Term Sense Disambiguation Using a Domain-Specific Thesaurus

Diana Maynard and Sophia Ananiadou
Dept. of Computing & Mathematics
Manchester Metropolitan University
Manchester, M1 5GD, U.K.

Abstract

Term extraction is important for many information systems applications. Although terms should be monoreferential, in reality they exhibit a high degree of ambiguity. This paper describes a method for automatic term sense disambiguation based on the identification of relevant contextual information from sublanguage corpora. We combine basic semantic roles derived from the corpus with domain-specific semantic categories from UMLS, a specialised medical thesaurus, and use an EBMT-based matching algorithm to compare related terms.

1 Introduction

The increasing volumes of electronic textual data and the development of new applications have given rise to a variety of disambiguation techniques. Term ambiguity can greatly reduce the efficiency of term extraction applications such as information extraction and retrieval, dictionary construction, automatic indexing, machine translation and hypertext linking.

Traditional methods of sense disambiguation involve comparing the distribution of the words in a context window with typical distributions of the word's senses to find the closest match [Gale et al.1992]. The use of lexical semantic information has largely been ignored for term disambiguation, although it is increasingly being used for other NLP applications.

We first extract potential terms and the relevant contextual information surrounding them. A linguistic filter is used to identify word classes which are expected to convey salient information about the term. Then semantic information is acquired from UMLS, a specialised electronic medical thesaurus. Finally, a semantics-based matching technique is used to compare the semantic features of term and context. This enables terms to be clustered together according to their semantics, and for different meanings of ambiguous terms to be identified.

2 Ambiguity in Terms

Terms are theoretically the embodiment of specialised concepts [Sager1990]. Ideally, each term should therefore refer to a specific concept, and each concept should be represented by only one term. In practice, however, terms exhibit a high degree of both polysemy and synonymy. The level of ambiguity can be reduced substantially by restricting the domain to a sublanguage, but ambiguous terms are by no means eliminated. The meaning of a concept should be characterised by its position in a network. However, this is not always the case in practice, since almost all hierarchies still suffer from the problem of multidimensionality, whereby the same concept can be classified in more than one way.

2.1 Types of Ambiguity

Some ambiguous terms are simply homographs, e.g. *tear*. This clearly has two unrelated meanings and represents two different concepts (a "rip" and "ocular fluid"). Other terms can be considered as hyponyms of different terms, and thus be represented in different parts of the hierarchy, even though they only refer to a single concept. For example, the term *hydroxycalciferol* only represents one concept, but can be a hyponym of either *steroid* or *vitamin*, and thus be assigned two different positions in the network.

Although multidimensionality is an inherent problem of hierarchical networks, it is really the result of a more cognitive issue - the way in which we view the world. For example, the term *football boot* can be seen as either an item of footwear or as a piece of sports equipment. Under which category should it therefore be placed in a hierarchical representation? Most probably, it would be placed under both, or there would be a link from one position to the other. How the term is best categorised might depend on the individual circumstance. For example, in a text about football it might be more useful to know that a football boot is an item of footwear; in a text about footwear it might be more useful to know that it is a piece of sports equipment, since this provides more specific information about the term. Since it is virtually impossible to avoid a multidimensional representation, we need to find some means of accessing the term in its most relevant location according to the situation.

2.2 Term sense disambiguation

In many ways, term sense disambiguation (TSD) is similar to the problem of word sense disambiguation (WSD). However, it differs in two main aspects. Firstly, terms are domain-specific, which means that general language resources and techniques may not be appropriate. Some work has been carried out on tuning lexicons to specific domains, e.g. [Basili et al.1995]. Secondly, the majority of technical terms consist of more than one word (the average length of NP terms is approximately 1.91 words, depending on the type of corpus). The ambiguity of multiword terms is generally not caused by different senses of the individual components of the term, but by different senses of the term as a whole. The different meanings of the term may be linked to different domains, but they may equally be present within a specific domain. Although domain-specific text is likely to contain the meanings of a term related to that domain, it does not rule out the possibility of the general meaning of that term being used as well. For example, the term *drug* found in a medical text could refer to either an illegal substance or simply to a kind of medicine. Similarly, the term *desktop* in a business text could refer to part of the furniture or to the layout of windows on a computer. Because of the differences between word sense ambiguity and term sense ambiguity, techniques applied to WSD are not always appropriate for TSD, as discussed in the next section.

The aim of our disambiguation procedure is not specifically to differentiate between different dictionary senses of terms, as is the case with most WSD systems. It is rather to attempt to establish different meanings or views of terms according to the context or situation in which they are being used.

3 Related Work

There has been an abundance of literature on different approaches to WSD, which we shall therefore not discuss in detail here. For pointers see [Ng and Zelle1997], [Chandrasekar and Bangalore1998], [Kilgarriff1997], [Resnik1997]. A variety of methods has been used, such as supervised learning from large corpora, e.g. [Yarowsky1995], multilingual approaches using the difference between languages. e.g. [Dagan et al.1991], statistical models, e.g. [Bruce and Wieber1994], and using linguistic and contextual information, e.g. [Ng and Lee1996].

Previous approaches to the extraction of semantic information have tended to use general language dictionaries or thesauri. One common method is to compare the content of the dictionary definition of a word with the words in the surrounding context. Lesk [Lesk1986] used an online dictionary to calculate the overlap between dictionary definitions of word senses. Smeaton and Quigley [Smeaton and Quigley1996] calculated word-word similarity between related words in a thesaurus. Yarowsky [Yarowsky1992] captured the meanings of word senses from semantic categories in Roget's Thesaurus.

The application of these methods to word sense disambiguation has been fairly successful. There are, however, two main reasons why they are not entirely suitable for our work. Firstly, they are used for word sense rather than term sense disambiguation. Secondly, they are designed to deal with general rather than domain-specific language. Technical terms are not covered sufficiently in any general language dictionary or thesaurus, and we propose that they really require a more specialised source of information.

An alternative approach to the acquisition of semantic information involves the use of a disambiguated corpus for training. Riloff and Lehnert [Riloff and Lehnert1992] developed an algorithm to derive relevancy cues from training texts, which they used for information extraction. Soderland et al. [Soderland et al.1995] also developed a system to identify concepts in a text, by means of linguistic features which reliably identified the conceptual content of a phrase. Grefenstette [Grefenstette1994] adopted a hybrid, knowledge-poor approach to the automatic extraction of semantic information from large corpora, using only syntactic information. The advantage of these techniques is that they require no dictionary or thesaurus and provide information specific to the corpus. However, they do require suitable large-scale corpora, which can be a problem.

4 Extracting Contextual Information

Contextual information has previously been used to identify synonyms, e.g. [Grefenstette1994], since two terms that share the same contextual environment are likely to have the same meaning. Conversely, different meanings of a term can also be identified in this way. If the contexts contain words belonging to different semantic families, it is likely that the term also has meanings belonging to different semantic families [Wolff-Terroine1977]. For example, the term *mass* can belong to a subset of medical terms (along with words such as *tumour* and *growth*), or to a subset of engineering terms (along with words such as *volume* and *weight*). The contexts in which the term is found can be compared in order to identify the different meanings of the term.

Previous work on identifying relevant contextual information has mostly been restricted to syntactic information [Frantzi and Ananiadou1997]. We propose that NLP of sublanguage texts really requires the incorporation of lexical semantic knowledge. Whilst semantic information has been extracted from dictionaries and corpora for applications such as information extraction, dictionary construction and knowledge acquisition [Riloff and Lehnert1992], [Grefenstette1994], [Reinert1986], it has largely been ignored for term disambiguation.

A method for term sense disambiguation relies on having a suitable means of extracting terms. Our approach uses a method for multi-word automatic term recognition called NC-value [Frantzi and Ananiadou1997]. Potential terms are first extracted from a corpus and ranked using the C-value method [Frantzi and Ananiadou1996], which is based on frequency of occurrence and term length. Contextual information is then incorporated into the algorithm, in the form of weights based on the statistical characteristics of the context words. The NC-value combines the original C-value with this set of weights to produce a reranking of the list of terms.

4.1 Using a Linguistic Filter

The performance of term extraction techniques has been improved by incorporating a linguistic filter to restrict the context window [Daille et al.1994], [Frantzi and Ananiadou1997], even if this is nothing more than a very simple filter based on shallow a priori knowledge.

In the same way, our technique might be optimised by identifying and comparing only the salient parts of the context. Radford et al.'s dynamic matching algorithm [Radford et al.1995] takes into account the order of the words within the context window, but not the type. Whilst this improves the efficiency, it is computationally expensive, even though semantic information is ignored.

When comparing contexts, it is evident that some types of information are more useful than others. The fact that two nouns are similar is generally far more interesting than the fact that two determiners are similar. Some matching techniques take this into account by assigning a weighting to different parts of speech [Zhao1996]. For our purposes, "empty" words such as determiners may not be useful at all, and can therefore be discarded before the matching process in order to maximise efficiency. We initially adopt the filter used in the NC-Value, i.e. the only context words allowed are adjectives, nouns and verbs.

5 Acquisition of Semantic Knowledge

There are two main ways of acquiring semantic information: by extracting it from dictionaries or thesauri, or by using knowledge-based or statistical techniques to extract it directly from the corpus. The latter has the advantages that real examples of language are collected, the information will be tailored to the domain for a specific sublanguage, and statistical information can be gathered which may be used to determine relevance. However, the techniques used may be unreliable or simply provide inadequate coverage, frequency information may be misleading, particularly in small or domain-specific corpora, and the type of information sought may not be available or easily extracted. Using dictionaries, on the other hand, ensures that the information is accurate, that the coverage is neutral, and that the information is unambiguous. There may be problems, though, in finding a suitable dictionary containing the necessary information, particularly for a sublanguage, and in actually extracting the relevant information.

Our solution to the problem of acquiring semantic knowledge is to combine the two methods, by extracting general information about semantic roles from the corpus, and augmenting this with domain-specific information about semantic categories from a specialised thesaurus. The first part has not yet been fully implemented, but is achieved by using an HPSG parser to extract information about semantic roles from the corpus. For the second part, we use the UMLS medical thesaurus to assign semantic categories to instances of concepts and to identify relationships between them.

5.1 Acquiring Information from the UMLS Thesaurus

UMLS, the Unified Medical Language System [NLM1997], is a set of knowledge sources containing information about medical terminology, organised in a hierarchical structure. Not only does it provide a classification system for the terms, but it also contains morphological, syntactic and semantic information. The Semantic Network contains additional information about the relations between the semantic classes to which the terms are assigned.

We have modified the hierarchy slightly in order to produce a more coarse-grained set of semantic types. This enables us to make more generalisations about terms. We use a corpus tagged with syntactic categories (using the Brill tagger [Brill1992]) and extract terms using Frantzi et al.'s ATR method [Frantzi and Ananiadou1997]. Each term and context word is then tagged with its semantic type from the UMLS Metathesaurus. For example, the term *actinic keratosis* will be tagged with the semantic type *acquired abnormality*.

```
ACTINIC KERATOSIS|ACQUIRED ABNORMALITY
```

Relational information about the semantic types can also be added. This information is generic rather than specific, i.e. details are provided about classes of terms rather than about individual terms, and do not necessarily hold for every member of that class. For example, a relationship exists between the semantic classes of *disease or syndrome* and *acquired abnormality*:

```
DISEASE OR SYNDROME|RESULT OF|ACQUIRED ABNORMALITY
```

However, although *acne* is classified as a *disease or syndrome*, and *actinic keratosis* is classified as an *acquired abnormality*, it does not necessarily mean that there is a relationship between *acne* and *actinic keratosis* such that *acne* is the result of *actinic keratosis*. All that can be inferred is that diseases or syndromes in general can be the result of acquired abnormalities.

6 Calculating Word Similarity

Methods for calculating similarity are mainly based on the principle of interchangeability [Uramoto1994]. If two words or terms are interchangeable within a particular context, they should have a high degree of similarity. We extend this idea by assuming that instances of a term occurring within similar linguistic environments will have the same meaning and will represent a single instance of a concept in a hierarchy. Similarity measures have been used in a range of applications such as EBMT, information retrieval, information extraction and word sense disambiguation. Similarity may be measured according to different criteria, such as syntactic category, subcategorisation information, semantic features or even pragmatic features. So two words might be syntactically interchangeable in that they share the same part of speech but have different meanings, or semantically interchangeable in that they have the same meaning but different parts of speech.

6.1 EBMT methods

EBMT aims to select the most similar example for a given problem, by defining a set of relevant examples and comparing them with the problem in terms of context and configuration. Most of these comparison methods involve using some kind of semantic distance measure, based on the relative positions of the two items within a hierarchical network. Our approach is modelled on techniques used in EBMT [Zhao1996], [Sumita and Iida1991]. By defining a hierarchically-based representation, we can use a similarity measurement based on semantic distance. Like Grefenstette, we measure similarity in terms of shared attributes, although we extend this idea by comparing the similarity of the attributes rather than just calculating the number of attributes in common.

In its simplest form, similarity in a hierarchy is measured by edge-counting – the shorter the distance between the words, the greater their similarity. A more reliable measurement is the Most Specific Common Abstraction (MSCA) method [Kolodner1993], which involves tracing the respective paths of the two words back up the hierarchy until a common ancestor is found, and then measuring the average distance from node to MSCA. The shorter the distance to the MSCA, the more similar the two words. The position of the MSCA within the hierarchy is also important. The lower down in the hierarchy the MSCA, the more specific it is and therefore the more information is shared by the two concepts, thus making them more similar.

6.2 A Measure for Semantic Distance

Our method combines both of these factors by defining a **positional** weight, the vertical position of the MSCA within the hierarchy, and a **commonality** weight, which is in essence the horizontal distance between the two nodes, in other words the number of shared common ancestors. It is intuitive that the greater the horizontal and/or vertical distance between nodes, the less similar they are. The importance of the position of the MSCA is less immediately apparent, but consider the following. Intuitively we would say that *apple* and *pear* are more similar than *fruit* and *vegetable*. The reason for this is that the first pair of words is more specific than the second pair, i.e. there is more information relevant to both of them. This in turn means that the two words share more attributes and therefore are more similar. The fact that the words are more specific is represented by their depth in the hierarchy.

Similarity is calculated in a very simple but effective way, multiplying the commonality weight by the number of words being considered, and dividing this by the positional weight. This produces a number between 0 and 1, where 1 is the case where the two nodes are identical. A similarity of 0 would only occur in the case where there is no common ancestor to both nodes. We prevent this from occurring by simulating a unique root node, because we believe that there is a degree of relatedness between all nodes in the hierarchy, and that it should not be divided into separate unrelated sections, as in the original UMLS hierarchy.

The commonality weight is calculated by the number of nodes in common, whilst the positional weight is the sum of the depths of the nodes. Because of the way in which the nodes are coded in the UMLS Semantic Network, this makes calculation very easy since the number of digits in the code represents the depth of that

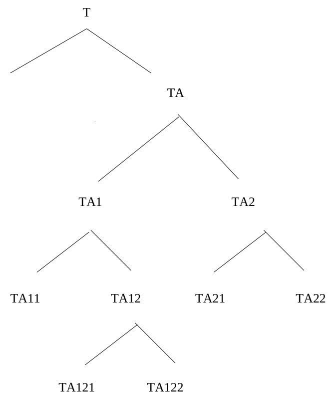

Figure 1: Fragment of the Semantic Network

node. By adding an extra digit to the code, we create an artificial unique root node, as mentioned earlier. The formula is thus the following:

$$\text{sim}(w_1...w_n) = \frac{n(\text{com}(w_1...w_n))}{\text{pos}(w_1...w_n)} \tag{1}$$

where
n is the number of words being compared
$\text{com}(w_1...w_n)$ is the commonality weight of words 1...n
$\text{pos}(w_1...w_n)$ is the positional weight of words 1...n

Suppose we have the term *adherent leucoma* occurring with the context term *haemorrhage*. We first find the semantic type of each and the hierarchy codes associated with the semantic types:

```
adherent leucoma|anatomical abnormality|TA122
haemorrhage|sign or symptom|TA22|
```

Then we calculate the similarity between the two:
$n = 2$
$\text{com}(w_1, w_2) = 2$(nodes T and TA are common to both)
$\text{pos}(w_1, w_2) = 5 + 4 = 9$
$\text{sim}(w_1, w_2) = 2$ x $2/9 = 0.444444$

7 Results

In the UMLS Metathesaurus, ambiguous terms are generally allocated more than one semantic type (except in the case where the two meanings are very closely related and fall into the same general semantic category). Using the similarity measure allows us to differentiate between these meanings, since the semantic type of one reading of the term will generally be closer than the other to the semantic type of the context word. This is reflected in the total weights allocated to the terms. The total weight is the sum of the similarity scores between that term and each context word occurring with it. So in general, ambiguous terms should be allocated a different total weight for each meaning of the term. If the two meanings are closely related, they may end up with the same or a similar total weight. Some examples of these are shown in Table 1.

Term	Weight	Term	Weight
cholesterol granuloma(1)	2.000001	cholesterol granuloma(2)	0.600000
cirsoid aneurysm(1)	2.012121	cirsoid aneurysm(2)	1.248484
corneal opacities(1)	9.416671	corneal opacities(2)	2.822222
corneal opacity(1)	13.916670	corneal opacity(2)	6.088888
degenerative change(1)	20.727782	degenerative change(2)	14.723231
graft rejection(1)	7.022220	graft rejection(2)	2.888888
hypermature cataract(1)	5.666666	hypermature cataract(2)	6.666666
intraocular lens(1)	2.16429	intraocular lens(2)	2.16429
macular corneal dystrophy(1)	1.400000	macular corneal dystrophy(2)	1.400000
peter's anomaly(1)	2.800000	peter's anomaly(2)	3.200000
plasma cell infiltration(1)	0.666667	plasma cell infiltration(2)	0.200000
retinal dysplasia(1)	6.000000	retinal dysplasia(2)	3.200000
retinal haemorrhage(1)	29.322229	retinal haemorrhage(2)	13.638381
senile keratosis(1)	14.446464	senile keratosis(2)	8.808078
vascular engorgement(1)	1.777778	vascular engorgement(2)	2.200000

Table 1: Similarity weights of some ambiguous terms

8 Discussion and Conclusion

Although the results we have presented are only preliminary, our aim is to show how we can use semantic information to help identify different meanings of multi-word terms. By using a similarity measure based on a thesaural hierarchy, we measure the closeness of instances of terms to information in the context which is considered relevant to that term. Further work remains to be carried out on the incorporation of more linguistic information into the algorithm, and complete evaluation of our method.

We have presented here the outline of a semantic approach to term sense disambiguation. Domain-specific information is extracted from a medical thesaurus and a similarity measure is then used to compare the linguistic features of the contexts in which instances of terms are found. This enables a weight to be produced for each meaning of the term. The objective is not so much to perform disambiguation in terms of dictionary senses, but rather in terms of conceptual views, in order to access the terms in their most relevant position in a hierarchy.

The approach adopted is domain-specific and knowledge-rich, although the techniques used are portable to other domains. Using a sublanguage corpus ensures a higher concentration of terms, and imposes greater restrictions at both semantic and syntactic levels. Whilst this narrows the interpretation of linguistic patterns and reduces ambiguity to a certain degree, domain-specific resources are still necessary since general language ones do not contain enough specialised information. Although knowledge-poor approaches have been successful for general language corpora, we propose that the incorporation of semantic knowledge is a useful component for term sense disambiguation in sublanguages.

References

[Basili et al.1995] R. Basili, M. Della Rocca, M.T. Pazienza, and P. Velardi. 1995. Contexts and categories: Tuning a general purpose verbclassification to sublanguages. In *Proc. of International Conference on Recent Advances in Natural Language Processing*, Tzigov Chark, Bulgaria.

[Brill1992] E. Brill. 1992. A simple rule-based part of speech tagger. In *Proc. of 3rd Conference of Applied Natural Language Processing*.

[Bruce and Wieber1994] R. Bruce and J. Wieber. 1994. Word-sense disambiguation using decomposable models. In *Proc. of 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico, USA.

[Chandrasekar and Bangalore1998] R. Chandrasekar and S. Bangalore. 1998. Knowing a word by the company it keeps using local information in a maximum entropy model for word sense disambiguation. Submitted for publication.

[Dagan et al.1991] I. Dagan, A. Itai, and U. Schwall. 1991. Two languages are more informative than one. In *Proc. of the Association for Computational Linguistics (ACL)*, pages 130–137.

[Daille et al.1994] B. Daille, E. Gaussier, and J.M. Langé. 1994. Towards automatic extraction of monolingual and bilingual terminology. In *Proc. of COLING 94*, pages 515–521.

[Frantzi and Ananiadou1996] K.T. Frantzi and S. Ananiadou. 1996. A hybrid approach to term recognition. In *Proceedings of NLP+IA 96*, volume 1, pages 93–98, Moncton, Canada, June.

[Frantzi and Ananiadou1997] K.T. Frantzi and S. Ananiadou. 1997. Automatic term recognition using contextual cues. In *Proceedings of 3rd DELOS Workshop.*

[Gale et al.1992] W.A. Gale, K.W. Church, and D. Yarowsky. 1992. Work on statistical methods for word sense disambiguation. In *Probabilistic Approaches to Natural Language: Papers from the 1992 Fall Symposium*, pages 54–60, Cambridge, Massachussetts. AAAI Press.

[Grefenstette1994] G. Grefenstette. 1994. *Explorations in Automatic Thesaurus Discovery.* Kluwer Academic Publishers.

[Kilgarriff1997] A. Kilgarriff. 1997. Evaluating word sense disambiguation programs: progress report. In *Proc. of SALT Workshop in Speech and Language Technology*, pages 114–120, Sheffield University, UK.

[Kolodner1993] J.L. Kolodner. 1993. *Case-Based Reasoning.* Morgan Kaufmann Publishers Inc., San Mateo, California.

[Lesk1986] M. Lesk. 1986. Automatic sense disambiguation: how to tell a pine cone from an ice cream cone. In *Proc. of the SIGDOC Conference*, pages 24–26.

[Ng and Lee1996] H.T. Ng and H.B. Lee. 1996. Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach. In *Proc. of 34th Annual Meeting of the Association for Computational Linguistics*, pages 40–47, Santa Cruz, California.

[Ng and Zelle1997] H.T. Ng and J. Zelle. 1997. Corpus-based approaches to semantic interpretation in NLP. *AI Magazine*, 18(4):45–64.

[NLM1997] NLM, U.S. Dept. of Health and Human Services, 1997. *UMLS Knowledge Sources*, 8th edition, January.

[Radford et al.1995] I. Radford, V.M. Arranz, S. Ananiadou, and J. Tsujii. 1995. Dynamic context matching for knowledge acquisition from small corpora. In S. Bolasco, L. Lebart, and A. Salem, editors, *Analisi Statistica dei Dati Testuali*, volume 1, pages 321–239, Rome, Italy. CISU.

[Reinert1986] M. Reinert. 1986. Un logiciel d'analyse lexicale (alceste). *Cahiers de l'Analyse des Donnees*, 4.

[Resnik1997] P. Resnik. 1997. A perspective on word sense disambiguation methods and their evaluation. In *Proc. of ACL-SIGLEX Workshop "Tagging Text with Lexical Semantics: Why, What and How?"*, Washington, D.C.

[Riloff and Lehnert1992] E. Riloff and W. Lehnert. 1992. Classifying texts using relevancy signatures. In *Proc. of AAAI.*

[Sager1990] J.C. Sager. 1990. *A Practical Course in Terminology Processing.* John Benjamins.

[Smeaton and Quigley1996] A. Smeaton and I. Quigley. 1996. Experiments on using semantic distances between words in image caption retrieval. In *Proc. of the 19th International Conference on Research and Development in Information Retrieval*, Zurich, Switzerland.

[Soderland et al.1995] S. Soderland, D. Fisher, J. Aseltine, and W. Lehnert. 1995. CRYSTAL: Inducing a conceptual dictionary. In *Proc. of the 14th International Joint Conference on Artificial Intelligence.*

[Sumita and Iida1991] E. Sumita and H. Iida. 1991. Experiments and prospects of example-based machine translation. In *Proc. of 29th Annual Meeting of the Association for Computational Linguistics*, pages 185–192, Berkeley, California.

[Uramoto1994] N. Uramoto. 1994. A best-match algorithm for broad-coverage example-based disambiguation. In *Proc. of 15th International Conference on Computational Linguistics*, volume 2, pages 717–721, Kyoto, Japan.

[Wolff-Terroine1977] M. Wolff-Terroine. 1977. Terminology and nomenclatures. In W. Schneider and A.L. Sagvall Hein, editors, *Compuational Linguistics in Medicine*, pages 55–61. North Holland Publishing Company.

[Yarowsky1992] D. Yarowsky. 1992. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In *Proc. of 14th International Conference on Computational Linguistics*, pages 454–460.

[Yarowsky1995] D. Yarowsky. 1995. *Three Machine Learning Algorithms for Lexical Ambiguity Resolution.* Ph.D. thesis, University of Pennsylvania.

[Zhao1996] G. Zhao. 1996. *Analogical Translator: Experience-Guided Transfer in Machine Translation.* Ph.D. thesis, Dept. of Language Engineering, UMIST, Manchester, England.