

Finding Predominant Word Senses in Untagged Text

Diana McCarthy & Rob Koeling & Julie Weeds & John Carroll

Department of Informatics,

University of Sussex

Brighton BN1 9QH, UK

{*dianam,robk,juliewe,johnca*}@sussex.ac.uk

Abstract

In word sense disambiguation (WSD), the heuristic of choosing the most common sense is extremely powerful because the distribution of the senses of a word is often skewed. The problem with using the predominant, or first sense heuristic, aside from the fact that it does not take surrounding context into account, is that it assumes some quantity of hand-tagged data. Whilst there are a few hand-tagged corpora available for some languages, one would expect the frequency distribution of the senses of words, particularly topical words, to depend on the genre and domain of the text under consideration. We present work on the use of a thesaurus acquired from raw textual corpora and the WordNet similarity package to find predominant noun senses automatically. The acquired predominant senses give a precision of 64% on the nouns of the SENSEVAL-2 English all-words task. This is a very promising result given that our method does not require any hand-tagged text, such as SemCor. Furthermore, we demonstrate that our method discovers appropriate predominant senses for words from two domain-specific corpora.

1 Introduction

The first sense heuristic which is often used as a baseline for supervised WSD systems outperforms many of these systems which take surrounding context into account. This is shown by the results of the English all-words task in SENSEVAL-2 (Cotton et al., 1998) in figure 1 below, where the first sense is that listed in WordNet for the PoS given by the Penn TreeBank (Palmer et al., 2001). The senses in WordNet are ordered according to the frequency data in the manually tagged resource SemCor (Miller et al., 1993). Senses that have not occurred in SemCor are ordered arbitrarily and after those senses of the word that have occurred. The figure distinguishes systems which make use of hand-tagged data (using HTD) such as SemCor, from those that do not (without HTD). The high per-

formance of the first sense baseline is due to the skewed frequency distribution of word senses. Even systems which show superior performance to this heuristic often make use of the heuristic where evidence from the context is not sufficient (Hoste et al., 2001). Whilst a first sense heuristic based on a sense-tagged corpus such as SemCor is clearly useful, there is a strong case for obtaining a first, or predominant, sense from untagged corpus data so that a WSD system can be tuned to the genre or domain at hand.

SemCor comprises a relatively small sample of 250,000 words. There are words where the first sense in WordNet is counter-intuitive, because of the size of the corpus, and because where the frequency data does not indicate a first sense, the ordering is arbitrary. For example the first sense of *tiger* in WordNet is **audacious person** whereas one might expect that **carnivorous animal** is a more common usage. There are only a couple of instances of *tiger* within SemCor. Another example is *embryo*, which does not occur at all in SemCor and the first sense is listed as **rudimentary plant** rather than the anticipated **fertilised egg** meaning. We believe that an automatic means of finding a predominant sense would be useful for systems that use it as a means of backing-off (Wilks and Stevenson, 1998; Hoste et al., 2001) and for systems that use it in lexical acquisition (McCarthy, 1997; Merlo and Leybold, 2001; Korhonen, 2002) because of the limited size of hand-tagged resources. More importantly, when working within a specific domain one would wish to tune the first sense heuristic to the domain at hand. The first sense of *star* in SemCor is **celestial body**, however, if one were disambiguating popular news **celebrity** would be preferred.

Assuming that one had an accurate WSD system then one could obtain frequency counts for senses and rank them with these counts. However, the most accurate WSD systems are those which require manually sense tagged data in the first place, and their accuracy depends on the quantity of training examples (Yarowsky and Florian, 2002) available. We

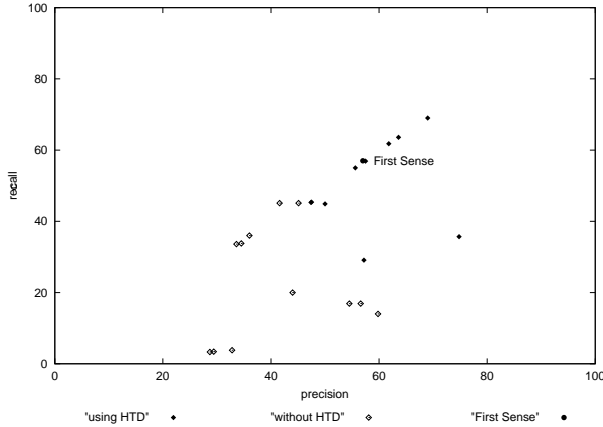


Figure 1: The first sense heuristic compared with the SENSEVAL-2 English all-words task results

are therefore investigating a method of automatically ranking WordNet senses from raw text.

Many researchers are developing thesauruses from automatically parsed data. In these each target word is entered with an ordered list of “nearest neighbours”. The neighbours are words ordered in terms of the “distributional similarity” that they have with the target. Distributional similarity is a measure indicating the degree that two words, a word and its neighbour, occur in similar contexts. From inspection, one can see that the ordered neighbours of such a thesaurus relate to the different senses of the target word. For example, the neighbours of *star* in a dependency-based thesaurus provided by Lin¹ has the ordered list of neighbours: *superstar*, *player*, *teammate*, *actor* early in the list, but one can also see words that are related to another sense of *star* e.g. *galaxy*, *sun*, *world* and *planet* further down the list. We expect that the quantity and similarity of the neighbours pertaining to different senses will reflect the dominance of the sense to which they pertain. This is because there will be more relational data for the more prevalent senses compared to the less frequent senses. In this paper we describe and evaluate a method for ranking senses of nouns to obtain the predominant sense of a word using the neighbours from automatically acquired thesauruses. The neighbours for a word in a thesaurus are words themselves, rather than senses. In order to associate the neighbours with senses we make use of another notion of similarity, “semantic similarity”, which exists between senses, rather than words. We experiment with several WordNet Similarity measures (Patwardhan and Pedersen, 2003) which aim to capture semantic relatedness within

¹Available at <http://www.cs.ualberta.ca/~lindek/demos/depsim.htm>

the WordNet hierarchy. We use WordNet as our sense inventory for this work.

The paper is structured as follows. We discuss our method in the following section. Sections 3 and 4 concern experiments using predominant senses from the BNC evaluated against the data in SemCor and the SENSEVAL-2 English all-words task respectively. In section 5 we present results of the method on two domain specific sections of the Reuters corpus for a sample of words. We describe some related work in section 6 and conclude in section 7.

2 Method

In order to find the predominant sense of a target word we use a thesaurus acquired from automatically parsed text based on the method of Lin (1998). This provides the k nearest neighbours to each target word, along with the distributional similarity score between the target word and its neighbour. We then use the WordNet similarity package (Patwardhan and Pedersen, 2003) to give us a semantic similarity measure (hereafter referred to as the WordNet similarity measure) to weight the contribution that each neighbour makes to the various senses of the target word.

To find the first sense of a word (w) we take each sense in turn and obtain a score reflecting the prevalence which is used for ranking. Let $N_w = \{n_1, n_2, \dots, n_k\}$ be the ordered set of the top scoring k neighbours of w from the thesaurus with associated distributional similarity scores $\{dss(w, n_1), dss(w, n_2), \dots, dss(w, n_k)\}$. Let $senses(w)$ be the set of senses of w . For each sense of w ($ws_i \in senses(w)$) we obtain a ranking score by summing over the $dss(w, n_j)$ of each neighbour ($n_j \in N_w$) multiplied by a weight. This weight is the WordNet similarity score ($wnss$) between the target sense (ws_i) and the sense of n_j ($ns_x \in senses(n_j)$) that maximises this score, divided by the sum of all such WordNet similarity scores for $senses(w)$ and n_j . Thus we rank each sense $ws_i \in senses(w)$ using:

$$Prevalence\ Score(ws_i) = \sum_{n_j \in N_w} dss(w, n_j) \times \frac{wnss(ws_i, n_j)}{\sum_{ws_{i'} \in senses(w)} wnss(ws_{i'}, n_j)} \quad (1)$$

where:

$$wnss(ws_i, n_j) = \max_{ns_x \in senses(n_j)} (wnss(ws_i, ns_x))$$

2.1 Acquiring the Automatic Thesaurus

The thesaurus was acquired using the method described by Lin (1998). For input we used grammatical relation data extracted using an automatic

parser (Briscoe and Carroll, 2002). For the experiments in sections 3 and 4 we used the 90 million words of written English from the BNC. For each noun we considered the co-occurring verbs in the direct object and subject relation, the modifying nouns in noun-noun relations and the modifying adjectives in adjective-noun relations. We could easily extend the set of relations in the future. A noun, w , is thus described by a set of co-occurrence triples $\langle w, r, x \rangle$ and associated frequencies, where r is a grammatical relation and x is a possible co-occurrence with w in that relation. For every pair of nouns, where each noun had a total frequency in the triple data of 10 or more, we computed their distributional similarity using the measure given by Lin (1998). If $T(w)$ is the set of co-occurrence types (r, x) such that $I(w, r, x)$ is positive then the similarity between two nouns, w and n , can be computed as: $ds_s(w, n) =$

$$\frac{\sum_{(r,x) \in T(w) \cap T(n)} (I(w, r, x) + I(n, r, x))}{\sum_{(r,x) \in T(w)} I(w, r, x) + \sum_{(r,x) \in T(n)} I(n, r, x)}$$

where:

$$I(w, r, x) = \log \frac{P(x|w \cap r)}{P(x|r)}$$

A thesaurus entry of size k for a target noun w is then defined as the k most similar nouns to w .

2.2 The WordNet Similarity Package

We use the WordNet Similarity Package 0.05 and WordNet version 1.6.² The WordNet Similarity package supports a range of WordNet similarity scores. We experimented using six of these to provide the $wnss$ in equation 1 above and obtained results well over our baseline, but because of space limitations give results for the two which perform the best. We briefly summarise the two measures here; for a more detailed summary see (Patwardhan et al., 2003). The measures provide a similarity score between two WordNet senses (s_1 and s_2), these being synsets within WordNet.

lesk (Banerjee and Pedersen, 2002) This score maximises the number of overlapping words in the gloss, or definition, of the senses. It uses the glosses of semantically related (according to WordNet) senses too.

jcn (Jiang and Conrath, 1997) This score uses corpus data to populate classes (synsets) in the WordNet hierarchy with frequency counts. Each

²We use this version of WordNet since it allows us to map information to WordNets of other languages more accurately. We are of course able to apply the method to other versions of WordNet.

synset, is incremented with the frequency counts from the corpus of all words belonging to that synset, directly or via the hyponymy relation. The frequency data is used to calculate the ‘‘information content’’ (IC) of a class $IC(s) = -\log(p(s))$. Jiang and Conrath specify a distance measure: $D_{jcn}(s_1, s_2) = IC(s_1) + IC(s_2) - 2 \times IC(s_3)$, where the third class (s_3) is the most informative, or most specific, superordinate synset of the two senses s_1 and s_2 . This is transformed from a distance measure in the WN-Similarity package by taking the reciprocal:

$$jcn(s_1, s_2) = 1/D_{jcn}(s_1, s_2)$$

3 Experiment with SemCor

In order to evaluate our method we use the data in SemCor as a gold-standard. This is not ideal since we expect that the sense frequency distributions within SemCor will differ from those in the BNC, from which we obtain our thesaurus. Nevertheless, since many systems performed well on the English all-words task for SENSEVAL-2 by using the frequency information in SemCor this is a reasonable approach for evaluation.

We generated a thesaurus entry for all polysemous nouns which occurred in SemCor with a frequency > 2 , and in the BNC with a frequency ≥ 10 in the grammatical relations listed in section 2.1 above. The **jcn** measure uses corpus data for the calculation of IC. We experimented with counts obtained from the BNC and the Brown corpus. The variation in counts had negligible affect on the results.³ The experimental results reported here are obtained using IC counts from the BNC corpus. All the results shown here are those with the size of thesaurus entries (k) set to 50.⁴

We calculate the accuracy of finding the predominant sense, when there is indeed one sense with a higher frequency than the others for this word in SemCor (PS_{acc}). We also calculate the WSD accuracy that would be obtained on SemCor, when using our first sense in all contexts (WSD_{sc}).

3.1 Results

The results in table 1 show the accuracy of the ranking with respect to SemCor over the entire set of 2595 polysemous nouns in SemCor with

³Using the default IC counts provided with the package did result in significantly higher results, but these default files are obtained from the sense-tagged data within SemCor itself so we discounted these results.

⁴We repeated the experiment with the BNC data for **jcn** using $k = 10, 30, 50$ and 70 however, the number of neighbours used gave only minimal changes to the results so we do not report them here.

measure	PS_{acc} %	WSD_{sc} %
lesk	54	48
jcn	54	46
baseline	32	24

Table 1: SemCor results

the **jcn** and **lesk** WordNet similarity measures. The random baseline for choosing the predominant sense over all these words ($\sum_{w \in Words} \frac{1}{|senses(w)|}$) is 32%. Both WordNet similarity measures beat this baseline. The random baseline for WSD_{sc} ($\sum_{t \in SemCor\ tokens} \frac{1}{|senses(t)|}$) is 24%. Again, the automatic ranking outperforms this by a large margin. The first sense in SemCor provides an upper-bound for this task of 67%.

Since both measures gave comparable results we restricted our remaining experiments to **jcn** because this gave good results for finding the predominant sense, and is much more efficient than **lesk**, given the precompilation of the IC files.

3.2 Discussion

From manual analysis, there are cases where the acquired first sense disagrees with SemCor, yet is intuitively plausible. This is to be expected regardless of any inherent shortcomings of the ranking technique since the senses within SemCor will differ compared to those of the BNC. For example, in WordNet the first listed sense of *pipe* is **tobacco pipe**, and this is ranked joint first according to the Brown files in SemCor with the second sense **tube made of metal or plastic used to carry water, oil or gas etc....** The automatic ranking from the BNC data lists the latter **tube** sense first. This seems quite reasonable given the nearest neighbours: *tube, cable, wire, tank, hole, cylinder, fitting, tap, cistern, plate....* Since SemCor is derived from the Brown corpus, which predates the BNC by up to 30 years⁵ and contains a higher proportion of fiction⁶, the high ranking for the **tobacco pipe** sense according to SemCor seems plausible.

Another example where the ranking is intuitive, is *soil*. The first ranked sense according to SemCor is the **filth, stain: state of being unclean** sense whereas the automatic ranking lists **dirt, ground, earth** as the first sense, which is the second ranked

⁵The text in the Brown corpus was produced in 1961, whereas the bulk of the written portion of the BNC contains texts produced between 1975 and 1993.

⁶6 out of the 15 Brown genres are fiction, including one specifically dedicated to detective fiction, whilst only 20% of the BNC text represents imaginative writing, the remaining 80% being classified as informative.

sense according to SemCor. This seems intuitive given our expected relative usage of these senses in modern British English.

Even given the difference in text type between SemCor and the BNC the results are encouraging, especially given that our WSD_{sc} results are for polysemous nouns. In the English all-words SENSEVAL-2, 25% of the noun data was monosemous. Thus, if we used the sense ranking as a heuristic for an “all nouns” task we would expect to get precision in the region of 60%. We test this below on the SENSEVAL-2 English all-words data.

4 Experiment on SENSEVAL-2 English all Words Data

In order to see how well the automatically acquired predominant sense performs on a WSD task from which the WordNet sense ordering has not been taken, we use the SENSEVAL-2 all-words data (Palmer et al., 2001).⁷ This is a hand-tagged test suite of 5,000 words of running text from three articles from the Penn Treebank II. We use an all-words task because the predominant senses will reflect the sense distributions of all nouns within the documents, rather than a lexical sample task, where the target words are manually determined and the results will depend on the skew of the words in the sample. We do not assume that the predominant sense is a method of WSD in itself. To disambiguate senses a system should take context into account. However, it is important to know the performance of this heuristic for any systems that use it.

We generated a thesaurus entry for all polysemous nouns in WordNet as described in section 2.1 above. We obtained the predominant sense for each of these words and used these to label the instances in the noun data within the SENSEVAL-2 English all-words task. We give the results for this WSD task in table 2. We compare results using the first sense listed in SemCor, and the first sense according to the SENSEVAL-2 English all-words test data itself. For the latter, we only take a first-sense where there is more than one occurrence of the noun in the test data and one sense has occurred more times than any of the others. We trivially labelled all monosemous items.

Our automatically acquired predominant sense performs nearly as well as the first sense provided by SemCor, which is very encouraging given that

⁷In order to do this we use the mapping provided at <http://www.lsi.upc.es/~nlp/tools/mapping.html> (Daudé et al., 2000) for obtaining the SENSEVAL-2 data in WordNet 1.6. We discounted the few items for which there was no mapping. This amounted to only 3% of the data.

	precision	recall
Automatic	64	63
SemCor	69	68
SENSEVAL-2	92	72

Table 2: Evaluating predominant sense information on SENSEVAL-2 all-words data.

our method only uses raw text, with no manual labelling. The performance of the predominant sense provided in the SENSEVAL-2 test data provides an upper bound for this task. The items that were not covered by our method were those with insufficient grammatical relations for the tuples employed. Two such words, *today* and *one*, each occurred 5 times in the test data. Extending the grammatical relations used for building the thesaurus should improve the coverage. There were a similar number of words that were not covered by a predominant sense in SemCor. For these one would need to obtain more sense-tagged text in order to use this heuristic. Our automatic ranking gave 67% precision on these items. This demonstrates that our method of providing a first sense from raw text will help when sense-tagged data is not available.

5 Experiments with Domain Specific Corpora

A major motivation for our work is to try to capture changes in ranking of senses for documents from different domains. In order to test this we applied our method to two specific sections of the Reuters corpus. We demonstrate that choosing texts from a particular domain has a significant influence on the sense ranking. We chose the domains of SPORTS and FINANCE since there is sufficient material for these domains in this publically available corpus.

5.1 Reuters Corpus

The Reuters corpus (Rose et al., 2002) is a collection of about 810,000 Reuters, English Language News stories. Many of the articles are economy related, but several other topics are included too. We selected documents from the SPORTS domain (topic code: GSPO) and a limited number of documents from the FINANCE domain (topic codes: ECAT and MCAT).

The SPORTS corpus consists of 35317 documents (about 9.1 million words). The FINANCE corpus consists of 117734 documents (about 32.5 million words). We acquired thesauruses for these corpora using the procedure described in section 2.1.

5.2 Two Experiments

There is no existing sense-tagged data for these domains that we could use for evaluation. We therefore decided to select a limited number of words and to evaluate these words qualitatively. The words included in this experiment are not a random sample, since we anticipated different predominant senses in the SPORTS and FINANCE domains for these words.

Additionally, we evaluated our method quantitatively using the Subject Field Codes (SFC) resource (Magnini and Cavaglia, 2000) which annotates WordNet synsets with domain labels. The SFC contains an **economy** label and a **sports** label. For this domain label experiment we selected all the words in WordNet that have at least one synset labelled **economy** and at least one synset labelled **sports**. The resulting set consisted of 38 words. We contrast the distribution of domain labels for these words in the 2 domain specific corpora.

5.3 Discussion

The results for 10 of the words from the qualitative experiment are summarized in table 3 with the WordNet sense number for each word supplied alongside synonyms or hypernyms from WordNet for readability. The results are promising. Most words show the change in predominant sense (PS) that we anticipated. It is not always intuitively clear which of the senses to expect as predominant sense for either a particular domain or for the BNC, but the first senses of words like *division* and *goal* shift towards the more specific senses (**league** and **score** respectively). Moreover, the chosen senses of the word *tie* proved to be a textbook example of the behaviour we expected.

The word *share* is among the words whose predominant sense remained the same for all three corpora. We anticipated that the **stock certificate** sense would be chosen for the FINANCE domain, but this did not happen. However, that particular sense ended up higher in the ranking for the FINANCE domain.

Figure 2 displays the results of the second experiment with the domain specific corpora. This figure shows the domain labels assigned to the predominant senses for the set of 38 words after ranking the words using the SPORTS and the FINANCE corpora. We see that both domains have a similarly high percentage of factotum (domain independent) labels, but as we would expect, the other peaks correspond to the **economy** label for the FINANCE corpus, and the **sports** label for the SPORTS corpus.

Word	PS BNC	PS FINANCE	PS SPORTS
<i>pass</i>	1 (accomplishment)	14 (attempt)	15 (throw)
<i>share</i>	2 (portion, asset)	2	2
<i>division</i>	4 (admin. unit)	4	6 (league)
<i>head</i>	1 (body part)	4 (leader)	4
<i>loss</i>	2 (transf. property)	2	8 (death, departure)
<i>competition</i>	2 (contest, social event)	3 (rivalry)	2
<i>match</i>	2 (contest)	7 (equal, person)	2
<i>tie</i>	1 (neckwear)	2 (affiliation)	3 (draw)
<i>strike</i>	1 (work stoppage)	1	6 (hit, success)
<i>goal</i>	1 (end, mental object)	1	2 (score)

Table 3: Domain specific results

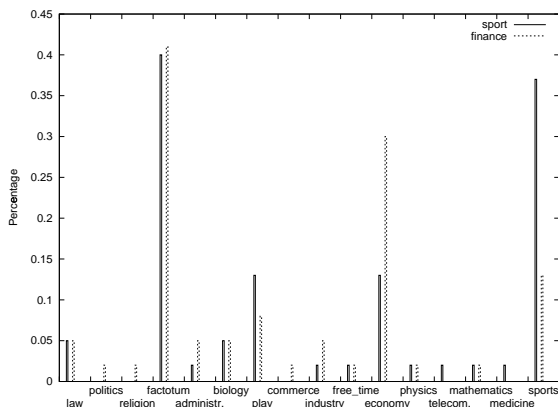


Figure 2: Distribution of domain labels of predominant senses for 38 polysemous words ranked using the SPORTS and FINANCE corpus.

6 Related Work

Most research in WSD concentrates on using contextual features, typically neighbouring words, to help determine the correct sense of a target word. In contrast, our work is aimed at discovering the predominant senses from raw text because the first sense heuristic is such a useful one, and because hand-tagged data is not always available.

A major benefit of our work, rather than reliance on hand-tagged training data such as SemCor, is that this method permits us to produce predominant senses for the domain and text type required. Buitelaar and Sacaleanu (2001) have previously explored ranking and selection of synsets in GermaNet for specific domains using the words in a given synset, and those related by hyponymy, and a term relevance measure taken from information retrieval. Buitelaar and Sacaleanu have evaluated their method on identifying domain specific concepts using human judgements on 100 items. We have evaluated our method using publically available resources, both for balanced and domain spe-

cific text. Magnini and Cavaglia (2000) have identified WordNet word senses with particular domains, and this has proven useful for high precision WSD (Magnini et al., 2001); indeed in section 5 we used these domain labels for evaluation. Identification of these domain labels for word senses was semi-automatic and required a considerable amount of hand-labelling. Our approach is complementary to this. It only requires raw text from the given domain and because of this it can easily be applied to a new domain, or sense inventory, given sufficient text.

Lapata and Brew (2004) have recently also highlighted the importance of a good prior in WSD. They used syntactic evidence to find a prior distribution for verb classes, based on (Levin, 1993), and incorporate this in a WSD system. Lapata and Brew obtain their priors for verb classes directly from subcategorisation evidence in a parsed corpus, whereas we use parsed data to find distributionally similar words (nearest neighbours) to the target word which reflect the different senses of the word and have associated distributional similarity scores which can be used for ranking the senses according to prevalence.

There has been some related work on using automatic thesauruses for discovering word senses from corpora Pantel and Lin (2002). In this work the lists of neighbours are themselves clustered to bring out the various senses of the word. They evaluate using the **lin** measure described above in section 2.2 to determine the precision and recall of these discovered classes with respect to WordNet synsets. This method obtains precision of 61% and recall 51%. If WordNet sense distinctions are not ultimately required then discovering the senses directly from the neighbours list is useful because sense distinctions discovered are relevant to the corpus data and new senses can be found. In contrast, we use the neighbours lists and WordNet similarity measures to im-

pose a prevalence ranking on the WordNet senses. We believe automatic ranking techniques such as ours will be useful for systems that rely on WordNet, for example those that use it for lexical acquisition or WSD. It would be useful however to combine our method of finding predominant senses with one which can automatically find new senses within text and relate these to WordNet synsets, as Ciaramita and Johnson (2003) do with unknown nouns.

We have restricted ourselves to nouns in this work, since this PoS is perhaps most affected by domain. We are currently investigating the performance of the first sense heuristic, and this method, for other PoS on SENSEVAL-3 data (McCarthy et al., 2004), although not yet with rankings from domain specific corpora. The **lesk** measure can be used when ranking adjectives, and adverbs as well as nouns and verbs (which can also be ranked using **jcn**). Another major advantage that **lesk** has is that it is applicable to lexical resources which do not have the hierarchical structure that WordNet does, but do have definitions associated with word senses.

7 Conclusions

We have devised a method that uses raw corpus data to automatically find a predominant sense for nouns in WordNet. We use an automatically acquired thesaurus and a WordNet Similarity measure. The automatically acquired predominant senses were evaluated against the hand-tagged resources SemCor and the SENSEVAL-2 English all-words task giving us a WSD precision of 64% on an all-nouns task. This is just 5% lower than results using the first sense in the manually labelled SemCor, and we obtain 67% precision on polysemous nouns that are not in SemCor.

In many cases the sense ranking provided in SemCor differs to that obtained automatically because we used the BNC to produce our thesaurus. Indeed, the merit of our technique is the very possibility of obtaining predominant senses from the data at hand. We have demonstrated the possibility of finding predominant senses in domain specific corpora on a sample of nouns. In the future, we will perform a large scale evaluation on domain specific corpora. In particular, we will use balanced and domain specific corpora to isolate words having very different neighbours, and therefore rankings, in the different corpora and to detect and target words for which there is a highly skewed sense distribution in these corpora.

There is plenty of scope for further work. We want to investigate the effect of frequency and choice of distributional similarity measure (Weeds

et al., 2004). Additionally, we need to determine whether senses which do not occur in a wide variety of grammatical contexts fare badly using distributional measures of similarity, and what can be done to combat this problem using relation specific thesauruses.

Whilst we have used WordNet as our sense inventory, it would be possible to use this method with another inventory given a measure of semantic relatedness between the neighbours and the senses. The **lesk** measure for example, can be used with definitions in any standard machine readable dictionary.

Acknowledgements

We would like to thank Siddharth Patwardhan and Ted Pedersen for making the WN Similarity package publically available. This work was funded by EU-2001-34460 project MEANING: Developing Multilingual Web-scale Language Technologies, UK EPSRC project Robust Accurate Statistical Parsing (RASP) and a UK EPSRC studentship.

References

- Satanjeev Banerjee and Ted Pedersen. 2002. An adapted Lesk algorithm for word sense disambiguation using WordNet. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-02)*, Mexico City.
- Edward Briscoe and John Carroll. 2002. Robust accurate statistical annotation of general text. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*, pages 1499–1504, Las Palmas, Canary Islands, Spain.
- Paul Buitelaar and Bogdan Sacaleanu. 2001. Ranking and selecting synsets by domain relevance. In *Proceedings of WordNet and Other Lexical Resources: Applications, Extensions and Customizations, NAACL 2001 Workshop*, Pittsburgh, PA.
- Massimiliano Ciaramita and Mark Johnson. 2003. Supersense tagging of unknown nouns in WordNet. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2003)*.
- Scott Cotton, Phil Edmonds, Adam Kilgarriff, and Martha Palmer. 1998. SENSEVAL-2. <http://www.sle.sharp.co.uk/senseval2/>.
- Jordi Daudé, Lluís Padró, and German Rigau. 2000. Mapping wordnets using structural information. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, Hong Kong.

- Véronique Hoste, Anne Kool, and Walter Daelemans. 2001. Classifier optimization and combination in the English all words task. In *Proceedings of the SENSEVAL-2 workshop*, pages 84–86.
- Jay Jiang and David Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In *International Conference on Research in Computational Linguistics*, Taiwan.
- Anna Korhonen. 2002. Semantically motivated subcategorization acquisition. In *Proceedings of the ACL Workshop on Unsupervised Lexical Acquisition*, Philadelphia, USA.
- Mirella Lapata and Chris Brew. 2004. Verb class disambiguation using informative priors. *Computational Linguistics*, 30(1):45–75.
- Beth Levin. 1993. *English Verb Classes and Alternations: a Preliminary Investigation*. University of Chicago Press, Chicago and London.
- Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of COLING-ACL 98*, Montreal, Canada.
- Bernardo Magnini and Gabriela Cavaglià. 2000. Integrating subject field codes into WordNet. In *Proceedings of LREC-2000*, Athens, Greece.
- Bernardo Magnini, Carlo Strapparava, Giovanni Pezzuli, and Alfio Gliozzo. 2001. Using domain information for word sense disambiguation. In *Proceedings of the SENSEVAL-2 workshop*, pages 111–114.
- Diana McCarthy, Rob Koeling, Julie Weeds, and John Carroll. 2004. Using automatically acquired predominant senses for word sense disambiguation. In *Proceedings of the ACL SENSEVAL-3 workshop*.
- Diana McCarthy. 1997. Word sense disambiguation for acquisition of selectional preferences. In *Proceedings of the ACL/EACL 97 Workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 52–61.
- Paola Merlo and Matthias Leybold. 2001. Automatic distinction of arguments and modifiers: the case of prepositional phrases. In *Proceedings of the Workshop on Computational Language Learning (CoNLL 2001)*, Toulouse, France.
- George A. Miller, Claudia Leacock, Randee Tengi, and Ross T Bunker. 1993. A semantic concordance. In *Proceedings of the ARPA Workshop on Human Language Technology*, pages 303–308. Morgan Kaufman.
- Martha Palmer, Christiane Fellbaum, Scott Cotton, Lauren Delfs, and Hoa Trang Dang. 2001. English tasks: All-words and verb lexical sample. In *Proceedings of the SENSEVAL-2 workshop*, pages 21–24.
- Patrick Pantel and Dekang Lin. 2002. Discovering word senses from text. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 613–619, Edmonton, Canada.
- Siddharth Patwardhan and Ted Pedersen. 2003. The cpan wordnet::similarity package. <http://search.cpan.org/author/SID/WordNet-Similarity-0.03/>.
- Siddharth Patwardhan, Satanjeev Banerjee, and Ted Pedersen. 2003. Using measures of semantic relatedness for word sense disambiguation. In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2003)*, Mexico City.
- Tony G. Rose, Mary Stevenson, and Miles Whitehead. 2002. The Reuters Corpus volume 1 - from yesterday's news to tomorrow's language resources. In *Proc. of Third International Conference on Language Resources and Evaluation*, Las Palmas de Gran Canaria.
- Julie Weeds, David Weir, and Diana McCarthy. 2004. Characterising measures of lexical distributional similarity.
- Yorick Wilks and Mark Stevenson. 1998. The grammar of sense: using part-of speech tags as a first step in semantic disambiguation. *Natural Language Engineering*, 4(2):135–143.
- David Yarowsky and Radu Florian. 2002. Evaluating sense disambiguation performance across diverse parameter spaces. *Natural Language Engineering*, 8(4):293–310.