

The SENSEVAL-2 Panel on Domains, Topics and Senses

Paul Buitelaar

DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbruecken, Germany
paulb@dfki.de

1 Introduction: Why Domains Matter in Sense Disambiguation

An important aspect of sense disambiguation is the wider semantic space (domain, topic) in which the ambiguous word occurs. This may be most clearly illustrated by some cross-lingual examples, as they would appear in (machine) translation. Consider for instance the English word *housing*. In a more general “sense”, this translates in German into *Wohnung*. In an engineering setting however it translates into *Gehäuse*. Also verbs may be translated differently (i.e. have a different sense) according to the semantic space in which they occur. For instance, English *warming up* translates into *erhitzen* in a more general sense, but into *aufwärmen* in the sports domain.

Because of the apparent relevance then of domains or topics on sense disambiguation, a panel was organized at SENSEVAL-2 to discuss some current and previous work in this area. The paper presents a more extended overview based on the relevant literature, besides giving a summary of the discussion that developed after the panel presentations.

2 Domains, Topics and Senses

2.1 Subject Codes

A semantic space may be indicated in a dictionary by use of a so-called “subject code”. In LDOCE for instance, subject codes like MD, for the medical domain, or ML, for meteorology are used to define which senses of a word are used in which domains. Three of the senses of the word *high* for instance correspond to three different domains: music (a high tone), drugs (the experience of being high) and meteorology (a high pressure area).

Subject codes can be used to detect the topic of a text segment by simply counting their frequency over all content words (Walker and Amsler 1986). At the same time, however, subject codes can be used in sense disambiguation by constructing topic specific context models (Guthrie et. al 1991). Such “neighborhoods” can be constructed by taking into account all words in the definitions and in sample sentences of all words in the dictionary that share the same subject code. For instance, the word *bank* has the following neighborhoods for the financial and medical domains:

<i>write</i>	<i>safe</i>	<i>sum</i>
<i>account</i>	<i>person</i>	<i>put</i>
<i>take</i>	<i>money</i>	<i>order</i>
<i>keep</i>	<i>pay</i>	<i>supply</i>
<i>paper</i>	<i>draw</i>	<i>cheque</i>

Table 1: Financial neighborhood of *bank*

<i>medicine</i>	<i>product</i>	<i>hold</i>
<i>origin</i>	<i>place</i>	<i>human</i>
<i>treatment</i>	<i>blood</i>	<i>hospital</i>
<i>use</i>	<i>store</i>	
<i>organ</i>	<i>comb</i>	

Table 2: Medical neighborhood of *bank*

Using subject codes in sense disambiguation has been shown to be fruitful, relative to using other sources of knowledge. As reported in (Stevenson and Wilks 1999), the performance of using only subject codes (79% precision) was much better than that of using only dictionary definition words (65%), or selection restrictions (44%). Given these results it seems worthwhile to identify also the semantic space of WordNet synsets more explicitly by the introduction of subject codes (Magnini and Cavaglià 2000). This allows for grouping together synsets across part-of-speech, as in the medical domain

(doctor#1, hospital#1; operate#7) and across sub-hierarchies, as in the sports domain (life_form#1: athlete#1; physical_object#1: game_equipment#1; act#2: sport#1; location#1: playing_field#1).

2.2 Topic Signatures and Variation

The topic specific context models as constructed by (Guthrie et al. 1991) can be viewed as “signatures” of the topic in question. Such topic signatures can, however, be constructed even without the use of subject codes by generating them (semi-) automatically from a lexical resource and then validating them on topic specific corpora (Hearst and Schütze 1993).

An extension of this idea is to treat senses, or rather WordNet synsets, as topics for which a signature can be constructed. One approach to this is to retrieve relevant documents through search engines on the web by defining queries for each synset (Agirre et al. 2000, Agirre et al. 2001). For instance, the following query can be defined for the first WordNet sense of *boy*:

- #1 (*boy* AND (*altar boy* OR *ball boy* OR ...))
- #2 AND NOT (*man* OR ... OR *broth of a boy* OR
- #3 *son* OR ... OR *mama's boy* OR
- #4 *nigger* OR ... OR *black*)

The document collections retrieved are then analysed and a list of the most relevant words for each synset is generated as its topic signature. Examples (abridged) for the first three senses of *boy* are:

Sense 1	Sense 2	Sense 3
child	gay	human
Child	reference	son
person	tpd-results	Human

Constructing topic signatures for senses implies that a dominant sense can be identified given a certain topic or domain. This may be true for clearly ambiguous words (i.e in the case of homonymy). For instance, *sentence* will be dominant in the judicial sense in the law domain and in the syntactic sense in the linguistics domain. However, for words with related senses (i.e in the case of systematic polysemy) the topic signatures will overlap, as with the results on *boy* in sense 1: *young male person* and sense 3: *son*. This has been shown also from a somewhat different viewpoint in reaction to (Gale et al. 1992), in which it was stated that one sense will

be uniquely used within a discourse (which we can equate with a topic or domain for our purposes here). Instead, many words have overlapping senses that will be used simultaneously throughout one discourse (Krovetz 1998).

The main question that remains now is, what exactly constitutes a discourse / subject / topic / domain? We can get closer at answering this question by looking at some empirical sense disambiguation results that involve a variation of topic. More specifically, we can observe some effects of topic variation by training a sense disambiguation system on one topic and applying it to another. For instance, training on Wall Street Journal while testing on SemCor and vice versa shows a degrading of 12% and 19% in precision (Escudero et al. 2000). On the other hand, applying context information (collocations) extracted from Wall Street Journal to a financial text in SemCor shows significantly higher precision than on texts in other domains in SemCor (Martinez and Agirre 2000).

These results therefore suggest that a discourse / subject / topic / domain corresponds to a larger or smaller chunk of text (a corpus, a text or a text segment) with a homogeneous distribution of senses and corresponding collocations.

2.3 Tuning

But even with a clearly defined domain, it is far from certain that any general sense inventory will be appropriate. “The usual scenario ... has been that the word senses are taken from a general purpose dictionary, ... whereas the material to be disambiguated is ... Wall Street Journal. ... So, the profiles [Signatures, Collocations] ... will be for general English senses according to the WSJ ...” (Kilgarriff 1998). Instead, a general sense inventory needs to be tuned to the domain at hand. This involves selecting only those senses that are most appropriate for the domain, as well as extending the sense inventory with novel words (terms) and novel senses, specific to the domain (Basili et al. 1997; Cucchiarelli and Velardi 1998; Turcato et al. 2000; Buitelaar and Sacaleanu 2001; Vossen 2001).

According to the method described in (Cucchiarelli and Velardi 1998), a domain specific sense inventory that is balanced (even distribution of words to senses) and at the right

level of abstraction (ambiguity vs. generalization) can be selected automatically given the following criteria: "Generality", "Discrimination Power", "Domain Coverage" and "Average Ambiguity." Applying these criteria in a quantitative way to a general sense inventory (i.e the WordNet hierarchy) and a given domain specific corpus automatically selects a set of relevant categories (i.e. top level synsets). For instance, this method selects following categories for the financial domain:

person, individual,...
instrumentality,...
written_communication,...
possession,...

Only senses that are subsumed by these categories are included in the domain specific sense inventory. For instance, for the word *stock*, only 5 out of 16 senses are selected:

- #1 capital > asset > possession
- #2 support > device > instrumentality
- #4 document > ... > written_communication
- #5 accumulation > asset > possession
- #6 ancestor > relative > person,individual

Senses that are discarded include:

- #7 soup > ...
- #9 plant_part > ...
- #12 lineage,line,line_of_descent > ...
- #14 lumber,timber > ...

The method described above uses a top down approach that propagates the domain relevance of certain top level synsets down through the (WordNet) hierarchy. A somewhat different approach would be to assign a domain relevance to each concept (i.e. word sense, synset) from the bottom up (Buitelaar and Sacaleanu 2001). This method determines the domain specific relevance of (WordNet, GermaNet) synsets on the basis of the relevance of their constituent synonyms that co-occur within representative domain corpora.

Next to selecting domain relevant concepts from the general sense inventory, novel terms (those not covered by the sense inventory) need to be accounted for also. This includes adding morphological and syntactic variants of known terms (Vossen 2001) as well as extending the inventory with semantically

related terms through classification and/or clustering.

3 Panel Discussion

In the panel presentations most of the issues discussed above were addressed. Central to the discussion were the following two questions:

- Is generic sense disambiguation possible?
- Is sense disambiguation always necessary?

The first question concerns the influence of the semantic space (topic, domain, etc.) on the disambiguation process. Unlike with PoS tagging, it seems hard and perhaps even theoretically impossible to define a 'general' training corpus and sense inventory for sense disambiguation. Instead, it seems necessary to tightly connect sense disambiguation to topic detection or text classification in order to recognize the wider semantic space of ambiguous words. The second question is concerned with the even more fundamental observation that sense disambiguation is unnecessary if one sense (or more than one, in the case of systematic polysemy) can be assigned unambiguously within a certain semantic space. The disambiguation problem then shifts towards an appropriate modelling of such semantic spaces (i.e. domain modelling). In summary, it may not be feasible to separate sense disambiguation from the domain in which it operates, which in turn implies that modelling this domain is the first priority for sense disambiguation. In the discussion, however, several arguments were raised against such a view of sense disambiguation.

First of all, such an approach drives us back to earlier domain specific methods. These were not very robust and required major efforts in adapting to new domains. As a counter argument to this point, it was noted that there are now many robust, machine-learning based methods available for lexical acquisition, which would allow for a rapid adaptation of the disambiguation resources to a new domain. The second main issue raised was that, from an evaluation point of view, it is important to evaluate the performance of different algorithms, independent from a specific domain or application. As a counter argument to this, the question was asked what such an evaluation

would then prove. Sense disambiguation evaluated without a particular (application) domain can only show an artificial result which is hard to interpret and to generalize over. This is illustrated in particular by low interannotator agreement scores obtained when disambiguating without the context of a certain domain.

The discussion did not reach a consensus on these points, although there was general agreement that future evaluation efforts in sense disambiguation should take applications (and hence certain domains) into account. The following table gives an overview of those teams that participated at SENSEVAL-2 and declared to be using domains, topical context or the „One Sense per Discourse“ heuristic.

Team	Domain Information	Topical Context	One Sense / Discourse
Lexical Sample Task (English)			
<i>IRST</i>	✓		
<i>TALP</i>	✓	✓	
<i>BCU-EHU</i>		✓	
<i>KUNLP</i>		✓	
All Words Task (English)			
<i>IRST</i>	✓		
<i>BCU-EHU</i>		✓	
<i>Sheffield</i>		✓	
<i>Sussex</i>			✓
<i>UCLA</i>			✓

On the lexical sample task, *KUNLP* and *TALP* had both high precision and recall, while *BCU-EHU* and *IRST* reached the highest precision of all participating systems, but at a low recall. On the all words task, all teams in the table scored average to low, except for *IRST*, which reached again a very high precision at a low recall.

These results are unfortunately still inconclusive about the general merit of domain and topic information. Only the anomalous results of *IRST* may indicate the advantage of domain information for reaching a high precision in sense disambiguation.

4 Acknowledgements

Many thanks to Eneko Agirre, Nancy Ide, Bernardo Magnini and Piek Vossen for their contributions to the panel, and to the SENSEVAL-2 audience for their active participation in the discussion. This research has in part been supported by EC/NSF grant IST-1999-11438 for the MUCHMORE project.

References

- Agirre E., Ansa O., Hovy E., Martinez D. *Enriching very large ontologies using the WWW*. In: Proceedings of the Ontology Learning Workshop ECAI 2000.
- Agirre E., Ansa O., Martinez D., Hovy E. *Enriching WordNet concepts with topic signatures*. In: Proceedings NAACL WordNet Workshop, 2001.
- Basili R., Della Rocca M., Pazienza M.-T. *Contextual Word Sense Tuning and Disambiguation*. Applied Artificial Intelligence, vol. 11, 1997.
- Buitelaar P., Sacaleanu B. *Ranking and Selecting Synsets by Domain Relevance*. In: Proceedings NAACL WordNet Workshop, 2001.
- Cucchiarelli A., Velardi P. *Finding a Domain-Appropriate Sense Inventory for Semantically Tagging a Corpus*. In: Journal of Natural Language Engineering, 1998
- Escudero G., Màrquez L., Rigau G. *An Empirical Study of the Domain Dependence of Supervised Word Sense Disambiguation Systems*. In: EMNLP 2000.
- Gale W., Church K., Yarowsky D. *One Sense per Discourse*. In: Proceedings of the 4th DARPA Speech and Natural Language Workshop, 1992.
- Hearst M., Schütze H. *Customizing a Lexicon to Better Suit a Computational Task*. In: Proceedings ACL SIGLEX Workshop 1993.
- Guthrie J. A., Guthrie I., Wilks Y., Aidinejad H. *Subject Dependent Co-Occurrence and Word Sense Disambiguation*. In: Proceedings of ACL 1991.
- Kilgarriff A. *Bridging the gap between lexicon and corpus: convergence of formalisms*. In: Proceedings of LREC Workshop on Adapting Lexical Resources, 1998.
- Krovetz R. *More than one sense per discourse*. NEC Research Memorandum, 1998.
- Magnini B., Cavaglia G. *Integrating Subject Field Codes into WordNet*. In: Proceedings LREC 2000.
- Martinez D., Agirre E. *One Sense per Collocation and Genre/Topic Variations*. In: Proceedings EMNLP 2000.
- Stevenson M., Wilks Y. *Combining Weak Knowledge Sources for Sense Disambiguation*. In: Proceedings IJCAI 1999.
- Turcato D., Popowich F., Toole J., Fass D., Nicholson D., Tisher G. *Adapting a synonym database to specific domains*. In: Proceedings of the ACL workshop on recent advances in NLP and IR. Hong Kong, 2000.
- Vossen P. *Extending, Trimming and Fusing WordNet for Technical Documents*. In: Proceedings NAACL WordNet Workshop, 2001.
- Walker D., Amsler R. *The use of machinereadable dictionaries in sublanguage analysis*. In: Analyzing Language in Restricted Domains, 1986.