

Predicting Strong Associations on the Basis of Corpus Data

Yves Peirsman

Research Foundation – Flanders &
QLVL, University of Leuven
Leuven, Belgium
yves.peirsman@arts.kuleuven.be

Dirk Geeraerts

QLVL, University of Leuven
Leuven, Belgium
dirk.geeraerts@arts.kuleuven.be

Abstract

Current approaches to the prediction of associations rely on just one type of information, generally taking the form of either word space models or collocation measures. At the moment, it is an open question how these approaches compare to one another. In this paper, we will investigate the performance of these two types of models and that of a new approach based on compounding. The best single predictor is the log-likelihood ratio, followed closely by the document-based word space model. We will show, however, that an ensemble method that combines these two best approaches with the compounding algorithm achieves an increase in performance of almost 30% over the current state of the art.

1 Introduction

Associations are words that immediately come to mind when people hear or read a given cue word. For instance, a word like *pepper* calls up *salt*, and *wave* calls up *sea*. Aitchinson (2003) and Schulte im Walde and Melinger (2005) show that such associations can be motivated by a number of factors, from semantic similarity to collocation. Current computational models of association, however, tend to focus on one of these, by using either collocation measures (Michelbacher et al., 2007) or word space models (Sahlgren, 2006; Peirsman et al., 2008). To this day, two general problems remain. First, the literature lacks a comprehensive comparison between these general types of models. Second, we are still looking for an approach that combines several sources of information, so as to correctly predict a larger variety of associations.

Most computational models of semantic relations aim to model semantic similarity in particu-

lar (Landauer and Dumais, 1997; Lin, 1998; Padó and Lapata, 2007). In Natural Language Processing, these models have applications in fields like query expansion, thesaurus extraction, information retrieval, etc. Similarly, in Cognitive Science, such models have helped explain neural activation (Mitchell et al., 2008), sentence and discourse comprehension (Burgess et al., 1998; Foltz, 1996; Landauer and Dumais, 1997) and priming patterns (Lowe and McDonald, 2000), to name just a few examples. However, there are a number of applications and research fields that will surely benefit from models that target the more general phenomenon of association. For instance, automatically predicted associations may prove useful in models of *information scent*, which seek to explain the paths that users follow in their search for relevant information on the web (Chi et al., 2001). After all, if the visitor of a web shop clicks on *music* to find the prices of iPods, this behaviour is motivated by an associative relation different from similarity. Other possible applications lie in the field of models of text coherence (Landauer and Dumais, 1997) and automated essay grading (Kakkonen et al., 2005). In addition, all research in Cognitive Science that we have referred to above could benefit from computational models of association in order to study the effects of association in comparison to those of similarity.

Our article is structured as follows. In section 2, we will discuss the phenomenon of association and introduce the variety of relations that it is motivated by. Parallel to these relations, section 3 presents the three basic types of approaches that we use to predict strong associations. Section 4 will first compare the results of these three approaches, for a total of 43 models. Section 5 will then show how these results can be improved by the combination of several models in an ensemble. Finally, section 6 wraps up with conclusions and an outlook for future research.

cue	association
amfibie ('amphibian')	kikker ('frog')
peper ('pepper')	zout ('salt')
roodborstje ('robin')	vogel ('bird')
granaat ('grenade')	oorlog ('war')
helikopter ('helicopter')	vliegen ('to fly')
werk ('job')	geld ('money')
acteur ('actor')	film ('film')
cello ('cello')	muziek ('music')
kruk ('stool')	bar ('bar')

Table 1: Examples of cues and their strongest association.

2 Associations

There are several reasons why a word may be associated to its cue. According to Aitchinson (2003), the four major types of associations are, in order of frequency, co-ordination (co-hyponyms like *pepper* and *salt*), collocation (like *salt* and *water*), superordination (*insect* as a hypernym of *butterfly*) and synonymy (like *starved* and *hungry*). As a result, a computational model that is able to predict associations accurately has to deal with a wide range of semantic relations. Past systems, however, generally use only one type of information (Wettler et al., 2005; Sahlgren, 2006; Michelbacher et al., 2007; Peirsman et al., 2008; Wandmacher et al., 2008), which suggests that they are relatively restricted in the number of associations they will find.

In this article, we will focus on a set of Dutch cue words and their single strongest association, collected from a large psycholinguistic experiment. Table 1 gives a few examples of such cue–association pairs. It illustrates the different types of linguistic phenomena that an association may be motivated by. The first three word pairs are based on similarity. In this case, strong associations can be hyponyms (as in *amphibian–frog*), co-hyponyms (as in *pepper–salt*) or hypernyms of their cue (as in *robin–bird*). The next three pairs represent semantic links where no relation of similarity plays a role. Instead, the associations seem to be motivated by a topical relation to their cue, which is possibly reflected by their frequent co-occurrence in a corpus. The final three word pairs suggest that morphological factors might play a role, too. Often, a cue and its association form the building blocks of a compound, and it is possible that one part of a compound calls up the other.

The examples show that the process of compounding can go in either direction: the compound may consist of cue plus association (as in *cellomuziek* ‘cello music’), or of association plus cue (as in *filmacteur* ‘film actor’). While it is not clear if it is the compounds themselves that motivate the association, or whether it is just the topical relation between their two parts, they might still be able to help identify strong associations.

3 Approaches

Motivated by the three types of cue–association pairs that we identified in Table 1, we study three sources of information (two types of distributional information, and one type of morphological information) that may provide corpus-based evidence for strong associatedness: collocation measures, word space models and compounding.

3.1 Collocation measures

Probably the most straightforward way to predict strong associations is to assume that a cue and its strong association often co-occur in text. As a result, we can use collocation measures like point-wise mutual information (Church and Hanks, 1989) or the log-likelihood ratio (Dunning, 1993) to predict the strong association for a given cue. **Point-wise mutual information** (PMI) tells us if two words w_1 and w_2 occur together more or less often than expected on the basis of their individual frequencies and the independence assumption:

$$PMI(w_1, w_2) = \log_2 \frac{P(w_1, w_2)}{P(w_1) * P(w_2)}$$

The **log-likelihood ratio** compares the likelihoods L of the independence hypothesis (i.e., $p = P(w_2|w_1) = P(w_2|\neg w_1)$) and the dependence hypothesis (i.e., $p_1 = P(w_2|w_1) \neq P(w_2|\neg w_1) = p_2$), under the assumption that the words in a text are binomially distributed:

$$\log \lambda = \log \frac{L(P(w_2|w_1); p) * L(P(w_2|\neg w_1); p)}{L(P(w_2|w_1); p_1) * L(P(w_2|\neg w_1); p_2)}$$

3.2 Word Space Models

A respectable proportion (in our data about 18%) of the strong associations are motivated by semantic similarity to their cue. They can be synonyms, hyponyms, hypernyms, co-hyponyms or

antonyms. Collocation measures, however, are not specifically targeted towards the discovery of semantic similarity. Instead, they model similarity mainly as a side effect of collocation. Therefore we also investigated a large set of computational models that were specifically developed for the discovery of semantic similarity. These so-called word space models or distributional models of lexical semantics are motivated by the distributional hypothesis, which claims that semantically similar words appear in similar contexts. As a result, they model each word in terms of its contexts in a corpus, as a so-called *context vector*. Distributional similarity is then operationalized as the similarity between two such context vectors. These models will thus look for possible associations by searching words with a context vector similar to the given cue.

Crucial in the implementation of word space models is their definition of context. In the current literature, there are basically three popular approaches. **Document-based models** use some sort of textual entity as features (Landauer and Dumais, 1997; Sahlgren, 2006). Their context vectors note what documents, paragraphs, articles or similar stretches of text a target word appears in. Without dimensionality reduction, in these models two words will be distributionally similar if they often occur together in the same paragraph, for instance. This approach still bears some similarity to the collocation measures above, since it relies on the direct co-occurrence of two words in text. Second, **syntax-based** models focus on the syntactic relationships in which a word takes part (Lin, 1998). Here two words will be similar when they often appear in the same syntactic roles, like *subject of fly*. Third, **word-based models** simply use as features the words that appear in the context of the target, without considering the syntactic relations between them. Context is thus defined as the set of n words around the target (Sahlgren, 2006). Obviously, the choice of context size will again have a major influence on the behaviour of the model. Syntax-based and word-based models differ from collocation measures and document-based models in that they do not search for words that co-occur directly. Instead, they look for words that often occur together with the same context words or syntactic relations. Even though all these models were originally developed to model semantic sim-

ilarity relations, syntax-based models have been shown to favour such relations more than word-based and document-based models, which might capture more associative relationships (Sahlgren, 2006; Van der Plas, 2008).

3.3 Compounding

As we have argued before, one characteristic of cues and their strong associations is that they can sometimes be combined into a compound. Therefore we developed a third approach which discovers for every cue the words in the corpus that in combination with it lead to an existing compound. Since in Dutch compounds are generally written as one word, this is relatively easy. We attached each candidate association to the cue (both in the combination cue+association and association+cue), following a number of simple morphological rules for compounding. We then determined if any of these hypothetical compounds occurred in the corpus. The possible associations that led to an observed compound were then ranked according to the frequency of that compound.¹ Note that, for languages where compounds are often spelled as two words, like English, our approach will have to recognize multiword units to deal with this issue.

3.4 Previous research

In previous research, most attention has gone out to the first two of our models. Sahlgren (2006) tries to find associations with word space models. He argues that document-based models are better suited to the discovery of associations than word-based ones. In addition, Sahlgren (2006) as well as Peirsman et al. (2008) show that in word-based models, large context sizes are more effective than small ones. This supports Wandmacher et al.'s (2008) model of associations, which uses a context size of 75 words to the left and right of the target. However, Peirsman et al. (2008) find that word-based distributional models are clearly outperformed by simple collocation measures, particularly the log-likelihood ratio. Such collocation measures are also used by Michelbacher et al. (2007) in their classification of asymmetric associations. They show the chi-square metric to be a robust classifier of associations as either symmetric or asymmetric, while a measure based on conditional probabilities is particularly suited to model

¹ If both compounds cue+association and association+cue occurred in the corpus, their frequencies were summed.

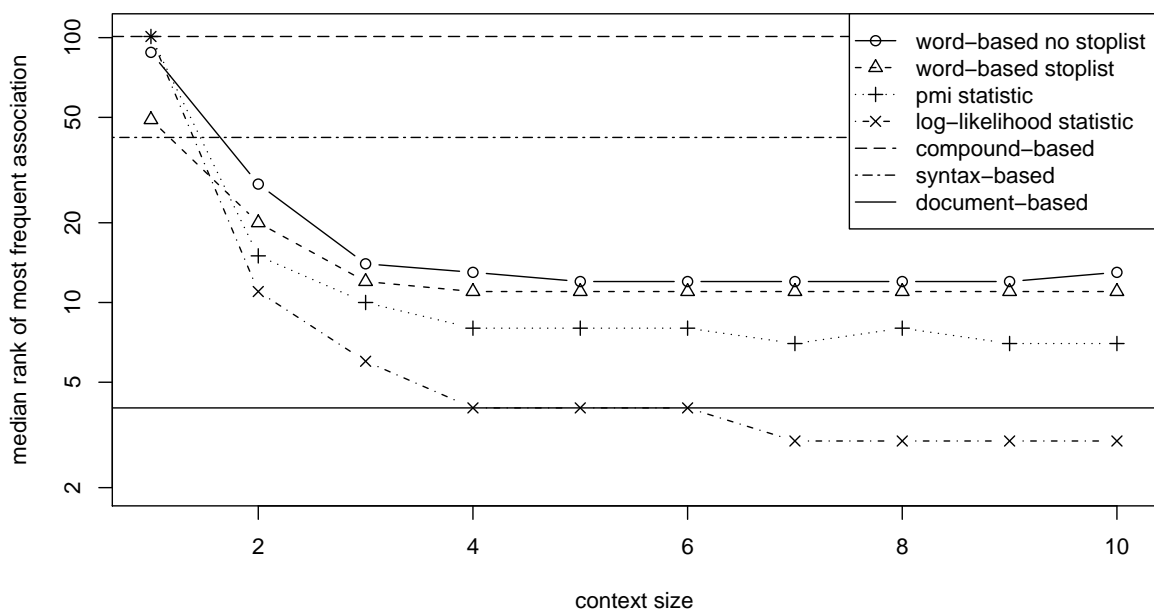


Figure 1: Median rank of the strong associations.

the magnitude of asymmetry. In a similar vein, Wettler et al. (2005) successfully predict associations on the basis of co-occurrence in text, in the framework of associationist learning theory. Despite this wealth of systems, it is an open question how their results compare to each other. Moreover, a model that combines several of these systems might outperform any basic approach.

4 Experiments

Our experiments were inspired by the association prediction task at the ESSLLI-2008 workshop on distributional models. We will first present this precise setup and then go into the results and their implications.

4.1 Setup

Our data was the Twente Nieuws Corpus (TwNC), which contains 300 million words of Dutch newspaper articles. This corpus was compiled at the University of Twente and subsequently parsed by the Alpino parser at the University of Groningen (van Noord, 2006). The newspaper articles in the corpus served as the contextual features for the document-based system; the dependency triples output by Alpino were used as input for the syntax-based approach. These syntactic

features of the type `subject of fly` covered eight syntactic relations — subject, direct object, prepositional complement, adverbial prepositional phrase, adjective modification, PP postmodification, apposition and coordination. Finally, the collocation measures and word-based distributional models took into account context sizes ranging from one to ten words to the left and right of the target.

Because of its many parameters, the precise implementation of the word space models deserves a bit more attention. In all cases, we used the context vectors in their full dimensionality. While this is somewhat of an exception in the literature, it has been argued that the full dimensionality leads to the best results for word-based models at least (Bullinaria and Levy, 2007). For the syntax-based and word-based approaches, we only took into account features that occurred at least two times together with the target. For the word-based models, we experimented with the use of a stoplist, which allowed us to exclude semantically “empty” words as features. The simple co-occurrence frequencies in the context vectors were replaced by the pointwise mutual information between the target and the feature (Bullinaria and Levy, 2007; Van der Plas, 2008). The similarity between two vectors was operationalized as the cosine of the angle be-

models	similar			related, not similar		
	mean	med	rank1	mean	med	rank1
pmi context 10	16.4	4	23%	25.2	9	10%
log-likelihood ratio context 10	12.8	2	41%	18.0	3	31%
syntax-based	16.3	4	22%	61.9	70	2%
word-based context 10 stoplist	10.7	3	27%	36.9	17	12%
document-based	10.1	3	26%	20.2	4	26%
compounding	80.7	101	5%	51.9	26	12%

Table 2: Performance of the models on semantically similar cue-association pairs and related but not similar pairs.

med = median; rank1 = number of associations at rank 1

tween them. This measure is more or less standard in the literature and leads to state-of-the-art results (Schütze, 1998; Padó and Lapata, 2007; Bullinaria and Levy, 2007). While the cosine is a symmetric measure, however, association strength is asymmetric. For example, *snelheid* (‘speed’) triggered *auto* (‘car’) no fewer than 55 times in the experiment, whereas *auto* evoked *snelheid* a mere 3 times. Like Michelbacher et al. (2007), we solve this problem by focusing not on the similarity score itself, but on the rank of the association in the list of nearest neighbours to the cue. We thus expect that *auto* will have a much higher rank in the list of nearest neighbours to *snelheid* than vice versa.

Our Gold Standard was based on a large-scale psycholinguistic experiment conducted at the University of Leuven (De Deyne and Storms, 2008). In this experiment, participants were asked to list three different associations for all cue words they were presented with. Each of the 1425 cues was given to at least 82 participants, resulting in a total of 381,909 responses. From this set, we took only noun cues with a single strong association. This means we found the most frequent association to each cue, and only included the pair in the test set if the association occurred at least 1.5 times more often than the second most frequent one. This resulted in a final test set of 593 cue-association pairs. Next we brought together all the associations in a set of candidate associations, and complemented it with 1000 random words from the corpus with a frequency of at least 200. From these candidate words, we had each model select the 100 highest scoring ones (the nearest neighbours). Performance was then expressed as the median and mean rank of the strongest association in this list. Associations absent from the list auto-

matically received a rank of 101. Thus, the lower the rank, the better the performance of the system. While there are obviously many more ways of assembling a test set and scoring the several systems, we found these all gave very similar results to the ones reported here.

4.2 Results and discussion

The median ranks of the strong associations for all models are plotted in Figure 1. The means show the same pattern, but give a less clear indication of the number of associations that were suggested in the top n most likely candidates. The most successful approach is the log-likelihood ratio (median 3 with a context size of 10, mean 16.6), followed by the document-based model (median 4, mean 18.4) and point-wise mutual information (median 7 with a context size of 10, mean 23.1). Next in line are the word-based distributional models with and without a stoplist (highest medians at 11 and 12, highest means at 30.9 and 33.3, respectively), and then the syntax-based word space model (median 42, mean 51.1). The worst performance is recorded for the compounding approach (median 101, mean 56.7). Overall, corpus-based approaches that rely on direct co-occurrence thus seem most appropriate for the prediction of strong associations to a cue. This is probably a result of two factors. First, collocation itself is an important motivation for human associations (Aitchinson, 2003). Second, while collocation approaches in themselves do not target semantic similarity, semantically similar associations are often also collocates to their cues. This is particularly the case for co-hyponyms, like *pepper* and *salt*, which score very high both in terms of collocation and in terms of similarity.

Let us discuss the results of all models in a bit

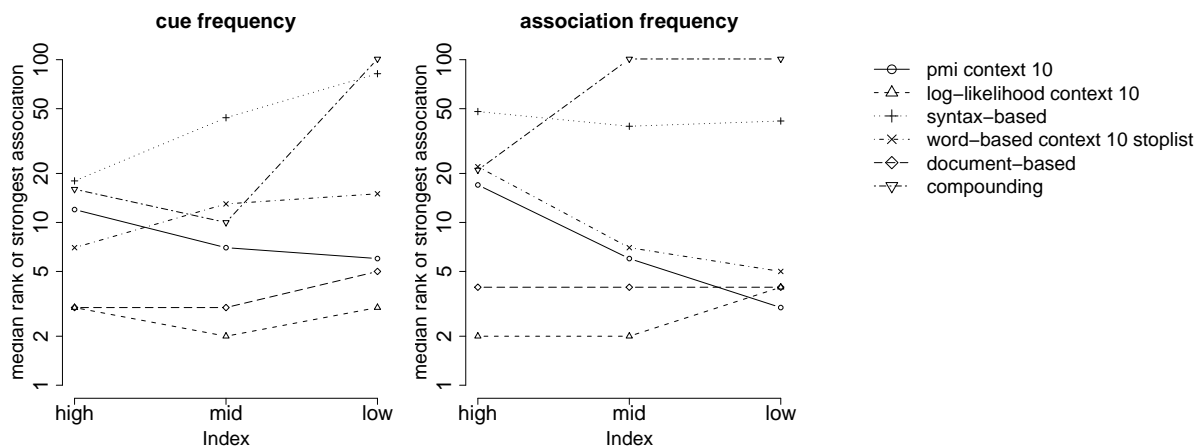


Figure 2: Performance of the models in three cue and association frequency bands.

more detail. A first factor of interest is the difference between associations that are similar to their cue and those which are related but not similar. Most of our models show a crucial difference in performance with respect to these two classes. The most important results are given in Table 2. The log-likelihood ratio gives the highest number of associations at rank 1 for both classes. Particularly surprising is its strong performance with respect to semantic similarity, since this relation is only a side effect of collocation. In fact, the log-likelihood ratio scores better at predicting semantically similar associations than related but not similar associations. Its performance moreover lies relatively close to that of the word space models, which were specifically developed to model semantic similarity. This underpins the observation that even associations that are semantically similar to their cues are still highly motivated by direct co-occurrence in text. Interestingly, only the compounding approach has a clear preference for associations that are related to their cue, but not similar.

A second factor that influences the performance of the models is frequency. In order to test its precise impact, we split up the cues and their associations in three frequency bands of comparable size. For the cues, we constructed a band for words with a frequency of less than 500 in the corpus (low), between 500 and 2,500 (mid) and more than 2,500 (high). For the associations, we had bands for words with a frequency of less than 7,500 (low), between 7,500 and 20,000 (mid) and more than 20,000 (high). Figure 2 shows the performance of the most important models in

these frequency bands. With respect to cue frequency, the word space models and compounding approach suffer most from low frequencies and hence, data sparseness. The log-likelihood ratio is much more robust, while point-wise mutual information even performs better with low-frequency cues, although it does not yet reach the performance of the document-based system or the log-likelihood ratio. With respect to association frequency, the picture is different. Here the word-based distributional models and PMI perform better with low-frequency associations. The document-based approach is largely insensitive to association frequency, while the log-likelihood ratio suffers slightly from low frequencies. The performance of the compounding approach decreases most. What is particularly interesting about this plot is that it points towards an important difference between the log-likelihood ratio and point-wise mutual information. In its search for nearest neighbours to a given cue word, the log-likelihood ratio favours frequent words. This is an advantageous feature in the prediction of strong associations, since people tend to give frequent words as associations. PMI, like the syntax-based and word-based models, lacks this characteristic. It therefore fails to discover mid- and high-frequency associations in particular.

Finally, despite the similarity in results between the log-likelihood ratio and the document-based word space model, there exists substantial variation in the associations that they predict successfully. Table 3 gives an overview of the top ten associations that are predicted better by one model than the other, according to the difference be-

model	cue–association pairs
document-based model	<i>cue–billiards, amphibian–frog, fair–doughnut ball, sperm whale–sea, map–trip, avocado–green, carnivore–meat, one-wheeler–circus, wallet–money, pinecone–wood</i>
log-likelihood ratio	<i>top–toy, oven–hot, sorbet–ice cream, rhubarb–sour, poppy–red, knot–rope, pepper–red, strawberry–red, massage–oil, raspberry–red</i>

Table 3: A comparison of the document-based model and the log-likelihood ratio on the basis of the cue–target pairs with the largest difference in log ranks between the two approaches.

tween the models in the logarithm of the rank of the association. The log-likelihood ratio seems to be biased towards “characteristics” of the target. For instance, it finds the strong associative relation between *poppy*, *pepper*, *strawberry*, *raspberry* and their shared colour *red* much better than the document-based model, just like it finds the relatedness between *oven* and *hot* and *rhubarb* and *sour*. The document-based model recovers more associations that display a strong topical connection with their cue word. This is thanks to its reliance on direct co-occurrence within a large context, which makes it less sensitive to semantic similarity than word-based models. It also appears to have less of a bias toward frequent words than the log-likelihood ratio. Note, for instance, the presence of *doughnut ball* (or *smoutebol* in Dutch) as the third nearest neighbour to *fair*, despite the fact it occurs only once (!) in the corpus. This complementarity between our two most successful approaches suggests that a combination of the two may lead to even better results. We therefore investigated the benefits of a committee-based or ensemble approach.

5 Ensemble-based prediction of strong associations

Given the varied nature of cue–association relations, it could be beneficial to develop a model that relies on more than one type of information. Ensemble methods have already proved their effectiveness in the related area of automatic thesaurus extraction (Curran, 2002), where semantic similarity is the target relation. Curran (2002) explored three ways of combining multiple ordered sets of words: (1) **mean**, taking the mean rank of each word over the ensemble; (2) **harmonic**, taking the harmonic mean; (3) **mixture**, calculating the mean similarity score for each word. We will study only the first two of these approaches, as the different metrics of our models cannot simply be combined

in a mean relatedness score. More particularly, we will experiment with ensembles taking the (harmonic) mean of the natural logarithm of the ranks, since we found these to perform better than those working with the original ranks.²

Table 4 compares the results of the most important ensembles with that of the single best approach, the log-likelihood ratio with a context size of 10. By combining the two best approaches from the previous section, the log-likelihood ratio and the document-based model, we already achieve a substantial increase in performance. The mean rank of the association goes from 3 to 2, the mean from 16.6 to 13.1 and the number of strong associations with rank 1 climbs from 194 to 223. This is a statistically significant increase (one-tailed paired Wilcoxon test, $W = 30866$, $p = .0002$). Adding another word space model to the ensemble, either a word-based or syntax-based model, brings down performance. However, the addition of the compound model does lead to a clear gain in performance. This ensemble finds the strongest association at a median rank of 2, and a mean of 11.8. In total, 249 strong associations (out of a total 593) are presented as the best candidate by the model — an increase of 28.4% compared to the log-likelihood ratio. Hence, despite its poor performance as a simple model, the compound-based approach can still give useful information about the strong association of a cue word when combined with other models. Based on the original ranks, the increase from the previous ensemble is not statistically significant ($W = 23929$, $p = .31$). If we consider differences at the start of the neighbour list more important and compare the logarithms of the ranks, however, the increase becomes significant ($W = 29787.5$, $p = 0.0008$). Its precise impact should thus further be investigated.

²In the case of the harmonic mean, we actually take the logarithm of $\text{rank}+1$, in order to avoid division by zero.

systems	mean			harmonic mean		
	med	mean	rank1	med	mean	rank1
loglik ₁₀ (baseline)	3	16.6	194			
loglik ₁₀ + doc	2	13.1	223	3	13.4	211
loglik ₁₀ + doc + word ₁₀	3	13.8	182	3	14.2	187
loglik ₁₀ + doc + syn	3	14.4	179	4	14.7	184
loglik ₁₀ + doc + comp	2	11.8	249	2	12.2	221

Table 4: Results of ensemble methods.

loglik₁₀ = log-likelihood ratio with context size 10;

doc = document-based model;

word₁₀ = word-based model with context size 10 and a stoplist;

syn = syntax-based model;

comp = compound-based model;

med = median; rank1 = number of associations at rank 1

Let us finally take a look at the types of strong associations that still tend to receive a low rank in this ensemble system. The first group consists of adjectives that refer to an inherent characteristic of the cue word that is rarely mentioned in text. This is the case for *tennis ball–yellow*, *cheese–yellow*, *grapefruit–bitter*. The second type brings together polysemous cues whose strongest association relates to a different sense than that represented by its corpus-based nearest neighbour. This applies to Dutch *kant*, which is polysemous between *side* and *lace*. Its strongest association, *Bruges*, is clearly related to the latter meaning, but its corpus-based neighbours *ball* and *water* suggest the former. The third type reflects human encyclopaedic knowledge that is less central to the semantics of the cue word. Examples are *police–blue*, *love–red*, or *triangle–maths*. In many of these cases, it appears that the failure of the model to recover the strong associations results from corpus limitations rather than from the model itself.

6 Conclusions and future research

In this paper, we explored three types of basic approaches to the prediction of strong associations to a given cue. Collocation measures like the log-likelihood ratio simply recover those words that strongly collocate with the cue. Word space models look for words that appear in similar contexts, defined as documents, context words or syntactic relations. The compounding approach, finally, searches for words that combine with the target to form a compound. The log-likelihood ratio with a large context size emerged as the best predictor of strong association, followed closely by the

document-based word space model. Moreover, we showed that an ensemble method combining the log-likelihood ratio, the document-based word space model and the compounding approach, outperformed any of the basic methods by almost 30%.

In a number of ways, this paper is only a first step towards the successful modelling of cue–association relations. First, the newspaper corpus that served as our data has some restrictions, particularly with respect to diversity of genres. It would be interesting to investigate to what degree a more general corpus — a web corpus, for instance — would be able to accurately predict a wider range of associations. Second, the models themselves might benefit from some additional features. For instance, we are curious to find out what the influence of dimensionality reduction would be, particularly for document-based word space models. Finally, we would like to extend our test set from strong associations to more associations for a given target, in order to investigate how well the discussed models predict relative association strength.

References

- Jean Aitchinson. 2003. *Words in the Mind. An Introduction to the Mental Lexicon*. Blackwell, Oxford.
- John A. Bullinaria and Joseph P. Levy. 2007. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behaviour Research Methods*, 39:510–526.
- Curt Burgess, Kay Livesay, and Kevin Lund. 1998. Explorations in context space: Words, sentences, discourse. *Discourse Processes*, 25:211–257.

- Ed H. Chi, Peter Pirolli, Kim Chen, and James Pitkow. 2001. Using information scent to model user information needs and actions on the web. In *Proceedings of the ACM Conference on Human Factors and Computing Systems (CHI 2001)*, pages 490–497.
- Kenneth Ward Church and Patrick Hanks. 1989. Word association norms, mutual information and lexicography. In *Proceedings of ACL-27*, pages 76–83.
- James R. Curran. 2002. Ensemble methods for automatic thesaurus extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2002)*, pages 222–229.
- Simon De Deyne and Gert Storms. 2008. Word associations: Norms for 1,424 Dutch words in a continuous task. *Behaviour Research Methods*, 40:198–205.
- Ted Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19:61–74.
- Peter W. Foltz. 1996. Latent Semantic Analysis for text-based research. *Behaviour Research Methods, Instruments, and Computers*, 29:197–202.
- Tuomo Kakkonen, Niko Myller, Jari Timonen, and Erkki Sutinen. 2005. Automatic essay grading with probabilistic latent semantic analysis. In *Proceedings of the 2nd Workshop on Building Educational Applications Using NLP*, pages 29–36.
- 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.
- Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of COLING-ACL98*, pages 768–774, Montreal, Canada.
- Will Lowe and Scott McDonald. 2000. The direct route: Mediated priming in semantic space. In *Proceedings of COGSCI 2000*, pages 675–680. Lawrence Erlbaum Associates.
- Lukas Michelbacher, Stefan Evert, and Hinrich Schütze. 2007. Asymmetric association measures. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP-07)*.
- Tom M. Mitchell, Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malva, Robert A. Mason, and Marcel Adam Just. 2008. Predicting human brain activity associated with the meanings of nouns. *Science*, 320:1191–1195.
- 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 Geeraerts. 2008. Size matters. Tight and loose context definitions in English word space models. In *Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics*, pages 9–16.
- Magnus Sahlgren. 2006. *The Word-Space Model. Using Distributional Analysis to Represent Syntagmatic and Paradigmatic Relations Between Words in High-dimensional Vector Spaces*. Ph.D. thesis, Stockholm University, Stockholm, Sweden.
- Sabine Schulte im Walde and Alissa Melinger. 2005. Identifying semantic relations and functional properties of human verb associations. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 612–619.
- Hinrich Schütze. 1998. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–124.
- Lonneke Van der Plas. 2008. *Automatic Lexico-Semantic Acquisition for Question Answering*. Ph.D. thesis, University of Groningen, Groningen, The Netherlands.
- Gertjan van Noord. 2006. At last parsing is now operational. In Piet Mertens, Cédric Fairon, Anne Disster, and Patrick Watrin, editors, *Verbum Ex Machina. Actes de la 13e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, pages 20–42.
- Tonio Wandmacher, Ekaterina Ovchinnikova, and Theodore Alexandrov. 2008. Does Latent Semantic Analysis reflect human associations? In *Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics*, pages 63–70.
- Manfred Wetzler, Reinhard Rapp, and Peter Sedlmeier. 2005. Free word associations correspond to contiguities between words in texts. *Journal of Quantitative Linguistics*, 12(2/3):111–122.