

# Extending Sense Collocations in Interpreting Noun Compounds

Su Nam Kim, Meladel Mistica and Timothy Baldwin

CSSE and NICTA VRL

University of Melbourne

VIC 3010, Australia

{snkim, mmistica, tim}@csse.unimelb.edu.au

## Abstract

This paper investigates the task of noun compound interpretation, building on the sense collocation approach proposed by Moldovan et al. (2004). Our primary task is to evaluate the impact of similar words on the sense collocation method, and decrease the sensitivity of the classifiers by expanding the range of sense collocations via different semantic relations. Our method combines hypernyms, hyponyms and sister words of the component nouns, based on WordNet. The data used in our experiments was taken from the nominal pair interpretation task of SEMEVAL-2007 (4th International Workshop on Semantic Evaluation 2007). In the evaluation, we test 7-way and 2-way class data, and show that the inclusion of hypernyms improves the performance of the sense collocation method, while the inclusion of hyponym and sister word information leads to a deterioration in performance.

## 1 Introduction

This paper investigates the automatic interpretation of noun compounds (i.e. NCs), such as *paper submission* and *computer science department*. NC interpretation is a well-known problem that aims to predict the **semantic relation** (i.e. SR) between a head noun and its modifying nominal(s) (i.e. modifiers). SRs, simply put, encapsulate how the head noun and the other nominals in a noun compound

are related. As English noun phrases are right-headed, the head noun occurs after all modifying nouns. For example, *brick house* is interpreted as a *house* that is modified by the word *brick*, which exhibits a PRODUCT-PRODUCER relationship between the two nouns in the compound. In contrast, the modifier and head in *house brick* exhibits a PART-WHOLE relationship, which is interpreted as a brick from a house, rather than the former interpretation of a house made of bricks. The set of SRs that we are concerned with in this paper is defined in Section 5.1.

Research on NCs can be categorised into four main groups: defining SRs, disambiguating the syntax of NCs, disambiguating the semantics of NCs, and interpreting NCs via SRs. Each task is detailed in Section 8. Interpreting NCs has received much attention of late, and the problem has been addressed in areas of machine translation (MT), information extraction (IE), and applications such as question-answering (QA). Processing NCs exhibits many challenges due to the following issues associated with the task (Lapata, 2002): (1) the compounding process is extremely productive; (2) the semantic relationship between the head noun and its modifier(s) is implicit; and (3) the interpretation of an NC can vary due to contextual and pragmatic factors. Due to these challenges, current NC interpretation methods are too error-prone to employ directly in NLP applications without any human intervention or preprocessing.

In this paper, we investigate the task of NC interpretation based on sense collocation. It has been shown that NCs with semantically similar compo-

nents share the same SR (Kim and Baldwin, 2007); this is encapsulated by the phrase coined as **sense collocation** in Moldovan et al. (2004). For example, NCs such as *apple pie* have the same interpretation as *banana cake*, where the modifiers of both NCs are semantically similar (they are both classified as fruit), and the head nouns of both NCs are a type of baked edible concoction. Given that the modifier is a *fruit* and the head noun is a *baked concoction*, then we can interpret NCs with this sense collocation as having the PRODUCT-PRODUCER SR. However, unlike the method of Kim and Baldwin (2007), where new data was induced by substituting the components of the NCs with semantically similar terms, our approach adds related terms as features for the classifier. The related terms we add are the NC components' hypernyms, hyponyms, and sister words. We expect that these similar words can contribute to disambiguate the SR when classifying NCs.

The remainder of this paper is structured as follows. In Section 2 we present previous research on NC interpretation. Section 3 and Section 4 describe our motivation and method, respectively. We describe the data used in our experiments in Section 5 and present the results of our experiments in Sections 6 and 7. In Section 8, we briefly outline some of the issues associated with NC research. Finally, we conclude our work in Section 9.

## 2 Related Work

A majority of research undertaken in interpreting NCs have been based on two statistical methods: SEMANTIC SIMILARITY (Barker and Szpakowicz, 1998; Rosario, 2001; Moldovan et al., 2004; Kim and Baldwin, 2005; Nastase, 2006; Girju, 2007; Kim and Baldwin, 2007) and SEMANTIC INTERPRETABILITY (Vanderwende, 1994; Lapata, 2002; Kim and Baldwin, 2006; Nakov, 2006). Our work, based on an extension of the sense collocation approach, corresponds to the semantic similarity category.

A significant contribution to this area is by Moldovan et al. (2004), who used the sense collocation (i.e. pair-of-word-senses) as their primary feature in disambiguating NCs. Many subsequent studies have been based on this sense collocation method, with the addition of other performance-

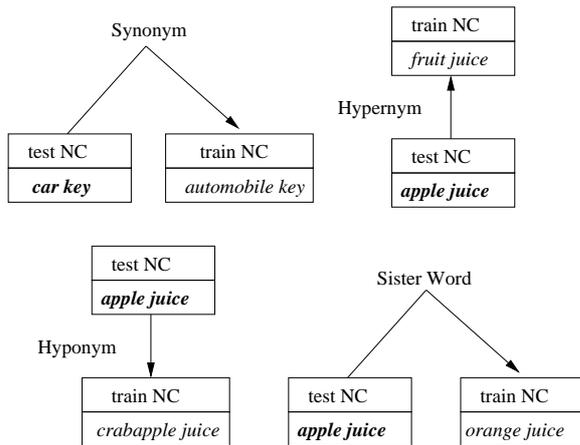
improving features. For example, Girju (2007) added contextual information (e.g. the grammatical role and POS) and cross-lingual information from 5 European languages as features to her model. In contrast, Kim and Baldwin (2007) utilise sense collocations in a different way: instead of adding additional features in their model, they increase the size of their training data by substituting components of existing training instances to generate additional training instances (which is assumed to have the same SR as the original). For an SR to be preserved the newly-generated NC must be semantically similar and hence maintain the same sense collocation as the original NC on which it was based. Rosario (2001; Kim and Baldwin (2005; Nastase (2006) attempted to interpret NCs by applying implicit sense collocations. In particular, they used various ways to retrieve sense collocations instead of manually assigning word senses to the NCs. Hence, their methods do not require direct use of word senses. Rosario (2001) retrieved the sense pairs in the context of a hierarchical class set for the biomedical domain. Kim and Baldwin (2005) used the similarity measure to express the sense collocation of NCs. Nastase (2006) listed the hypernyms of components as sense features.

## 3 Motivation

As mentioned above, Moldovan et al. (2004) showed that the sense collocation of NCs is a key feature when interpreting NCs. Further research in this area has shown that not only synonymous NCs share the same SR, but NCs whose components are replaced with similar words also have the same SR as the original NCs (Kim and Baldwin, 2007). For example, *car factory*, *automobile factory* and *truck factory* substituted with a synonym, hypernym and sister word, respectively, share the same SR of PRODUCT-PRODUCER.

Figure 3 shows an example of semantic neighbours for the two NCs *car key* and *apple pie*. *Car key* can be interpreted as PRODUCT-PRODUCER by referring to the training NC *automobile key*, since they have the same sense collocation. With *apple juice*, the sense collocation method tries to locate matching sense collocations in the training data, and finds that *fruit juice* matches closely, with the mod-

*apple pie (SR=MAKE)*



ifier being a hypernym of *apple*. From this, we can hope to correctly interpret *apple juice* as having the SR PRODUCT-PRODUCER. In order to achieve this, we require some means of comparing nouns taxonomically, both vertically to capture hypernyms and hyponyms, and horizontally to capture sister words.

As intimated above, our motivation in conducting this research is to be able to include hypernym, hyponym and sister word information without using direct substitution over the training instances, but still preserving the essence of the sense collocation approach. The disadvantage of the method employed by Kim and Baldwin (2007) is that noise will inevitably infect the training data, skewing the classifier performance. The original method described in Moldovan et al. (2004) only relies on observed sense collocations. The components of the NCs are represented as specific synsets in WordNet, and the model does not capture related words. Hence, in this paper, we aimed to develop a model that can take advantage of relatedness between WordNet synsets via hypernyms, hyponyms and sister words, without the risk of losing semantic granularity or introducing noisy training data. Note that in Kim and Baldwin (2007), we used synonyms, hypernyms and sister words. As synonyms have the identical sense collocation (i.e. pairing of synsets) within WordNet, we ignore it in this research. Instead, we add hyponyms as a means of broadening the range of sense collocation.

## 4 Method

At first, we describe the principal idea of **sense collocation method** on NC interpretation and the probability model proposed in (Moldovan et al., 2004). Then we present our method using hypernyms, hyponyms and sister words in order to extend sense collocation method.

### 4.1 Sense Collocation

The basic idea behind sense collocation method in Moldovan et al. (2004) was based on the "pair-of-word-senses" from the component nouns in noun compounds as features of the classifier. They also introduced a probability model called **semantic scattering**, as detailed in Equations 1 and 2 below as a supervised classification technique. In essence, the probability  $P(r|f_i f_j)$  (simplified to  $P(r|f_{ij})$ ) of a modifier and head noun with word sense  $f_i$  and  $f_j$ , respectively, occurring with SR  $r$  is calculated based on simple maximum likelihood estimation:

$$P(r|f_{ij}) = \frac{n(r, f_{ij})}{n(f_{ij})} \quad (1)$$

The preferred SR  $r^*$  for the given sense combination is that which maximizes the probability:

$$\begin{aligned} r^* &= \operatorname{argmax}_{r \in R} P(r|f_{ij}) \\ &= \operatorname{argmax}_{r \in R} P(f_{ij}|r)P(r) \end{aligned} \quad (2)$$

### 4.2 Adding Similar Words

We extend the approach of Moldovan et al. (2004) by adding similar words as features focusing on hypernyms, hyponyms and sister words of the modifier and head noun.

We accumulate the features for semantic relations based on different taxonomic relation types, from which we construct a feature vector to build a classifier over. The features of each taxonomic relation types are listed below. The first is features used in the original sense collocation method. The second, third and fourth are our experimental features added hypernyms, hyponyms and sister words respectively.

1.  $\langle WS_{mod}, WS_{head} \rangle$
2.  $\langle WS_{mod}, H_{mod}^i, WS_{head}, H_{head}^i \rangle$

Semantic relation	Definition	Examples
Cause-Effect (CE)	$N_1$ is the cause of $N_2$	<i>virus flu, hormone growth, inhalation death</i>
Instrument-Agency (IA)	$N_1$ is the instrument of $N_2$ , $N_2$ uses $N_1$	<i>laser printer, ax murderer, sump pump drainage</i>
Product-Producer (PP)	$N_1$ is a product of $N_2$ , $N_2$ produces $N_1$	<i>honey bee, music clock, supercomputer business</i>
Origin-Entity (OE)	$N_1$ is the origin of $N_2$	<i>bacon grease, desert storm, peanut butter</i>
Theme-Tool (TT)	$N_2$ is intended for $N_1$	<i>reorganization process, copyright law, work force</i>
Part-Whole (PW)	$N_1$ is part of $N_2$	<i>table leg, daisy flower, tree forest</i>
Content-Container (CC)	$N_1$ is store or carried inside $N_2$	<i>apple basket, wine bottle, plane cargo</i>

Table 1: The set of 7 semantic relations, where  $N_1$  is the head noun and  $N_2$  is a modifier

3.  $\langle WS_{mod}, O_{mod}, WS_{head}, O_{head} \rangle$

4.  $\langle WS_{mod}, S_{mod}, WS_{head}, S_{head} \rangle$

where *mod* is the modifier, *head* is the head noun,  $WS_{mod}$  is the WordNet synset of the modifier,  $WS_{head}$  is the WordNet synset of the head,  $H^i$  is an  $i$ th-degree ancestor (with direct hypernyms corresponding to  $H^1$ ),  $O$  is a hyponym and  $S$  is a sister word. We include up to the 7th-degree ancestor (i.e.  $H^7$ )<sup>1</sup> in our model. Note that while a given synset has a unique hypernym in WordNet (assuming no cycles, or the ability to remove cycles by precompiling a tree structure), it can have arbitrarily many hyponyms and sister words. Here, we take the cross product of the different hyponym and sister word candidates for a given synset.

We build our final classifier with TiMBL v5.16.0, a memory-based learner (Daelemans et al., 2004).

## 5 Data

### 5.1 Semantic Relation

The SR between a head and its modifier(s) in a NC tells us how to (default) interpret the NC. For example *door knob* corresponds to the PART-WHOLE relation, which means we can interpret *knob* as being part of a *door*. We sidestep the considerable challenge of developing an optimal set of semantic relation categories by using the set of SRs and data from the SEMEVAL-2007 nominal pair interpretation task (Girju et al., 2007). The SRs defined for the task are: CAUSE-EFFECT(CE), CONTENT-CONTAINER(CC), INSTRUMENT-AGENCY(IA), ORIGIN-ENTITY(OE), PART-WHOLE(PW), PRODUCT-PRODUCER(PP) and THEME-TOOL(TT). Table 1 provides a definition of each SR along with example NCs.

<sup>1</sup>We adopted 7th-degree ancestor from the previous study in (Nastase, 2006).

### 5.2 Data Collection

From the SEMEVAL-2007 annotated data (Girju et al., 2007), we collect two sets of data: a 2-class dataset and a 7-class dataset. The 2-class dataset is taken from the original SEMEVAL-2007 task, and comprises a set of positive and negative instances for each of the 7 SRs. The 7-class dataset is derived from this, by combining all positive NCs across the 7 SRs. The taxonomic relations are derived from WordNet3.0. In each of the two sets, we use each of hypernyms, hyponyms and sister words. Table 2 shows the number of hyponyms and sister words in each dataset.

class	Hyponym		Sister word	
	<i>mod</i>	<i>head</i>	<i>mod</i>	<i>head</i>
7-classes	4866	4708	7167	7456
2-classes (CE)	1272	774	3220	2043
2-classes (IA)	955	1804	1726	3722
2-classes (PP)	1526	1688	3058	3009
2-classes (OE)	2394	1730	3861	2907
2-classes (TT)	1383	812	2767	1698
2-classes (PW)	1403	1770	2900	4117
2-classes (CC)	1598	820	2620	1909

Table 2: Total number of hyponym- and sister word-based NCs

## 6 7-way classification experiment

We ran our first experiment over the 7-class dataset. The baseline was computed using a *Zero-R* classifier (i.e. majority vote).<sup>2</sup> The performance of the original method proposed in Moldovan et al. (2004) is considered as a benchmark. Table 3 shows the performance of the original sense collocation method

<sup>2</sup>Majority class was PRODUCT-PRODUCER.

	B	M+	H <sup>1</sup>	H <sup>2</sup>	H <sup>3</sup>	H <sup>4</sup>	H <sup>5</sup>	H <sup>6</sup>	H <sup>7</sup>	O	S
Accuracy	.217	.496	.544	.552	.573	.562	<b>.588</b>	.568	.557	.197	.142

Table 3: Results for the 7-way classification task:  $B$  = baseline,  $M+$  = Moldovan et al. (2004) method,  $H^i$  =  $i$ th-order Hypernym,  $O$  = Hyponym and  $S$  = Sister word; the best performing system is indicated in **boldface**

	B	M+	H <sup>1</sup>	H <sup>2</sup>	H <sup>3</sup>	H <sup>4</sup>	H <sup>5</sup>	H <sup>6</sup>	H <sup>7</sup>	O	S
CE	.547	.547	.533	.573	.600	.606	.586	.607	<b>.630</b>	.467	.453
IA	.507	.581	.595	.608	.649	<b>.671</b>	.653	.629	.645	.500	.500
PP	.655	.667	.679	.691	.679	<b>.737</b>	.700	.690	.687	.655	.655
OE	.558	.636	.623	.610	.662	.645	.662	.625	<b>.712</b>	.558	.558
TT	.636	.697	.727	.712	.742	<b>.766</b>	.732	.717	.650	.515	.394
PW	.634	.620	.690	.690	.629	.657	.585	<b>.731</b>	.630	.633	.634
CC	.514	.676	<b>.703</b>	.689	.689	.676	.667	.647	.698	.446	.514
All	.579	.632	.649	.653	.662	<b>.679</b>	.654	.661	.667	.541	.534

Table 4: Results for each of the 2-way classification tasks:  $B$  = baseline,  $M+$  = Moldovan et al. (2004) method,  $H^i$  =  $i$ th-order Hypernym,  $O$  = Hyponym and  $S$  = Sister word; the best performing system is indicated in **boldface**

and that of extended sense collocation as proposed in this paper.

Table 3 shows that our method, combined with hypernyms outperforms the original sense collocation method, with the highest accuracy of .5880 achieved with 5th-degree ancestors of the head noun and modifier. This confirms that hypernyms are valuable in extending the range of sense collocation for NC interpretation.

In stark contrast to the results for hypernyms, the results for hyponyms and sister words significantly reduced the accuracy. The reason for this anomaly is that hypernyms are able to generalize the sense collocation without losing key discriminative features (i.e. the hypernyms always, by definition, subsume the original semantic information), while hyponyms and sister words add many sense collocations for which we have no direct evidence (i.e. we indiscriminately specialise the semantics without any motivation). Hence, hyponyms and sister words drastically blur the sense collocation .

The reason that the accuracy of the hypernym method drops in the result tables a certain level is that the semantic collocations start to blend in together, and lose their power of discrimination.

## 7 2-way classification experiment

Our second experiment was to run the systems over the original data from SEMEVAL-2007, in the form of a binary classifier for each of the 7 SRs. The performance of each of the 2-way classification tasks is shown in Table 4.

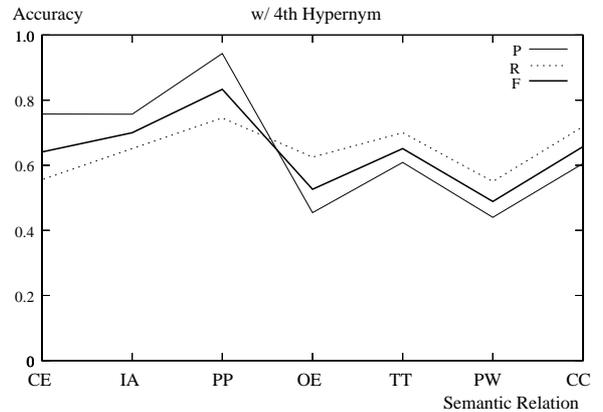


Figure 1: Positive Task with Hypernym<sub>4th</sub>

As we can see in Table 4, the basic pattern of the results is the same as for the 7-way classification task in Table 3. Adding hypernyms enhances performance, peaking for 4th-degree ancestors in this case at .6786. As with the 7-way classification task,

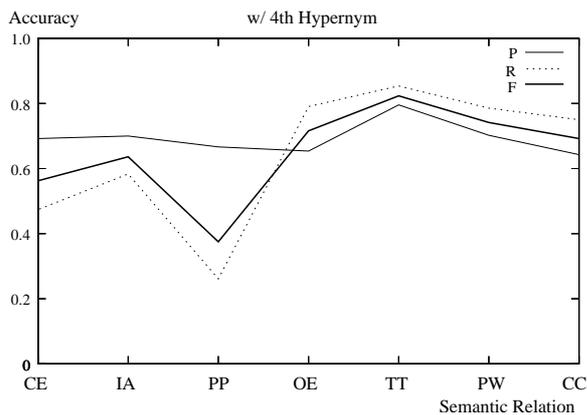


Figure 2: Negative Task with Hypernym<sub>4th</sub>

hyponyms and sister words degraded performance, for the same reasons as before.

Looking at the performance of each SR, we found that some SRs are more easy to interpret than others. Notably, PRODUCT-PRODUCER and THEME-TOOL were high performers, while CAUSE-EFFECT was considerably harder to classify. These trends coincide with the system results for the SEMEVAL-2007 task. Girju et al. (2007) analyze this effect in terms of the intrinsic semantic complexity of the different SRs, and also the relative size of the training data for each SR. These effects are also observable in the breakdown of precision and recall of each SR in Figure 1.

As we used the same data for SEMEVAL-2007 task, we are able to directly compare the performance of our method with the competition results. The Table 5 shows the three baselines provided by the SEMEVAL-2007 organisers. We also present the best-performing system and the average performance within group B from the SEMEVAL-2007 task as our method uses only word senses. The methods for computing the baselines are described in (Girju et al., 2007). *All True* is computed to guess "true" for all relations which maximizes recall. *Probability* is computed by randomly assigning "true"(or "false") with probability matching the distribution of "true"(or "false") in the testing dataset for the given relation. It intends to balance precision and recall. *Majority* is computed by assigning "true" or "false" to which ever is the majority in the testing dataset for the given relation.

As shown in Table 5, our performance using hy-

Method	P	R	F	A
All True	.485	1.00	.648	.485
Probability	.485	.485	.485	.517
Majority	.813	.429	.308	.570
Best	.797	.698	.724	<b>.763</b>
Average	.650	.637	.631	.656

Table 5: Results of 2-way classification ( $P$ =precision,  $R$ =recall,  $F$ =F-score,  $A$ =accuracy)

pernyms outperformed all of three baselines. The performances using hyponyms and sister words only exceeded the baselines computed by "all true" and "probability" measure as these taxonomies tend to decrease the performances. The interesting point here is that although the method is meant for general-purpose interpretation not for binary decision task, our proposed method with hypernyms simply achieved better than the baselines and relatively performed well compared other systems in the competition (average accuracy of group B = .656 vs. our best = .679). Therefore, we conclude that sense collocation integrated with hypernyms is a potential to extend sense collocation as well as improve the performance for NC interpretation task.

## 8 NC research primer

In this section, apart from previous works related to NC interpretation task described in Section ??, we present a short description of the computational tasks and previous attempts to provide the broad comprehension of tasks related to NCs.

The major tasks related to NCs involve syntactic and semantic disambiguation. Related to the task of semantic disambiguation is the task of defining what relations exist in NCs. This has gained much attention in recent decades, as well as controversy, (Downing, 1977; Levi, 1979; Finin, 1980). In the study conducted by Levi (1979), it was claimed that there were 9 distinct SRs, which could be discretely defined and interpreted within NCs, while Finin (1980) claimed an unlimited number of SRs. The problems surrounding this task involve the issue of granularity versus coverage, which to date remains widely debated.

Syntactic disambiguation (called bracketing) is required when NCs are composed of more than 2

components, such as in the case of *computer science department*, introducing the need for phrasal disambiguation (Lauer, 1995; Nakov, 2005). Lauer (1995) proposed probabilistic models (based on dependency and adjacency analyses of the data). Later Nakov (2005) built upon this by adding linguistic features into these probabilistic models.

Methods employed in word sense disambiguation (WSD) have also been used to enhance NC interpretation; the noun components that comprise the NCs are disambiguated using these WSD techniques (SparckJones, 1983; Kim and Baldwin, 2007). Kim and Baldwin (2007) carried out experiments on automatically modeling WSD and attested the usefulness of conducting analysis of the word senses in the NC in determining its SR.

In the automatic interpretation of NCs, many claims have been made for the increase in performance, but these works make their own assumptions for interpretation (Barker and Szpakowicz, 1998; Moldovan et al., 2004; Kim and Baldwin, 2005; Girju, 2007; Kim and Baldwin, 2007; Seaghdha, 2007).

## 9 Conclusion

In this paper, we have investigated the impact of using different taxonomic relations to expand a sense collocation method of NC interpretation. That is, we experimented with the integration of similar terms into a sense collocation model. We added up to 7 hypernyms, direct hyponyms and direct sister words terms as features to a classifier. We ran experiments on 7-way and 2-way classification tasks using data from SEMEVAL-2007, and found that the inclusion of hypernym information significantly improved accuracy, while hyponyms and sister words degraded performance by arbitrarily overspecialising the sense information.

Our study has shown that adding features broadens the scope of sense collocation. While intuitively all of hypernyms, hyponyms and sister words would appear to provide rich features for a sense collocation method, further research is needed to develop ways of successfully incorporating hyponyms and sister words into the NC interpretation task.

## References

- Ken Barker and Stan Szpakowicz. 1998. Semi-automatic recognition of noun modifier relationships. In *Proc. of the 17th International Conference on Computational Linguistics*, pp. 96–102, Montreal, Canada.
- Walter Daelemans, Jakub Zavrel, Ko van der Sloot, and Antal van den Bosch. 2004. *TiMBL: Tilburg Memory Based Learner, version 5.1, Reference Guide*. ILK Technical Report 04-02.
- Pamela Downing. 1977. On the Creation and Use of English Compound Nouns. *Language*, 53(4):810–842.
- Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, USA.
- Timothy W. Finin. 1980. *The Semantic Interpretation of Compound Nominals*. Ph.D. thesis, University of Illinois.
- Roxana Girju. 2007. Improving the Interpretation of Noun Phrases with Cross-linguistic Information. In *Proc. of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 568–575, Prague, Czech Republic.
- Roxana Girju, Preslav Nakov, Vivi Nastase, Stan Szpakowicz, Peter Turney and Deniz Yuret. 2007. SemEval-2007 Task 04: Classification of Semantic Relations between Nominals. In *Proc. of Semantic Evaluation Workshop*, Prague, Czech Republic.
- Su Nam Kim and Timothy Baldwin. 2005. Automatic interpretation of Noun Compounds using WordNet similarity. In *Proc. of the 2nd International Joint Conference On Natural Language Processing*, pp. 945–956, JeJu, Korea.
- Su Nam Kim and Timothy Baldwin. 2006. Interpreting Semantic Relations in Noun Compounds via Verb Semantics. In *Proc. of the 44th Annual Meeting of the Association for Computational Linguistics and 21st International Conference on Computational Linguistics (COLING/ACL-2006)*, pp. 491–498, Sydney, Australia.
- Su Nam Kim and Timothy Baldwin. 2007. Disambiguating Noun Compound. In *Proc. of the 22nd AAAI conference on Artificial Intelligence (AAAI-07)*, 2007, pp.901-906, British Columbia, Canada.
- Su Nam Kim and Timothy Baldwin. 2007. Interpreting Noun Compound Using Bootstrapping and Sense Collocation. In *Proc. of the Pacific Association for Computational Linguistics (PACLING)*, Melbourne, Australia.
- Maria Lapata. 2002. The disambiguation of nominalizations. *Computational Linguistics*, 28(3):357–388.

- Mark Lauer. Designing Statistical Language Learners: Experiments on Noun Compounds, Ph.D thesis. 1995, Macquarie University.
- Judith Levi. 1979. The syntax and semantics of complex nominals. In *The Syntax and Semantics of Complex Nominals*. New York:Academic Press.
- Dan Moldovan, Adriana Badulescu, Marta Tatu, Daniel Antohe, and Roxana Girju. 2004. Models for the semantic classification of noun phrases. In *Proc. of the HLT-NAACL 2004 Workshop on Computational Lexical Semantics*, pp. 60–67, Boston, USA.
- Preslav Nakov and Marti Hearst. 2005. Search Engine Statistics Beyond the n-gram:Application to Noun Compound Bracketing. In *Proc. of the 9th Conference on Computational Natural Language Learning (CoNLL-2005)*, pp. 17-24, Ann Arbor, Michigan, USA.
- Preslav Nakov and Marti Hearst. 2006. Using Verbs to Characterize Noun-Noun Relations. In *Proc. of the 12th International Conference on Artificial Intelligence: Methodology, Systems, Applications (AIMSA)*, Bularia.
- Vivi Nastase and Jelber Sayyad-Shiarabad and Marina Sokolova and Stan Szpakowicz. 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)*, pp. 781-787, Boston, MA, USA.
- Barbara Rosario and Hearst Marti. 2001. Classifying the Semantic Relations in Noun Compounds via a Domain-Specific Lexical Hierarchy. In *In Proc. of EMNLP*, 82–90
- Karen Sparck Jones. 1983. Compound noun interpretation problems. *Computer Speech Processing*, Prentice-Hall, Englewood Clifles, NJ, USA.
- Diarmuid Seaghdha and Ann Copestake. 2007. Co-occurrence Contexts for Noun Compound Interpretation. In *Proc. of the ACL-2007 Workshop on A Broader Perspective on Multiword Expressions*, Prague, Czech Republic.
- Lucy Vanderwende. 1994. Algorithm for automatic interpretation of noun sequences. In *Proc. of the 15th Conference on Computational linguistics*, pp. 782–788.