider range of syntactic constructions. Extending the system to deal with more parts of speech and types of phrases is part of our future work.

One possible limitation of our approach is that it is seed-dependent, which makes the recall of the system questionable. Thus, another important future research avenue is the creation of a more diverse seed set. We expect that a set of expressions representative of the whole variety of common metaphorical mappings, already described in linguistics literature, would enable the system to attain a very broad coverage of the corpus. Master Metaphor List (Lakoff et al., 1991) and other existing metaphor resources could be a sensible starting point on a route to such a dataset.

## Acknowledgments

We are very grateful to our anonymous reviewers for their useful feedback on this work and the volunteer annotators for their interest, time and help. This research is funded by generosity of Cambridge Overseas Trust (Katia Shutova), Dorothy Hodgkin Postgraduate Award (Lin Sun) and the Royal Society, UK (Anna Korhonen).

## References

- Andersen, O. E., J. Nioche, E. Briscoe, and J. Carroll. 2008. The BNC parsed with RASP4UIMA. In *Proceedings of LREC 2008*, Marrakech, Morocco.
- Bergsma, S., D. Lin, and R. Goebel. 2008. Discriminative learning of selectional preference from unlabeled text. In *Proceedings of the EMNLP*.
- Birke, J. and A. Sarkar. 2006. A clustering approach for the nearly unsupervised recognition of nonliteral language. In *In Proceedings of EACL-06*, pages 329–336.
- Brew, C. and S. Schulte im Walde. 2002. Spectral clustering for German verbs. In *Proceedings of EMNLP*.
- Briscoe, E., J. Carroll, and R. Watson. 2006. The second release of the rasp system. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 77–80.
- Burnard, L. 2007. *Reference Guide for the British National Corpus (XML Edition)*.

- Chen, J., D. Ji, C. Lim Tan, and Z. Niu. 2006. Un-supervised relation disambiguation using spectral clustering. In *Proceedings of COLING/ACL*.
- Fass, D. 1991. met\*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.
- Fellbaum, C., editor. 1998. *WordNet: An Electronic Lexical Database (ISBN: 0-262-06197-X)*. MIT Press, first edition.
- Fillmore, C. J., C. R. Johnson, and M. R. L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16(3):235–250.
- Gedigan, M., J. Bryant, S. Narayanan, and B. Ciric. 2006. Catching metaphors. In *In Proceedings of the 3rd Workshop on Scalable Natural Language Understanding*, pages 41–48, New York.
- Joanis, E., S. Stevenson, and D. James. 2008. A general feature space for automatic verb classification. *Natural Language Engineering*, 14(3):337–367.
- Karov, Y. and S. Edelman. 1998. Similarity-based word sense disambiguation. *Computational Linguistics*, 24(1):41–59.
- Kingsbury, P. and M. Palmer. 2002. From TreeBank to PropBank. In *Proceedings of LREC-2002*, Gran Canaria, Canary Islands, Spain.
- Kipper, K., A. Korhonen, N. Ryant, and M. Palmer. 2006. Extensive classifications of English verbs. In *Proceedings of the 12th EURALEX International Congress*.
- Korhonen, A., Y. Krymolowski, and Z. Marx. 2003. Clustering polysemic subcategorization frame distributions semantically. In *Proceedings of ACL 2003*, Sapporo, Japan.
- Korhonen, A., Y. Krymolowski, and T. Briscoe. 2006. A large subcategorization lexicon for natural language processing applications. In *Proceedings of LREC 2006*.
- Krishnakumaran, S. and X. Zhu. 2007. Hunting elusive metaphors using lexical resources. In *Proceedings of the Workshop on Computational Approaches to Figurative Language*, pages 13–20, Rochester, NY.
- Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago.
- Lakoff, G., J. Espenson, and A. Schwartz. 1991. The master metaphor list. Technical report, University of California at Berkeley.
- Levin, B. 1993. *English Verb Classes and Alternations*. University of Chicago Press, Chicago.
- Lin, D. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the 17th international conference on Computational linguistics*, pages 768–774.
- Martin, J. H. 1988. Representing regularities in the metaphoric lexicon. In *Proceedings of the 12th conference on Computational linguistics*, pages 396–401.
- Meila, M. and J. Shi. 2001. A random walks view of spectral segmentation. In *AISTATS*.
- Meila, M. 2001. The multicut lemma. Technical report, University of Washington.
- Pantel, P. and D. Lin. 2002. Discovering word senses from text. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 613–619. ACM.
- Pereira, F., N. Tishby, and L. Lee. 1993. Distributional clustering of English words. In *Proceedings of ACL-93*, pages 183–190, Morristown, NJ, USA.
- Pragglejaz Group. 2007. MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22:1–39.
- Preiss, J., T. Briscoe, and A. Korhonen. 2007. A system for large-scale acquisition of verbal, nominal and adjectival subcategorization frames from corpora. In *Proceedings of ACL-2007*, volume 45, page 912.
- Resnik, P. 1993. *Selection and Information: A Class-based Approach to Lexical Relationships*. Ph.D. thesis, Philadelphia, PA, USA.
- Resnik, P. 1997. Selectional preference and sense disambiguation. In *ACL SIGLEX Workshop on Tagging Text with Lexical Semantics*, Washington, D.C.
- Rooth, M., S. Riezler, D. Prescher, G. Carroll, and F. Beil. 1999. Inducing a semantically annotated lexicon via EM-based clustering. In *Proceedings of ACL 99*, pages 104–111.
- Schulte im Walde, S. 2006. Experiments on the automatic induction of German semantic verb classes. *Computational Linguistics*, 32(2):159–194.
- Shutova, E. and S. Teufel. 2010. Metaphor corpus annotated for source - target domain mappings. In *Proceedings of LREC 2010*, Malta.
- Shutova, E. 2010. Automatic metaphor interpretation as a paraphrasing task. In *Proceedings of NAACL 2010*, Los Angeles, USA.
- Siegel, S. and N. J. Castellan. 1988. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill Book Company, New York, USA.