Modelling Word Similarity.
An Evaluation of Automatic Synonymy Extraction Algorithms.

Kris Heylen*, Yves Peirsman*†, Dirk Geeraerts*, Dirk Speelman*

*QLVL - University of Leuven, †Research Foundation - Flanders
Blijde-Inkomststraat 21, PO Box 3038, 3000 Leuven (Belgium)
{kris.heylen, yves.peirsman, dirk.geeraerts, dirk.speelman}@arts.kuleuven.be

Abstract

Vector-based models of lexical semantics retrieve semantically related words automatically from large corpora by exploiting the property that words with a similar meaning tend to occur in similar contexts. Despite their increasing popularity, it is unclear which kind of semantic similarity they actually capture and for which kind of words. In this paper, we use three vector-based models to retrieve semantically related words for a set of Dutch nouns and we analyse whether three linguistic properties of the nouns influence the results. In particular, we compare results from a dependency-based model with those from a 1st and 2nd order bag-of-words model and we examine the effect of the nouns' frequency, semantic speficity and semantic class. We find that all three models find more synonyms for high-frequency nouns and those belonging to abstract semantic classses. Semantic specificty does not have a clear influence.

1. Introduction

One of the major challenges in the development of computational language resources is the modelling of natural language semantics. The retrieval of semantically similar words is essential for the automatic extraction or extension of thesauri, but also for query expansion modules in information retrieval applications. Within statistical NLP, the task of finding meaning related words has been succesfully approached with so-called vector-based models of lexical semantics. 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 cull statistics about the co-occurence of a word with a large number of context features from large corpora and put these into a vector. The semantic similarity between two words is then quantified as the similarity between their respective context vectors.

Although all vector 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*. The first models were so-called document-based models with Latent Semantic Analysis (Landauer and Dumais, 1997) as best known example. They used whole documents or paragraphs as contexts so that words that often co-occurred in documents appeared as semantically similar. For extracting tight semantic relations like synonyms, word-based models have become more popular in recent years and they will be the focus of this study. Word-based models restrict contexts to the words in near proximity to the target words for which they try to find related words. For these models, two words will be similar if they often co-occur with the same context words, but unlike document-based models, they do not expect target words to co-occur regularly with each other. Even within the broad class of word-based models, different definitions of context are used. One option is simply to look at the context words that appear in a pre-defined window around the target word. The context features are in this case the so-called first order co-occurrences of the target (Levy and Bullinaria, 2001). Because these models do not differentiate between the context words within the context window, they are often called bag-of-words models. A second order bag-of-words-model (Schütze, 1998) makes use of the second order co-occurrences, i.e. the context words of the first order co-occurrences, and by doing so should allow to generalise over meaning related context words and avoid data sparseness. Finally, a third option (Lin, 1998) (Padó and Lapata, 2007) only takes those context words into account that stand in a specific syntactic dependency relation to the target word. In this case, only context words like verbs governing a target noun in its subject function, or adjectives modifying the target are counted as context features.

It is clear that these different types of context features are likely to capture different kind of semantic information. However, so far, little is known about the influence of the context definition on the semantic information present in the vector spaces. While most researchers choose one specific vector space model and apply it to their task, "comparisons between the ... models have been few and far between in the literature" (Padó and Lapata, 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 (Peirsman et al., 2007) and (Peirsman et al., 2008), we evaluated the overall performance of the three types of word-based models mentioned above and we analysed the semantic relations that they retrieved. However, in these studies we did not take into account that models might behave differently for different classes of target words. Yet, both for thesaurus extraction and query expansion it is vital to know whether the algorithms work equally well for all types of target words. In this study we therefore take a closer look at three properties of the target words, viz. frequency, semantic specificity and semantic class, and we analyse whether they influence the performance of the three models.

Section 2. discusses the data and parameter settings for the three vector-space models. Section 3. first presents the evaluation measures we used and than dicusses, consecutively, the influence of frequency, semantic specificity and semantic class on the peformance of the three models. In section 4. we wrap up with conclusions and some suggestions for future research.

2. Set-up

Our experiments compare three types of word-based vector space models: a first-order and second-order bag-of-words model and a dependency-based model. The data for these models consists of the 300 million word Twente Nieuws Corpus of Dutch lemmatised and parsed newspaper text [1]. We extracted from the lemmatised corpus the 10,000 most frequent nouns and their context vectors and then calculated for each target noun the single most similar noun among the remaining 9,999 possibilities, which we will designate as the target's *nearest neighbour*. The parameters of the three models were set as follows:

Dimensionality Only the 4,000 most frequent features were used.

Context Window For the first and second order co-occurrence approaches, a context window of three words on either side of the target word was used.

Weighting scheme Context vectors contained the point-wise mutual information between the feature and the target, rather than raw frequency.

Frequency cut-off With second-order and first-order co-occurrences, only those features that occurred at least five times together with the target word were counted. For the syntactic model no such a cut-off was used to avoid data sparseness.

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

Stop list For the 1st and 2nd order co-occurrence models, semantically empty words were not feauture candidates. For the syntactic model, no such stop list had to be used because of the imposed dependency relation.

Syntactic relations For the syntactic model, 8 syntactic relations were taken into account [2]:

1. subject of verb v
2. direct object of verb v
3. prepositional complement of verb v introduced by preposition p
4. head of an adverbial PP of verb v introduced by preposition p
5. modified by adjective a
6. postmodified by a PP with head n, introduced by preposition p
7. modified by an apposition with head n
8. coordinated with head n

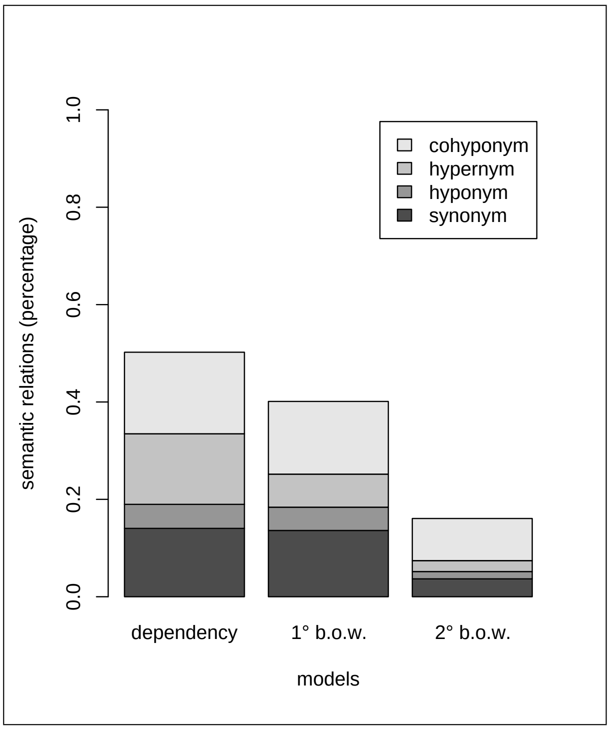

Figure 1: General performance of the three models

3. Results and discussion

For all three models, we performed two types of evaluation: overall performance and specific semantic relations retrieved. As a gold standard, the Dutch part of EuroWordNet (Vossen, 1998) was used. Like its English counterpart, EuroWordNet is a lexical database structured as a hierarchical network of concepts. The Dutch section of EuroWordNet contains 44K synsets, which is a fair bit sparser than English WordNet (117K synsets) and which, as we will see, will influence the results. Our evaluations are based on the target-neighbour pairs retrieved by the three models and of course only pairs with both the target and nearest neighbour present in EuroWordNet can be assessed. For the dependency-based model, 7479 pairs out of the potential 10,000 were present in the database and could be retained for further analysis. For the first and second order bag-of-words models this figure was 6776 and 6727 pairs respectively.

Following previous studies (van der Plas and Bouma, 2005), (Van de Cruys, 2006) and (Peirsman et al., 2007), we analyse the overall performance of the three 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 (Wu and Palmer, 1994) which has become somewhat of a standard for measuring simililarity in lexical taxonomies. Equation 1 shows that the score divides twice the depth of the lowest

[1] Parsing was done at the University of Groningen with the Alpino dependency parser for Dutch (van Noord, 2006)

[2] Each specific instantiation of the variables v, p, a, or n led to a new context feature

shared hypernym (h_l) of two words by the sum of these two words' pathlength to the top of the hierarchy. The score ranges from 0 (no similarity) to 1 (perfect similarity) [3].

$$s_{WP}(w_1, w_2) = \frac{2 \times depth(h_l)}{D_{PL}(w_1, h_l) + D_{PL}(w_2, h_l) + 2 \times depth(h_l)} \quad (1)$$

To evaluate the specific semantic relations that are retrieved by the three models, we checked which semantic relation, if any, a nearest neighbour entertains 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 very strict definition of semantic relatedness by only allowing hyponyms and hypernyms that are one step and cohyponyms that are two steps removed in the hierarchy[4].

Our previous studies (Peirsman et al., 2007) and (Peirsman et al., 2008) showed that the dependency-based model generally outperformed the other models, both in terms of overall performance and in terms of the relative frquency of specific semantic relations retrieved. The average Wu & Palmer score for the dependency-based model was 0.62, compared to 0.52 and 0.31 for the first and second order bag-of-words models. Figure 1 shows that the dependency-based model finds a tightyly related neighbour for 50% of the target words and a true synonym for 14%. The first bag-of-words model does a bit less well with 40% and 13% respectively and the second order bag-of-words model performs rightout poorly with 16% related neighbours and only 3% synonyms retrieved. In this study, we will now investigate whether this performance is influenced by the frequency, semantic specificity and semantic class of the target nouns, and whether the influence is the same for all three models.

[3]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 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

[4]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.

	Frequency	Specificity	Sem. class
depend.	0.10	0.24	0.25
1^o b.o.w.	0.18	0.19	0.18
2^o b.o.w.	0.22	0.22	0.23

Table 1: Correlation between average Wu & Palmer similarity score and log frequency, semantic specificity and semantic class (regression R).

3.1. Influence of frequency

For each of the target words we counted their frequency in our 300 million word corpus. To measure the influence of a target's frequency on overall performance, we calculated[5] the correlation between the natural logarithm of a target's frequency and the average Wu & Palmer score of the target-neighbour pairs. The first column in Table 1 shows a modest but significant positive correlation for all three models ($t_{dep} = 8.6$, $t_{1bow} = 15.3$, $t_{2bow} = 18.1$, all three with $p < .01$), meaning that the nearest neighbours of high-frequency nouns are on average more semantically similar to their targets than nearest neighbours of low-frequency nouns. The 95% confidence intervals show that the correlation is significantly stronger for the first $(0.16 - 0.21)$ and second $(0.19 - 0.24)$ order bag-of-words models than for the dependency based one $(0.8 - 0.12)$. Apparently, the overall best performing dependency-based model is less sensitive to frequency differences between target nouns.

Looking at the specific semantic relations retrieved by the three models, we see the same influence of frequency re-occurring. Figure 2 shows these relations for 5 log frequency bands. For all three models, the nearest neighbours of high-frequency nouns are more often tightly related to their target than those of low frequency nouns. If we look at the dependency-based model, we can see that the percentage of nearest neighbours entertaining one of the four semantic relations to their target steadily increases with frequency from 46% in the lowest band (log-freq. 6 to 7 = freq. 403 to 1096) to 66% in the highest (log-freq. 9 to 13 = freq. 8103 to 442,413). Focussing on synonyms, however, we see that the increase is not completely monotonous. The percentage increases from 13% in the lowest band to 19% in the fourth band, but then drops again slightly for the highest frequency words. However, this small drop in synonyms is accompanied by a big leap in the number of hyponyms, which represent the second most similar semantic relation. A similar observation can be made for both bag-of-words models. Note also that the first order bag-of-words models is slightly better at finding synonyms for the mid-frequency bands 7 to 8 (15%) and 8 to 9 (15.5%) than the dependency based model with respectively 14% and 14.7%. To assess whether log frequency has indeed a significant effect on the specific semantic relations retrieved, we use a multinomial logistic regression that models the influence of log frequency on the odds of encountering a synonym, hyponym, hypernym or cohyponym rather than no relation. We fit one regression model for each of the three

[5]All statistical analyses were carried out with R

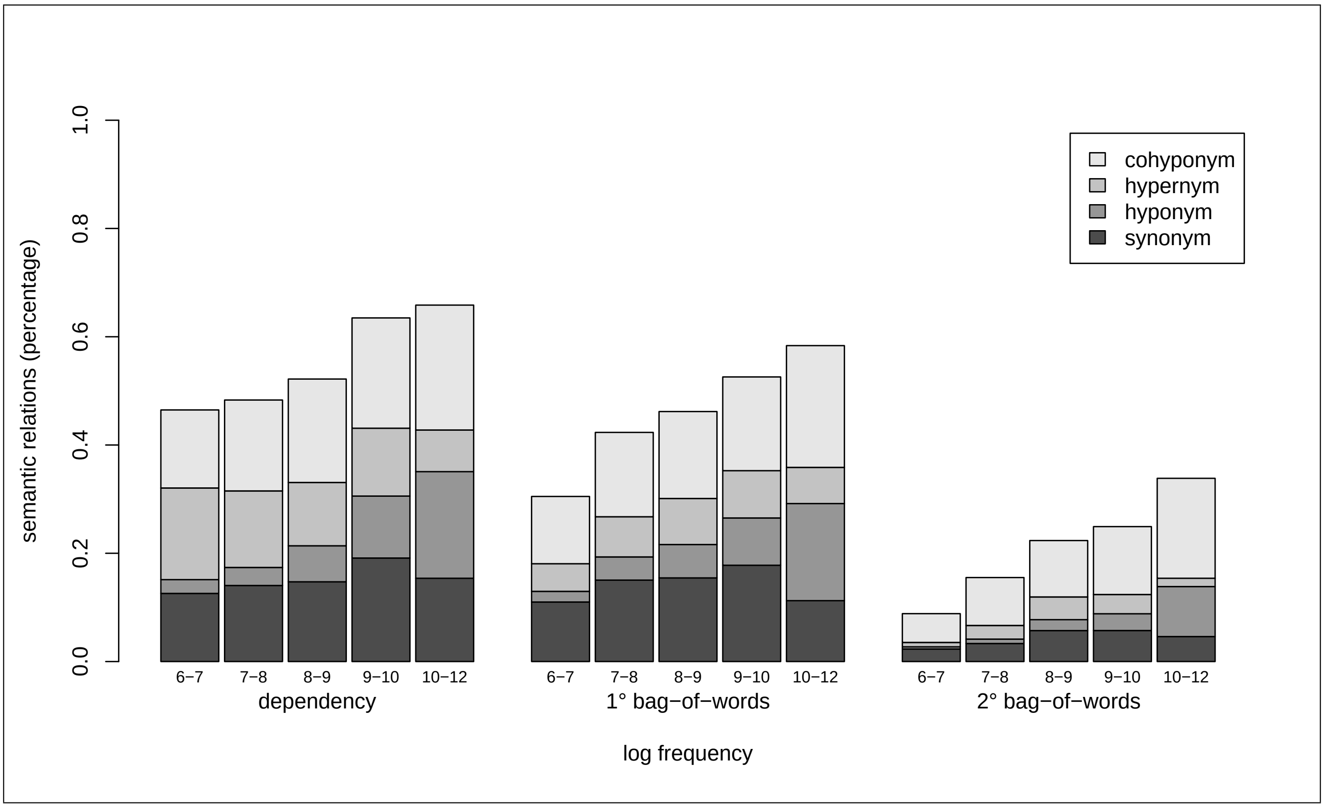

Figure 2: Effect of Frequency

systems. All three regression models are highly significant ($LR_{dep} = 321$, $LR_{1bow} = 297$, $LR_{2bow} = 282$, all three with $p < .01$) and show significant ($p < .01$) higher odds in favour of synonyms ($+22\%_{dep}$, $+26\%_{1bow}$, $+40\%_{2bow}$), hyponyms ($+90\%_{dep}$, $+91\%_{1bow}$, $+135\%_{2bow}$) and cohyponyms ($+26\%_{dep}$, $+33\%_{1bow}$, $+45\%_{2bow}$) with each 1 unit increase in log frequency. For hypernyms, the effect was not significant for the dependency-based model, but the odds did increase significanly for the first ($+31\%$) and second ($+47\%$) order bag-of-words model. Comparison of the 95% confidence intervals learns that the impact of frequency on finding synonyms was significantly stronger for the second order bag-of-words model than for the dependency based model. As we already saw for overall performance, the best performing model, viz. the dependency-based model, seems less sensitive to the frequency of the target words.

3.2. Influence of semantic specificity

Words can differ in terms of their semantic specificity: A word like *craft* is a much more general term than say *hydroplane* or *space shuttle*. To assess whether semantic specificity influences the overal performance of our three models, we calculated the correlation between the Wu & Palmer score for target-neighbour pairs and the target's depth in the conceptual hierarchy of EuroWordNet. Since this hierarchy goes down from general concepts to more specific ones, the depth in the hierarchy is an indicator of semantic specificity[6]. In EuroWordNet this depth varies from 1 to 13. The second column in Table 1 shows a moderate positive, but significant correlation between the Wu & Palmer score and the depth of a target word for all three models ($t_{dep} = 21.4$, $t_{1bow} = 16.3$, $t_{2bow} = 18.2$; $p < .05$). At first sight, it is easier to find semantically related words for more specific target words than for more general ones. However, when we look at the specific semantic relations retreived, a more complicated picture emerges. Figure 3 shows the percentage of relations found at different depths for the three models and this time, there is no clear tendency to discerned. The total percentage of nearest neighbours displaying one of the four relations does not show a linear trend. Rather, the percentage decreases at first, then goes up again for depths 7 to 9 and ends with a small drop for the most specific nouns deep down in the hierarchy. The same holds for the percentage of synonyms and hyponyms found: the most general nouns seem to have relatively more synonyms and hyponyms among their nearest neighbours and so do the more specific target words, especially at depths 7 to 9. Looking at the data, nouns at depth 7 to 9 seem to refer mainly to persons like *teacher*, *villain* or *opponent*. However, it is not immediately clear why it should be easier to find synonyms for person designations. The respective multinomial logistic regression models show effects that are only border significant and very small. In other words, the odds of finding a specific semantic relation do not seem to be influenced by semantic specificity, at least not in a linear fashion and measured through hierarchy depth. Possibly, the presence of base level concepts in the structure of the lexicon might

[6] If a target word occurred at different depths due to polysemy, we took the minimal depth as a conservative estimate for semantic specificity.

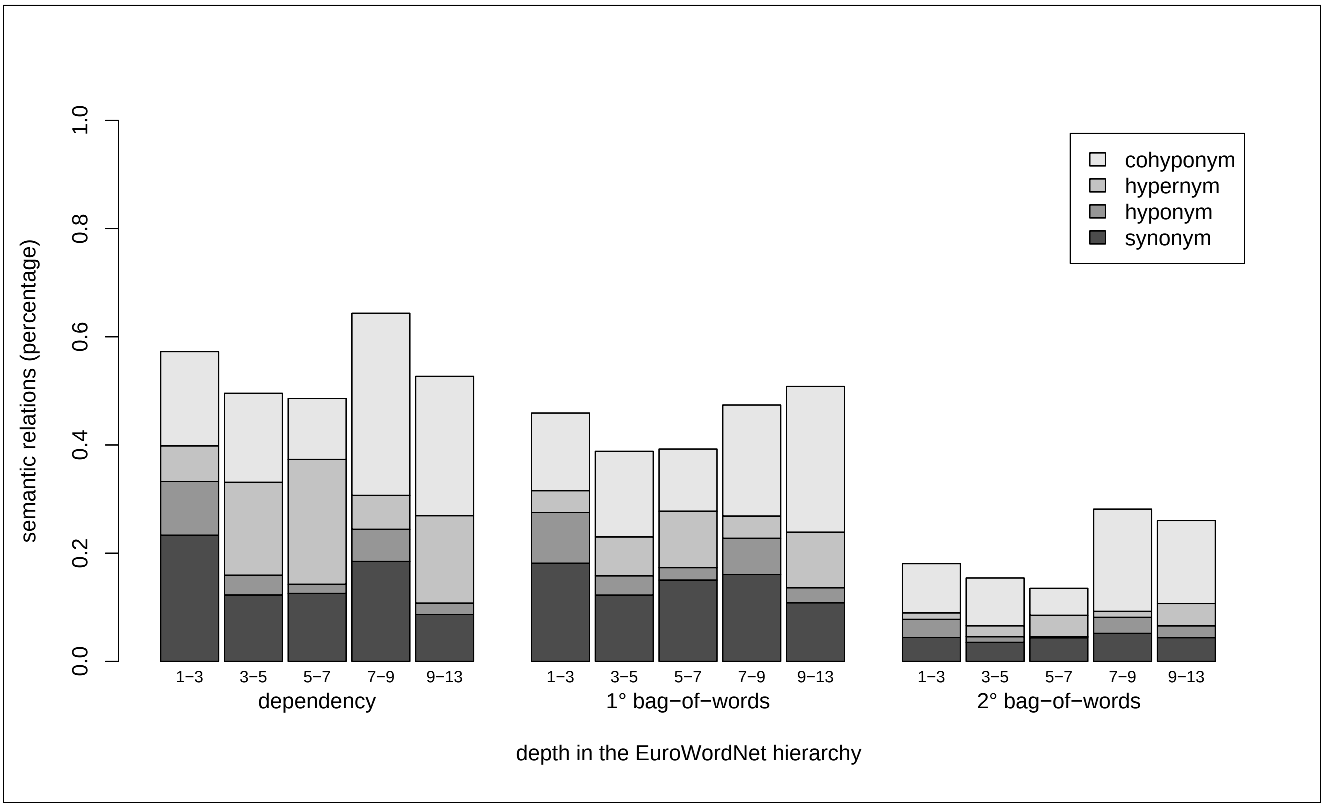

Figure 3: Effect of Semantic Specificity

explain this non-linearity. These will probably occur at intermediate levels of specifity, could have a higher number of synonyms and hyponyms due to their special status and migth explain the peaks at depths 7 to 9. However, this hypothesis certainly needs further investigation.

Given the obvious lack of a linear relation it is actually rather remarkable that we still find a fairly strong correlation between the Wu & Palmer score and hierarchy depth, a correlation which is even slightly stronger than the one for log frequency, which did show a clear linear effect. This might be an artefact of the normalization for hierarchy depth in the the Wu & Palmer score as can be seen in the denominater of formula 1. Wu & Palmer's reason for this normalization was the intuition that words close to each other deep down in the hierarchy are more related than those higher up. For example, *hydroplane* and *jetplane* are more related than *craft* and *machine* although their relative distance in the hierarchy is similar. As a consequence, the Wu & Palmer score partially reflects how deep a target-neighbour pair is situated in the hierarchy and this might explain the fairly high correlation. Indeed, if we use inverse path length as a similarity measure, which does not normalise for hierarchy depth, the correlation completely disappears: it is then a non-significant -0.03, -0.01 and -0.01 for the dependency-based, 1st order and 2nd order bag-of-word model respectively. We can therefore conclude that semantic specificity does not have a clear influence on the performance of the three models.

3.3. Influence of semantic class

Nouns can refer to different types of concepts: objects, events, properties, locations etc. To assess whether our models behave differently for target words from distinct semantic classes, we annotated each target word in our sample automatically with its highest but one ancestor in the EuroWordnet hierarchy[7]. This resulted in 69 different semantic classes, of which we only took the 10 most frequent into account . These were: *object*, *event*, *property*, *situation*, *group*, *part*, *utterance*, *substance*, *location* and *thought*. For target words belonging to more than one semantic class due to polysemy, only the most frequently occurring class was taken into account. To test whether the overall performance of the three models was different for different semantic classes, we performed a one-way Analysis of Variance (anova) of the average Wu & Palmer score over the 10 classes. The third column in Table 1 shows the R-measure of explained variance for these three anova models[8]. For all three models, the anova's were highly significant ($F_{dep} = 51.3, F_{1bow} = 24.8, F_{2bow} = 41.7$, all three $p < .01$) so that semantic class membership can be said to account for a substantial share of the variation in Wu & Palmer score. The anova models also showed that the average Wu & Palmer score was significantly higher for *objects* than for all other categories ($p < .01$). For example, the average Wu & Palmer score for *objects* is 25% higher than for *thoughts*.

Let us then look at the specific semantic relations. Figure 4 shows the semantic relations retrieved for five of the

[7] The highest ancester is always *thing*.

[8] For simple linear regression, this R-measure corresponds to the pearson correlation so that the R-measure here can be said to be on the same scale as the correlations found for frequency and semantic specificity

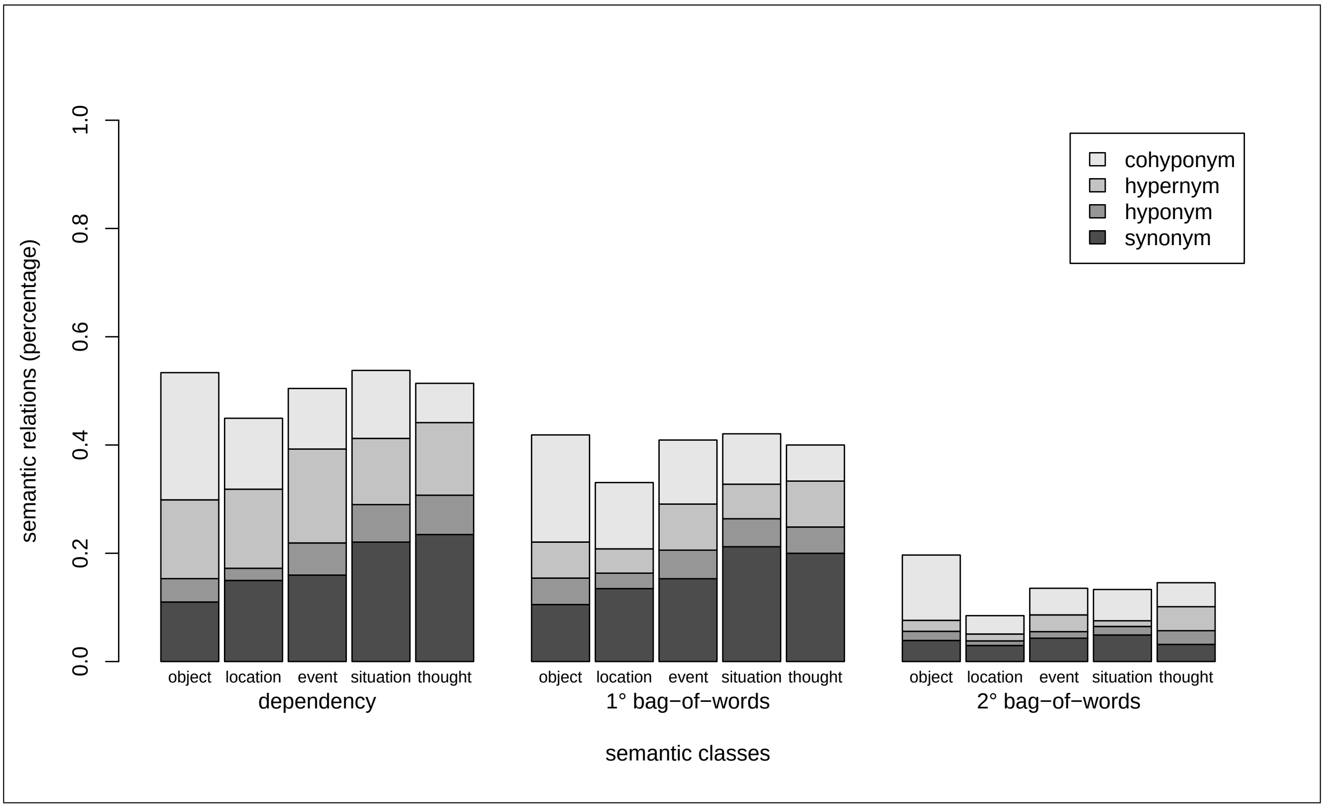

Figure 4: Effect of Semantic Class.

semantic classes. Whereas the total percentage of related neighbours does not differ markedly between the semantic classes, the relative frequency of synonyms, cohyponyms and, to a lesser extent, hyponyms does. If we zoom in on the dependency-based model and go from left to right in the figure, we see the percentage of synonyms increase steadily from 11% to 23% and the share of hyponyms from 4% to 7%, whereas the percentage of cohyponyms drops from 23% to 7%. A multinomial logistic regression (model significance: $LLR = 178$, $p < .01$) confirms the significance of these differences between the semantic classes. For example, the odds of finding synonyms for *thoughts* is 104% higher than those for *objects* whereas the odds of finding cohyponyms for *thoughts* is 70% lower than those for *objects*. The odds of finding hypernyms did not significantly differ from one semantic class to another. Am ultinomial logistic regression for the first-order bag-of-words model shows effects of semantic class that are similar, but somewhat weaker, although the difference with the dependency-model is not significant. Finally, the multinomial logistic regression model for the second-order bag-of-word model shows only a significant decrease in the odds of finding cohyponyms going from *objects* to *thoughts* and no effect for the other semantic relations. Comparison of the 95% confidence intervals shows that the effect differences between the three models are not significant. In other words, semantic class membership has a comparable effect on all three models.

Looking more closely at the share of synonyms and hyponyms when going from *objects* over *locations*, *events* and *situations* to *thoughts*, there seems to be an upward trend that follows a cline from concrete to abstract semantic classes. In other words, relatively more synonyms are found for abstract semantic classes like *situations* and *thoughts* than for concrete ones like *objects* and *locations*.

As with the analysis of semantic specificity in the previous section, we see a remarkabe discrepancy between the evaluation of overall performance and the evaluation of the specific semantic relations retrieved. The anova models showed that the average Wu & Palmer scores for *objects* is higher than for the 9 other semantic classes, e.g. 25% higher than for *thoughts*. Yet the *thought* class has relatively more synonyms and hyponyms for roughly the same total amount of related neighbours retrieved. One would expect the overall similarity score to go up when relatively more tighter semantic relations like synonymy and hyponymy are found, not down! Could this again be an artefact of the Wu & Palmer measure's bias for hierarchy depth? It turns out that the average depth of *object* nouns is 6.76 whereas the average for nouns from the *thought* class is only 3.4. This might not have so much to do with a real difference in semantic specificity between between *object* and *thought* nouns as with a difference in granularity of the EuroWordNet hierarchy for these two semantic classes. The subdivision in subconcepts is just much more fine-grained for *objects* than for *thoughts* and this makes the Wu & Palmer score with its depth normalization unsuitable for comparing semantic similarities between these two classes. If we do an Analysis of Variance using non-normalised inverse path length, the difference in average semantic similarity between the semantic classes is as expected with the nearest neighbours for *thoughts* having a higher average similarity to their target than those for *objects*. This shows again that caution is needed when interpreting results based on similarity measurements from a conceptual hierarchy with limited coverage like Dutch EuroWordNet.

4. Conclusions

Vector-space models have become the de facto standard for dealing with lexical semantics in many NLP applications. They are unsupervised, easily scalable and have proven to be succesful for many tasks involving word meaning. However, still fairly little is known about their exact linguistic properties and exactly this is vital for choosing the type of model that is best suited for a particular task. In this paper we have compared three models for finding semantically related nouns in Dutch and we have analysed how three properties of the target nouns influence the performance of the models. The three models we have looked at used different definitions of context for generating context vectors. In particular, we compared a model using syntactic dependencies as context features with a 1st and 2nd order bag-of-words model. For these three models, we analysed the influence of the frequency, the semantic specificity and the semantic class of the target nouns on the semantically most related words they returned.

Our analysis clearly showed that all three models returned significantly more semantically related words for high-frequency nouns and especially more synonyms and hyponyms. The generally best performing model, viz. the dependency-based model, was somewhat less sensitive to this effect of frequency. Although the dependency-based model remained the best performing model overall, the first-order bag-of-words model was slightly better at retrieving synonyms for mid-frequency nouns.

The effect of semantic specificity on the retrieval of semantically related words was not straight forward. Although greater depth of a target word in the EuroWordNet hierarchy seemed to result in a higher semantic relatedness of the words retrieved, an analysis of the specific semantic relations could not reveal a clear trend for any of the three models. It appeared that a bias for semantically specific words in the Wu & Palmer similarity measure made the assessment of overall performance unreliable.

Although semantic class membership did not have a marked influence on the total amount of semantically related words returned by the three models, there was a clear difference in the specific semantic relations retrieved. The systems found significantly more synonyms and hyponyms for nouns referring to *thoughts* or *situations* than for those belonging to classes like *object* or *location*. In general, there seemed to be a tendency that tighter semantic relations were found for nouns belonging to abstract semantic classes than for those belonging to concrete semantic classes.

We hope to have shown that there are important differences in the performance of vector-space models relative to the properties of the target words for which they have to find semantically related words. For applications like query expansion or automatic thesaurus extraction it should be taken into account that vector-space models do not work equally well for all types of words. We therefore recommend that evaluations of these applications should also take the possible effect of word properties into consideration.

5. References

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):211–240.

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, editors, *Connectionist Models of Learning, Development and Evolution: Proceedings of the Sixth Neural Computation and Psychology Workshop*, pages 273–282. Springer.

Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the 17th international conference on Computational linguistics*, pages 768–774.

Sebastian Padó and Mirella Lapata. 2007. Dependency-based construction of semantic space models. *Computational Linguistics*, 33(2):161–199.

Yves Peirsman, Kris Heylen, and Dirk Speelman. 2007. Finding semantically related words in dutch. co-occurrences versus syntactic contexts. In Marco Baroni, Alessandro Lenci, and Magnus Sahlgren, editors, *Proceedings of the 2007 Workshop on Contextual Information in Semantic Space Models: Beyond Words and Documents*, pages 9–16.

Yves Peirsman, Kris Heylen, and Dirk Speelman. 2008. Putting things in order. first and second order contexts models for the calculation of semantic similarity. In *Actes des 9ièmes Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008)*, pages 907–916.

Hinrich Schütze. 1998. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–124.

Tim Van de Cruys. 2006. The application of singular value decomposition to dutch noun-adjective matrices. In Piet Mertens, Cedrick Fairon, Anne Dister, and Patrick Watrin, editors, *TALN06. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles*.

Lonneke van der Plas and Gosse Bouma. 2005. Syntactic contexts for finding semantically related words. In *Proceedings of CLIN 04*.

Gertjan van Noord. 2006. At last parsing is now operational. In Piet Mertens, Cedrick Fairon, Anne Dister, and Patrick Watrin, editors, *TALN06. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles*.

Piek Vossen, editor. 1998. *EuroWordNet: a multilingual database with lexical semantic networks for European Languages*. Kluwer, Dordrecht.

Zhibiao Wu and Martha Palmer. 1994. Verb semantics and lexical selection. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 133–138.