

# Automatic Synonymy Extraction

## A Comparison of Syntactic Context Models

*Kris Heylen, Yves Peirsman and Dirk Geeraerts*  
QLVL - University of Leuven, Belgium<sup>1</sup>

### Abstract

Distributional models of lexical semantics identify semantically similar words through contextual similarity. Previous studies have shown that syntactic contexts are especially good at finding (near) synonyms. In this paper, we compare models based on eight different syntactic dependency relations and we evaluate their separate and combined performance on a test set of Dutch nouns. Firstly, we analyze to what extent their results overlap. Secondly, we assess the overall performance of the models by looking at the average similarity of the words they return. And thirdly, we compare the specific semantic relations retrieved by the models. The analyses show that although models based on the subject and object relation give the most consistent results, it is the model based on adjective modification that gives the best results. It even outperforms the combined model at finding true synonyms.

### 7.1 Introduction

The automatic retrieval of semantically similar words has become an important task in NLP research with applications in Information Retrieval, Word Sense Dis-

---

<sup>1</sup>Yves Peirsman is a Ph.D. Fellow of the Research Foundation - Flanders (FWO)

ambiguation, Thesaurus Extraction, Anaphora Resolution and even Parsing. So-called distributional models of lexical semantics have turned out to be one of the most promising approaches for modelling semantic similarity. They rely on the assumption that words with a similar meaning tend to occur in similar contexts and, consequently, that a word's meaning can be modelled as a function of the contexts it occurs in. In practice, these models gather statistics about the co-occurrence of a word with a large number of context features from corpora and put these into a vector. The semantic similarity between two words is then measured as the distributional similarity between their respective context vectors.

Over the last decade, various implementations of the distributional approach have been developed. They mainly differ with respect to the restrictions they impose on what counts as *context* for modelling the meaning of a target word. Some variants allow all the other words in the same document, others only take words in a predefined window around the target word into account, and still others use only context words in a specific syntactic relation to the target word. It is clear that these different types of context features are likely to capture different kinds of semantic information. However, until recently, little was known about the influence of the context definition on the semantic information present in these word vector spaces. While most researchers choose one specific model and apply it to their task, "comparisons between the ... models have been few and far between in the literature" (Padó, Sebastian and Lapata, Mirella 2007). Yet, without any knowledge of the linguistic characteristics of the models, it is impossible to know which approach is best suited for a particular task, and why. In previous studies (Yves Peirsman and Kris Heylen and Dirk Speelman 2007, Yves Peirsman and Kris Heylen and Dirk Speelman 2008, Kris Heylen and Yves Peirsman and Dirk Geeraerts and Dirk Speelman 2008), we compared the performance of models using a pre-defined context window and those relying on syntactically related words. These studies showed that the syntactic model with its very strict context definition outperformed the other models in finding semantically similar nouns for Dutch. That syntactic model was based on eight syntactic dependency relations. However, our evaluation did not differentiate between these eight relations. In this paper, we will focus on models that are based on each of the eight syntactic relations separately. On the basis of a test set of Dutch nouns, we will compare the models both among each other, as well as with the combined model. This way, we want to find out which relation is most informative for detecting semantic similarity and contributes most to the overall success of the syntactic model. More specifically, we will first investigate to what extent the results of the different models overlap. Then we analyse the overall quality of the results in terms of their average semantic similarity to the test set, and finally, we will assess how well the models can detect specific semantic relations like synonymy.

In Section 7.2, we first situate our approach in the broader field of research into distributional models and we discuss previous studies into the properties of syntactic models. Section 7.3 discusses the data and parameter settings we used in our test set-up. Section 7.4 first presents the evaluation scheme we applied and then discusses, consecutively, the overlap of the models, their overall performance

and their ability to detect specific semantic relations. Finally, in section 7.5, we wrap up with conclusions and some suggestions for future research.

## 7.2 Related work

### 7.2.1 Distributional models

Distributional models, a.k.a. word space models, semantic space models or vector-based models of lexical semantics, rely on insights that were already formulated in the 1950's by Zelig Sabbettai Harris (1954), Warren Weaver (1955) and John Rupert Firth (1957), viz. that a word's meaning can be induced from the contexts it appears in. However, it was not before the 1990's and the advent of large corpora and increased computational power that a practical implementation became feasible. Distributional models actually became popular first within cognitive psychology as a way to model lexical learning and word memory (Kevin Lund and Curt Burgess 1996, Thomas K. Landauer and Susan T. Dumais 1997, Will Lowe and Scott McDonald 2000) but were soon enthusiastically adopted by the NLP community.

Although all word space models are based on the same underlying assumption, they do come in many different flavours. The main difference between them lies in how they define the central notion of *context*. Figure 7.1 offers a classification according to context definition. Document-based models use whole documents or paragraphs as contexts so that words that often co-occur in documents appear as semantically similar. Latent Semantic Analysis (Thomas K. Landauer and Susan T. Dumais 1997) is probably the best known example of this type. Because they use documents as context, they are especially useful for the grouping of topic identifying terms in document classification.

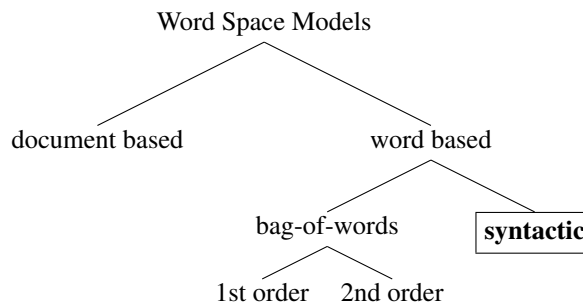


Figure 7.1: Syntactic models within the family of Distributional Models

For extracting tight semantic relations like synonymy, however, word-based models have turned out to be more suitable. They restrict contexts to the words in near proximity to the target word and treat two words as similar if these often co-occur with the same context words. Unlike document-based models, they do not expect target words to co-occur regularly with each other. Within the class of

word-based models, we can make a distinction between bag-of-word and syntactic models. Bag-of-word models simply look at the context words that appear in a pre-defined window around the target word. They are called bag-of-words models because they do not differentiate between the words within the context window. Context words with different POS values or syntactic functions are all treated on a par. A further subdivision can be made between first-order and second-order bag-of-words models. In the case of first-order models, the context features are the words that directly co-occur with the target word in the pre-defined window (Joseph P. Levy and John A. Bullinaria 2001), whereas second order bag-of-word models make use of the second order co-occurrences, i.e. the context words of the first order co-occurrences. By doing so, they should allow to generalize over meaning related context words and avoid data sparseness. These properties make that second-order models have been primarily used in Word Sense Disambiguation for grouping word tokens in sense clusters (Hinrich Schütze 1998), rather than for the clustering of word types.

Finally, we come to the syntactic or dependency-based models. Unlike the bag-of-word models they do impose a specific relation between the target and its context words. They only take context words into account that stand in a pre-defined syntactic dependency relation to the target word. Context features are then words like verbs governing the target word in its subject function or adjectives modifying the target, plus the respective dependency relation. These models have proven to be especially apt at finding words with tightly related meanings and they will be the focus of this study.

### 7.2.2 Syntactic models

Syntactic models owe their success to the fact that not all context words are equally informative for inferring a word's meaning. For example, verbs subcategorize for subject or object nouns from specific semantic classes. Consequently, a subject or object noun's governing verb tells a lot more about that noun's meaning than another randomly chosen context word. The same holds for adjectives modifying a given noun. Researchers have tried to capitalize on this link between semantics and morpho-syntactic structure to various degrees by taking into account part of speech tags (Dominic Widdows 2003, Klaus Rothenhäusler and Hinrich Schütze 2007), shallow syntactic analyses (Gregory Grefenstette 1994, Lillian Lee 1999, James R. Curran and Marc Moens 2002) and full syntactic parses (Dekang Lin 1998). Only the latter approach fully exploits the semantic information contained within syntactic dependency relations but also requires most resources. However, thanks to the advent of robust automatic parsers, the use of full-blown dependency information has become a viable option that we will also pursue here.

In their excellent overview article, Padó, Sebastian and Lapata, Mirella (2007) propose a general framework for implementing distributional models that integrates the additional parameters necessary to describe dependency-based models. One of those parameters is the *context selection function*, which specifies the relation that must hold between target and context words. For bag-of-words models,

this relation is simply one of occurrence within the context window. In the case of syntactic models, it is a bit more complex. The relation is then stated in terms of the possible paths through a syntactic dependency tree that can connect a context word with a target word. These paths can be of length one, e.g. between a target noun and its governing context verb, but they can also be longer. In the NP "the man with the hat", the context word *hat* is a postmodifier to the target noun *man* linked by a path of length two with connections from *man* to *with* and from *with* to *hat*. If the dependency tree is regarded as a graph with the words as nodes and the dependency relations as edges, the context selection function specifies, what length a dependency path can have, which parts-of-speech the start, end and intermediate nodes are allowed to be, and of what type the edges can be. It is this context selection parameter that we will vary in our study, while keeping the other parameters constant.

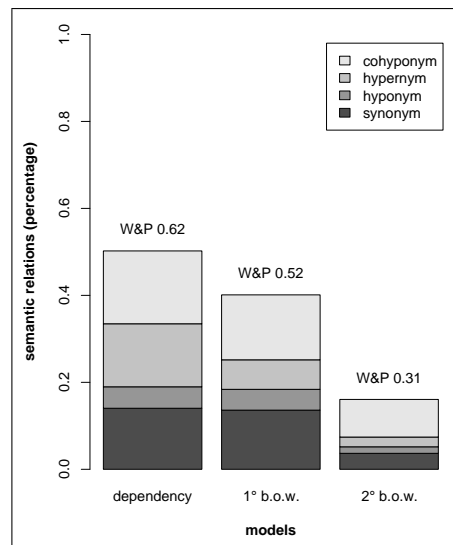


Figure 7.2: Performance of the syntactic model compared to the first and second order bag-of-words models (Wu & Palmer score and semantic relations for single nearest neighbours)

For Dutch, distributional models have been applied to Semantic Class Induction (Van de Cruys, Tim 2005), Multi Word Expression Extraction (Van de Cruys, Tim and Villada Moirón, Begoña 2007), Word Sense Discrimination (Van de Cruys, Tim 2007), Cognitive Word Association Modelling (De Deyne, Simon and Storms, Gert 2008) and Question Answering (Lonneke van der Plas and Gosse Bouma 2005a). Van der Plas and Tiedemann have compared the performance of distributional models based on monolingual and parallel corpora (Lonneke van der Plas and Jörg Tiedemann 2006). Our own studies (Yves Peirsman and Kris Heylen and Dirk Speelman 2007, Yves Peirsman and Kris Heylen and Dirk Speelman 2008, Kris Heylen and Yves Peirsman and Dirk Geeraerts and Dirk

Speelman 2008) have compared bag-of-word and syntactic models, and found, as Figure 7.2 shows, that a syntactic model outperforms the bag-of-word models both in terms of overall performance and specific semantic relations retrieved<sup>2</sup>. As the best performing model, we will focus on the syntactic model here. In this regard, our study links up directly with previous work by van der Plas and Bouma (Lonneke van der Plas and Gosse Bouma 2005b). They too compared the overall performance of syntactic models based on six different dependency relations and section 7.4.2 will partially be a replication of their experiments. However, this study will look at two additional dependency relations and will not only assess overall performance, but also the overlap of the models and the specific syntactic relations they retrieve.

Table 7.1: Paths and examples for the eight dependency relations

Rel.	Path	Example
su	noun $\xrightarrow{su}$ verb	Het <i>meisje</i> slaapt (The girl sleeps)
obj	noun $\xrightarrow{obj^1}$ verb	Hij eet een <i>appel</i> (He eats an apple)
pc	noun $\xrightarrow{obj^1}$ preposition $\xrightarrow{pc}$ verb	Ze luistert naar de <i>radio</i> (She listens to the radio)
advPP	noun $\xrightarrow{obj^1}$ preposition $\xrightarrow{mod}$ verb	Hij woont in een <i>dorp</i> (He lives in a village)
pmPP	noun $\xleftarrow{mod}$ preposition $\xleftarrow{obj^1}$ noun	Het <i>meisje</i> met de jurk (The girl with the dress)
adj	noun $\xleftarrow{mod}$ adjective	De <i>gelaarsde kat</i> (The booted cat)
app	noun $\xleftarrow{app}$ noun	De <i>koningin</i> , een wijze vrouw (The queen, a wise woman)
cnj	noun $\xleftrightarrow{cnj}$ noun	De <i>krekel</i> en de mier (The cricket and the ant)

### 7.3 Set-up

The data for our experiments consists of the 300 million word Twente Nieuws Corpus of Dutch lemmatised and parsed newspaper text (Roeland J.F. Ordelman 2002)<sup>3</sup>. Parsing was done at the University of Groningen with the Alpino dependency parser for Dutch (Gertjan van Noord 2006). Based on Alpino's parsing scheme, we selected eight types of syntactic dependency relations for constructing our word spaces. The target word's function in these relations was one of the following:

1. **su**: subject of verb  $v$
2. **obj**: direct object<sup>4</sup> of verb  $v$
3. **pc**: prepositional complement of verb  $v$  introduced by preposition  $p$

<sup>2</sup>See section 7.4 for a description of these evaluation measures

<sup>3</sup>Publication years 1999 up to 2002 of *Algemeen Dagblad*, *NRC*, *Parool*, *Trouw* and *Volkskrant*

<sup>4</sup>This includes the subjects of passive verbs

4. **advPP**: head of an adverbial PP of verb  $v$  introduced by preposition  $p$
5. **pmPP**: postmodified by a PP with head  $n$ , introduced by preposition  $p$
6. **adj**: modified by adjective  $a$
7. **app**: modified by an apposition with head  $n$
8. **cnj**: coordinated (via a conjunct) with head  $n$

Each specific instantiation of the variables  $v$ ,  $p$ ,  $a$ , or  $n$  led to a new context feature. Table 7.1 shows the dependency paths as they were extracted from Alpino's output along with some examples. In *su*, *obj* and *pc*, the target noun is the dependent word in the relation, and in *cnj* there is a symmetrical relation between the target and context word. In the other cases, the target word is the head. Three relations (*pc*, *advPP*, *pmPP*) have a path length of two. The others are direct relations of path length one. We extracted from the lemmatised corpus the 10,000 most frequent nouns and recorded for each of them with which specific instantiations of the eight relation types they occurred and how often. Based on this information, we constructed nine word spaces: one for each of the eight dependency relations separately, and one for the combination of relations. For the models based on separate dependency relations, only the 1000 most frequent features were used to guarantee computational feasibility. For the combined model, we kept the original dimensionality from our previous studies and took the 4000 most frequent features into account<sup>5</sup>. The remaining parameters were kept constant and were set as follows:

**Weighting scheme** Context vectors contained the point-wise mutual information (Kenneth Ward Church and Patrick Hanks 1990) between the feature and the target, rather than raw frequency.

**Similarity metric** The cosine of the angle described by two context vectors was used to measure the similarity between these vectors.

Using these models, we calculated for each target noun the 100 most similar nouns among the remaining 9999 possibilities, which we will designate as the target's *nearest neighbours*. To give an idea of the output, Table 7.2 gives the first nearest neighbour found by the nine models for a random sample of 10 target nouns.

## 7.4 Results and Discussion

We performed three types of evaluation: overlap between the models, overall performance in terms of the nearest neighbours' average semantic similarity to their targets, and specific semantic relations retrieved. The two latter types of evaluation use a gold standard, which we will discuss briefly first.

We evaluate our models against the Dutch part of EuroWordNet (Vossen, Piek 1998). Like its English counterpart, EuroWordNet is a lexical database structured

<sup>5</sup>Distribution over relations: *adj* 843; *advPP* 431; *app* 57; *cnj* 524; *obj* 675; *pc* 336; *pmPP* 412; *su* 684

Table 7.2: Random sample of target words and their nearest neighbours

TARGET	adj	advPP	app	cnj	all
aanzien	uitstraling	aanleiding	vliegverkeer	prestige	prestige
capriool	frats	steunpilaar	beschadiging	ommekeer	stunt
flirt	uitstapje	oogst	azijn	drank	avontuur
grondlegger	opsteller	spil	coup	oprichter	oprichter
grot	kelder	kelder	voetbalveld	ravijn	dorp
kerkgebouw	kerk	kerk	zuivel	synagoge	kerk
opdrachtgever	financier	werkgever	aandeelhouder	ontwerper	klant
thee	koffie	koffie	zwijn	koffie	koffie
toilet	wc	strand	bushalte	douche	wc
trots	held	stelligheid	maatschappij	liefde	vreugde

TARGET	objl	pc	pmPP	su	all
aanzien	prestige	populariteit	imago	populariteit	prestige
capriool	frats	ophef	zonschijn	bijval	stunt
flirt	hoogmoed	provocatie	omgang	rechtszaak	avontuur
grondlegger	schilder	exponent	geschiedenis	schepper	oprichter
grot	paleis	restaurant	gehucht	hiernamaals	dorp
kerkgebouw	pand	autosnelweg	supermarkt	hangar	kerk
opdrachtgever	eigenaar	werkgever	uitvoerder	werkgever	klant
thee	koffie	koffie	koffie	koffie	koffie
toilet	wc	wc	wc	keuken	wc
trots	zelfvertrouwen	ontzag	imago	opluchting	vreugde

as a hierarchical network of concepts, each represented as the set of synonyms (synset) that refers to it. The Dutch section of EuroWordNet contains 44K synsets, which is a fair bit sparser than English WordNet (117K synsets). Our evaluations are based on the target-neighbour pairs retrieved by the models and, of course, only pairs with both the target and nearest neighbour present in EuroWordNet can be assessed. To allow for a fair comparison, we wanted to evaluate all the models on exactly the same set of target words. Therefore, we could only take target words into account that were themselves present in EuroWordNet, and for which each of the first nearest neighbours retrieved by the 9 models was also present. Because of the relative sparseness of EuroWordNet, this led to a drastic reduction of the data set to 2749 target words. Needless to say, this means the results below should be interpreted with some caution.

#### 7.4.1 Overlap

As a first evaluation, we want to know to what extent the models retrieve the same related words for our test set of Dutch nouns. To do so, we use the overlap metric developed by Sahlgren (Magnus Sahlgren 2006). The metric reflects how similar the results of two models are, simply by calculating the overlap between the nearest neighbours found for each target word<sup>6</sup>: For each of the 10,000 nouns in

<sup>6</sup>This overlap is generally very low. Sahlgren, for instance, found a maximum of around 10% overlap between the document-based and bag-of-word models.



our test set<sup>7</sup>, we took the 100 nearest neighbours and then calculated how many neighbours the models shared. The total number of shared neighbours divided by 10,000 then gives an average overlap between each of the models expressed in percentages<sup>8</sup>. For the combined model, the resulting similarity matrix showed, unsurprisingly, that the degree of overlap with separate relation models follows closely the relative frequency of these relations in the dimensions of the combined model. The real interesting question is to what extent the separate relation models, based on different information, overlap. We therefore took the similarity matrix for these models, turned it into a dissimilarity matrix by taking  $1 - \text{overlap}$ , and fed this into a hierarchical cluster analysis using complete linkage as a cluster criterion. Figure 7.3 shows that the results of the models based on the direct object and subject relation are most similar with an overlap of 24% (and thus a distance of 0.76). This overlap is actually quite high and might be explained by the fact that in both relations the target noun is directly dependent on a context verb. The next subcluster consists of the adjective and coordination-based model. These two relations do not have a lot in common that might explain their clustering. Such a communality does exist for the next subcluster consisting of the prepositional complement and adverbial PP models. Both relations connect the target noun to a governing context verb via a preposition in a two step dependency path. In fact, the distinction between prepositional complements and adverbial PP's is sometimes arbitrary in Alpino's parsing scheme. For example, *radio* in *luisteren naar de radio* (listen to the radio) is parsed as a prepositional complement, whereas *televisie* in *kijken naar de televisie* (watch television) is regarded as an adverbial PP. Finally, the dendrogram shows that the nearest neighbours retrieved by the post-modifying and apposition models overlap least with those of the other models. We now know how similar the results of the different models are. However, this doesn't tell us anything about the quality of these results. That is what we will look at in the next two sections.

#### 7.4.2 Overall performance

Following previous studies (Lonneke van der Plas and Gosse Bouma 2005b, Van de Cruys, Tim 2006, Yves Peirsman and Kris Heylen and Dirk Speelman 2007) we analyse the overall performance of our models by measuring the average semantic similarity of the nearest neighbours to their targets as recorded in EuroWordNet. To do so, we use the Wu & Palmer similarity score (Zhibiao Wu and Martha Palmer 1994) which has become somewhat of a standard for measuring similarity in lexical taxonomies. Wu & Palmer basically measure how far two words are removed in the hierarchy and apply a normalization for the relative hierarchy depth, which then results in a score ranging from 0 (no similarity) to 1 (perfect similarity)<sup>9</sup>.

<sup>7</sup>Since we do not have to rely on EuroWordNet here, we can use the complete data set.

<sup>8</sup>Note that this overlap measure does not take into account the rank of nearest neighbours.

<sup>9</sup>If a word occurs at different places in the hierarchy because of polysemy, only the highest Wu & Palmer score is taken into account. When the system returns, say, *depository* for a polysemous word

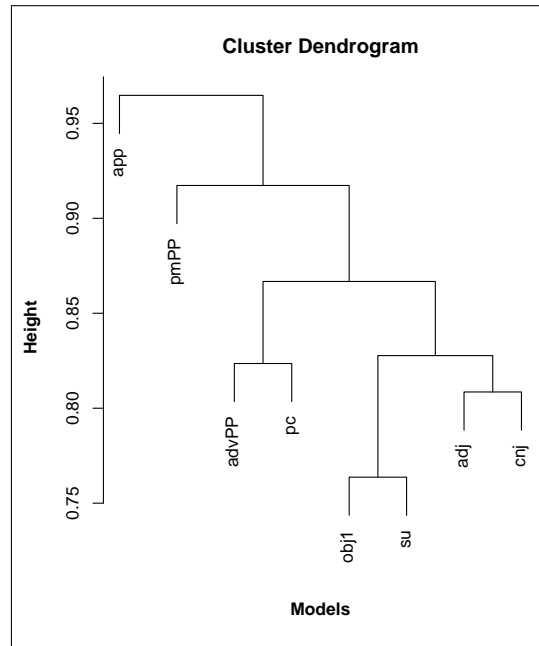


Figure 7.3: Clustering of the models based on their results

Figure 7.4 shows this average similarity for the 1st, 10, 50 and 100 most related words retrieved by the models<sup>10</sup>. We see that the combined model performs best with an average Wu & Palmer score of 0.65 for the first nearest neighbour. This shows that it is worthwhile to combine different dependency relations. With two relations extra (advPP and pmPP), we do even slightly better than van der Plas and Bouma's (Lonneke van der Plas and Gosse Bouma 2005b) maximum score of 0.60 on their test set, although that might also be due to the bigger size of our training corpus. Note that our adjective model on its own also scores a surprising 0.61, which is almost as high as the combined model. This result confirms that adjectives are highly informative for modelling the semantics of the nouns they modify. The adjective model is closely followed by the one based on the direct-object relation with a 0.59 score. The four next models (su, pc, cnj, advPP) perform a bit less well with scores around 0.50 for the first nearest neighbour, but this difference gets less pronounced when more nearest neighbours are taken into account. The worst models are clearly those based on the post-modifying PP and apposition relation. The pmPP model still scores reasonably well for the first nearest neighbour, but

like *bank*, it seems fair to assume that the identified similarity is to the financial meaning of bank rather than to the river side meaning.

<sup>10</sup>As explained in section 7.4 we required that at least the 1st nearest neighbour was present in EuroWordNet for all models. For nearest neighbours up to rank 100, we only took those into account that were present in the database and simply ignored the others for calculating the average similarity.

then deteriorates quickly. The apposition model performs right out poorly across the whole spectrum with an average similarity as low as 0.23 for the first nearest neighbour.

Interestingly, the ranking of the models based on average similarity does not completely correspond to their relative degree of overlap: the subject and object models gave the most consistent results according to the overlap measure, but it is actually the adjective model that performs best when evaluated against a gold standard. On the other hand, the poor performance of the apposition model is reflected in its low overlap with the other models. Finally, we also point out that our results do not completely match those found by van der Plas and Bouma (Lonneke van der Plas and Gosse Bouma 2005b). In their experiment, the object-based model performed best, although the adjective model came in second. The main difference lies in the performance of the apposition model, which still scored a fairly good 0.51 in van der Plas and Bouma's experiment. These differences might be due to the fact that a different test set was used.

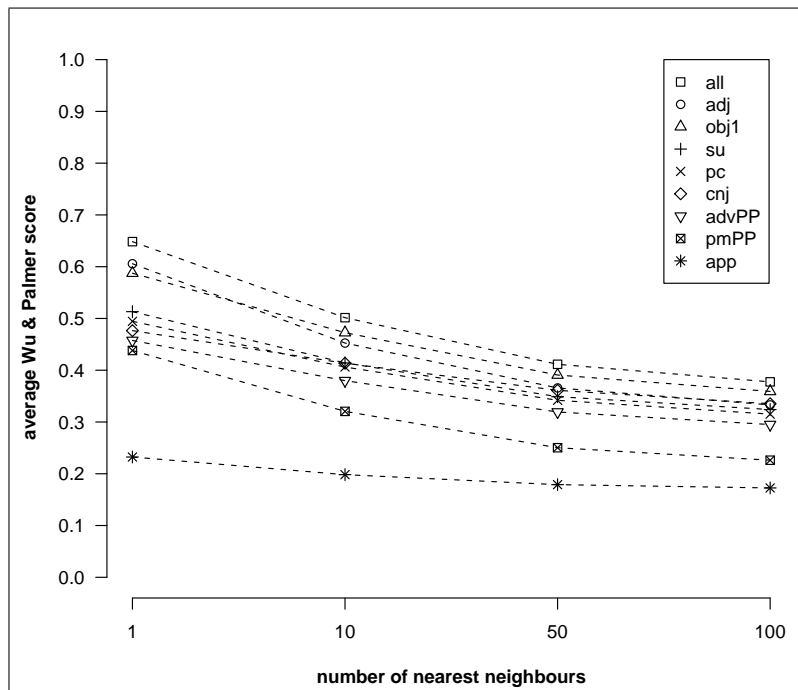


Figure 7.4: Overall performance of the 9 models

### 7.4.3 Specific semantic relations

Although the average semantic similarity of the nearest neighbours might give a good idea of overall performance, many applications of distributional models, like thesaurus extraction or query expansion, are foremost interested in finding specific semantic relations. To evaluate the models' performance on this task, we checked which semantic relation, if any, each first nearest neighbour entertained with its target according to EuroWordNet. Four semantic relations were taken into account (in decreasing order of semantic relatedness): synonymy, hyponymy, hypernymy and co-hyponymy. They were defined as follows:

**synonym** word occurring in the same synset as the target

**hyponym** word occurring in a synset that is a direct daughter of the target's synset

**hypernym** word occurring in a synset that is a direct mother of the target's synset

**cohyponym** word occurring in a synset that is a direct daughter of the target's hypernymic synset

Note that we use a strict definition of semantic relatedness by only allowing a minimal number of steps between a target and its neighbour in the hierarchy<sup>11</sup>.

Figure 7.5 shows the relative frequency of the four semantic relations among the most related words retrieved by the nine models. The aggregated percentages confirm by and large the results of the overall performance analysis: the combined model performs best with 55% of its results displaying one of the four semantic relations. The adjective and object model follow at 48% and 45% respectively. The apposition-based model performs worst with only 10% of the retrieved words having a semantic relation to the target. However, if we look at the individual semantic relations, we do see interesting differences between the models that were not revealed by the overall performance analysis. First of all, the adjective model is slightly better than the combined model at finding true synonyms (15.4% versus 14.7%). Secondly, in the group of average performing models, we see that the model based on post-modifying PP's is also notably better at finding true synonyms than its overall performance score would predict. Both observations suggest that modifiers are clearly good indicators of a noun's precise meaning, resulting in a better retrieval of tightly related words. If strict synonymy for nouns is what is needed for an application, distributional models based on adjective modification alone might be the best option to go with.

## 7.5 Conclusions and future work

Dependency-based distributional models were already known to perform quite well on the task of finding semantically related words when compared to bag-of-word competitors. Our detailed analysis of dependency-based models in this

<sup>11</sup>As with the Wu & Palmer score, only the shortest connection in the hierarchy was taken into account for target-neighbour pairs with multiple connections due to polysemy.

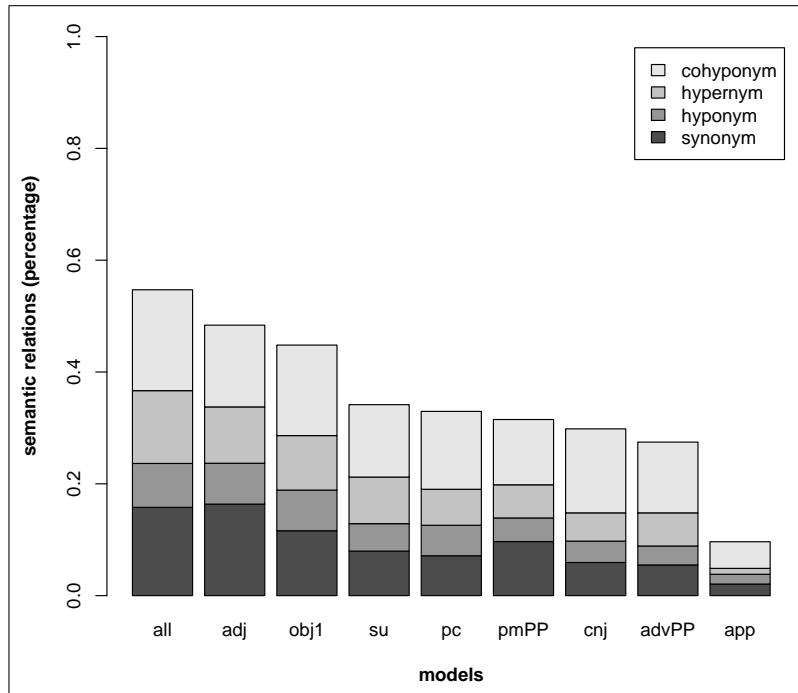


Figure 7.5: Semantic relations retrieved by the 9 models

paper has shed more light on the question why these models are more successful. More specifically, we were able to tease out their sources of information for noun semantics by comparing models that used specific types of dependency relations as context features. Our three evaluation schemes were able to highlight different aspects of their performance. By comparing the overlap in results, we saw that models with context features based on direct verb dependency, i.e. object and subject relations, give the most consistent results. However, our overall performance analysis showed that these are not necessarily the best results: those were generated by the model using adjectives as context features. The evaluation of overall performance also showed that a combination of dependency relations allows to retrieve words with a slightly higher average similarity than the best performing individual relation. Finally, the analysis of specific semantic relations retrieved showed that models based on modifiers, either adjectives or post-modifying PP's, are especially good at finding true synonyms. The adjective-based model even outperformed the combined model on this task.

We have only looked at eight types of dependency relations so far and in a next step we would like to extend this set. In particular, an analysis of relations with longer path lengths could reveal whether post-modification by certain relative clauses, say, those headed by specific verbs is also informative for noun seman-

tics. Apart from analysing which dependency relation works best individually, we would also like to find out which combination is best suited to retrieve semantically similar words. We have looked at one combination, but for eight relations there are 248 possible combinations. By studying how individual dependency relations interact, we want to get a systematic overview of the links between a word's meaning and the syntactic contexts it occurs in.

## References

- De Deyne, Simon and Storms, Gert (2008), Word associations: Network and semantic properties, *Behavior Research Methods* **40** (1), pp. 213–231.
- Dekang Lin (1998), Automatic retrieval and clustering of similar words, *Proceedings of the 17th international conference on Computational linguistics*, pp. 768–774.
- Dominic Widdows (2003), Unsupervised methods for developing taxonomies by combining syntactic and statistical information, *Proceedings of the Joint Human Language Technology Conference and Annual Meeting of the North American Chapter of the Association for Computational Linguistics*, pp. 197–204.
- Gertjan van Noord (2006), At Last Parsing Is Now Operational, in Piet Mertens and Cedrick Fairon and Anne Dister and Patrick Watrin, editor, *TALN06. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles*.
- Gregory Grefenstette (1994), Corpus-derived first, second and third-order word affinities, *Proceedings of the Sixth EURALEX International Congress*.
- Hinrich Schütze (1998), Automatic word sense discrimination, *Computational Linguistics* **24** (1), pp. 97–124.
- James R. Curran and Marc Moens (2002), Scaling Context Space, *Proceedings of the 40th annual meeting of the Association for Computational Linguistics (ACL)*, pp. 231–238.
- John Rupert Firth (1957), A synopsis of linguistic theory 1930-1955, *Studies in Linguistic Analysis*, Oxford Philological Society, pp. 1–32.
- Joseph P. Levy and John A. Bullinaria (2001), Learning Lexical Properties from Word Usage Patterns: Which Context Words Should be Used?, in R.M. French and J.P. Sougne, editor, *Connectionist Models of Learning, Development and Evolution: Proceedings of the Sixth Neural Computation and Psychology Workshop*, Springer, pp. 273–282.
- Kenneth Ward Church and Patrick Hanks (1990), Word Association Norms, Mutual Information, and Lexicography, *Computational Linguistics* **16** (1), pp. 22–29.
- Kevin Lund and Curt Burgess (1996), Producing high-dimensional semantic spaces from lexical co-occurrence, *Behavior Research Methods, Instruments, and Computers* **28** (2), pp. 203–208.

- Klaus Rothenhäusler and Hinrich Schütze (2007), Part of Speech Filtered Word Spaces, in Marco Baroni and Alessandro Lenci and Magnus Sahlgren, editor, *Proceedings of the 2007 Workshop on Contextual Information in Semantic Space Models: Beyond Words and Documents*, pp. 25–32.
- Kris Heylen and Yves Peirsman and Dirk Geeraerts and Dirk Speelman (2008), Modelling Word Similarity: An Evaluation of Automatic Synonymy Extraction Algorithms, *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Lillian Lee (1999), Measures of Distributional Similarity, *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pp. 25–32.
- Lonneke van der Plas and Gosse Bouma (2005a), Automatic Acquisition of Lexico-Semantic Knowledge for QA, *Proceedings of the IJCNLP workshop on Ontologies and Lexical Resources*, pp. 76–84.
- Lonneke van der Plas and Gosse Bouma (2005b), Syntactic Contexts for Finding Semantically Related Words, *Proceedings of CLIN 04*, pp. 173–186.
- Lonneke van der Plas and Jörg Tiedemann (2006), Finding Synonyms Using Automatic Word Alignment and Measures of Distributional Similarity, *Proceedings of the COLING/ACL*, pp. 866–873.
- Magnus Sahlgren (2006), *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*, PhD thesis, Institutionen för lingvistik, Stockholms Universitet.
- Padó, Sebastian and Lapata, Mirella (2007), Dependency-Based Construction of Semantic Space Models, *Computational Linguistics* **33** (2), pp. 161–199.
- Roeland J.F. Ordelman (2002), Twente Nieuws Corpus (TwNC), *Technical report*, Parlevink Language Technology Group. University of Twente.
- Thomas K. Landauer and Susan T. Dumais (1997), A Solution to Plato’s Problem: The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge, *Psychological Review* **104** (2), pp. 211–240.
- Van de Cruys, Tim (2005), Semantic Clustering in Dutch: Automatically inducing semantic classes from large-scale corpora, in Khalil Simaan and Maarten de Rijke and Remko Scha and Rob van Son, editor, *Proceedings of CLIN 05*, pp. 17–32.
- Van de Cruys, Tim (2006), The Application of Singular Value Decomposition to Dutch Noun-Adjective Matrices, in Piet Mertens and Cedrick Fairon and Anne Dister and Patrick Watrin, editor, *TALN06. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles*.
- Van de Cruys, Tim (2007), Exploring Three Way Contexts for Word Sense Discrimination, in Marco Baroni and Alessandro Lenci and Magnus Sahlgren, editor, *Proceedings of the 2007 Workshop on Contextual Information in Semantic Space Models: Beyond Words and Documents*, pp. 33–41.
- Van de Cruys, Tim and Villada Moirón, Begoña (2007), Lexico-Semantic Multiword Expression Extraction, in Peter Dirix and Ineke Schuurman and Vin-

- cent Vandeghinste and Frank Van Eynde, editor, *Proceedings of the 17th Meeting of Computational Linguistics in the Netherlands*, pp. 175–190.
- Vossen, Piek, editor (1998), *EuroWordNet: a multilingual database with lexical semantic networks for European Languages*, Kluwer, Dordrecht.
- Warren Weaver (1955), Translation, in W.N. Locke and D.A. Booth, editor, *Machine Translation of Languages*, MIT press, pp. 15–23.
- Will Lowe and Scott McDonald (2000), The direct route: Mediated priming in semantic space, *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, pp. 675–680.
- Yves Peirsman and Kris Heylen and Dirk Speelman (2007), Finding semantically related words in Dutch. Co-occurrences versus syntactic contexts., in Marco Baroni and Alessandro Lenci and Magnus Sahlgren, editor, *Proceedings of the 2007 Workshop on Contextual Information in Semantic Space Models: Beyond Words and Documents*, pp. 9–16.
- Yves Peirsman and Kris Heylen and Dirk Speelman (2008), Putting things in order. First and second order contexts models for the calculation of semantic similarity, *Actes des 9ièmes Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008)*, pp. 907–916.
- Zelig Sabbettai Harris (1954), Distributional structure, *Word* **10** (21), pp. 146–162.
- Zhibiao Wu and Martha Palmer (1994), Verb semantics and lexical selection, *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pp. 133–138.