

Combining Word Patterns and Discourse Markers for Paradigmatic Relation Classification

Michael Roth

ILCC, School of Informatics
University of Edinburgh
mroth@inf.ed.ac.uk

Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
schulte@ims.uni-stuttgart.de

Abstract

Distinguishing between paradigmatic relations such as synonymy, antonymy and hypernymy is an important prerequisite in a range of NLP applications. In this paper, we explore discourse relations as an alternative set of features to lexico-syntactic patterns. We demonstrate that statistics over discourse relations, collected via explicit discourse markers as proxies, can be utilized as salient indicators for paradigmatic relations in multiple languages, outperforming patterns in terms of recall and F₁-score. In addition, we observe that markers and patterns provide complementary information, leading to significant classification improvements when applied in combination.

1 Introduction

Paradigmatic relations (such as synonymy, antonymy and hypernymy; cf. Murphy, 2003) are notoriously difficult to distinguish automatically, as first-order co-occurrences of the related words tend to be very similar across the relations. For example, in *The boy/girl/person loves/hates the cat*, the nominal co-hyponyms *boy*, *girl* and their hypernym *person* as well as the verbal antonyms *love* and *hate* occur in identical contexts, respectively. Vector space models, which represent words by frequencies of co-occurring words to enable comparisons in terms of *distributional similarity* (Schütze, 1992; Turney and Pantel, 2010), hence perform below their potential when inferring the type of relation that holds between two words. This distinction is crucial, however, in a range of tasks: in sentiment analysis, for example, words of the same and opposing polarity need to be distinguished; in textual entailment, systems further need to identify hypernymy because of directional inference requirements.

Accordingly, while there is a rich tradition on identifying word pairs of a single paradigmatic relation, there is little work that has addressed the distinction between two or more paradigmatic relations (cf. Section 2 for details). In more general terms, previous approaches to distinguishing between several semantic relations have predominantly relied on manually created knowledge sources, or lexico-syntactic patterns that can be automatically extracted from text. Each option comes with its own shortcomings: knowledge bases, on the one hand, are typically developed for a single language or domain, meaning that they might not generalize well; word patterns, on the other hand, are noisy and can be sparse for infrequent word pairs.

In this paper, we propose to strike a balance between availability and restrictedness by making use of *discourse markers*. This approach has several advantages: markers are frequently found across genres (Webber, 2009), they exist in many languages (Jucker and Yiv, 1998), and capture various semantic properties (Hutchinson, 2004). We implement discourse markers within a vector space model that aims to distinguish between the three paradigmatic relations *synonymy*, *antonymy* and *hypernymy* in German and in English, across the three word classes of nouns, verbs, adjectives. We examine the performance of discourse markers as vector space dimensions in isolation and also explore their contribution in combination with lexical patterns.

2 Related Work

As mentioned above, there is a rich tradition of research on identifying a single paradigmatic relations. Work on *synonyms* includes Edmonds and Hirst (2002), who employed a co-occurrence network and second-order co-occurrence, and Curran (2003), who explored word-based and syntax-based co-occurrence for thesaurus construction.

Van der Plas and Tiedemann (2006) compared a standard distributional approach against cross-lingual alignment; Erk and Padó (2008) defined a vector space model to identify synonyms and the substitutability of verbs. Most computational work on *hypernyms* was performed for nouns, cf. the lexico-syntactic patterns by Hearst (1992) and an extension of the patterns by dependency paths (Snow et al., 2004). Weeds et al. (2004), Lenci and Benotto (2012) and Santus et al. (2014) identified hypernyms in distributional spaces. Computational work on *antonyms* includes approaches that tested the co-occurrence hypothesis (Charles and Miller, 1989; Fellbaum, 1995), and approaches driven by text understanding efforts and contradiction frameworks (Harabagiu et al., 2006; Mohammad et al., 2008; de Marneffe et al., 2008).

Among the few approaches that distinguished *between* paradigmatic semantic relations, Lin et al. (2003) used patterns and bilingual dictionaries to retrieve distributionally similar words, and relied on clear antonym patterns such as ‘either X or Y’ in a post-processing step to distinguish synonyms from antonyms. The study by Mohammad et al. (2013) on the identification and ranking of opposites also included synonym/antonym distinction. Yih et al. (2012) developed an LSA approach incorporating a thesaurus, to distinguish the same two relations. Chang et al. (2013) extended this approach to induce vector representations that can capture multiple relations. Whereas the above mentioned approaches rely on additional knowledge sources, Turney (2006) developed a corpus-based approach to model relational similarity, addressing (among other tasks) the distinction between synonyms and antonyms. More recently, Schulte im Walde and Köper (2013) proposed to distinguish between the three relations antonymy, synonymy and hyponymy based on automatically acquired word patterns.

Regarding pattern-based approaches to identify and distinguish lexical semantic relations in more general terms, Hearst (1992) was the first to propose lexico-syntactic patterns as empirical pointers towards relation instances, focusing on hyponymy. Girju et al. (2003) applied a single pattern to distinguish pairs of nouns that are in a causal relationship from those that are not, and Girju et al. (2006) extended the work towards part-whole relations, applying a supervised, knowledge-intensive approach. Chklovski and Pantel (2004) were the first to apply pattern-

based relation extraction to verbs, distinguishing five non-disjoint relations (*similarity, strength, antonymy, enablement, happens-before*). Pantel and Pennacchiotti (2006) developed *Espresso*, a weakly-supervised system that exploits patterns in large-scale web data to distinguish between five noun-noun relations (*hypernymy, meronymy, succession, reaction, production*). Similarly to Girju et al. (2006), they used generic patterns, but relied on a bootstrapping cycle combined with reliability measures, rather than manual resources. Whereas each of the aforementioned approaches considers only one word class and clearly disjoint categories, we distinguish between paradigmatic relations that can be distributionally very similar and propose a unified framework for nouns, verbs and adjectives.

3 Baseline Model and Data Set

The task addressed in this work is to distinguish between synonymy, antonymy and hypernymy. As a starting point, we build on the approach and data set used by Schulte im Walde and Köper (2013, henceforth just S&K). In their work, frequency statistics over automatically acquired co-occurrence patterns were found to be good indicators for the paradigmatic relation that holds between two given words of the same word class. They further experimented with refinements of the vector space model, for example, by only considering patterns of a specific length, weighting by pointwise mutual information and applying thresholds based on frequency and reliability.

Baseline Model. We re-implemented the best model from S&K with the same setup: word pairs are represented by vectors, with each entry corresponding to one out of almost 100,000 patterns of lemmatized word forms (e.g., *X affect how you Y*). Each value is calculated as the log frequency of the corresponding pattern occurring between the word pairs in a corpus, based on exact match. For English, we use the ukWaC corpus (Baroni et al., 2009); for German, we rely on the COW corpus instead of deWaC, as it is larger and better balanced (Schäfer and Bildhauer, 2012).

Data Set. The evaluation data set by S&K is a collection of target and response words in German that has been collected via Amazon Mechanical Turk. The data contains a balanced amount of instances across word categories and relations, also taking into account corpus frequency, degree of ambiguity and semantic classes. In total, the

	S&K			Reimplemented		
	P	R	F ₁	P	R	F ₁
Nouns						
SYN-ANT	77.4	65.0	70.7	76.7	62.2	68.7
SYN-HYP	75.0	57.0	64.8	73.3	59.5	65.7
Verbs						
SYN-ANT	70.6	40.0	51.1	84.6	36.7	51.2
SYN-HYP	42.0	26.7	32.6	52.6	33.3	40.8
Adjectives						
SYN-ANT	88.9	66.7	76.2	94.1	66.7	78.0
SYN-HYP	68.4	54.2	60.5	65.0	54.2	59.1

Table 1: 2-way classification results by Schulte im Walde and Köper (2013) and our re-implementation. All numbers in percent.

data set consists of 692 pairs of instances, distributed over three word classes (nouns, verbs, adjectives) and three paradigmatic relations (synonymy, antonymy, hypernymy).

Intermediate Evaluation. We compare our re-implementation to the model by S&K using their 80% training and 20% test split, focusing on 2-way classifications involving synonymy. The results, summarized in Table 1, confirm that our re-implementation achieves similar results. Observed differences are probably an effect of the distinct corpora applied to induce patterns and counts.

We notice that the performance of both models strongly depends on the affected pair of relations and word category. For example, precision varies in the 2-way classification between synonymy and antonymy from 70.6% to 94.1%. Given the small amount of test data, some of the 80/20 splits might be better suited for the model than others. To avoid resulting bias effects, we perform our final evaluation using 5-fold cross-validation on a merged set of all training and test instances. To illustrate the performance of models in multiple languages, we further conduct experiments on a data set for English relation pairs that has been collected by Giulia Benotto and Alessandro Lenci, following the same methodology as the German collection. The English data set consists of 648 pairs of instances, also distributed over nouns, verbs, adjectives, and covering synonymy, antonymy, hypernymy.

4 Markers for Relation Classification

The aim of this work is to establish corpus statistics over discourse relations as a salient source of

CONTRAST	but, although, rather ...
RESTATEMENT	indeed, specifically, ...
INSTANTIATION	(for) example, instance, ...

Table 2: Examples of discourse relations/markers.

information to distinguish between paradigmatic relations. Our approach is motivated by linguistic studies that indicated a connection between discourse relations and lexical relations of words occurring in the respective discourse segments: Murphy et al. (2009) have shown, for example, that antonyms frequently serve as indicators for contrast relations in English and Swedish. More generally, pairs of word tokens have been identified as strong features for classifying discourse relations when no explicit discourse markers are available (Pitler et al., 2009; Biran and McKeown, 2013).

Whereas word pairs have frequently been used as features for disambiguating discourse relations, to the best of our knowledge, our approach is novel in that we are the first to apply discourse relations as features for classifying lexical relations. One reason for this might be that discourse relations in general are only available in manually annotated corpora. Previous work has shown, however, that such relations can be classified reliably given the presence of explicit discourse markers.¹ We hence rely on such markers as proxies for discourse relations (for examples, cf. Table 2).

4.1 Model and Hypothesis

We propose a vector space model that represents pairs of words using as features the discourse markers that occur between them. The underlying hypothesis of this model is as follows: if two phrases frequently co-occur with a specific discourse marker, then the discourse relation expressed by the corresponding marker should also indicate the relation between the words in the affected phrases. Following this hypothesis, contrast relations might indicate antonymy, whereas elaborations may indicate synonymy or hyponymy. Although such relations will not hold between every pair of words in two connected discourse segments, we hypothesize that correct instances (of all considered word classes) can be identified based on high relative frequency.

In our model, frequency statistics are computed over sentence-internal co-occurrences of

¹Pitler et al. (2008) report an accuracy of up to 93%.

word pairs and discourse markers. Since discourse relations are typically directed, we take into consideration whether a word occurs to the left or to the right of the respective marker. Accordingly, the features of our model are special cases of single-word patterns with an arbitrary number of wild card tokens (e.g., the marker feature ‘though’ corresponds to the pattern “ $X * \textit{though} * Y$ ”). Yet, our specific choice of features has several advantages: Whereas strict and potentially long patterns can be rare in text, discourse markers such as “however”, “for example” and “additionally” are frequently found across genres (Webber, 2009). Although combinations of tokens could also be replaced by wild cards in any automatically acquired pattern, this would generally lead to an exponentially growing feature space. In contrast, the set of discourse markers in our work is fixed: for English, we use 61 markers annotated in the Penn Discourse TreeBank 2.0 (Prasad et al., 2008); for German, we use 155 one-word translations of the English markers, as obtained from an online dictionary.^{2,3} Taking directionality into account, our vector space model consists of 2×61 and 2×155 features, respectively.

4.2 Development Set and Hyperparameters

We select the hyperparameters of our model using an independent development set, which we extract from the lexical resource GermaNet (Hamp and Feldweg, 1997). For each considered word category, we extract instances of synonymy, antonymy and hypernymy. In total, 1502 instances are identified, with 64 of them overlapping with the evaluation data set described in Section 3. Note though that the development set is not used for evaluation but only to select the following hyperparameters.

We experimented with different vector values (absolute frequency, log frequency, pointwise mutual information (PMI)), distance measures (cosine, euclidean) and normalization schemes. In contrast to S&K, who did not observe any improvements using PMI, we found it to perform best, combined with euclidean distance and no additional normalization. This finding might be an immediate effect of discourse markers being

²<http://dict.leo.org>

³We also experimented with larger sets of markers, including conjunctions and adverbials in sentence-initial positions, but did not notice any considerable effect. Future work could use manual sets of markers, e.g. those by Pasch et al. (2003), though such sets are only available in few languages.

generally more frequent than strict word patterns, which also leads to more reliable PMI values.

5 Evaluation

In our evaluation, we assess the performance of the marker-based model and demonstrate the benefits of incorporating discourse markers into a pattern-based model, which we apply as a baseline. We evaluate on several data sets: the collection of target-response pairs in German from previous work, and a similar data set that was collected for English target words (cf. Section 3); for comparison reasons, we also apply our models to the balanced data set of related and unrelated noun pairs by Yap and Baldwin (2009).⁴ We perform 3-way and 2-way relation classification experiments, using 5-fold cross-validation and a nearest centroid classifier (as applied by S&K).

Results. The 3-way classification results of the baseline and our marker-based model are summarized in Table 3, with best results for each setting marked in bold. On the German data set, our model always outperforms a random baseline (33% F_1 -score). The results on the English data set are overall a bit lower, possibly due to corpus size. In almost all classification tasks, our marker-based model achieves a higher recall and F_1 -score than the pattern-based approach. The precision results of the marker-based model are overall below the pattern-based model. This drop in performance does not come as a surprise though, considering that the model only makes use of 122 and 310 features, in comparison to tens of thousands of features in the pattern approach.

A randomized significance test over classified instances (cf. Yeh, 2000) revealed that only two differences in results are significant. We hypothesize that one reason for this outcome might be that both models cover complementary sets of instances. To verify this hypothesis, we apply a combined model, which is based on a weighted linear combination of distances computed by the two individual models.⁵ As displayed in Table 3, this combined model yields further improvements

⁴Note that we could, in principle, also apply our models to unbalanced data. Our main focus lies however on examining the direct impact of different feature sets. We hence decided to keep the evaluation setup simple and used a classifier that does not take into account class frequency.

⁵We determined the best weights on the development set and found these to be 0.9 and 0.1 for the output of the pattern-based and marker-based model, respectively.

		Nouns			Verbs			Adjectives		
		P	R	F ₁	P	R	F ₁	P	R	F ₁
German	Patterns	55.6	40.8	47.0	55.6	35.6	43.4	53.5	41.1	46.5
	Markers	42.6	38.7	40.5	48.4	46.2**	47.3	51.1	48.6	49.9
	Combined	50.4	45.7*	48.0	52.6	50.2**	51.4**	53.4	50.8**	52.1
English	Patterns	46.4	28.0	34.9	44.7	28.5	34.8	56.6	32.1	41.0
	Markers	39.0	34.3	36.5	38.3	36.3	37.2	50.0	41.2**	45.2
	Combined	43.0	37.8**	40.3*	41.8	39.6**	40.7*	53.5	44.4**	48.5**

Table 3: 3-way classification results using 5-fold cross-validation. All numbers in percent. Asterisks indicate significant differences to the pattern-based baseline model (* p<0.10, ** p<0.05).

Combined model	German			English		
	P	R	F ₁	P	R	F ₁
Nouns						
SYN-ANT	61.7	55.7	58.5	52.9	44.2	48.2
SYN-HYP	66.5	60.4	63.3	62.2	58.6	60.4
ANT-HYP	70.9	64.6	67.6	59.1	50.6	54.5
Verbs						
SYN-ANT	58.9	55.0	56.8	49.6	45.8	47.6
SYN-HYP	67.6	64.0	65.8	66.4	63.0	64.7
ANT-HYP	67.3	66.4	66.9	62.9	60.7	61.8
Adjectives						
SYN-ANT	74.8	69.4	72.0	67.0	56.6	61.3
SYN-HYP	58.0	56.1	57.0	56.4	46.0	50.7
ANT-HYP	73.7	70.7	72.2	69.8	57.8	63.2

Table 4: 2-way results of the combined model. Bold numbers indicate improvements over both individual models. All numbers in percent.

in recall and F₁-score, leading to the best 3-way classification results. All gains in recall are significant, confirming that the single models indeed contribute complementary information. For example, only the pattern-based model classifies “intentional”–“accidental” as antonyms, and only the marker-based model predicts the correct relation for “double”–“multiple” (hypernymy). The combined model classifies both pairs correctly.

Table 4 further assesses the strength of the combined model on the 2-way classifications. The table highlights results indicating improvements over *both* individual models. We observe that the combined model achieves the best recall and F₁-score in 15 out of 18 cases.

Relation	SYN	ANT	HYP
Patterns	0.97	0.97	0.94
Markers	0.77*	0.82*	0.91*
Combined	0.93*	0.98	0.96*

Table 5: Results in F₁-score on the balanced data set by Yap and Baldwin (* p<0.05).

A final experiment is performed on the data set by Yap and Baldwin (2009) to see whether our models can also distinguish word pairs of individual relations from unrelated pairs of words. The results, listed in Table 5, show that the marker-based model cannot perform this task as well as the pattern-based model. The combined model, however, outperforms both individual models in 2 out of 3 cases. Despite their simplicity, our models achieve results close to the F₁-scores reported by Yap and Baldwin (0.98–0.99), who employed syntactic pre-processing and an SVM-based classifier, and experimented with different corpora.

6 Conclusions

In this paper, we proposed to use discourse markers as indicators for paradigmatic relations between words and demonstrated that a small set of such markers can achieve higher recall than a pattern-based model with tens of thousands of features. Combining patterns and markers can further improve results, leading to significant gains in recall and F₁. As our new model only relies on a raw corpus and a fixed list of discourse markers, it can easily be extended to other languages.

Acknowledgments

The research presented in this paper was funded by the DFG grant SCHU-2580/2-1 and the DFG Heiselberg Fellowship SCHU-2580/1-1.

References

- Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3):209–226.
- Or Biran and Kathleen McKeown. 2013. Aggregated word pair features for implicit discourse relation disambiguation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 69–73, Sofia, Bulgaria, August.
- Kai-Wei Chang, Wen-tau Yih, and Christopher Meek. 2013. Multi-relational latent semantic analysis. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1602–1612, Seattle, Washington, USA, October.
- Walter G. Charles and George A. Miller. 1989. Contexts of antonymous adjectives. *Applied Psycholinguistics*, 10(3):357–375.
- Tim Chklovski and Patrick Pantel. 2004. VerbOcean: Mining the Web for fine-grained semantic verb relations. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain, 25–26 July 2004, pages 33–40.
- James Curran. 2003. *From Distributional to Semantic Similarity*. Ph.D. thesis, Institute for Communication and Collaborative Systems, School of Informatics, University of Edinburgh.
- Marie-Catherine de Marneffe, Anna N. Rafferty, and Christopher D. Manning. 2008. Finding contradictions in text. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1039–1047, Columbus, Ohio, USA.
- Philip Edmonds and Graeme Hirst. 2002. Near-synonymy and lexical choice. *Computational Linguistics*, 28(2):105–144.
- Katrin Erk and Sebastian Padó. 2008. A structured vector space model for word meaning in context. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Waikiki, Honolulu, Hawaii, 25-27 October 2008.
- Christiane Fellbaum. 1995. Co-occurrence and antonymy. *International Journal of Lexicography*, 8(4):281–303.
- Roxana Girju, Adriana Badulescu, and Dan Moldovan. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, Edmonton, Alberta, Canada, 27 May –1 June 2003, pages 80–87.
- Roxana Girju, Adriana Badulescu, and Dan Moldovan. 2006. Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1):83–135.
- Birgit Hamp and Helmut Feldweg. 1997. GermaNet - a lexical-semantic net for German. In *Proceedings of the Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications at ACL/EACL-97*, Madrid, Spain, 12 July 1997, pages 9–15.
- Sanda Harabagiu, Andrew Hickl, and Finley Lacatusu. 2006. Negation, contrast and contradiction in text processing. In *Proceedings of the 21st National Conference on Artificial Intelligence*, pages 755–762.
- Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 15th International Conference on Computational Linguistics*, Nantes, France, 23-28 August 1992, pages 539–545.
- Ben Hutchinson. 2004. Acquiring the meaning of discourse markers. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain, 21–26 July 2004, pages 685–692.
- Andreas H. Jucker and Zael Yiv, editors. 1998. *Discourse Markers: Descriptions and Theory*, volume 57 of *Discourse & Beyond New Series*. John Benjamin Publishing Company.
- Alessandro Lenci and Giulia Benotto. 2012. Identifying hypernyms in distributional semantic spaces. In *Proceedings of the First Joint Conference on Lexical and Computational Semantic*, pages 75–79.
- Dekang Lin, Shaojun Zhao, Lijuan Qin, and Ming Zhou. 2003. Identifying synonyms among distributionally similar words. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pages 1492–1493. Morgan Kaufmann Publishers Inc.
- Saif M. Mohammad, Bonnie Dorr, and Graeme Hirst. 2008. Computing word-pair antonymy. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 982–991, Honolulu, Hawaii, USA.
- Saif M. Mohammad, Bonnie J. Dorr, Graeme Hirst, and Peter D. Turney. 2013. Computing lexical contrast. *Computational Linguistics*, 39(3):555–590.
- M. Lynne Murphy, Carita Paradis, Caroline Willners, and Steven Jones. 2009. Discourse functions of antonymy: A cross-linguistic investigation of Swedish and English. *Journal of Pragmatics*, 41(11):2159–2184.
- M. Lynne Murphy. 2003. *Semantic relations and the lexicon*. Cambridge University Press.

- Patrick Pantel and Marco Pennacchiotti. 2006. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, Sydney, Australia, 17–21 July 2006, pages 113–120.
- Renate Pasch, Ursula Brausse, Eva Breindl, and Ulrich Wassner. 2003. *Handbuch der deutschen Konnektoren*. Walter de Gruyter, Berlin.
- Emily Pitler and Ani Nenkova. 2008. Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 186–195, Honolulu, Hawaii, October.
- Emily Pitler, Annie Louis, and Ani Nenkova. 2009. Automatic sense prediction for implicit discourse relations in text. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 683–691, Suntec, Singapore, August.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltakaki, Livio Robaldo, Aravind K Joshi, and Bonnie L Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC-2008)*, Marrakesh, Morocco, May.
- Enrico Santus, Alessandro Lenci, Qin Lu, and Sabine Schulte im Walde. 2014. Chasing hypernyms in vector spaces with entropy. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 38–42, Gothenburg, Sweden.
- Roland Schäfer and Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pages 486–493, Istanbul, Turkey, May.
- Sabine Schulte im Walde and Maximilian Köper. 2013. Pattern-based distinction of paradigmatic relations for German nouns, verbs, adjectives. In *Language Processing and Knowledge in the Web*, pages 184–198. Springer.
- Hinrich Schütze. 1992. Dimensions of meaning. In *In Proceedings of Supercomputing*, pages 787–796.
- Rion Snow, Daniel Jurafsky, and Andrew Y Ng. 2004. Learning syntactic patterns for automatic hypernym discovery. In *Advances in Neural Information Processing Systems*, volume 17, pages 1297–1304.
- Peter D. Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37(1):141–188.
- Peter D. Turney. 2006. Similarity of semantic relations. *Computational Linguistics*, 32(3):379–416.
- Lonneke Van der Plas and Jörg Tiedemann. 2006. Finding synonyms using automatic word alignment and measures of distributional similarity. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 866–873.
- Bonnie Webber. 2009. Genre distinctions for discourse in the Penn TreeBank. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 674–682, Suntec, Singapore, August.
- Julie Weeds, David Weir, and Diana McCarthy. 2004. Characterising measures of lexical distributional similarity. In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 1015–1021.
- Willy Yap and Timothy Baldwin. 2009. Experiments on pattern-based relation learning. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pages 1657–1660.
- Alexander Yeh. 2000. More accurate tests for the statistical significance of result differences. In *Proceedings of the 18th International Conference on Computational Linguistics*, pages 947–953, Saarbrücken, Germany, August.
- Wen-tau Yih, Geoffrey Zweig, and John Platt. 2012. Polarity inducing latent semantic analysis. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1212–1222, Jeju Island, Korea, July.