Word sense disambiguation using automatically acquired verbal preferences

John Carroll (johnca@cogs.susx.ac.uk) and Diana McCarthy (dianam@cogs.susx.ac.uk)
Cognitive & Computing Sciences, University of Sussex

Abstract.
The selectional preferences of verbal predicates are an important component of a computational lexicon. They have frequently been cited as being useful for WSD, alongside other sources of knowledge. We evaluate automatically acquired selectional preferences on the level playing field provided by SENSEVAL to examine to what extent they help in WSD.

Keywords: selectional preferences

Abbreviations: WSD – word sense disambiguation; ATCM – Association Tree Cut Model;POS – part-of-speech; SCF – subcategorization frame

1. Introduction

Selectional preferences have frequently been cited as being a useful source of information for WSD. It has however been noted that their use is limited (Resnik, 1997) and that additional sources of knowledge are required for full and accurate WSD. This paper outlines the use of automatically acquired preferences for WSD and evaluation of these at the SENSEVAL workshop.

The preferences are automatically acquired from raw text using the system described in section 2. The target data is disambiguated as described in section 2.4.

1.1. Scope

The preferences are obtained for the argument slots of verbal predicates where those slots involve noun phrases, i.e. subject, direct object and prepositional phrases. Preferences were not obtained in this instance for indirect objects since these are less common. The system has not at this stage been adapted for other relationships. For this reason disambiguation was only attempted on nouns occurring as argument heads in these slot positions. Moreover, preferences are only obtained where there is sufficient training data for the verb, (using a threshold of 10 instances). Disambiguation only takes place where the preferences

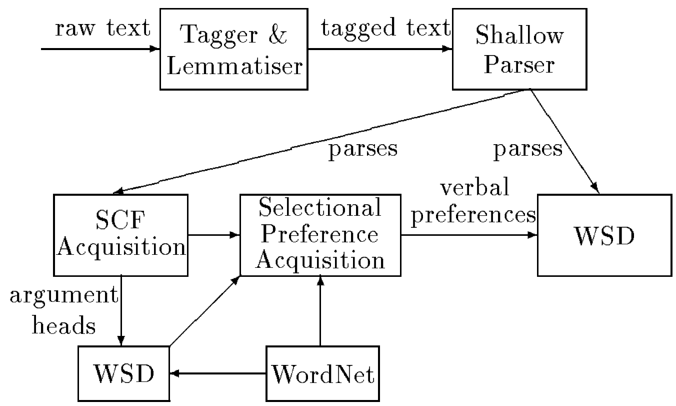

Figure 1. System Overview

are strong enough (above a threshold on the score representing preference strength) and where the preferences can discriminate between the senses. Proper nouns were neither used nor disambiguated. Some minor identification of multi-word expressions was performed since these items are easy to disambiguate and we would not want to use the preferences in these cases.

2. System Description

The system for acquisition is depicted in figure 1. Raw text is tagged and lemmatised and fed into the shallow parser. The output from this is then fed into the SCF acquisition system which produces argument head data alongside the SCF entries. From this argument head tuples consisting of the slot, verb (and preposition for prepositional phrase slots) and noun are fed to the preference acquisition module. To obtain the selectional preferences, 10.8 million words of parsed text from the BNC were used as training data. Some rudimentary WSD is performed on the nouns before preference acquisition. The selectional preference acquisition system then produces preferences for each verb and slot. These preferences are disjoint sets of WordNet (Miller et al., 1993b) noun classes, covering all WordNet nouns with a preference score attached to each class. The parser is then used on the target data and disambiguation is performed on target instances in argument head position. All these components are described in more detail below.

2.1. Shallow parser and SCF acquisition

The shallow parser takes text (re-)tagged by an HMM tagger (Elworthy, 1994), using the CLAWS-2 tagset (Garside et al., 1987), lemmatised with an enhanced version of the GATE system morphological analyser (Cunningham et al., 1995). The shallow parser and SCF acquisition are described in detail by Briscoe & Carroll 1997; briefly, the POS tag sequences are analysed by a definite clause grammar over POS and punctuation labels, the most plausible syntactic analysis (with respect to a training treebank derived from the SUSANNE corpus (Sampson, 1995)) being returned. Subject and (nominal and prepositional) complement heads of verbal predicates are then extracted from successful parses, and from parse failures sets of possible heads are extracted from any partial constituents found.

2.2. WSD of the argument head data

WSD of the input data seems to help preference acquisition itself (Ribas, 1995; McCarthy, 1997). We use a cheap and simple method using frequency data from the SemCor project (Miller et al., 1993a). The first sense of a word is selected provided that a) the sense has been seen more than three times, b) the predominant sense is seen more than twice as often as the second sense and c) the noun is not one of those identified as 'difficult' by the human taggers.

2.3. Selectional Preference Acquisition

The preferences are acquired using Abe and Li's method (Abe and Li, 1996) for obtaining preferences as sets of disjoint classes across the WordNet noun hypernym hierarchy. These classes are each assigned 'association scores' which indicate the degree of preference between the verb and class given the specified slot. The ATCM is collectively the set of classes with association scores provided for a verb. The association scores are given by $\frac{p(c|v)}{p(c)}$, where c is the class and v the verb. A small portion of an ATCM for the direct object slot of *eat* is depicted in figure 2. The verb forms are not disambiguated. The ambiguity of a verb form is reflected in the preferences given on the ATCM.

The models are produced using the minimum description length principle (Rissanen, 1978). This makes a compromise between a simple model and one which describes the data efficiently. To obtain the models the hypernym hierarchy is populated with frequency information from the data and the estimated probabilities are used for the calculations that compare the cost (in bits) of the model and the data when encoded in the model.

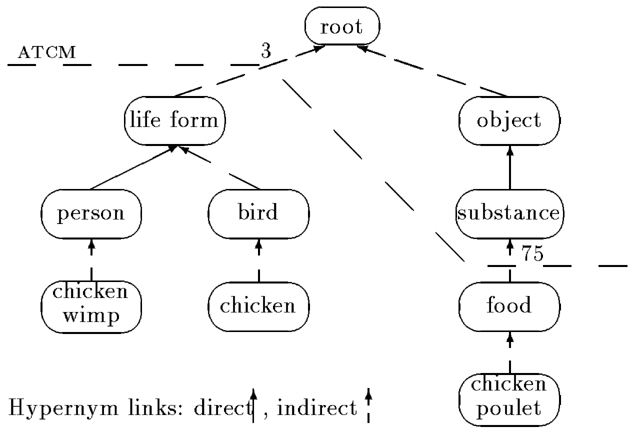

Figure 2. ATCM for *eat* Direct Object

2.4. Word Sense Disambiguation using Selectional Preferences

WSD using the ATCMs simply selects all senses for a noun that fall under the node in the cut with the highest association score with senses for this word. For example the sense of *chicken* under **FOOD** would be preferred over the senses under **LIFE FORM**, when occurring as the direct object of *eat*. The granularity of the WSD depends on how specific the cut is.

Target instances are disambiguated to a WordNet sense level. Each WordNet sense was mapped to the Hector senses required for SENSEVAL, using the mapping provided by the organisers.

3. Results

The preferences were only applied to nouns. For the all-nouns task fine-grained precision is 40.8% and recall 12.5%. The low recall is to be expected since many of the test items occur outside the argument head positions that we use. Coarse-grained precision is 56.2% and recall 17.2%. Performance is better when we look at the items which do not need disambiguation for POS. For these, coarse grained precision is 69.4% and recall 20.2%.

An important advantage of our approach is that our preferences do not require sense tagged data and so can perform the untrainable-nouns task. On the fine-grained untrainable-nouns task our system obtains 69.1% precision and 20.5% recall.

3.1. SOURCES OF ERROR

1. POS errors – These affect the parser. POS errors also contribute to the errors on the all-nouns task, where many of the items require POS disambiguation. 30% of the errors for *shake* were due to POS errors.

2. Parser errors – Preference acquisition in the training phase is subject to parser errors in identifying SCFs, although some of these would be filtered out as 'noise'. Errors in parsing the target data are more serious, since they might result in heads being identified incorrectly. Lack of coverage is also a problem: only 59% of the sentences in the target data were parsed successfully. Empirically, the grammar covers around 70–80% of general corpus text (Carroll and Briscoe, 1996), but the current disambiguation component appears to be rather inefficient since 15% of sentences fail due to being timed out. Data from parse failures is of lower quality since sets of possible heads are returned for each predicate, rather than just a single head.

3. multi-word expression identification – Many of the multi-word expressions were not detected due to easily correctable errors. This resulted in the preferences being applied where inappropriate.

4. errors arising from the mapping between WordNet and Hector.

5. thresholding – WordNet classes with a low prior probability are removed in the course of preference acquisition. Because of this, some senses are omitted from the outset.

6. preference errors – Other contextual factors should be taken into consideration as well as preferences. Our system does comparably (in terms of precision and recall) with other systems using verbal preferences alone.

4. Discussion

The results from SENSEVAL indicate that selectional preferences are not a panacea for WSD. A fully fledged system needs other knowledge sources. We contend that selectional preferences can help in situations where there are no other salient cues and the preference of the predicate for the argument is sufficiently strong.

One advantage of automatically acquired selectional preferences is that they do not require supervised training data. Although our system does use sense ranking from SemCor when acquiring the preferences, it can be used without this. Another advantage is that domain-specific preferences can be acquired without any manual intervention if further text of the same type as the target text is available.

SENSEVAL has allowed different WSD strategies to be compared on a level playing field. What is now needed is further comparative work to see the relative strengths and weaknesses of different approaches and to identify when and how complementary knowledge sources can be combined.

References

Abe, N. and H. Li: 1996, 'Learning Word Association Norms Using Tree Cut Pair Models.'. In: *Proceedings of the 13th International Conference on Machine Learning ICML*. pp. 3–11.

Briscoe, T. and J. Carroll: 1997, 'Automatic Extraction of Subcategorization from Corpora.'. In: *Fifth Applied Natural Language Processing Conference*. pp. 356–363.

Carroll, J. and E. Briscoe: 1996, 'Apportioning development effort in a probabilistic LR parsing system through evaluation'. In: *Proceedings of the 1st ACL SIGDAT Conference on Empirical Methods in Natural Language Processing*. pp. 92–100.

Cunningham, H., R. Gaizauskas, and Y. Wilks: 1995, 'A general architecture for text engineering (GATE) — a new approach to language R&D'. Technical Report CS-95-21, University of Sheffield, UK, Department of Computer Science.

Elworthy, D.: 1994, 'Does Baum-Welch re-estimation help taggers?'. In: *4th ACL Conference on Applied Natural Language Processing*.

Garside, R., G. Leech, and G. Sampson: 1987, *The computational analysis of English: A corpus-based approach*. Longman, London.

McCarthy, D.: 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*. pp. 52–61.

Miller, George, A., C. Leacock, R. Tengi, and R. T. Bunker: 1993a, 'A semantic concordance.'. In: *Proceedings of the ARPA Workshop on Human Language Technology*. pp. 303–308.

Miller, G., R. Beckwith, C. Felbaum, D. Gross, and K. Miller: 1993b, 'Introduction to WordNet: An On-Line Lexical Database'. ftp//clarity.princeton.edu/pub/WordNet/5papers.ps.

Resnik, P.: 1997, 'Selectional Preference and Sense Disambiguation.'. In: *Proceedings of Workshop Tagging Text with Lexical Semantics: Why What and How?* p. ??

Ribas, F.: 1995, 'On Acquiring Appropriate Selectional Restrictions from Corpora Using a Semantic Taxonomy.'. Ph.D. thesis, University of Catalonia.

Rissanen, J.: 1978, 'Modeling by Shortest Data Description.'. *Automatica* **14**, 465–471.

Sampson, G.: 1995, *English for the computer*. Oxford University Press.