Unsupervised approaches to metonymy recognition

Yves Peirsman

University of Leuven
Quantitative Lexicology and Variational Linguistics
yves.peirsman@arts.kuleuven.be

Résumé

Jusqu'à présent, la question de la reconnaissance automatique de métonymies a souvent été abordée avec des approches supervisées. Toutefois, ces approches nécessitent l'annotation d'un nombre important d'occurrences d'apprentissage et, dès lors, elles empêchent le développement d'un système de reconnaissance de métonymies à grande échelle. Cet article étudie la possibilité de résoudre ce problème du goulot d'étranglement de l'acquisition des connaissances en recourant à des techniques d'apprentissages non supervisées. Bien que la technique en question, l'algorithme de Schütze (1998), soit souvent appliquée en désambiguïsation sémantique, je montrerai qu'elle s'avère trop peu solide pour le cas spécifique de la reconnaissance de métonymies. À cet effet, je propose d'étudier l'influence de quatre variables sur les performances de la technique non supervisée, à savoir le type de données, la taille de la fenêtre d'observation, l'application de la décomposition en valeurs singulières (SVD) et le type de sélection de propriétés.

Mots-clés : reconnaissance de métonymies, désambiguïsation sémantique, apprentissage par machine non supervisé.

Abstract

To this day, the automatic recognition of metonymies has generally been addressed with supervised approaches. However, these require the annotation of a large number of training instances and hence, hinder the development of a wide-scale metonymy recognition system. This paper investigates if this knowledge acquisition bottleneck in metonymy recognition can be resolved by the application of unsupervised learning. Although the investigated technique, Schütze's (1998) algorithm, enjoys considerable popularity in Word Sense Disambiguation, I will show that it is not yet robust enough to tackle the specific case of metonymy recognition. In particular, I will study the influence on its performance of four variables — the type of data set, the size of the context window, the application of SVD and the type of feature selection.

Keywords: metonymy recognition, word sense disambiguation, unsupervised machine learning.

1. Introduction

Metonymy is traditionally defined as a figure of speech that uses "one entity to refer to another that is related to it" (Lakoff and Johnson, 1980, p.35). In example (1), for instance, *West Germany* stands for the government of that country :

(1) This morning in Bonn, Dr Kohl will preside at an emergency cabinet meeting to discuss how *West Germany* should respond to the events of recent days. (BNC)

These days, metonymy is a hot topic in theoretical linguistics as well as in Natural Language Processing (NLP). Throughout the whole spectrum of linguistic disciplines, researchers seem to have realized that this figure of speech is ubiquitous in everyday language. As a result, a wide variety of applications in NLP need to be able to recognize and interpret metonymical words.

This automatic recognition and interpretation of metonymical words is generally referred to as the task of metonymy resolution (Fass, 1997). Originally, it was thought that metonymies could be recognized almost trivially, as words that violated selectional restrictions (Pustejovsky, 1995). *West Germany* above, for instance, violates the restriction that *respond* requires a human subject. Therefore any algorithm that has access to a database of selectional restrictions should have no problem identifying it as metonymical.

It was soon realized, however, that such algorithms would fail to deal appropriately with the freedom with which metonymies occur. For instance, there are metonymies that do respect the selectional restrictions of their head verb : *Nixon* in example (2) does not violate the selectional restrictions of the verb *to bomb*, and yet, it metonymically refers to the army under Nixon's command. In addition, some verbs have very weak selectional restrictions : in example (3) it is world knowledge rather than linguistic knowledge that makes us interpret *Shakespeare* in a metonymical way.

(2) *Nixon* bombed Hanoi.

(3) Daniel likes *Shakespeare*.

Examples like these demonstrate that metonymy recognition should not consist of the application of rigid rules. Rather, it should take into account statistical information about the semantic and grammatical context in which the target word occurs.

The existing statistical metonymy recognition algorithms are mainly supervised (see section 2). This means that they need a large number of training instances that are annotated with sense labels. This data is used to learn the correlation between the sense labels and the observed values of the semantic or grammatical features. On the basis of these relations, the classifiers then build a model for an entire semantic class of possibly metonymical words. Their need for a large number of annotated training data, however, hinders the development of a wide-scale metonymy recognition system, since each extension to a new class requires the labelling of a large training corpus.

This paper will therefore investigate if this so-called 'knowledge acquisition bottleneck' can be tackled by the application of Schütze's (1998) unsupervised word sense discrimination algorithm, an approach that does not need a single annotated example. In section 2, I will discuss the most important literature on corpus-based metonymy recognition, and introduce the notion of unsupervised learning. In section 3, I will highlight the research questions behind my experiments. Section 4 will present the experimental results, while section 5 wraps up with the conclusions of this study.

2. Previous work

2.1. Metonymy recognition

Markert and Nissim (2002a) were the first to notice that metonymy recognition can be seen as a special case of Word Sense Disambiguation (WSD). Since possibly metonymical words are polysemous, the recognition of a metonymy boils down to the automatic assignment of a sense label to a polysemous word. Neither of the differences between metonymy and WSD pose a threat to such an approach. First, although the number of possible metonymical readings of a word is in theory unlimited, a quick glance at Markert and Nissim's (2002b) corpus data shows that in practice, almost all observed metonymies belong to a small number of metonymical patterns that can be identified a priori. Second, while classic WSD builds a separate model for each polysemous word, metonymy recognition algorithms can be applied to an entire semantic class of words. After all, all words in the same class may undergo the same metonymical shifts.

Markert and Nissim focused their investigation on two such semantic classes: country names (Markert and Nissim, 2002b) and organization names (Nissim and Markert, 2005). For each of these classes, they extracted a set of occurrences from the British National Corpus (BNC). This paper will use their corpora of country names — one with about 1,000 instances of the name *Hungary* and one with 1,000 mixed country names.[1] In the annotation of these corpora, Markert and Nissim made a distinction between the metonymical patterns *place-for-people, place-for-event* and *place-for-product*. In addition, they used the label *mixed* for examples with more than one reading, and *othermet* for instances that did not belong to any of the pre-defined metonymical patterns.

Nissim and Markert's (2003; 2005) algorithms were evaluated on the basis of their accuracy — the number of instances that they classified correctly divided by the total number of instances — and the F-score on the metonymical class — the harmonic mean between precision and recall on that class. The best-performing algorithm for the mixed country data, for instance, obtained an accuracy of 87 % and an F-score of 62.7 %. The results for the *Hungary* data set were similar.

Promising though these results may be, the algorithms have their disadvantages. First, they are rather complex: Nissim and Markert's (2003) most successful algorithm involves smoothed probabilities and iterative searches through Dekang Lin's (1998) thesaurus of semantically similar words, for instance. Second, and most importantly, they rely on the annotation of a large number of training instances.

In an earlier paper, I addressed both these problems with a combination of example-based learning (which reduces complexity) and active learning (which reduces the number of labelled training examples) (Peirsman, forthc). My experiments showed that TiMBL, an example-based classifier, achieved similar performance to Nissim and Markert's (2003) algorithm, even without semantic information. Table 1 gives its results for the country and *Hungary* data, which were reached on the basis of grammatical information only (*i.e.*, the grammatical function(s) and head(s) of the target word). In addition, TiMBL benefited clearly from the application of active learning, a type of learning where the algorithms automatically choose the training examples that should be annotated. A few simple distance-based algorithms, which selected those training examples that were most representative of the data and most informative to the classifier, proved able to reduce the number of labelled training instances drastically.

In the present paper, I will again address the knowledge acquisition bottleneck in metonymy recognition. This time, however, I will take a more radical approach. I will test if the automatic recognition of metonymical words can proceed on the basis of an unsupervised WSD algorithm that does not need any annotated training examples.

[1] The data from this study is publicly available and can be downloaded from http://homepages.inf.ed.ac.uk/mnissim/mascara.

	Acc	P	R	F
mixed	86.59%	80.17%	49.47%	61.18%
Hungary	84.74%	80.39%	51.90%	63.08%

Table 1. TiMBL*'s results for the country and* Hungary *data*

word	Acc	baseline	word	Acc	baseline
capital	94%	64%	space	76%	56%
interest	93%	58%	suit	95%	57%
motion	87%	55%	tank	92%	90%
plant	70%	54%	train	74%	74%
ruling	91%	60%	vessel	98%	69%

Table 2. The results of Schütze's (1998) Word Sense Discrimination

2.2. Unsupervised learning

The unsupervised algorithm that I will use was first applied to the automatic discrimination of word senses by Schütze (1998). Following Miller and Charles' (1991) observation that humans rely on contextual similarity in order to determine semantic similarity, Schütze (1998) hypothesized that there must be a correlation between contextual similarity and word meaning as well: "a sense is a group of contextually similar occurrences of a word" (Schütze, 1998, p.99).

The resulting algorithm takes a number of steps:

1. First, all words in the training corpus are mapped onto so-called *word vectors*. These contain the frequencies of all observed co-occurrents of the word.
2. Then the algorithm focuses on the target word, as it builds a vector representation for each of its contexts. This is done by adding up the word vectors of all its co-occurrents in that context.
3. Next, the dimensionality of the vector space may be reduced with a technique such as Singular Value Decomposition (Golub and Van Loan, 1989).
4. As a final step in the learning stage, the context vectors are grouped into a pre-defined number of clusters. Each of these clusters is assumed to represent one of the senses of the target, and therefore their centroids are called *sense vectors*.
5. Finally, the classification of a test word proceeds by assigning it to the sense vector that lies nearest to its context vector.

Table 2 shows that, with about 8,000 training instances on average, this algorithm obtains promising results, compared to a baseline system that automatically assigns the most frequent sense label.

Schütze's (1998) approach is implemented in SenseClusters (Purandare and Pedersen, 2004), a software package that also incorporates some interesting variations on and extensions to the algorithm. Two of these will be used in this paper. First, like Purandare and Pedersen (2004), I will use bigrams instead of simple co-occurrences in order to construct the context vectors. Bigrams are "ordered pairs of words that co-occur within five positions of each other" (Purandare and Pedersen, 2004, p.2). Second, I will apply the clustering algorithm of Repeated Bisections, since this was found to perform particularly well with small data sets, like Markert and Nissim's (2002b). A final mention should go to the evaluation procedure. A word sense discrimination technique only outputs a number of clusters, and does not tell us which cluster corresponds to which meaning. Therefore the best match between clusters and meanings is the one which leads to the fewest misclassifications — the confusion matrix that maximizes the diagonal sum.

3. Research questions

The combination of metonymy recognition with unsupervised learning algorithms brings with it a number of questions, which the experiments in the next section are meant to answer.

- **Can unsupervised algorithms deal with a set of mixed target words as well as with a single word ?**
  Markert and Nissim's (2002b) corpora contain one set with the name *Hungary* only and one with a variety of country names. To my knowledge, the investigated unsupervised algorithm has not yet been applied to such a set of mixed target words. After all, a set of mixed target words will normally have more different co-occurrences, so that the resulting models might be less robust. By applying the same algorithms to both data sets, the experiments in the next section will help investigate if this hypothesis is correct.

- **Are smaller context windows better than large ones ?**
  In metonymy recognition, preference is generally given to smaller context windows. When using co-occurrence features, Markert and Nissim (2002a) observed that precision as well as recall increased considerably when the window around the target word was reduced from ten words on either side to about three. Unsupervised clustering, in contrast, generally proceeds with large context windows. Schütze's (1998) results above were reached with 25 words on either side of the target, for instance.
  Underlying this question is a crucial difference between metonymy recognition and the sense discrimination problems to which unsupervised clustering is normally applied. With its large context windows, Schütze's (1998) method is particularly well-suited for the discrimination of meanings that depend on the topic of a text. When *plant* occurs in a text about botany, for instance, it will usually refer to the biological organism ; when the text is about economy, pollution, or the like, *plant* will generally have its industrial sense. However, this observation does not hold for metonymies. In a text about politics, say, *Brussels* may have either its literal or a metonymical sense. As Markert and Nissim (2002a) argued, the reading of a possibly metonymical word depends more on words in its immediate vicinity than on the topic of the text. Therefore smaller context windows can be expected to lead to better results.

- **Does Singular Value Decomposition result in better performance ?**
  Schütze (1998) found that his algorithm clearly performs better with SVD than without. Still, there are reasons for investigating if this is also the case with metonymies. SVD is a technique that reduces the dimensionality of the vector space, and is therefore very useful in cases of data sparseness. Moreover, SVD helps tackle vocabulary issues such as synonymy (where two dimensions represent the same concept) and polysemy (where one dimension incorporates several concepts). It is said to abstract away from word dimensions, and to discover topical dimensions instead. However, as I observed above, text topics may not be very relevant to metonymy recognition. I therefore hypothesize that metonymy recognition will work better in word dimensions than in topical ones.

	+LL, +SVD		+LL, -SVD		-LL, +SVD		-LL, -SVD	
	Acc	*F*	*Acc*	*F*	*Acc*	*F*	*Acc*	*F*
15	55.73	37.74***	63.16	30.14	58.93	33.22*	59.86	31.63
12	56.45	37.39***	57.89	32.23	58.41	36.13**	57.69	31.67
10	58.31	37.07***	58.72	35.28**	56.45	34.06*	58.20	34.57*
7	55.01	37.89***	55.01	36.44**	56.35	37.70***	64.50	34.85***
5	55.01	26.85	55.62	23.49	63.78	36.30***	65.12	33.98**(*)

Table 3. Results of the unsupervised algorithm on the Hungary data

+LL : statistical feature selection
-LL : frequency-based feature selection
+SVD : dimensionality reduction with SVD
-SVD : no dimensionality reduction
* : indicates if the clusters differ significantly from the random baseline
* : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$

– **Should features be selected on the basis of a statistical test ?**
This question is of a less theoretical nature than the ones above. It simply asks whether features (*i.e.*, bigrams) should be selected on the basis of their frequency, or on the basis of their statistical relation to the target word. Schütze (1998, p.102) hypothesized that “candidate words whose occurrence depends on whether the ambiguous word occurs will be indicative of one of the senses of the ambiguous word and hence useful for disambiguation”. He also observed, however, that statistical selection is outperformed by frequency-based selection when SVD is not used. The experiments in the next section will investigate if this is also the case with metonymy recognition.

4. Experiments

I will try to answer the research questions above by an application of unsupervised word sense discrimination to Markert and Nissim’s (2002b) *Hungary* and mixed country data. Here are some specifics of the experimental setup:

- The number of pre-defined clusters was set to two.
- SVD, when used, reduced the number of dimensions to 300.
- Statistical feature selection, when used, selected only those bigrams with a log-likelihood score of 3.841 or more (following Purandare and Pedersen 2004).
- Initial experiments on the mixed country data indicated that a stoplist (a list of frequent words that are automatically excluded from feature selection) decreased performance, so it was left out.
- All metonymical labels were collapsed into one category, and *mixed* instances were ignored.
- All algorithms were evaluated with ten-fold cross-validation.

4.1. Hungary data

In the first round of experiments, I applied Purandare and Pedersen’s (2004) algorithm to Markert and Nissim’s (2002b) *Hungary* data, the set that I anticipated to be easier. 76.99% of the

	cluster 1	cluster 2
literal	518	228
metonymical	123	100

Table 4. Confusion matrix of the (-LL,+SVD*) algorithm with a context size of five words*

instances in this set are literal, which means that a majority baseline system that assigns all instances to the literal class will achieve an accuracy of 76.99%. From table 3, it is immediately clear that none of the investigated algorithms beats this baseline. Moreover, the results of the unsupervised approach are also much lower than those of its supervised competitors. Accuracy, for instance, lies more than 20 % below the accuracy reached by TiMBL.

However, these low accuracy values do not necessarily mean that the unsupervised clustering algorithm is completely insensitive to the meaning distinctions studied here. Consider the confusion matrix in table 4 as an example. This matrix represents the results of the (-LL,+SVD) algorithm with a context size of five words on either side of the target. Although the accuracy of this particular algorithm (63.77 %) does not beat the majority baseline, it is clear that the system is not entirely unsuccessful at distinguishing metonymical from literal meanings. The proportion of metonymical meanings in cluster 2 (30.49 %) is much higher than that in cluster 1 (19.19 %).

Therefore, we can compare the output of the algorithms with another baseline — one that divides the test instances among its two clusters randomly. Such a random baseline would have the same proportion of metonymies in both clusters. What we want to investigate is thus if the proportion of metonymies is higher in the smaller cluster, or, in other words, if there exists a correlation between the cluster of a test instance and its label. To this goal, we can make use of a t-test or a χ^2-test. In all but one case, these give the same broad level of significance for the results, as indicated in table 3.[2] In the example above, for instance, a Welch two-sample t-test indeed confirms that the proportion of metonymies in the second cluster is significantly different (in this case, higher) than that in the first cluster ($t = -3.7866, df = 576.065, p = 0.0001687$). The results of this test are therefore much more useful than the accuracy values for an evaluation of the different algorithms in the light of the experimental questions.

Since the first question (concerning mixed data sets) cannot yet be answered, I will start with the second one: should context windows be small or large ? The results of my experiments confirm Markert and Nissim's (2002a) claim that an approach based on co-occurrences (or bigrams) gives the best results with smaller context windows. Nevertheless, these context windows should not be too small, either, since they have to allow the algorithm to find sufficient useful features. In my experiments on the *Hungary* data, a window size of seven words on either side of the target led to the best results.

The next question concerned the use of Singular Value Decomposition. I hypothesized that metonymy recognition works best in word dimensions, and that SVD can therefore be skipped. This hypothesis is contradicted by the experimental results. In fact, in nine out of ten cases, the algorithms without SVD gave a lower F-score than those with SVD. Apparently, the disadvantage of trading word dimensions for more topical dimensions is outweighed by the advantage of SVD's ability to deal with data sparseness.

[2] For the (-LL,-SVD) algorithm, p < 0.001 according to the χ^2-test and p < 0.01 according to the t-test. The same is true for the mixed country data below.

	+LL, +SVD		+LL, -SVD		-LL, +SVD		-LL, -SVD	
	Acc	F	Acc	F	Acc	F	Acc	F
15	57.47	29.76	58.35	24.35	58.79	24.55	60.00	23.21
12	59.23	33.63***	58.68	33.57***	64.95	24.23	63.74	21.43
10	55.71	32.04*	59.34	38.33***	62.31	20.79	66.70	18.77
7	58.90	35.74***	62.75	35.43***	61.32	25.42	67.69	26.13
5	60.55	32.90***	67.14	36.25***	62.31	31.26**	68.46	30.51**(*)

Table 5. Results of the unsupervised algorithm on the mixed country data.

Finally, I wanted to investigate if features should be selected on the basis of a statistical test or not. If we look at the results of the algorithms with SVD, the usefulness of statistical selection seems to depend on context size. Large contexts, and the resulting high number of possible features, clearly benefit from statistical feature selection. However, when the context is small and does not contain too many possibly features, the log-likelihood test is best dropped in favour of frequency-based selection.

In summary, on the basis of this first round of experiments, we have already been able to answer most of the experimental questions above. Although it was shown that unsupervised algorithms do not make for robust metonymy recognition systems, they are often able to identify two clusters that correlate with the literal and metonymical senses of the target word. In general, smaller context sizes proved more successful in this respect than larger ones, and systems with SVD scored better than those without. Statistical feature selection, finally, is best used in combination with large context windows.

4.2. Mixed country data

In the second round of experiments, I applied the algorithms to Markert and Nissim's (2002b) mixed country data. Although the majority baseline for this mixed data set lies even higher than that above, at an accuracy of 80.99 %, neither Nissim and Markert's (2003) classifier nor TiMBL (Peirsman, forthc) had any problem dealing with it. However, I anticipated that the results of the unsupervised algorithms would be lower than those on the *Hungary* data set, because a mixed set of target words brings with it a larger number of different co-occurrences or bigrams.

The actual results in table 5 do not fully bear out this expectation. There may be fewer F-scores that beat the random baseline than with the *Hungary* data, but the results are surprisingly consistent. As a rule of thumb, any algorithm with statistical feature selection and a window size smaller than fifteen words is able to identify two clusters that correlate significantly with the two senses of the target words.

With respect to the second research question, it is again smaller context windows that are most successful. This time five words seems to be the ideal window size overall, but when statistical feature selection is applied, the algorithms can deal with larger context sizes as well. This observation, then, confirms the above conclusion that larger contexts are best used in combination with statistical feature selection. With frequency-based feature selection, in contrast, the F-scores on this mixed data set are dramatically lower than those on the single-word set above. This can be explained by the larger number of bigrams in the mixed data set: statistical selection can help the algorithm distinguish informative features from uninformative ones.

The only conclusion from the previous section that is not corroborated by the present results is the usefulness of SVD. In fact, the algorithms seem rather insensitive to the presence or absence of the dimensionality reduction stage. As long as statistical feature selection is performed and the context size is not too large, the algorithms are guaranteed to beat the random baseline.

In short, even though the results of the unsupervised approach again lie much lower than the majority baseline, they are on a par with those on the *Hungary* data. This is promising, since it indicates that mixed data sets are not necessarily more difficult for unsupervised clustering than single-word sets. In order to deal with the larger number of possible features, however, the features should be selected on the basis of a statistical test and the context should not be too large.

5. Conclusions

This paper has investigated if the knowledge acquisition bottleneck in metonymy recognition can be solved by an application of Schütze's (1998) and Purandare and Pedersen's (2004) unsupervised WSD algorithm. The experimental results proved that this was not yet the case. None of the investigated variations on the original algorithm was able to beat the accuracy of a baseline algorithm that classifies all instances as literal. Nevertheless, the clusters that were identified did often correlate with the literal and metonymical senses of the target word. In 32 out of 40 experiments (80 %), the smaller cluster indeed contained a larger proportion of metonymical examples. In 24 cases (60 %), this difference was significant.

The algorithm was able to output such clusters that significantly correlated with the target readings for both the single-word data set and the mixed data set. A few variables influenced its rate of success. First of all, small contexts were generally more successful than larger ones. Large contexts contain a wide variety of possible features, and are best used in combination with statistical feature selection, the second variable. For the mixed data set, this statistical feature selection seems almost a necessary condition to reach good results. Finally, the role of SVD was less easy to determine. For the single-word set I concluded that the classifier worked best in word dimensions, but it appeared insensitive to the type of dimensions (word or topic) with the mixed data set.

The failure of the investigated unsupervised learning algorithm to beat the majority baseline can be attributed to a few factors. First, unsupervised machine learning is generally outperformed by supervised algorithms, so the lower results should come as no surprise. Second, the studied algorithm takes a bag-of-words approach to the data, and is therefore blind to all structural or syntactic information. This syntactic information, however, is extremely important in metonymy recognition, where the interpretation of a target word often depends on its head (Markert and Nissim, 2002a). Without this information, any metonymy recognition algorithm is necessarily handicapped.

From these findings, it is clear what directions future research should take. First, it is important to determine if the classifier benefits from a larger training corpus. Even though the average number of training instances used here is already larger than that in Purandare and Pedersen (2004), using more data from the BNC will probably make the context and sense vectors more robust. Yet, as I pointed out above, it is unlikely that the resulting classifiers will already be able to compete with state-of-the-art metonymy recognition systems. Before this is conceivable, research should focus on how syntactic information can be taken up into the feature vectors.

Acknowledgements

I would like to thank Mirella Lapata, Dirk Geeraerts, Dirk Speelman and Gert De Sutter for their feedback on this project. I am also very grateful to Katja Markert and Malvina Nissim for their helpful information about their research.

References

FASS D. (1997). *Processing Metaphor and Metonymy*. Stanford, CA: Ablex.

GOLUB G. H. and VAN LOAN C. F. (1989). *Matrix Computations*. London: The Johns Hopkins University Press.

LAKOFF G. and JOHNSON M. (1980). *Metaphors We Live By*. London: The University of Chicago Press.

LIN D. (1998). ”An information-theoretic definition of similarity”. In *Proceedings of the International Conference on Machine Learning*. Madison, USA.

MARKERT K. and NISSIM M. (2002a). ”Metonymy Resolution as a Classification Task”. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*. Philadelphia, USA.

MARKERT K. and NISSIM M. (2002b). ”Towards a Corpus Annotated for Metonymies: the Case of Location Names”. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*. Las Palmas, Spain.

MILLER G. A. and CHARLES W. G. (1991). ”Contextual correlates of semantic similarity”. In *Language and Cognitive Processes*, 6 (1), 1-28.

NISSIM M. and MARKERT K. (2003). ”Syntactic Features and Word Similarity for Supervised Metonymy Resolution”. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL-03)*. Sapporo, Japan.

NISSIM M. and MARKERT K. (2005). ”Learning to buy a Renault and talk to BMW: A supervised approach to conventional metonymy”. In H. Bunt(ed.), *Proceedings of the 6th International Workshop on Computational Semantics*. Tilburg, The Netherlands.

PEIRSMAN Y. (forthc.). ”Example-based metonymy recognition for proper nouns”. In *Proceedings of the Student Session of the Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006)*. Trento, Italy.

PURANDARE A. and PEDERSEN T. (2004). ”Word Sense Discrimination by Clustering Contexts in Vector and Similarity Spaces”. In *Proceedings of the Conference on Computational Natural Language Learning*. Boston, USA.

PUSTEJOVSKY J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

SCHÜTZE H. (1998). ”Automatic Word Sense Discrimination”. In *Computational Linguistics*, 24 (1), 97-124.