Unsupervised Metaphor Paraphrasing using a Vector Space Model

Ekaterina Shutova[1,2,3] Tim Van de Cruys[1,4] Anna Korhonen[1,2]
(1) DTAL, University of Cambridge, UK
(2) Computer Laboratory, University of Cambridge, UK
(3) ICSI and ICBS, University of California at Berkeley, USA
(4) IRIT, UMR 5505, CNRS, Toulouse, France
katia@icsi.berkeley.edu, tim.vandecruys@irit.fr, anna.korhonen@cl.cam.ac.uk

ABSTRACT
We present the first fully unsupervised approach to metaphor interpretation, and a system that produces literal paraphrases for metaphorical expressions. Such a form of interpretation is directly transferable to other NLP applications that can benefit from a metaphor processing component. Our method is different from previous work in that it does not rely on any manually annotated data or lexical resources. First, our method computes candidate paraphrases according to the context in which the metaphor appears, using a vector space model. It then uses a selectional preference model to measure the degree of literalness of the paraphrases. The system identifies correct paraphrases with a precision of 0.52 at top rank, which is a promising result for a fully unsupervised approach.

KEYWORDS: Metaphor, paraphrasing, lexical substitution, vector space model.

Proceedings of COLING 2012: Posters, pages 1121–1130,
COLING 2012, Mumbai, December 2012.

1 Introduction

Metaphor has traditionally been viewed as an artistic device that lends vividness and distinction to its author's style. This view was first challenged by Lakoff and Johnson (1980), who claimed that it is a productive phenomenon that operates at the level of mental processes. Humans often use metaphor to describe abstract concepts through reference to more concrete experiences.

Being a characteristic property of human thought and communication, metaphor becomes an important problem for natural language processing. Shutova and Teufel (2010) have shown in an empirical study that the use of metaphor is ubiquitous in natural language text (according to their data, on average every third sentence in general domain text contains a metaphorical expression). Due to this high frequency usage, a system capable of recognizing and interpreting metaphorical expressions in unrestricted text would become an invaluable component of many semantics-oriented NLP applications.

The majority of previous computational approaches to metaphor rely on manually created knowledge and thus operate on a limited domain and are expensive to build and extend. Hand-coded knowledge has proved useful for both *metaphor identification*, i.e. distinguishing between literal and metaphorical language in text (Fass, 1991; Martin, 1990; Krishnakumaran and Zhu, 2007; Gedigian et al., 2006) and *metaphor interpretation*, i.e. identifying the intended literal meaning of a metaphorical expression (Fass, 1991; Martin, 1990; Narayanan, 1997; Barnden and Lee, 2002). However, to be applicable in a real-world setting a metaphor processing system needs to be able to identify and interpret metaphorical expressions in unrestricted text. The recent metaphor paraphrasing approach of Shutova (2010) was designed with this requirement in mind and used statistical methods, but still relied on the WordNet (Fellbaum, 1998) database to generate the initial set of paraphrases. In this paper, we take the metaphor paraphrasing task a step further and present a fully unsupervised approach to this problem. In our method, candidate substitutes for the metaphorical term are generated using a vector space model. Vector space models have been previously used in the general lexical substitution task (Mitchell and Lapata, 2008; Erk and Padó, 2008, 2009; Thater et al., 2009, 2010; Erk and Padó, 2010; Van de Cruys et al., 2011). However, (to the best of our knowledge) they have not yet been deployed in tasks involving figurative meaning transfers, such as interpretation of metonymy or metaphor. In this paper, we address this problem and apply a vector space model of word meaning in context to metaphor paraphrasing, appropriately adapting it to the task.

In comparison to lexical substitution, metaphor paraphrasing presents an additional challenge, namely that of discriminating between literal and metaphorical substitutes. Shutova (2010) used a selectional preference-based model for this purpose, obtaining encouraging results in a supervised setting. We evaluate the capacity of our vector space model to discriminate between literal and figurative paraphrases on its own, as well as integrating it with a selectional preference-based model similar to that of Shutova (2010) and thus evaluating the latter in an unsupervised setting. Our system thus operates in two steps. It first computes candidate paraphrases according to a latent model of semantic similarity based on the context of the metaphorically used word, and then measures the literalness of the candidates using a selectional preference model.

We focus on paraphrasing metaphorical verbs and evaluate our system using the dataset of Shutova (2010) especially designed for this task. The comparison against a paraphrasing gold standard provided by Shutova (2010) is complemented by an evaluation against direct human judgements of system output.

2 Method

2.1 Generation of candidate paraphrases using a vector space model

Paraphrase candidates are generated by first computing the specific meaning of the metaphorical term in its context. The meaning of a word instance in context is computed by adapting its original (global) meaning vector according to the dependency relations in which the word instance participates. For this purpose, we build a factorization model in which words, together with their window-based context words and their dependency relations, are linked to latent dimensions. Both types of contexts are combined to be able to induce broad, topical semantics as well as tight, synonym-like semantics. The factorization model allows us to determine which dimensions are important for a particular context, and adapt the dependency-based feature vector of the word accordingly. The model uses non-negative matrix factorization (NMF) (Lee and Seung, 2000) in order to find latent dimensions, using the minimization of the Kullback-Leibler divergence as an objective function. A more detailed description of the factorization model can be found in Van de Cruys et al. (2011).

Our paraphrase generation model has been trained on part of the UKWAC corpus (Baroni et al., 2009), covering about 500M words. The corpus has been part of speech tagged and lemmatized with Stanford Part-Of-Speech Tagger (Toutanova and Manning, 2000; Toutanova et al., 2003), and parsed with MaltParser (Nivre et al., 2006), so that dependency triples could be extracted.

Using the latent distributions yielded by our factorization model, it is now possible to compute the meaning vector for a particular word in context, and subsequently the most similar words to this meaning vector, which will be our candidate paraphrases.

Intuitively, the contextual features of the word (i.e. the dependency-based context features) will highlight the important semantic dimensions of the particular instance, creating a probability distribution over latent factors $p(\mathbf{z}|d_j)$. Using this probability distribution, a new probability distribution is determined over dependency features given the context, following equation 1.

$$p(\mathbf{d}|d_j) = p(\mathbf{z}|d_j)p(\mathbf{d}|\mathbf{z}) \tag{1}$$

The last step is to weight the original probability vector of the word according to the probability vector of the dependency features given the word's context, by taking the pointwise multiplication of probability vectors $p(\mathbf{d}|w_i)$ and $p(\mathbf{d}|d_j)$.

$$p(\mathbf{d}|w_i, d_j) = p(\mathbf{d}|w_i) \cdot p(\mathbf{d}|d_j) \tag{2}$$

This final step is a crucial one in the model. The model is not just based on latent factors; rather, the latent factors are used to determine which of the features in the original word vector are the salient ones given a particular context. This allows us to compute an accurate adaptation of the original word vector in context.

As an example, take the metaphorical expression *reflect concern*. We want to compute the meaning vector for the verb *reflect* (w_i) in the context of its direct object, $concern_{dobj}$ (d_j). Using the probability distribution over latent factors given the dependency context $p(\mathbf{z}|d_j)$ (a result that comes out of the factorization), we can compute the probability of dependency features given the context – $p(\mathbf{d}|d_j)$.

The former step yields a general probability distribution over dependency features that tells us how likely a particular dependency feature is given the context $concern_{dobj}$ that the verb

appears in. Our last step is now to weight the original probability vector of the target word (the aggregate of dependency-based context features over all contexts of the target word) according to the new distribution given the context in which the verb appears. Features associated with *concern* (or more specifically, the dependency features associated with latent factors that are related to the feature $concern_{dobj}$) will be emphasized, while features associated with unrelated latent factors are leveled out. We can now return to our original matrix **A** and compute the top similar words for the adapted vector of *reflect* given the dependency feature $concern_{dobj}$, which yields the results presented in 1. If we instead compute the meaning vector for *reflect* given the dependency feature $light_{dobj}$ (as in the non-metaphorical expression *reflect light*), we get the results in 2.

1. **reflect**$_V$, *concern*$_{dobj}$: *address*, *highlight*, *express*, ...
2. **reflect**$_V$, *light*$_{dobj}$: *emit*, *shine*, *flash*, ...

The top 6 candidate paraphrases the model produces for some example metaphorical expressions are shown in Table 1.

similarity score	replacement
verb – direct object	
reflect concern	
0.1657	address
0.1638	highlight
0.1608	*express*
0.1488	focus
0.1473	outline
0.1415	comment
subject – verb	
campaign *surge*	
0.1492	subside
0.1214	*intensify*
0.1146	erupt
0.0967	plummet
0.0935	swell
0.0928	slump

Table 1: The list of paraphrases with the initial ranking. The correct paraphrases are printed in italic.

association	replacement
verb – direct object	
reflect concern	
0.1822	*express*
0.0809	nurture
0.0771	share
0.0522	reinforce
0.0088	demonstrate
0.0088	lack
subject – verb	
campaign *surge*	
0.0377	*intensify*
0.0028	sweep
0.0009	boom
≈ 0	grow
≈ 0	sweep
≈ 0	plunge

Table 2: Paraphrases re-ranked by the selectional preference model. Correct paraphrases are printed in italic.

2.2 Reranking of candidate paraphrases using a selectional preference model

The candidate lists which are generated by the vector space model contain a number of substitutes that retain the meaning of a metaphorical expression as closely as possible. However, due to the fact that the model favours the substitutes that are similar to the metaphorical verb, the highly-ranked substitutes are sometimes also metaphorically used. For example, "*speed up* change" is the top-ranked paraphrase for "*accelerate* change" and the literal paraphrase "*facilitate* change" appears only in rank 10. As the task is to identify the literal interpretation, this ranking still needs to be refined.

Following Shutova (2010), we use a selectional preference model to discriminate between literally and metaphorically used substitutes. Verbs used metaphorically are likely to demonstrate semantic preference for the source domain, e.g. *speed up* would select for MACHINES, or VEHICLES, rather than CHANGE (the target domain), whereas the ones used literally for the target domain, e.g. *facilitate* would select for PROCESSES (including CHANGE). We therefore expect that selecting the verbs whose preferences the noun in the metaphorical expression matches best should allow us to filter out non-literalness.

We automatically acquired selectional preference (SP) distributions of the candidate substitutes (for subject-verb and verb-object relations) from the British National Corpus (BNC) (Burnard, 2007) parsed by the RASP parser (Briscoe et al., 2006). We obtained SP classes by clustering the 2000 most frequent nouns in the BNC into 200 clusters using the algorithm of Sun and Korhonen (2009). We quantified selectional preferences using the association measure proposed by Resnik (1993). It represents SPs 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 irrespective of the identity of the verb. This difference then defines the *selectional preference strength* (SPS) of the verb, quantified in terms of Kullback-Leibler divergence as follows.

$$S_R(v) = D(P(c|v)||P(c)) = \sum_c P(c|v) \log \frac{P(c|v)}{P(c)}, \quad (3)$$

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. SPS measures how strongly the predicate constrains its arguments. Resnik then quantifies how well a particular argument class fits the verb using another measure called *selectional association*:

$$A_R(v,c) = \frac{1}{S_R(v)} P(c|v) \log \frac{P(c|v)}{P(c)} \quad (4)$$

We use selectional association as a measure of semantic fitness, i.e. literalness, of the paraphrases. The selectional preference model was applied to the top 20 substitutes suggested by the vector space model. The threshold of 20 substitutes was set experimentally on a small development set. The paraphrases were re-ranked based on their selectional association with the noun in the context. Those paraphrases that are not well suited or used metaphorically are dispreferred within this ranking. The new ranking (top 6 paraphrases) is shown in Table 2. The expectation is that the paraphrase in the first rank (i.e. the verb with which the noun in the context has the highest association) represents a literal interpretation.

3 Evaluation and Discussion

We compared the rankings of the initial candidate generation by the vector space model (**VS**) and the selectional preference-based reranking (**SP**) to that of an unsupervised paraphrasing baseline. We thus evaluated the ability of **VS** on its own to detect literal paraphrases, as well as the effectiveness of the **SP** model of Shutova (2010) in an unsupervised setting and in combination with **VS**.

3.1 Dataset

To our knowledge, the only metaphor paraphrasing dataset and gold standard available to date is that of Shutova (2010). We used this dataset to develop and test our system. Shutova (2010)

annotated metaphorical expressions in a subset of the BNC sampling various genres: literature, newspaper/journal articles, essays on politics, international relations and sociology, radio broadcast (transcribed speech). The dataset consists of 62 phrases that include a metaphorical verb and either a subject or a direct object. The metaphorical expressions in the dataset include e.g. *stir excitement, reflect enthusiasm, accelerate change, grasp theory, cast doubt, suppress memory, throw remark* (verb-object constructions) and *campaign surged, factor shaped [..], tension mounted, ideology embraces, changes operated, approach focuses, example illustrates* (subject-verb constructions). 10 phrases in the dataset were used during development to observe the behavior of the system, and the remaining 52 constituted the test set. 11 of them were subject-verb constructions and 41 were verb-direct object constructions.

3.2 Baseline system

The baseline system is also unsupervised and incorporates two methods: that of generating most similar substitutes for the metaphorical verb regardless of its context and a method for their re-ranking based on the likelihood of their co-occurrence with the noun in the metaphorical expression. Thus a list of most similar substitutes is first generated using a standard dependency-based vector space model (Padó and Lapata, 2007). The likelihood of a paraphrase is then calculated as a joint probability of the candidate substitutes and the noun in the context as follows:

$$L_v = P(v,n)) = P(v) \cdot P(n|v) = \frac{f(v)}{\sum_k f(v_k)} \cdot \frac{f(v,n)}{f(v)} = \frac{f(v,n)}{\sum_k f(v_k)} \quad (5)$$

where $f(v,n)$ is the frequency of the co-occurrence of the substitute with the context and $\sum_k f(v_k)$ is the total number of verbs in the corpus.

3.3 Evaluation method and results

We evaluated the paraphrases with the aid of human judges and against a human-created gold standard in two different experimental settings.

Setting 1 Human judges were presented with a set of sentences containing metaphorical expressions and their rank 1 paraphrases produced by **VS**, by **SP** and by the baseline, randomised. They were asked to mark the ones that have the same meaning as the metaphorically used term – and are used literally in the context of the paraphrase expression – as correct.

We had 4 volunteer annotators who were all native speakers of English and had no or sparse linguistic expertise. Their agreement on the task was $\kappa = 0.54$ ($n = 2, N = 115, k = 4$), whereby the main source of disagreement was the presence of highly conventionalised metaphorical paraphrases. We then evaluated the system performance against their judgements in terms of precision at rank 1, $P(1)$. Precision at rank 1 measures the proportion of correct literal interpretations among the paraphrases in rank 1. A paraphrase was considered correct if at least 3 judges out of 4 marked it as such. The results are demonstrated in Table 3. The **VS** model identifies literal paraphrases with $P(1) = 0.48$ and the **SP** model with $P(1) = 0.52$. Both models outperform the baseline that only achieves $P(1) = 0.40$.

Setting 2 We then also evaluated **VS**, **SP** and baseline rankings against a human-constructed paraphrasing gold standard. The gold standard was created by Shutova (2010) as follows. Five independent annotators were presented with a set of sentences containing metaphorical

relation	baseline $P(1)$	**VS** $P(1)$	**SP** $P(1)$	Shutova (2010) $P(1)$
verb – direct object	0.37	0.43	0.48	0.79
verb – subject	0.54	0.64	0.72	0.83
Average across dataset	0.40	0.48	0.52	0.81

Table 3: Results in the evaluation setting 1

expressions and asked to write down all suitable literal paraphrases for the metaphorical verbs. The annotators were all native speakers of English and had some linguistics background. Shutova (2010) then compiled a gold standard incorporating all of their annotations. For example, the gold standard for the phrase *brushed aside the accusations* contained the verbs *rejected, ignored, disregarded, dismissed, overlooked, discarded*.

However, given that the metaphor paraphrasing task is open-ended, it is hard to construct a comprehensive gold standard. For example, for the phrase *stir excitement* the gold standard includes the paraphrase *create excitement*, but not *provoke excitement* or *stimulate excitement*, which are more precise paraphrases. Thus the gold standard evaluation may unfairly penalise the system, which motivates our two-phase evaluation against both the gold standard and direct judgements of system output.

The system output was compared against the gold standard using *mean average precision* (MAP) as a measure. MAP is defined as follows:

$$MAP = \frac{1}{M}\sum_{j=1}^{M}\frac{1}{N_j}\sum_{i=1}^{N_j}P_{ji}, \tag{6}$$

where M is the number of metaphorical expressions, N_j is the number of correct paraphrases for the metaphorical expression, P_{ji} is the precision at each correct paraphrase (the number of correct paraphrases among the top i ranks). First, average precision is estimated for individual metaphorical expressions, and then the mean is computed across the dataset. This measure allows us to assess ranking quality beyond rank 1, as well as the recall of the system. As compared to the gold standard, MAP of **VS** is 0.40, MAP of **SP** is 0.41 and that of the baseline is 0.37.

3.4 Discussion

Our system consistently produces better results than the baseline, with an improvement of 12% in precision on our human evaluation (**SP**) and an improvement of 4% MAP on the gold standard (**SP**). At first sight, these improvements of our unsupervised system may not seem very high, in particular when compared to the results of the supervised system of Shutova (2010). Note, however, that our results are in line with the performance of unsupervised approaches on the lexical substitution task. Unsupervised approaches to lexical substitution perform well below their supervised counterparts (which are usually based on WordNet), and often have difficulties getting significant improvements over a baseline of a simple dependency-based vector space model of semantic similarity (Erk and Padó, 2008; Van de Cruys et al., 2011). We therefore think that the method presented here takes a promising step in the direction of unsupervised metaphor paraphrasing.

The **SP** re-ranking of the candidates yields an improvement over the **VS** model used on its own, as expected. Our data analysis has shown that **SP** produces higher quality top paraphrases with

respect to their literalness, however the two models perform similarly on the meaning retention task (according to our own judgements 55% of the top ranked paraphrases had a similar meaning to that of a metaphorical verb for both models). The difference in MAP scores of the two models is, however, not as high as that of the respective $P(1)$ scores. This can be explained by the fact that the **VS** model produces a number of antonymous candidates. The candidates are then re-ranked by the **SP** model which does not consider meaning retention, but rather the semantic fit of a candidate interpretation in the context. As a result, a number of antonymous paraphrases that are highly associated with the noun in the context get ranked above some of the correct literal paraphrases, lowering the method's MAP score. For example, the antonymous paraphrase *tension eased* for the metaphorical expression *tension mounted* is ranked higher than the correct paraphrase *tension intensified*. In general, antonymous paraphrasing was the most common type of error. Antonyms are known to attract high similarity scores within a distributional similarity framework. This is an issue that needs to be addressed in the future, in lexical substitution in general and metaphor paraphrasing in particular.

Although the **SP** model generally improves the initial **VS** ranking, there were some instances where this was not the case. One such example is the metaphorical expression *break agreement*. The top ranked paraphrases suggested in the first step, *breach* and *violate*, were overrun by the well matching paraphrases *ratify* and *sign*, that have a different – almost opposite – meaning.

The baseline tends to produce metaphorical paraphrases rather than literal ones. However, in a few cases the baseline suggests better rank 1 paraphrases than the system. For example, it interprets the expression *leak a report* as *circulate a report*, as opposed to *print a report* incorrectly suggested by the system. This is due to the fact that the paraphrase generation relies entirely on one single context word (in this case *report*); taking a broader context into account might alleviate this problem.

4 Conclusion

In this paper we presented the first fully unsupervised approach to metaphor interpretation. Our system produces literal paraphrases for metaphorical expressions in unrestricted text. Producing metaphorical interpretations in textual format makes our system directly usable by other NLP applications that can benefit from a metaphor processing component. The fact that, unlike all previous approaches to this problem, our system does not use any supervision makes it easily scalable to new domains and applications, as well as portable to a wider range of languages.

Our method identifies literal paraphrases for metaphorical expressions with a precision of 0.52 measured at top-ranked paraphrases. Given the unsupervised nature of our system and considering the state-of-the-art in unsupervised lexical substitution, we consider this a promising result. Following Shutova (2010), the current experimental design and test set focuses on subject-verb and verb-object metaphors only, but we expect the method to be equally applicable to other parts of speech and a wider range of syntactic constructions. Our context-based vector space model is suited to all part-of-speech classes and types of relations. Selectional preferences have been previously successfully acquired not only for verbs, but also for nouns, adjectives and even prepositions (Brockmann and Lapata, 2003; Zapirain et al., 2009; Ó Séaghdha, 2010). Extending the system to deal with further syntactic constructions is thus part of our future work.

Acknowledgements

The work in this paper was funded by the Royal Society (UK), the Isaac Newton Trust (Cambridge, UK), and EU grant 7FP-ITC-248064 'PANACEA'.

References

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

Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009). The wacky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226.

Briscoe, E., Carroll, J., and Watson, R. (2006). The second release of the rasp system. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 77–80.

Brockmann, C. and Lapata, M. (2003). Evaluating and combining approaches to selectional preference acquisition. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics - Volume 1*, EACL '03, pages 27–34, Budapest, Hungary.

Burnard, L. (2007). *Reference Guide for the British National Corpus (XML Edition)*.

Erk, K. and Padó, S. (2008). A structured vector space model for word meaning in context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 897–906, Waikiki, Hawaii, USA.

Erk, K. and Padó, S. (2009). Paraphrase assessment in structured vector space: Exploring parameters and datasets. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*, pages 57–65, Athens, Greece.

Erk, K. and Padó, S. (2010). Exemplar-based models for word meaning in context. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 92–97, Uppsala, Sweden.

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.

Gedigian, M., Bryant, J., Narayanan, S., and Ciric, B. (2006). Catching metaphors. In *In Proceedings of the 3rd Workshop on Scalable Natural Language Understanding*, pages 41–48, New York.

Krishnakumaran, S. and Zhu, X. (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 Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press, Chicago.

Lee, D. D. and Seung, H. S. (2000). Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems 13*, pages 556–562.

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

Mitchell, J. and Lapata, M. (2008). Vector-based models of semantic composition. *proceedings of ACL-08: HLT*, pages 236–244.

Narayanan, S. (1997). Knowledge-based Action Representations for Metaphor and Aspect (KARMA). Technical report, PhD thesis, University of California at Berkeley.

Nivre, J., Hall, J., and Nilsson, J. (2006). Maltparser: A data-driven parser-generator for dependency parsing. In *Proceedings of LREC-2006*, pages 2216–2219.

Ó Séaghdha, D. (2010). Latent variable models of selectional preference. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden.

Padó, S. and Lapata, M. (2007). Dependency-based construction of semantic space models. *Computational Linguistics*, 33(2):161–199.

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

Shutova, E. (2010). Automatic metaphor interpretation as a paraphrasing task. In *Proceedings of NAACL 2010*, Los Angeles, USA.

Shutova, E. and Teufel, S. (2010). Metaphor corpus annotated for source - target domain mappings. In *Proceedings of LREC 2010*, Malta.

Sun, L. and Korhonen, A. (2009). Improving verb clustering with automatically acquired selectional preferences. In *Proceedings of EMNLP 2009*, pages 638–647, Singapore.

Thater, S., Dinu, G., and Pinkal, M. (2009). Ranking paraphrases in context. In *Proceedings of the 2009 Workshop on Applied Textual Inference*, pages 44–47, Suntec, Singapore.

Thater, S., Fürstenau, H., and Pinkal, M. (2010). Contextualizing semantic representations using syntactically enriched vector models. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 948–957, Uppsala, Sweden.

Toutanova, K., Klein, D., Manning, C. D., and Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of HLT-NAACL 2003*, pages 252–259.

Toutanova, K. and Manning, C. D. (2000). Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000)*, pages 63–70.

Van de Cruys, T., Poibeau, T., and Korhonen, A. (2011). Latent vector weighting for word meaning in context. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1012–1022, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Zapirain, B., Agirre, E., and Màrquez, L. (2009). Generalizing over lexical features: selectional preferences for semantic role classification. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 73–76.