Disambiguating Noun Compounds

Su Nam Kim and Timothy Baldwin
CSSE
University of Melbourne
Australia
{snkim,tim}@csse.unimelb.edu.au

Abstract

This paper is concerned with the interaction between word sense disambiguation and the interpretation of noun compounds (NCs) in English. We develop techniques for disambiguating word sense specifically in NCs, and then investigate whether word sense information can aid in the semantic relation interpretation of NCs. To disambiguate word sense, we combine the one sense per collocation heuristic with the grammatical role of polysemous nouns and analysis of word sense combinatorics. We built supervised and unsupervised classifiers for the task and demonstrate that the supervised methods are superior to a number of baselines and also a benchmark state-of-the-art WSD system. Finally, we show that WSD can significantly improve the accuracy of NC interpretation.

Introduction

Word sense disambiguation (WSD) is the task of assigning predefined sense labels to word instances. It is one of the most challenging tasks in natural language processing (NLP), to the extent that it has been described as an *AI-complete* problem, i.e. a problem which requires human-level machine intelligence to solve. WSD is conventionally conceptualised as an intermediate task in NLP, although there is some controversy as to what applications it has utility in given current performance levels (Agirre & Edmonds 2006). Sanderson (1996) famously found that unreasonable levels of accuracy were required for WSD to enhance information retrieval performance, whereas Vickrey *et al.* (2005) and Carpuat & Wu (2005) got mixed results for WSD in the context of machine translation.

Ide & Veronis (1998) categorise WSD research into two primary categories: corpus-based and knowledge-based. Corpus-based methods use features based on neighbouring (content) words in a fixed word window of the target word, while knowledge-based methods extract features from lexical resources such as dictionaries. As examples of corpus-based WSD, Yarowsky (1995) used a bootstrap approach to disambiguate word senses, while McCarthy *et al.* (2004) proposed a method for learning the first sense of a word based on grammatical context, content words in a word window, and ontological semantics. As examples of knowledge-based WSD, Leacock & Chodorow (1998) acquired examples automatically to generate features based on identification of monosemous hypernyms and hyponyms of the target word, and Banerjee & Pedersen (2003) showed the usefulness of hypernyms in definition overlap-based WSD.

In this paper, we are interested in the applications of WSD to noun compounds (NCs) such as *apple pie* or *family car*;[1] for a given NC, we will refer to the first element (e.g. *apple*) as the modifier and the second element (e.g. *pie*) as the head noun. Specifically, we are interested in two tasks: (1) WSD of the individual elements of NCs, e.g. *apple* and *pie* in *apple pie*, and (2) the utility of WSD in interpreting the semantic relations in noun compounds, e.g. interpreting *apple pie* as corresponding to the MATERIAL semantic relation due to apple pies being made from apples. This second task is of interest to both (a) the WSD community in that we attempt to showcase a successful application of WSD to a real-world NLP task, and (b) the lexical semantics community in that NCs continue to be the target of active research on interpretation (Levi 1979; Moldovan *et al.* 2004; Kim & Baldwin 2006).

Our motivation in carrying out WSD over NCs is that sense-disambiguated elements in NCs provide stronger evidence to disambiguate the word sense of polysemous nouns than the context of usage of the NC. Also, in focusing on the elements of the NC, we can bring the one sense per collocation heuristic into play, in assuming that the elements in a given NC will always occur with the same sense, and further, that a given NC will always have the same semantic relation irrespective of context. We return to discuss this in detail in the Motivation section. We compare this approach with a standard corpus-based approach, using a state-of-the-art WSD system, and show that our disambiguation method based solely on word sense combinatorics is more successful at disambiguating word sense than existing methods.

Having disambiguated the word sense of the elements in a given NC, we can then test the second motivation that word sense can aid the interpretation of NCs, following the lead of Moldovan *et al.* (2004). Unlike Moldovan *et al.* (2004), however, we specifically compare word sense-sensitive and word sense-insensitive approaches to NC interpretation. In our experiments, we provide conclusive evidence that WSD can indeed boost the performance of NC interpretation.

Note that we do not dispute the claim that context influences NC interpretation (e.g., see Nastase *et al.* (2006)). Rather, for the our current purposes we focus exclusively on word sense at the type-level for NCs in isolation, and leave the harder task of token-level context modelling to future re-

[1]We focus exclusively on binary noun compounds in this paper.

Kim, Su Nam and Timothy Baldwin (2007) Disambiguating Noun Compounds, In Proceedings of the Twenty-Second Conference on Artificial Intelligence (AAAI-07), Vancouver, Canada, pp. 901-6.

sense	*day* mod	*day* head	*day* $*$	*art* mod	*art* head	*art* $*$
ws_1	.13	.04	**.41**	**.85**	**.62**	**.67**
ws_2	.02	.04	.20	.11	.04	.22
ws_3	**.80**	.00	.12	.00	.03	.08
ws_4	.00	**.91**	.20	.04	.31	.03
ws_5	.04	.01	.05	—	—	—
ws_6	.00	.00	.03	—	—	—
ws_7	.00	.00	.00	—	—	—
ws_8	.01	.00	.00	—	—	—
ws_9	.00	.00	.00	—	—	—
ws_{10}	.01	.00	.00	—	—	—

Table 1: Sense distribution for *day* and *art* as an NC modifier (mod), head noun (head) and overall in SEMCOR ($*$)

search.

The remainder of the paper is structured as follows. We first present the motivations underlying this research, before outlining the resources used in this research. Next, we propose supervised and unsupervised methods for WSD of NC elements and describe the procedure used to generate the dataset we evaluate our method over. Finally, we evaluate the proposed WSD methods and move onto the question of whether the output of these methods can enhance NC semantic relation interpretation, and conclude the paper with a discussion.

Motivation

Sense Distribution of Polysemous Elements in NCs

Our motivation is that the syntactic role of polysemous elements in NCs biases their sense distribution to varying degrees. Table 1 shows the distribution of word senses for two polysemous nouns, *art* and *day*, as the modifier and head in a range of NC types, and also overall in SEMCOR (a subsection of the Brown Corpus which has been manually sense-tagged with WORDNET senses). Note that here and throughout the paper, word senses and sense glosses are taken from WORDNET 2.1.

Day occurs most frequently with ws_3 ("daytime, daylight") when used as the modifier in an NC, but ws_4 ("a day assigned to a particular purpose or observance") when used as the head noun and ws_1 overall in SEMCOR. In the case of *art*, the predominant sense in all three cases is ws_1 ("the products of human creativity; works of art collectively"), but as a head noun, there are also significant numbers of ws_4 usages ("photographs or other visual representations in a printed publication") while the sense distibution for head noun instances is less skewed. In summary, we claim that the sense distribution of nouns is biased when they occur in NCs, with a different bias depending on whether they occur as the modifier or head noun.

Sense Restrictions in NCs

Moldovan *et al.* (2004) used word sense combinatorics to interpret semantic relations in NCs. They argued that the semantic relation of an NC can be predicted from the combination of word senses of the modifier (n_1) and head noun (n_2), based on simple conditional probability:

$$sr^* = \mathrm{argmax}_{sr_i} P(sr_i | ws(n_1), ws(n_2)) \qquad (1)$$

where $ws(n_*)$ returns the word sense of noun n_* and each sr_i is a semantic relation.

Based on this, we formulate our probabilistic model to disambiguate the word sense of polysemous element in NCs as follows:

$$ws^*(n_i) = \mathrm{argmax}_{ws(n_i)} P(ws(n_i) | ws(n_j), sr) \qquad (2)$$

where n_i is the element we are seeking to disambiguate, n_j is the remaining element in the NC, which we are assuming has already been disambiguated (see below for details), and sr is the semantic relation between the modifier and head noun. As we are attempting to use word sense to disambiguate the semantic relation, we instead use the syntactic role of n_i (i.e. modifier or head noun), as motivated above. Note that we could feasibly identify the semantic relation independent of word sense information, e.g. using the method of Kim & Baldwin (2005) based on nearest-neighbour matching over the union of senses of the modifier and head noun. Based on experimentation, however, the accuracy of standalone semantic relation disambiguation is inadequate, motivating the use of word sense information.

We have predicted that the word sense of an NC element is biased by the syntactic role in which it occurs and the element with which it co-occurs. Figure 1 shows an example of word sense restriction based on collocation: *art* as a modifier occurs only with sense 1 in combination with head nouns such as *exhibition*, but sense 4 in combination with head nouns such as *journal*.

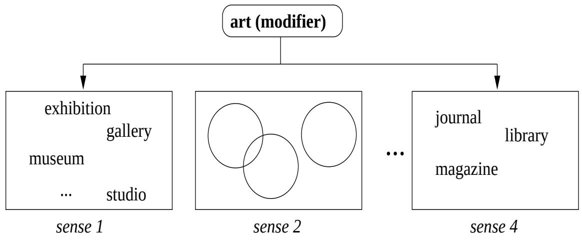

Figure 1: Sense restriction due to the semantics of the disambiguated element in NCs

One Sense per Collocation

Yarowsky (1993) originally proposed one sense per collocation as a general bootstrapping technique for WSD, in assuming that a word will occur with the same sense in combination with a given word collocation (specifically targeting NCs and adjective-noun collocations) across all token occurrences. Yarowsky demonstrated the effectiveness of the heuristic in achieving an accuracy of 90-99% over a range of binary disambiguation bootstrapping tasks. In earlier work, Yarowsky (1993) illustrated that for different word classes, different contexts were most effective in WSD. In the case of nouns, the best disambiguating context was directly adjacent adjectives or nouns, underlying the effectiveness of the heuristic for our work.

In our method, we apply the one sense per collocation heuristic to disambiguate the word sense of the NC elements. Our use of the heuristic differs slightly from that of Yarowsky, in that we are seeking to disambiguate both elements of the NC, rather than just one element based on what words it co-occurs with. Also, we are applying it to the full WORDNET sense inventory, rather than just coarse-grained binary distinctions, and thus don't expect to be able to reproduce the heady numbers of Yarowsky. Note that in using the one sense per collocation heuristic, we are not making any linguistic claims about the potential for a given NC to have different senses based on context. Rather, we are claiming that the majority of token occurrences of a given NC will conform to a given sense combination.

Resources

The set of semantic relations we use is based on that of Barker & Szpakowicz (1998), and made up of 20 relations including AGENT (e.g. *student protest*), CAUSE (e.g. *flood water*) and OBJECT (e.g. *horse doctor*).

As mentioned above, we use WORDNET 2.1 (Fellbaum 1998) as our sense inventory. WORDNET is a rich lexical ontology which is structured primarily using lexical relations such as synonymy, hypo/hypernymy and troponymy. Word senses in WORDNET are defined relative to synsets, that is sets of synonyms.

CORELEX (Buitelaar 1998) is a noun classification based on a unified approach to systematic polysemy and the semantic underspecification of nouns, and derives from WORDNET 1.5. It contains 45 basic types, systematic polysemous classes and 39,937 nouns with tags. We use CORELEX to obtain noun clusters to aid WSD. Note that CORELEX differs from WORDNET in that there is no hierarchical structure between clusters.

SENSELEARNER (Mihalcea & Faruque 2004) is an all (open-class) words WSD system, i.e. it disambiguates all words which are contained in WORDNET. The system uses minimal supervision, i.e. it uses a relatively small data set as training data. The system learns a global model for each word category (noun, verb, adjective and adverb) rather than learning a separate classifier for each target word. The system achieved 64.6% accuracy over the English all words task in Senseval-3, making it state-of-the-art at the time of writing. We use SENSELEARNER as a benchmark to compare our WSD method against.

The British National Corpus (BNC) is used as the source of corpus counts for our unsupervised WSD method.

Finally, we used the TIMBL 5.1 memory-based learner (Daelemans *et al.* 2004) to build all our supervised classifiers.

Approach

Supervised method

The one sense per collocation heuristic encodes the prediction that all occurrences of a noun in a given NC will have the same sense. We base determination of that sense on its co-occurring NC element, leading to the formulation in Equation 2. As features of the classifier, we use the role of the target noun and word sense of the second noun in the NC.

sense	substituted NCs
1	**craft/artifact** museum
2	**artistic production/creative activity** museum
3	**artistry/superior skill** museum
4	**artwork/graphics/visual communication** museum

Table 2: Substitution method for each sense of polysemous element in ***art*** *museum*

art:4	authority:7	bar:14	channel:8	child:4
circuit:6	day:10	nature:5	stress:5	

Table 3: Target noun set, and the polysemy of each

To determine the word sense of the collocating element (n_j), we experiment with the use of two types of semantics: noun classes from CORELEX, and the first sense from WORDNET.

Unsupervised method

As an unsupervised approach to NC WSD, we use lexical substitution, similarly to Mihalcea & Moldovan (1999) and Agirre & Martinez (2000). That is, we substitute the target element with synonyms from each of its WORDNET synsets, then compute the probability of each underlying word sense by calculating the frequency of the substituted NCs in a corpus. Since each word sense can have more than one synonym, we calculate the mean frequency across all synonyms. Finally, the word sense which has highest substitution-based frequency is assigned as the word sense of the target element. In the case that the target noun (n_1) is a modifier collocating with head noun n_2, this equates to the following calculation:

$$ws^*(n_1) = \text{argmax}_{s_i \in ws(n_1)} \frac{\sum_{n_j \in ss(s_i) \setminus \{s_i\}} freq(n_j, n_2)}{|ss(s_i) \setminus \{s_i\}|} \quad (3)$$

where each s_i is a word sense of n_1, and $ss(s_i)$ returns the synset containing sense s_i. The calculation in the case the target noun is a head noun is analogous, with the only change being in the calculation of the corpus frequency.

Table 2 illustrates the process for the target noun *art* in the context of *art museum*, over the four WORDNET senses of *art*.

Dataset

In order to evaluate the proposed method, we constructed a sense-annotated dataset of NCs. First, we retrieved 20 frequent polysemous nouns from SEMCOR and Senseval-2. We then identified binary NCs in the BNC incorporating those 20 nouns in either modifier or head noun position (but not both). From this, we extracted polysemous nouns which occurred as both the modifier and head noun over at least 50 NC token instances, out of which we selected the 9 nouns which occurred in the most NCs, as detailed in Table 3.

For the 9 target nouns, we then (fully) randomly selected 50 NCs in each of the modifier and head noun positions. As

art **lesson**

Sense 1 lesson
=> teaching, instruction,...
=> education
=> profession
=> ...
Sense 2 example, deterrent example..

Figure 2: Word sense of the collocating element *art*

a result, we have 100 NCs per target noun, and a total of 900 instances.

Two linguistically-trained human annotators sense-tagged all 900 NCs, based on analysis of their token occurrences in the BNC. In the case that a given NC was used over a range of sense assignments, the majority class assignment was selected for the NC type. The initial type-level inter-annotator agreement was 69.2%, and all instances of disagreement were resolved based on discussion. Those same 900 NCs were then annotated with a semantic relation by a single expert annotator, once again with reference to their token occurrences in the BNC.

To get the noun classes for the collocating elements to use in the supervised classifier, we used CoreLex. Among the 900 NC instances, an entry was found for 61.6% of the collocating elements in CoreLex. For the remaining elements a CoreLex class was manually assigned following the CoreLex clustering methodology.

We also determined the semantics of the collocating elements, based on the first sense prediction method of McCarthy *et al.* (2004). Here, we are predicting the predominant sense of each noun in the BNC in an unsupervised manner, and using this to disambiguate all instances of that noun. We adopt the method of Kim & Baldwin (2006), in supplementing the predominant sense with its three WordNet hypernyms, as illustrated in Figure 2 for the collocating head noun *lesson* in *art lesson* (target noun = *art*).

For the unsupervised classifier, we calculated the web count of each synonymy-substituted NC using Google. This was motivated by the findings of Lapata & Keller (2004) that the web provides reliable probability estimates for tasks including compound noun interpretation and bracketing. In our case, we generated both singular and plural forms of each NC using morph (Minnen, Carroll, & Pearce 2001) in calculating the frequency of a given NC. We did not use synonyms that included the target noun; in the instance that no usable synonyms could be found, we used hypernyms (again with the constraint on lexical overlap with the target noun). For example, for sense 1 of *art*, we selected *artifact* and *craft* as substitutes but not *fine art*.

Evaluation

In this section, we first evaluate the supervised and unsupervised WSD methods, and compare their performance to a range of baselines and a benchmark WSD system. We then return to the second question posed in this research, namely whether word sense is beneficial in NC interpretation.

WSD Evaluation

We evaluate the supervised WSD method via 10-fold cross-validation over the 900 NC instances[2] using TiMBL, and evaluate the unsupervised method over the same 900 instances. We use two unsupervised baselines: (1) random sense assignment, and (2) predicted first sense of the target noun, based on the full BNC (comprising NC and non-NC instances of a given target noun). We also have one supervised baseline in the form of a majority class classifier, based on 10-fold cross-validation over the 900 instances. Finally, in order to benchmark our results, we ran SenseLearner over our dataset using the pre-trained word class models, randomly selecting one of the original sentential contexts from the BNC for each NC and sense labelling it. The classification accuracy for the output of each WSD method over each target noun, broken down across the modifier and head noun positions, is shown in Table 4. The best-performing method in each row is indicated in boldface.

The best-performing classifier overall was the supervised classifier with CoreLex noun features (combined accuracy = 55%), although there was very little separating the equivalent classifier with WordNet features (combined accuracy = 54%). Given the relatively close results and the fact that the CoreLex-based classifier required considerable manual lexical expansion of CoreLex in order to get full coverage over all the collocating nouns, the WordNet-based classifier would be our preferred classifier for open-text applications. The majority-class baseline was actually the best-performing classifier for around half of the sub-experiments in Table 4, but was often only slightly better than the CoreLex and WordNet-based classifiers, whereas there were a number of instances of the majority-class baseline falling well behind the other two supervised classifiers. According to our analysis, the lower performance is due to a lack of training data (i.e. 50 instances for each target word). There was no significant difference in the performance for modifier vs. head noun WSD, but in secondary experimentation we verified that conditioning the disambiguation on the syntactic role improved accuracy.

The SenseLearner benchmark performed surprisingly poorly. That we should do better than a generic WSD system is not particularly surprising, as we have fine-tuned our method to the task, but that the combined accuracy of SenseLearner should be almost half that of the proposed method is a significant finding, especially when we consider that SenseLearner was trained over Senseval-2 and SemCor data. This shows that the one sense per collocation heuristic and word sense combinatorics are stronger predictors of noun sense in NCs than standard contextual features.

The unsupervised method performed well below the supervised methods (both the majority class baseline and the

[2] 10-fold cross-validation is performed individually for each of the 9 target words as either head noun or modifier. As a result, the 9 target words have 50 instances for each of the two syntactic positions.

Target noun	Role in NC	Baseline Random	First	Majority	Supervised CoreLex	WordNet	Unsupervised	SenseLearner
art	modifier	.25	.68	.68	.64	**.70**	.44	.54
	head noun	.25	**.54**	**.54**	.48	.51	.30	.50
	both	.25	**.61**	**.61**	.56	**.61**	.37	.52
authority	modifier	.14	.06	**.78**	.70	.77	.18	.06
	head noun	.14	.08	**.60**	.52	.54	.36	.08
	both	.14	.07	**.69**	.61	.65	.27	.07
bar	modifier	.07	.46	.46	**.54**	.47	.20	.46
	head noun	.07	.30	.24	**.46**	.40	.24	.28
	both	.07	.38	.35	**.50**	.43	.22	.37
channel	modifier	.13	**.24**	**.24**	**.24**	.18	.26	.22
	head noun	.13	.16	.26	**.28**	.24	.30	.12
	both	.13	.20	.25	**.26**	.21	.28	.17
child	modifier	.25	**.72**	**.72**	.50	.69	.24	.60
	head noun	.25	**.78**	**.78**	.76	.76	.38	.76
	both	.25	**.75**	**.75**	.63	.73	.31	.68
circuit	modifier	.17	**.68**	**.68**	.62	.61	.62	.66
	head noun	.17	.54	.54	.48	**.57**	.42	.52
	both	.17	**.61**	**.61**	.55	.59	.52	.59
day	modifier	.10	.18	**.68**	.64	.62	.24	.14
	head noun	.10	.06	**.90**	.88	.89	.16	.06
	both	.10	.12	**.79**	.76	.75	.20	.10
nature	modifier	.20	.04	**.70**	**.70**	**.70**	.30	.04
	head noun	.20	.34	.14	**.44**	.38	.20	.32
	both	.20	.19	.42	**.57**	.54	.25	.18
stress	modifier	.20	.02	.48	**.50**	.46	.30	.02
	head noun	.20	.08	.08	.24	**.27**	.28	.08
	both	.20	.05	.28	**.37**	.36	.29	.05
Total	modifier	.16	.34	**.60**	.59	.58	.31	.30
	head noun	.16	.32	.45	**.50**	**.50**	.29	.30
	both	.16	.33	.53	**.55**	.54	.30	.30

Table 4: WSD accuracy over each target noun in the modifier and head noun positions (the best-performing method in each row is indicated in **boldface**)

WordNet and CoreLex classifiers) and slightly below the first sense baseline, at the same combined accuracy as SenseLearner.

NC Interpretation Evaluation

In order to quantify the impact of word sense on NC interpretation, we build a classifier which takes the output of a WSD system and uses these sense features to classify the semantic relation of the NC. In order to investigate the impact of WSD performance on NC interpretation, we compare the predictions of our two supervised classifiers with hand-tagged sense data (for the target noun), to establish an upper bound for the task.

The method we use for NC interpretation is essentially that of Moldovan *et al.* (2004) as outlined above. However, instead of disambiguating the collocating element in the NC, we use semantic features from CoreLex and WordNet as before. This is combined with the WordNet sense of the target noun to form the input to our supervised classifier; as before, we perform 10-fold cross validation over the 900-instance dataset. We compare this directly to a first-sense disambiguation method, trained over the full BNC (i.e. the same as used for the first-sense baseline in WSD). The output of the first sense classifier is combined with the CoreLex and WordNet features of the collocating noun as above, producing a fully comparable classifier.

Method	CoreLex	WordNet
baseline	.273	.273
similarity	.346	.346
system-tagged	.402	.426
first-sense	.403	.425
hand-tagged	.447	.540

Table 5: Accuracy of interpreting semantic relations in NCs

As a benchmark for this task, we implemented the method proposed by Kim & Baldwin (2005), which is based on nearest-neighbour matching over the union of senses of the modifier and head noun, with distance defined by word-level similarity in WordNet (*similarity* in Table 5). That is, this method makes use of word sense information but does not attempt to perform explicit WSD. The initial type-level inter-annotator agreement of tagging semantic relations in the test data was 52.31%. The baseline for the task is a majority-class classifier (the majority class being TOPIC).

Table 5 shows the classification accuracy of NC interpretation using CoreLex and WordNet semantic features for the collocating noun, and the word sense of the target noun predicted by the WORDNET-based supervised method, the first sense (McCarthy *et al.* 2004) disambiguation method, and hand-tagged sense data. These are compared to the results of our majority-class baseline and the method of Kim & Baldwin (2005).

The performance of the WORDNET-based supervised classifier and the first sense disambiguation method are almost identical using both the WORDNET and CORELEX semantic representations of the collocating noun, but the upper bound classifier based on hand-tagged data is better than both these, particularly for the CORELEX representation of the collocating noun. This suggests that the features of the collocating nouns are weighted higher than the noisy word sense features of the target noun, and that to approach the upper bound accuracy, significantly higher WSD performance is required. Both automatic WSD-based methods clearly outperform both the baseline and the benchmark interpretation method, demonstrating that word sense can indeed boost the performance of NC interpretation. Note that since all of the NCs in our dataset contain polysemous nouns, the performance of the Kim & Baldwin (2005) method is considerably lower than that reported in the original paper.

To answer our original question about whether WSD can contribute to NC interpretation, the answer appears to be a resounding yes. This is significant both in documenting a task where WSD makes a positive impact, and in opening up a new research direction in the field of NC interpretation.

Conclusion

In this research, we proposed both supervised and unsupervised approaches to disambiguating the word sense of NC elements. Based on the one sense per collocation heuristic and semantic features of the collocating noun, we disambiguate a specific target word. We showed that our method was better than a number of baselines, and also vastly superior to a state-of-the-art WSD system over this specialised task. We then took the based of our WSD methods and applied it to the task of NC interpretation, and found that WSD significantly enhanced the accuracy of NC interpretation.

References

Agirre, E., and Edmonds, P. 2006. Word sense disambiguation: Algorithms and applications. In *Word Sense Disambiguation: Algorithms and Applications*. Dordrecht, Netherlands: Springer.

Agirre, E., and Martinez, D. 2000. Exploring automatic word sense disambiguation with decision lists and the web. In *In Proc. of COLING workshop on Semantic Annotation and Intelligent Content*.

Banerjee, S., and Pedersen, T. 2003. Extended gloss overlaps as a measure of semantic relatedness. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, 805–810.

Barker, K., and Szpakowicz, S. 1998. Semi-automatic recognition of noun modifier relationships. In *Proceedings of the 17th international conference on Computational linguistics*, 96–102.

Buitelaar, P. 1998. *CoreLex: Systematic Polysemy and Underspecification*. Ph.D. Dissertation, Brandeis University.

Carpuat, M., and Wu, D. 2005. Word sense disambiguation vs. staitstical machine translation. In *In Proceedings of 43rd Annual Meeting of the Association for Computational Linguistics*.

Daelemans, W.; Zavrel, J.; van der Sloot, K.; and van den Bosch, A. 2004. TiMBL: Tilburg Memory Based Learner, version 5.1, reference guide.

Fellbaum, C., ed. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, USA: MIT Press.

Ide, N., and Veronis, J. 1998. Word sense disambiguation: The state of the art. *Computational Linguistics* 24(1).

Kim, S. N., and Baldwin, T. 2005. Automatic interpretation of noun compounds using WordNet similarity. In *In Proc. of the 2nd International Joint Conference On Natural Language Processing*, 945–956.

Kim, S. N., and Baldwin, T. 2006. Automatic identification of English verb particle constructions using linguistic features. In *In Proc. of the Third ACL-SIGSEM Workshop on Prepositions*, 65–72.

Lapata, M., and Keller, F. 2004. The web as a baseline: Evaluating the performance of unsupervised web-based models for a range of NLP tasks. In *In Proc. of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 121–128.

Leacock, C., and Chodorow, N. 1998. Combining local context and wordnet similarity for word sense identification. In *Combining local context and WordNet similarity for word sense identification*.

Levi, J. 1979. The syntax and semantics of complex nominals. In *New York:Academic Press*.

McCarthy, D.; Koeling, R.; Weeds, J.; and Carroll, J. 2004. Finding predominant senses in untagged text. In *In Proc. of the 42nd Annual Meeting of the ACL*, 280–287.

Mihalcea, R., and Faruque, E. 2004. SenseLearner: Minimally supervised word sense disambiguation for all words in open text. In *In Proc. of ACL/SIGLEX Senseval-3*.

Mihalcea, R., and Moldovan, D. 1999. Au automatic method for generating sense tagged corpora. In *In Proc. of 16th Conference of the American Association of Artificial Intelligence*.

Minnen, G.; Carroll, J.; and Pearce, D. 2001. Applied morphological processing of English. *Natural Language Engineering* 7(3):207–223.

Moldovan, D.; Badulescu, A.; Tatu, M.; Antohe, D.; and Girju, R. 2004. Models for the semantic classification of noun phrases. In *HLT-NAACL 2004: Workshop on Computational Lexical Semantics*, 60–67.

Nastase, V.; Sayyad-Shiarabad, J.; Sokolova, M.; and Szpakowicz, S. 2006. Learning noun-modifier semantic relations with corpus-based and WordNet-based features. In *Proc. of the 21st Annual Conference on Artificial Intelligence (AAAI-06)*.

Sanderson, M. 1996. *Word sense disambiguation and Information retrieval*. Ph.D. Dissertation, University of Glasgow.

Vickrey, D.; Biewald, L.; Teyssier, M.; and Koller., D. 2005. Word-sense disambiguation for machine translation. In *In Proc. of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 771–778.

Yarowsky, D. 1993. One sense per collocation. In *In Proc. of ARPA Human Language Technology Workshop*, 266–271.

Yarowsky, D. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *In Proc. of the 33rd Annual Meeting of the ACL*, 189–196.