Automatic Metaphor Interpretation as a Paraphrasing Task

Ekaterina Shutova
Computer Laboratory
University of Cambridge
15 JJ Thomson Avenue
Cambridge CB3 0FD, UK
Ekaterina.Shutova@cl.cam.ac.uk

Abstract

We present a novel approach to metaphor interpretation and a system that produces literal paraphrases for metaphorical expressions. Such a representation is directly transferable to other applications that can benefit from a metaphor processing component. Our method is distinguished from the previous work in that it does not rely on any hand-crafted knowledge about metaphor, but in contrast employs automatically induced selectional preferences. Being the first of its kind, our system is capable of paraphrasing metaphorical expressions with a high accuracy (0.81).

1 Introduction

Metaphors arise when one concept is viewed in terms of the properties of the other. In other words it is based on *similarity* between the concepts. Similarity is a kind of association implying the presence of characteristics in common. Here are some examples of metaphor.

(1) News *travels* fast. (Lakoff and Johnson, 1980)

(2) How can I *kill* a process? (Martin, 1988)

(3) And then my heart with pleasure *fills*,
And *dances* with the daffodils.[1]

In metaphorical expressions seemingly unrelated features of one concept are associated with another concept. In the example (2) the *computational process* is viewed as something *alive* and, therefore, its forced termination is associated with the act of killing.

Metaphorical expressions represent a great variety, ranging from conventional metaphors, which we reproduce and comprehend every day, e.g. those in (1) and (2), to poetic and largely novel ones, such as (3). The use of metaphor is ubiquitous in natural language text and it is a serious bottleneck in automatic text understanding. In order to estimate the frequency of the phenomenon, we conducted a corpus study on a subset of the British National Corpus (BNC) (Burnard, 2007) representing various genres. We manually annotated metaphorical expressions in this data and found that 241 out of 761 sentences contained a metaphor or (rarely) an idiom. Due to such a high frequency of their use, a system capable of 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 recognition* (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 addressed in NLP.

To date the most influential account of metaphor recognition has been that of Wilks (1978). According to Wilks, metaphors represent a violation of selectional restrictions in a given context. Consider the following example.

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

[1] taken from the verse "I wandered lonely as a cloud" written by William Wordsworth in 1804.

Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, pages 1029–1037,
Los Angeles, California, June 2010.

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*.

Most approaches to metaphor interpretation rely on task-specific hand-coded knowledge (Fass, 1991; Martin, 1990; Narayanan, 1997; Narayanan, 1999; Feldman and Narayanan, 2004; Barnden and Lee, 2002; Agerri et al., 2007) and produce interpretations in a non-textual format. However, the ultimate objective of automatic metaphor processing is a type of interpretation that can be directly embedded into other systems to enhance their performance. Thus, we define metaphor interpretation as a *paraphrasing task* and build a system that automatically derives literal paraphrases for metaphorical expressions in unrestricted text.

In summary, our system (1) produces a list of all possible paraphrases for a metaphorical expression (induced automatically from a large corpus); (2) ranks the paraphrases according to their likelihood derived from the corpus; (3) discriminates between literal and figurative paraphrases by detecting selectional preference violation and outputs the literal ones; and (4) disambiguates the sense of the paraphrases using WordNet (Fellbaum, 1998) inventory of senses.

We tested our system on a collection of metaphorical expressions representing verb-subject and verb-object constructions, where the verb is used metaphorically. To compile this dataset we manually annotated such phrases in a subset of the BNC using the metaphor identification procedure (MIP) (Pragglejaz Group, 2007). We then evaluated the quality of paraphrasing with the help of human annotators and created a gold standard for this task.

2 Experimental Data

Since we focus on single-word metaphors expressed by a verb, our annotation task can be viewed as verb classification according to whether the verbs are used metaphorically or literally. However, some verbs have weak or no potential of being a metaphor and, thus, our study is not concerned with them. We excluded the following verb classes: (1) auxiliary verbs; (2) modal verbs; (3) aspectual verbs (e.g. *begin, start, finish)*; (4) light verbs (e.g. *take, give, put, get, make*).

2.1 The Corpus

Our corpus is a subset of the BNC. We sampled texts representing various genres: literature, newspaper/journal articles, essays on politics, international relations and history, radio broadcast (transcribed speech). The corpus contains 761 sentences and 13642 words.

2.2 Annotation Scheme

The annotation scheme we use is based on the principles of the metaphor identification procedure (MIP) developed by Pragglejaz Group (2007). We adopt their definition of *basic* sense of a word and their approach to distinguishing basic senses from the metaphorical ones. MIP involves metaphor annotation at the word level as opposed to identifying metaphorical relations (between words) or source–target domain mappings (between concepts or domains). Such annotation can be viewed as a form of word sense disambiguation with an emphasis on metaphoricity.

In order to discriminate between the verbs used metaphorically and literally we use the following procedure as part of our guidelines:

1. For each verb establish its meaning in context and try to imagine a more basic meaning of this verb on other contexts. As defined in the framework of MIP (Pragglejaz Group, 2007) basic meanings normally are: (1) more concrete; (2) related 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.

Consider the following example sentence:

(5) If he <u>asked</u> her to <u>post</u> a letter or <u>buy</u> some razor blades from the chemist, she was <u>transported</u> with pleasure.

In this sentence one needs to annotate four verbs that are underlined. The first 3 verbs are used in their basic sense, i.e. literally (*ask* in the context of “a person asking another person a question or a favour”;

post in the context of "a person posting/sending a letter"; *buy* in the sense of "making a purchase"). Thus, they are tagged as literal. The verb *transport*, however, in its basic sense is used in the context of "goods being transported/carried by a vehicle". The context in this sentence involves "a person being transported by a feeling", which contrasts the basic sense in that the agent of *transporting* is an EMOTION as opposed to a VEHICLE. Thus, we can infer that the use of *transport* in this sentence is metaphorical.

2.3 Annotation Reliability

We tested reliability of this annotation scheme using multiple annotators on a subset of the corpus. The rest of the annotation was done by a single annotator.

Annotators We had three independent volunteer annotators, who were all native speakers of English and had some linguistics background.

Material and Task All of them received the same text taken from the BNC containing 142 verbs to annotate. They were asked to classify verbs as metaphorical or literal.

Guidelines and Training The annotators received written guidelines (2 pages) and were asked to do a small annotation exercise (2 sentences containing 8 verbs in total). The goal of the exercise was to ensure they were at ease with the annotation format.

Interannotator Agreement We evaluate reliability of our annotation scheme by assessing interannotator agreement in terms of κ (Siegel and Castellan, 1988). The classification was performed with the agreement of 0.64 (κ), which is considered reliable. The main source of disagreement was the high conventionality of some expressions, i.e. cases where the metaphorical etymology could be clearly traced, but the senses are highly lexicalized.

2.4 Phrase Selection

Only the phrases that were tagged as metaphorical by all of the annotators were included in the test set. Here are some examples of such phrases: *memories were slipping away; hold the truth back; stirred an unfathomable excitement; factors shape results; mending their marriage; brushed aside the accusations* etc. In order to avoid extra noise we placed some additional criteria to select the test phrases: (1) exclude phrases where subject or object referent is unknown, e.g. containing pronouns such as in *in which they [changes] operated*; (2) exclude phrases whose metaphorical meaning is realised solely in passive constructions (e.g. *sociologists have been inclined to [..]*); (3) exclude phrases where the subject or object of interest are represented by a named entity (e.g. *Then Hillary leapt into the conversation*); (4) exclude multiword metaphors (e.g. *go on pilgrimage with Raleigh or put out to sea with Tennyson*). The resulting test set contains 62 metaphorical expressions.

3 The Method

The system takes phrases containing annotated single-word metaphors (where a verb is used metaphorically, its context is used literally) as input. It generates a list of possible paraphrases that can occur in the same context and ranks them according to their likelihood derived from the corpus. Subsequently it identifies shared features of the paraphrases and the metaphorical verb using WordNet hierarchy of concepts and removes the unrelated concepts. Among the related paraphrases it then identifies the literal ones given the context relying on the automatically induced selectional preferences.

3.1 The Model for Paraphrase Ranking

We model the likelihood of a particular paraphrase as a joint probability of the following events: the interpretation (another term to replace the one used metaphorically) i co-occurring with the other lexical items from its context $w_1, ..., w_N$ in the relations $r_1, ..., r_N$ respectively.

$$L_i = P(i, (w_1, r_1), (w_2, r_2), ..., (w_N, r_N)), \quad (1)$$

where $w_1, ..., w_N$ and $r_1, ..., r_N$ represent the fixed context of the term used metaphorically in the sentence. This context will be kept as part of the paraphrase, and the term used metaphorically will be replaced.

We take each relation of the term in a phrase to be independent from the other relations of this term in this phrase. E.g. for a verb in the presence of both the subject and the object the `Verb-Subject` and `Verb-Object` relations would be considered to be independent events within the model. This yields the following approximation:

$$P(i,(w_1,r_1),(w_2,r_2),...,(w_N,r_N)) = P(i)\cdot P((w_1,r_1)|i)\cdot ...\cdot P((w_N,r_N)|i). \quad (2)$$

We can calculate the probabilities using maximum likelihood estimation

$$P(i) = \frac{f(i)}{\sum_k f(i_k)}, \quad (3)$$

$$P(w_n, r_n|i) = \frac{f(w_n, r_n, i)}{f(i)}, \quad (4)$$

where $f(i)$ is the frequency of the interpretation on its own, $\sum_k f(i_k)$ is the number of times this part of speech is attested in the corpus and $f(w_n, r_n, i)$ - the frequency of the co-occurrence of the interpretation with the context word w_n in the relation r_n. By performing appropriate substitutions into (2) we obtain

$$P(i,(w_1,r_1),(w_2,r_2),...,(w_N,r_N)) = \frac{f(i)}{\sum_k f(i_k)}\cdot\frac{f(w_1,r_1,i)}{f(i)}\cdot ...\cdot\frac{f(w_N,r_N,i)}{f(i)} = \frac{\prod_{n=1}^{N} f(w_n,r_n,i)}{(f(i))^{N-1}\cdot\sum_k f(i_k)} \quad (5)$$

This model is then used to rank the possible replacements of the term used metaphorically in the fixed context according to the data.

3.2 Parameter Estimation

The parameters of the model were estimated from the British National Corpus that was parsed using the RASP parser of Briscoe et al. (2006). We used the grammatical relations (GRs) output of RASP for BNC created by Andersen et al. (2008). The same output of RASP was used to identify the GRs in the metaphorical expressions themselves, as the metaphor corpus from which they were extracted is a subset of the BNC. To obtain the counts for $f(w_n, r_n, i)$ we extracted all the terms appearing in the corpus in the relation r_n with w_n for each lexical item - relation pair. The initial list of replacements for the metaphorical term was constructed by taking an overlap of the lists of terms for each lexical item - relation pair.

3.3 Identifying Shared Meanings in WordNet

It should be noted that the context-based model described in 3.1 overgenerates and hence there is a need to further narrow the search space. It is acknowledged in the linguistics community that metaphor is to a great extent based on similarity between the concepts involved. We exploit this fact to refine paraphrasing. After obtaining the initial list of possible substitutes for the metaphorical term, we filter out the terms whose meaning does not share any common features with that of the metaphorical term. Consider a Computer Science metaphor *kill a process*, which stands for *terminate a process*. The basic sense of *kill* implies an *end* or *termination* of life. Thus, *termination* is the shared element of the metaphorical verb and its literal interpretation.

Such overlap of features can be identified using the hyponymy relations in the WordNet taxonomy. Within the initial list of paraphrases we select the terms that are a hypernym of the metaphorical term or share a common hypernym with it[2]. To maximize the accuracy we restrict the hypernym search to three level distance in the taxomomy. The filtered lists of metaphorical verb replacements for some of the phrases from our dataset together with their log-likelihood are demonstrated in Table 1. Selecting the highest ranked paraphrase from this list as a literal interpretation will serve as a baseline.

3.4 Filtering Based on Selectional Preferences

The obtained lists contain some irrelevant paraphrases (e.g. *contain the truth* for *hold back the truth*) and some paraphrases where the substitute is used metaphorically again (e.g. *suppress the truth*). However, the task is to identify the literal interpretation, therefore, these need to be removed.

One way of dealing with both problems at once is to take into account selectional preferences of the verbs in our list. The verbs used metaphorically are likely to demonstrate strong semantic preference for the source domain, e.g. *suppress* would select for *movements (political)* rather than *ideas*, or *truth*, (the target domain), whereas the ones used literally (e.g.,

[2]We excluded the expressions containing a term whose metaphorical sense is included in WordNet from the test set, to ensure that the system does not rely on this extra hand-coded knowledge about metaphor.

Log-likelihood	Replacement
Verb-DirectObject	
hold back truth:	
-13.09	contain
-14.15	conceal
-14.62	suppress
-15.13	hold
-16.23	keep
-16.24	defend
stir excitement:	
-14.28	create
-14.84	provoke
-15.53	make
-15.53	elicit
-15.53	arouse
-16.23	stimulate
-16.23	raise
-16.23	excite
-16.23	conjure
Subject-Verb	
report leak:	
-11.78	reveal
-12.59	issue
-13.18	disclose
-13.28	emerge
-14.84	expose
-16.23	discover

Table 1: The list of paraphrases with the initial ranking

Association	Replacement
Verb-DirectObject	
hold back truth:	
0.1161	conceal
0.0214	keep
0.0070	suppress
0.0022	contain
0.0018	defend
0.0006	hold
stir excitement:	
0.0696	provoke
0.0245	elicit
0.0194	arouse
0.0061	conjure
0.0028	create
0.0001	stimulate
≈ 0	raise
≈ 0	make
≈ 0	excite
Subject-Verb	
report leak:	
0.1492	disclose
0.1463	discover
0.0674	reveal
0.0597	issue
≈ 0	emerge
≈ 0	expose

Table 2: The list of paraphrases reranked using selectional preferences

conceal) would select for *truth*. This would potentially allow us to filter out non-literalness, as well as unrelated verbs, by selecting the verbs that the noun in the metaphorical expression matches best.

We automatically acquired selectional preference distributions of the verbs in the paraphrase lists (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 the algorithm of Sun and Korhonen (2009). The obtained clusters formed our selectional preference classes. We adopted the association 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* as follows.

$$S_R(v) = D(P(c|v)||P(c)) = \sum_c P(c|v) \log \frac{P(c|v)}{P(c)}, \tag{6}$$

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. Selectional preference strength measures how strongly the predicate constrains its arguments. In order to quantify how well a particular argument class fits the verb, Resnik defines another measure called *selectional association*:

$$A_R(v,c) = \frac{1}{S_R(v)} P(c|v) \log \frac{P(c|v)}{P(c)}. \tag{7}$$

We use this measure to rerank the paraphrases and filter out those not well suited or used metaphorically. The new ranking is demonstrated in Table 2. The expectation is that the paraphrase in the first rank (i.e. the verb with which the noun in question has the highest association) represents the literal interpretation.

3.5 Sense Disambiguation

Another feature of our system is that having identified literal interpretations, it is capable to perform their word sense disambiguation (WSD). Disambiguated metaphorical interpretations are potentially a useful source of information for NLP applications dealing with word senses.

We adopt WordNet representation of a sense. Disambiguation is performed by selecting WordNet nodes containing those verbs that share a common hypernym with the metaphorical verb. The list of disambiguated interpretations for a random selection of phrases from our dataset is demonstrated in Table 3. However, we did not evaluate the WSD of the paraphrases at this stage.

4 Evaluation and Discussion

We evaluated the paraphrases with the help of human annotators in two different experimental settings.

Setting 1: the annotators were presented with a set of sentences containing metaphorical expressions and their rank 1 paraphrases produced by the system and by the baseline. They were asked to mark the ones that have the same meaning as the term used metaphorically and are used literally in the context of the paraphrase expression as correct.

We had 7 volunteer annotators who were all native speakers of English (one bilingual) and had no or sparse linguistic expertise. Their agreement on the task was 0.62 (κ), whereby the main source of disagreement was the presence of highly lexicalised metaphorical paraphrases. We then evaluated the system performance against their judgments in terms of *accuracy*. Accuracy measures the proportion of correct literal interpretations among the paraphrases in rank 1. The results are demonstrated in Table 4, the final systems identifies literal paraphrases with the accuracy of 0.81.

Setting 2: the annotators were presented with a set of sentences containing metaphorical expressions and asked to write down all suitable literal paraphrases for the highlighted metaphorical verbs. We had 5 volunteer subjects for this experiment (note that these were people not employed in the previous setting); they were all native speakers of English and had some linguistics background. We then compiled a gold standard by incorporating all of the annotations. E.g. the gold standard for the phrase *brushed aside the accusations* contains the verbs *rejected, ignored, disregarded, dismissed, overlooked, discarded*.

We compared the system output against the gold standard using *mean reciprocal rank* (MRR) as a measure. MRR is traditionally used to evaluate the performance of Question-Answering systems. We adapted this measure in order to be able to assess ranking quality beyond rank 1 and the recall of our system. An individual metaphorical expression receives a score equal to the reciprocal of the rank at which the first correct literal interpretation (according to the human gold standard) is found among the top five paraphrases, or 0 if none of the five paraphrases contains a correct interpretation. Once the individual reciprocal ranks of metaphorical expressions are estimated their mean is computed across the dataset. The MRR of our system equals 0.63 and that of the baseline is 0.55. However, it should be noted that given that our task is open-ended, it is hard to construct a comprehensive gold standard. For example, for the phrase *stir excitement* most annotators suggested only one paraphrase *create excitement*, which is found in rank 3. However, the top ranks of the system output are occupied by *provoke* and *stimulate*, which are more precise paraphrases, although they have not occurred to the annotators. Such examples result in the system's MRR being significantly lower than its accuracy at rank 1.

The obtained results are promising, the selectional preference-based reranking yields a considerable improvement in accuracy (26%) over the baseline. However, for one of the phrases in the dataset, *mend marriage*, the new ranking overruns the correct top suggestion of the baseline, *improve marriage*, and outputs *repair marriage* as the most likely literal interpretation. This is due to both the conventionality of some metaphorical senses (in this case *repair*) and to the fact that some verbs, e.g. *improve*, expose a moderate selectional preference strength, i.e. they are equally associated with a large number of classes. This demonstrates potential drawbacks of the selectional preference-based solutions. Another

Met. Expression	Top Int.	Its WordNet Sense
Verb-DirectObject		
stir excitement	provoke	**(arouse-1 elicit-1 enkindle-2 kindle-3 evoke-1 fire-7 raise-10 provoke-1)** - call forth (emotions, feelings, and responses): "arouse pity"; "raise a smile"; "evoke sympathy"
inherit state	acquire	**(get-1 acquire-1)** - come into the possession of something concrete or abstract: "She got a lot of paintings from her uncle"; "They acquired a new pet"
reflect concern	manifest	**(attest-1 certify-1 manifest-1 demonstrate-3 evidence-1)** - provide evidence for; stand as proof of; show by one's behavior, attitude, or external attributes: "The buildings in Rome manifest a high level of architectural sophistication"; "This decision demonstrates his sense of fairness"
brush aside accusation	reject	**(reject-1)** - refuse to accept or acknowledge: "we reject the idea of starting a war"; "The journal rejected the student's paper"
Verb-Subject		
campaign *surged*	improve	**(better-3 improve-2 ameliorate-2 meliorate-2)** - to make better: "The editor improved the manuscript with his changes"
report *leaked*	disclose	**(unwrap-2 disclose-1 let_on-1 bring_out-9 reveal-2 discover-6 expose-2 divulge-1 break-15 give_away-2 let_out-2)** - make known to the public information that was previously known only to a few people or that was meant to be kept a secret: "The auction house would not disclose the price at which the van Gogh had sold"; "The actress won't reveal how old she is"
tension *mounted*	lift	**(rise-1 lift-4 arise-5 move_up-2 go_up-1 come_up-6 uprise-6)** - move upward: "The fog lifted"; "The smoke arose from the forest fire"; "The mist uprose from the meadows"

Table 3: Disambiguated paraphrases produced by the system

Relation	Baseline	System
Verb-DirectObject	0.52	0.79
Verb-Subject	0.57	0.83
Average	0.55	0.81

Table 4: Accuracy with the evaluation setting 1

controvertial example was the metaphorical expression *tension mounted*, for which the system produced a paraphrase *tension lifted* with the opposite meaning. This error is likely to have been triggered by the feature similarity component, whereby one of the senses of *lift* would stem from the same node in WordNet as the metaphorical sense of *mount*.

5 Related Work

According to Conceptual Metaphor Theory (Lakoff and Johnson, 1980) metaphor can be viewed as an analogy between two distinct domains - the *target* and the *source*. Consider the following example:

(6) He *shot down* all of my arguments. (Lakoff and Johnson, 1980)

A mapping of a concept of *argument* (target) to that of *war* (source) is employed here. The idea of such interconceptual mappings has been exploited in some NLP systems.

One of the first attempts to identify and interpret metaphorical expressions in text automatically is the approach of Fass (1991). It originates in the work of Wilks (1978) and utilizes hand-coded knowledge. Fass (1991) 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 (4) 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 argument and the desired argument and finds (*thing,use,energy_source*), which represents the metaphorical interpretation.

Almost simultaneously with the work of Fass (1991), Martin (1990) presents a Metaphor Interpretation, Denotation and Acquisition System (MIDAS). The idea behind this work is that the more specific conventional metaphors descend from the general ones. Given an example of a metaphorical expression, MIDAS searches its database for a corresponding metaphor that would explain the anomaly. If it does not find any, it abstracts from the example to more general concepts and repeats the search. If it finds a suitable general metaphor, it creates a mapping for its descendant, a more specific metaphor, based on this example. This is also how novel metaphors are acquired. MIDAS has been integrated with the Unix Consultant (UC), the system that answers users questions about Unix.

Another cohort of approaches relies on performing inferences about entities and events in the source and target domains for metaphor interpretation. These include the KARMA system (Narayanan, 1997; Narayanan, 1999; Feldman and Narayanan, 2004) and the ATT-Meta project (Barnden and Lee, 2002; Agerri et al., 2007). Within both systems the authors developed a metaphor-based reasoning framework in accordance with the theory of conceptual metaphor. The reasoning process relies on manually coded knowledge about the world and operates mainly in the source domain. The results are then projected onto the target domain using the conceptual mapping representation. The ATT-Meta project concerns metaphorical and metonymic description of mental states and reasoning about mental states using first order logic. Their system, however, does not take natural language sentences as input, but logical expressions that are representations of small discourse fragments. KARMA in turn deals with a broad range of abstract actions and events and takes parsed text as input.

Veale and Hao (2008) derive a "fluid knowledge representation for metaphor interpretation and generation", called Talking Points. Talking Points are a set of characteristics of concepts belonging to source and target domains and related facts about the world which the authors acquire automatically from WordNet and from the web. Talking Points are then organized in *Slipnet*, a framework that allows for a number of insertions, deletions and substitutions in definitions of such characteristics in order to establish a connection between the target and the source concepts. This work builds on the idea of *slippage* in knowledge representation for understanding analogies in abstract domains (Hofstadter and Mitchell, 1994; Hofstadter, 1995). Consider the metaphor *Make-up is a Western burqa*:

Make-up =>
$\equiv$ typically worn by women
$\approx$ expected to be worn by women
$\approx$ must be worn by women
$\approx$ must be worn by Muslim women
Burqa <=

By doing insertions and substitutions the system arrives from the definition *typically worn by women* to that of *must be worn by Muslim women*, and thus establish a link between the concepts of *make-up* and *burqa*. Veale and Hao (2008), however, did not evaluate to which extent their method is useful to interpret metaphorical expressions occurring in text.

6 Conclusions

We presented a novel approach to metaphor interpretation and a system that produces literal paraphrases for metaphorical expressions. Such a representation is directly transferable to other applications that can benefit from a metaphor processing component. Our method is distinguished from the previous work in that it does not rely on any hand-crafted knowledge, other than WordNet, but in contrast employs automatically induced selectional preferences.

Our system is the first of its kind and it is capable of paraphrasing metaphorical expressions with a high accuracy (0.81). Although we reported results on a test set consisting of verb-subject and verb-object metaphors only, we are convinced that the described interpretation techniques can be similarly applied to other parts of speech and a wider range of syntactic constructions. Extending the system to deal with more types of phrases is part of our future work.

Acknowledgments

I am very grateful to Anna Korhonen, Simone Teufel, Ann Copestake and the reviewers for their helpful feedback on this work, Lin Sun for sharing his noun clustering data and the volunteer annotators. My studies and, thus, this research are funded by generosity of Cambridge Overseas Trust.

References

R. Agerri, J.A. Barnden, M.G. Lee, and A.M. Wallington. 2007. Metaphor, inference and domain-independent mappings. In *Proceedings of International Conference on Recent Advances in Natural Language Processing (RANLP-2007)*, pages 17–23, Borovets, Bulgaria.

O. E. Andersen, J. Nioche, E. Briscoe, and J. Carroll. 2008. The BNC parsed with RASP4UIMA. In *Proceedings of the Sixth International Language Resources and Evaluation Conference (LREC'08)*, Marrakech, Morocco.

J.A. Barnden and M.G. Lee. 2002. An artificial intelligence approach to metaphor understanding. *Theoria et Historia Scientiarum*, 6(1):399–412.

E. Briscoe, 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.

L. Burnard. 2007. *Reference Guide for the British National Corpus (XML Edition)*. URL=http://www.natcorp.ox.ac.uk/XMLedition/URG/.

D. Fass. 1991. met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.

J. Feldman and S. Narayanan. 2004. Embodied meaning in a neural theory of language. *Brain and Language*, 89(2):385–392.

C. Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database (ISBN: 0-262-06197-X)*. MIT Press, first edition.

D. Hofstadter and M. Mitchell. 1994. The Copycat Project: A model of mental fluidity and analogy-making. In K.J. Holyoak and J. A. Barnden, editors, *Advances in Connectionist and Neural Computation Theory*, Ablex, New Jersey.

D. Hofstadter. 1995. *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. HarperCollins Publishers.

G. Lakoff and M. Johnson. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago.

J. H. Martin. 1988. Representing regularities in the metaphoric lexicon. In *Proceedings of the 12th conference on Computational linguistics*, pages 396–401.

J. H. Martin. 1990. *A Computational Model of Metaphor Interpretation*. Academic Press Professional, Inc., San Diego, CA, USA.

S. Narayanan. 1997. Knowledge-based action representations for metaphor and aspect (KARMA). Technical report, PhD thesis, University of California at Berkeley.

S. Narayanan. 1999. Moving right along: A computational model of metaphoric reasoning about events. In *In Proceedings of the National Conference on Artificial Intelligence (AAAI 99)*, pages 121–128, Orlando, Florida.

Pragglejaz Group (P. Crisp, R. Gibbs, A. Cienki, G. Low, G. Steen, L. Cameron, E. Semino, J. Grady, A. Deignan and Z. Kovecses). 2007. MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22:1–39.

P. Resnik. 1993. *Selection and Information: A Class-based Approach to Lexical Relationships*. Ph.D. thesis, Philadelphia, PA, USA.

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

S. Siegel and N. J. Castellan. 1988. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill Book Company, New York, USA.

L. Sun and A. Korhonen. 2009. Improving verb clustering with automatically acquired selectional preferences. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 638–647, Singapore, August.

T. Veale and Y. Hao. 2008. A fluid knowledge representation for understanding and generating creative metaphors. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 945–952, Manchester, UK.

Y. Wilks. 1978. Making preferences more active. *Artificial Intelligence*, 11(3):197–223.