Evaluation of GermanNet: Problems Using GermaNet for Automatic Word Sense Disambiguation

Jahn-Takeshi Saito, Joachim Wagner, Graham Katz, Philip Reuter, Michael Burke, Sabine Reinhard

Institute for Cognitive Science
University of Osnabrück
49074 Osnabrück
Germany

Abstract

WordNets such as GermaNet have been frequently used as an inventory of word-senses for word-sense disambiguation tasks. In the work described here we evaluate the adequacy of GermaNet for this task. That is we attempt to determine the degree to which GermaNet provides an adequate inventory of senses for word-sense annotation of running text. Our findings were on the whole very encouraging. GermaNet provides an appropriate sense for 83 % of the content words in our texts. More interestingly, an error analysis showed that simple morphological processing could significantly improve coverage.

1. Introduction

The use of WordNet for sense tagging of English is by now an established research program (Miller, et.al 1994; Resnik 1998; Landes, Leacock & Tengi 1998). With the advent of WordNet-style lexical resources for languages other than English (Bloksma, Díez-Orzas & Vossen 1996) the application of these resources, to sense-tagging for these languages is a natural evolution. A number of questions arise in this context, however. While the original WordNet has been used with success for English, there is no guarantee that this experience generalizes to other WordNets for other languages. Both the language itself and the particular WordNet developed for it may present problems that were not present in the WordNet/English case. Our goal here was to evaluate how useful GermaNet is as a resource for word sense tagging for German.

Our task, then, was to annotate a corpus of German text using GermaNet and to determine how close to the ideal of providing an appropriate sense tag for all content words in the corpus GermaNet is. This is of interest both as an evaluation of GermaNet itself, and also because German and English differ in ways that, *a priori*, might indicate that German would be a difficult language to sense-tag (Hamp & Feldweg 1997). German has, for example, highly productive word-formation processes and a rich derivational morphology.

In form, however, our work was very similar to that done for English by Landes, Leacock & Tengi (1998) in that we simply set out to manually disambiguated words in a corpus, tagging each appearance of a content word in the corpus. In contrast to their work, we developed our own (German language) corpus and used GermaNet as our repository of word senses. Additionally, work was separate from the development of GermaNet and we did not have contact with GermaNet lexicographers.

2. GermaNet

GermaNet is a lexical-semantic net based on the WordNet example (Kunze & Wagner 1999a). It is intended to cover the basic vocabulary of German. Although GermaNet relies on the design principles and shares the same database structure as the Princeton WordNet (Miller 1990), it is build from scratch and features some modifications. In contrast to WordNet, GermaNet includes non-lexicalized *artificial* concepts to fill lexical gaps (e.g. to provide the missing antonym for *thirsty*) and to avoid unjustified co-hyponomy. Additionally, cross-classification of concepts, which is seldom used in WordNet, is an essential feature of GermaNet, and regular polysemy is integrated via a special relation between synsets. There are also some particular differences with respect to the way parts of speech are handled. Adjectives in GermaNet, for example, are hierarchically structured (in contrast to a clustering approach in WordNet). It wasn't clear that any of these differences affected the usefulness of GermaNet for sense-tagging, however.

More important was GermaNets coverage. Although GermaNet is comparable in size to WordNet, it is significantly smaller, as indicated in Table 1.

	GermaNet	WordNet 1.7
Noun	27824	74488
Verb	8810	12754
Adjective	5141	18523
Adverb	2	3612
Total:	41777	109377

Table 1. GermaNet vs. WordNet

Although GermaNet has been integrated into EuroWordNet (Kunze & Wagner 1999b), the version we used for our research was the stand-alone GermaNet.

3. The annotation task

As a preliminary to the development of an automatic sense tagger for German we hand-tagged eleven small German texts. We used these hand tagged texts to evaluate the applicability of GermaNet to large-scale sense tagging applications. The procedure we used for annotation was fairly straightforward. We automatically lemmatized the words and tagged them for part of speech using the Stuttgart TreeTagger. To actually carry out sense tagging, we developed a software tool for presenting words in texts along with their GermaNet synsets, which was used by five annotators to annotate the texts. The texts were annotated on a word-by-word basis, with each token that had been tagged either as a verb, a noun or an adjective presented for word-sense tagging. For words that could not be annotated with GermaNet synsets, the problem that the word appeared to pose was noted by the annotator, if one was apparent. These error-annotations were used to classify the types of words that presented difficulties for sense-tagging using GermaNet synsets.

3.1. Corpus preparation

As there is not yet a standard representative German corpus, we choose to develop our own corpus. The corpus consisted of eight short excerpts from novels for children and young people and three articles taken from German newspapers. The total number of words in our corpus was 5625 and the individual subcorpora varied in size from 257 to 1021 words.

The entire corpus was both lemmatized and tagged for part of speech by the IMS TreeTagger (Schmid 1994). These lemmata were then used to automatically compile a list of GermaNet synsets for each token in the corpus that appeared in GermaNet. For each lemma, the complete set of GermaNet synsets associated with the lemma by GermaNet was stored alongside the lemma. The POS information was **not** used in this step for filtering, so as to exclude this as a source of error. As indicated in Table 2, GermaNet assigned a synset to more than 90% of the content words (noun, verb or adjective tagged words) in the texts. Strikingly, the percentage of content words not assigned an appropriate synset by GermaNet is lower for the newspaper corpora (about 80%) then for the childrens fiction corpora (about 85%).

3.2. Corpus Annotation

For purposes of annotation, the eight short-novel corpora were split up randomly into 24 pieces which were recombined into equal-sized subcorpora and distributed among our five annotators. The pieces were systematically permutated in order to minimize the influence of inter annotator differences. After annotation was complete the pieces were reordered, so that statistics could be obtained on a per corpus basis. At a later stage the newspaper subcorpora News 1, News 2, and News 3 were annotated. Although the annotation procedure was the same, these subcorpora were annotated by a single annotator.

The actual annotation was carried out as follows. The five annotators – all native speakers of German – were provided with a software tool and a set of files to be tagged. The software tool (see fig. 1) presented the annotator with each occurrence of a lemma for which GermaNet provided synsets.

Figure 1. The TAZAN annotation tool

The annotator task was to mark the appropriate synset, if there was one. In addition to the textual context the word appeared in, i.e. the sentence, annotators were shown the set of synsets for the lemma and the basic characterization provided by GermaNet for these synsets. These contained brief descriptions of the synset, examples of typical uses of that sense of the word and an indication of where the synset was located in the GermaNet hierarchy. For verbs the syntactic frame associated with the sense was also indicated. The synsets were presented to the annotator grouped by POS. In choosing a synset, annotators also implicitly indicated what they took to be the correct POS for the word in contexts.

For the lemma *essen*, for example, the following information was presented, with three noun senses and one verbal sense.

[nomen essen Sense 1] Essen, Mahl, Mahlzeit --
('Einnahme von Speisen')
[nomen essen Sense 2] Gericht, Speise, Essen --
('Speise, die für eine Mahlzeit zubereitet ist') =>

Corpus	Word Tokens	Content words	Synset Assigned	Marked	Marked (of Assigned)
Fiction	4330	1770	1658 (93.7%)	1497 (84.6%)	90.3%
Newspaper 1	257	143	129 (90.2%)	124 (86.7%)	86.7%
Newspaper 2	474	206	179 (86.9%)	161 (78.2%)	89.9%
Newspaper 3	564	270	233 (86.3%)	205 (75.9%)	76.3%

Table 2: Quantitative Characterization of the Corpora and Annotation Results

Nahrung, Nahrungsmittel, Lebensmittel, Esswaren, Eßwaren*o, Essen, Speisen
[nomen essen Sense 3] Nahrung, Nahrungsmittel, Lebensmittel, Esswaren, Eßwaren*o, Essen, Speisen => Objekt -- ('Entität mit räumlicher Ausdehnung')
[verb essen Sense 1] essen, futtern*s, nehmen -- ('etwas zu sich nehmen', "Er isst kein Fleisch."(NN.AN), "Er futtert wie ein Scheunendrescher."(NN.BR), "Sie nimmt viel Flüssigkeit zu sich."(NN.AN.PP), "Die Kinder futtern fleißig Schokolade."(NN.AN.BM) "Heute abend werde ich warm essen."(NN.BM)) => verzehren -- ('Ein Lebensmittel essen oder trinken, Perspektive auf Lebensmittel', "Auf der Weihnachstfeier haben die Mitarbeiter zehn Kilo Fleisch verzehrt.", "Sie verzehrte ihr Gemüse ohne Appetit.")

The annotators were also encouraged to use the GermaNet browser to locate additional information about a synset if a decision was difficult.

To annotate, the annotator simply selected (via check box) the appropriate sense(s) for the word as used in the context presented. They were able to move freely forwards and backwards through the corpus and to change their choice of synset at any time. The task was not an easy one. To fully annotate even one of our 24 small subcorpora took our annotators approximately an hour of annotation time. Typically, however, our annotators divided up the task into a number of sessions.

Note that annotators were instructed to mark all synsets considered appropriate. That means that the annotator could mark more than one of the senses GermaNet assigned to the word or reject all of them. This means that words which were not assigned at least one GermaNet synset were not presented for tagging at all. As indicated in the sixth column of Table 2, this was typically around 10% of the content words.

3.3. Results of annotation task

The results of our annotation exercise are indicated in the final columns of Table 2. This column indicates the percentage of the total number of content words (NVA tagged words) for which an annotator marked at least one of the supplied senses as correct and the percentage of the total number of words assigned a synset by GermNet for which at least one of the synsets assigned was marked by an annotator as being appropriate. This is a raw measure of how well GermaNet could be used to sense tag our corpora. That is, in about 90% of the cases, if a word appears in GermaNet, then the annotators found that GermaNet provided an appropriate sense for the word as used in the corpus. While not disappointing, the numbers may seem low. In fact they are misleadingly low, as a significant proportion of these errors are not due to GermaNet at all. In section 4 we will discuss these error factors extensively.

3.4. Inter annotator agreement

An important question, however, was the degree to which the judgement of our annotators varied. We made provision for evaluating inter annotator agreement by having all the annotators tag one small subset of the short novel corpus. This subcorpus contains 431 tokens and was annotated by all five annotators. Only 170 of these tokens were assigned a list of synsets by the GermaNet. So there were 170 points the annotators could disagree on. To evaluate inter annotator agreement, we looked at whether for each of these 170 tokens any synset was marked or not by the annotators. The number of tokens that were not marked as having any acceptable GermaNet assigned synset is shown in Table 3. All five numbers are in the 95% intervall [35, 57] of the binomial distribution with n = 170 and p = 46.0 / 170 = 0.271.

Annotator	1	2	3	4	5
Token with no synset marked	40	44	56	50	40
Mean	46.0				
Variance	38.4				
Standard deviation	6.2				

Table 3: Basic statistics of annotation

It is not, of course, correct to infer from this that the annotators agree on which tokens to mark. To evaluate the more narrow question of whether our annotators agree on this we compared our annotators pairwise. Table 4 shows how many tokens can be counted in the union and intersection of two annotators' annotation records filtered for tokens that have no marked synset and in which each token was prefixed with a unique token ID. The size of the intersection gives the number of tokens that they agree on and the difference to the size of the union gives the number of tokens they disagree on. If, for example, annotator 1 and 2 completely agreed, the number of tokens would be max(40,44) = 44 in the union and min(40,44) = 40 in the intersection. If they disagreed as often as possible, the numbers would be 40+44 = 84 and 0. Table 4 gives these numbers, with the possible ranges in square brackets. The numbers seem to show quite good agreement.

A way of measuring inter annotator agreement is provided by Cohen's (1960) kappa statistic. This measure indicates the degree to which the observed agreement rate differs from chance, and is given by:

$$\kappa = \frac{P_a - P_e}{1 - P_e}$$

where P_a is the observed agreement rate and P_e is the expected chance agreement. Numbers above 0.80 are generally considered to give evidence for a good agreement, whereas numbers below 0.67 indicate poor agreement (Carletta 1996). Our κ values – indicated in the final column of Table 4 – are between or even above these standard values, indicating acceptable agreement.

We did not analyze agreement of polysemy judgements, that is, agreement on what sense should be assigned to which word (c.f. Veronis 1998), because they are irrelevant to our study. Furthermore, token counts per type are too small to get significant results. It is important to keep in mind that we were primarily interested in whether GermaNet is a rich enough lexical resource, not with whether the annotators agreed exactly on how to use it.

Annotator pair	Token in Union	Intersection	κ
(1, 2)	48 [44, 84]	36 [0, 40]	0.81
(1, 3)	58 [56, 96]	38 [0, 40]	0.71
(1, 4)	51 [50, 90]	39 [0, 40]	0.82
(1, 5)	45 [40, 80]	35 [0, 40]	0.84
(2, 3)	61 [56, 100]	39 [0, 44]	0.69
(2, 4)	54 [50, 94]	40 [0, 44]	0.79
(2, 5)	50 [44, 84]	34 [0, 40]	0.75
(3, 4)	62 [56, 106]	44 [0, 50]	0.75
(3, 5)	58 [56, 96]	38 [0, 40]	0.71
(4, 5)	55 [50, 90]	35 [0, 40]	0.70
all five	64 [56, 107]	30 [0, 40]	0.75

Table 4: Inter annotator agreement

4. Error analysis

In order to analyze the quality and extent of GermaNet's coverage, then, we chose to further examine those tokens for which GermaNet should provide a synset, but for which no sysnset was marked by our annotators. These are the cases in which GermaNet fails to do its job. Our goal was to quantify this failure and to assess its most likely causes.

We take it to be the case that in the ideal case GermaNet would associate an appropriate sense for all occurrences of nouns, verbs, and adjectives. Given a perfect POS tagger a perfect lemmatizer, a perfect GermaNet and a perfect human annotator, every NVA-tagged word in our corpus should be marked by the annotators with at least one synset. (Perhaps *exactly* one would be more ideal; in our study we ignored this however. We were concerned that GermaNet be rich enough, not that it be too rich.)

In practice, of course, the results are not perfect. In the following we will discuss the degree to which our results deviated from the ideal. As we saw in Table 2, the number of content words which could be assigned a synset at all by GermaNet ranges from just over 83% to just under 94%. In only about 90% of the cases was one of the synsets assigned to a word by GermaNet marked as being the correct one by our annotators.

In fact, however, a large proportion of this error was introduced not by GermaNet, but by TreeTagger, which we used to lemmatize and tag our texts for part of speech. While errors in POS tagging could lead to suboptimal performance, POS tagging errors were fairly rare in our texts (as Schmid (1994) shows the tagger employed can reach an accuracy of about 97.5%). Furthermore, the kinds of errors that would be problematic in our task (mistagging of prepositions, adverbs or articles as nouns, verbs or adjectives) are the least common type. So POS tagging did not contribute significantly to the errors. Lemmatization errors, however, contribute significantly to the error rate, since every incorrectly lemmatized word resulted directly in an error: When a word is not properly lemmatized it is impossible for the human annotator to choose the correct synset, since this synset is not an available choice, as we have looked up the wrong word in GermaNet.

In order to evaluate GermaNet, then, we needed to classify our errors, so as to determine which errors were the result of GermaNet design or coverage problems and which, like lemmatization errors, were epiphenomenal.

4.1. The error classes

For purposes of our evaluation we took any NVA tagged token in our corpus to which no GermaNet synset was assigned to be an error, and we assigned each error occurrence to one of the following error classes: **lemma, particle**, **collocation, compound**, **derivation**, **auxiliary**, and **other**. The classification of errors was carried out by a single annotator (JS) using a Java-implemented GUI-tool. Each error was assigned to exactly one of the error classes. These classes were chosen because either they were a type of error that was particularly common, or because they were a type of error that the GermaNet developers had suggested might cause problems.

The error classes are described as follows:

Lemma. As mentioned, when a word is not properly lemmatized it is impossible for the human annotator to choose the correct synset, since it is not available for choice. An example of this kind of error is when the particle *mal* in "Mal wieder hat er es getan" is lemmatized as *malen*, the verb 'to draw'.

Particle. German seperable verbs, such as *vorschlagen*, contain prefixes which significantly alter the meaning of a verb (*schlagen* – "hit"; *vorschlagen* – "propose"). These verbs should be lemmatized as a single lexeme. Unfortunately in many contexts the prefix is not concatenated with the verb, as in:

Er *schlug* einen Kompromiss *vor*.
"He proposed a compromise."

This presents difficulties for lemmatizers. Very often the lemmatizer does not link the particle verb's root and prefix leading to a wrong lemmatized form, omitting the prefix (e.g. *schlagen* instead of *vorschlagen*).

Auxiliary. The verbs *sein* and *haben* (as well as certain modals and others) are also problematic. These verbs can be used simply as syntactic operators – auxiliaries – on the one hand, or as main verb on the other. As auxiliaries, there is a sense in which they should not be sense tagged (since they are not "open class"). In this group we mark those cases in which such a verb is not tagged but is recognized as being used as an auxiliary.

Strictly, speaking both **particle** and **auxiliary** errors can be thought of as lemmatization errors of a very specific type, and cannot really be attributed to GermaNet. In contrast to these we distinguished three types of errors that can be attributed to word-formation processes:

Collocation. Many words are used in a very specific sense in combination with other words (*ins Wasser fallen* to mean "cancelled", for example). In those cases in which the word to be tagged was recognized as forming part of a collocation, it was assigned to this class. While it is arguably not the task of a lexicon to account for collocations and idioms, we were interested in assessing the degree to which these are problematic.

Compound. Compounding – the formation of a new word from two or more existing words (for example

Errors class	Corpus			
	Fiction	News 1	News 2	News 3
Lemma	12.3	5.3	13.3	10.8
Particle	5.9	0	4.4	4.6
Auxiliary	25.3	21.1	22.2	21.5
Compound	11.5	31.1	11.1	32.3
Derivation	5.2	10.5	4.4	6.1
Collocation	2.2	5.3	2.2	1.5
Other	36.8	26.3	42.2	23.1
Total errors	269	19	46	62

Table 5: Distribution of errors by class and corpus (in percent)

Montagsauto) is a productive word formation process in German (as in English). As the sense to be associated with the compound is a fairly arbitrary function of the meaning of the constituent words (cf. Fanselow 1981), it is in principle difficult to provide appropriate synsets for words formed this way.

Derivation. The generation of nouns from verbs (for example *Vorbereitung* from *vorbereiten*) and the generation of diminutive forms (for example *Hündchen* from *Hund*) are productive process in German. These are somewhat more regular and might be accounted for by a GermaNet with sophisticated morphological processing (like that suggested by Kunze (1999) for particle verbs).

Finally there are the errors that fit into none of these classes:

Other. All other forms of derivation are covered by the "other coverage" default error class. The major component of this class is simply the set of words which are simply missing form GermaNet, i.e, those that should be and could be listed, but are not.

4.2. Results of error analysis

In Table 5 we present the distribution of the different type of errors by error class in each of our small corpora. It is clear there was significant variation across the corpora as to which error classes were predominant. The variation was particularly evident in the case of **lemma** and **compound** errors. The most significant class of errors was the **auxiliary** class. These were fairly uniform, accounting for between a quarter and a fifth of all errors in each of the corpora. The surprising fact that we noted in section 3, that the newspaper corpora appear to be better handled by GermaNet than the fiction corpus, gets a simple explanation: lemmatization-related errors were more pronounced in the newspaper corpus. In fact, looking only at non lemmatization-related errors, we see that the childrens fiction is, as we might expect, less error prone than the newspaper articles.

The newspaper corpora evidenced significantly more errors that were due to the use of productive morphology. The **compound** errors were the most prominent, particularly in the newspaper corpora, although was significant variation here as well. Other **derivation** errors, however, had a relatively small share. **Collocations** though they appear in most corpora, also play a minor role.

In Table 6 the distribution of errors by POS is displayed. It is obvious why **particle** and **auxiliary** errors would be limited to verbs, as they are verb-specific error types. More interesting is the fact that errors that could be attributed to productive morphology were essentially limited to nouns and adjectives. Essentially only nouns were involved in **derivation** errors, while for adjectives (other than **lemma** errors) essentially only **compound** errors were present

Error class	Part of Speech		
	Verb	Noun	Adjective
Lemma	3.5	23.7	13.9
Particle	13.2	0	1.3
Auxiliary	58.8	0	1.3
Compound	0	31.6	8.9
Derivation	1.8	17.1	1.3
Collocation	3.5	2.6	0
Other	19.3	25	73.4
Total errors	114	76	79

Table 6: Distributions of errors in Fiction corpus by class and part of speech (in percent)

5. Conclusion

Our results were very encouraging. On average 92% of the words which were tagged as verbs, nouns or adjectives were provided with at least one sense by GermaNet, and more than 83% were provided with at least one sense that was judged as the correct sense by our annotators. One of the major sources of error was, in fact, external to GermaNet: On average 15% of the content words were incorrectly lemmatized, leading to incorrect lookup. Additionally we found that many of the potential sources of coverage failure suggested by Hamp & Feldweg (1997) were indeed evident: productive morphological processes such as derivation and compounding as well as collocative uses of words accounted for a nearly 25% of the errors we noted. Particle verbs also presented problems for our annotators, as in some cases the verb was not lemmatized with its separable prefix. Clearly a more sophisticated lemmatizer could have eliminated some of these errors. In other cases productive combinations with main verbs gave rise to forms which were not covered by GermaNet. For nouns a predominant source of errors was the existence of a large number of nouns that were clearly derived via productive rules of derivation from verbs. These could, presumably, be looked up on the verbal hierarchy. Words formed via compounds were also a significant source of noun and adjective errors. Words that could not be properly tagged because they were used as part of a collocation accounted for only minority of the errors overall, however.

We also found that the effectiveness of GermaNet as used for the word-sense disambiguation task as well as the kinds of errors that were found was highly dependent on the variety of text to be disambiguated. This suggests that it is crucial that in WordNet evaluation both domain and text type be standardized, and that a variety of types be used.

Finally, many of the types of errors that we found were clearly German-language specific. This finding suggests that language-specific issues are quite important when evaluating the effectiveness of a particular WordNet and that simple cross-WordNet evaluation will likely lead to a incorrect evaluation of the value or coverage of a particular WordNet. With respect to GermaNet, our results suggest that sense-tagging using GermaNet, while quite good as it is, could be significantly improved by integrating additional morphological processing into the tagger. In particular, methods for dealing with compound words and derived words could lead to significant improvements.

6. References

Bloksma, L., P. Díez-Orzas, and P. Vossen, 1996. User requirements and functional specification of the EuroWordNet project. EuroWordNet (LE-4003), Deliverable D001, University of Amsterdam.

Carletta, Jean, 1996. Assessing agreement on classification tasks: the kappa statistics. *Computational Linguistics*, 22(2), 249-254.

Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Dutoit, Dominique, Laurent Catherin, and Andreas Wagner, 1998. Specification of German and French Wordnets. EuroWordNet (LE4-8328), Deliverable 2D002.

Fanselow, Gisbert, 1981. *Zur Syntax und Semantik der Nominalkomposition – Ein Versuch praktischer Anwendung der Montague-Grammatik auf die Wortbildung des Deutschen*. Tübingen: Niemeyer.

Hamp, Birgit and Helmut Feldweg, 1997. GermaNet - a lexical-semantic Net for German. In: P. Vossen et al. (eds.), *Proceedings of the ACL/EACL-97 workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, Madrid, pp. 9-15.

Kunze, Claudia and Andreas Wagner, 1999. Integrating GermaNet into EuroWordNet, a multilingual lexical-semantic database. *Sprache und Datenverarbeitung*.

Kunze, Claudia (ed.), 1999. Final wordnets for German, French, Estonian, and Czech. EuroWordNet (LE-8328), Deliverable 2D014.

Kunze, Claudia and Andreas Wagner, 1999. The German Wordnet. EuroWordNet (LE-8328), Deliverable 2D014.

Kunze, Claudia, 1999. Semantics of Verbs within GermaNet and EuroWordNet. In: V. Kordoni (ed.), *Proceedings of the ESSLLI-99 Workshop on 'Lexical Semantics and Linking in Constraint-Based Theories'*, pp. 189-200.

Landes, Shari, Claudia Leacock, and Randee I. Tengi, 1998. Building Semantic Concordances. In: Christiane Fellbaum, (ed.), *WordNet: an electronic lexical database*. MIT. Chapter 8, pp. 199-216.

Miller, G., M. Chodorow, S. Landes, C. Leacock, and R. Thomas, 1994. Using a semantic concordance for sense identification. In *Proceedings of the ARPA Human Language Technology Workshop*, San Francisco, Morgan Kaufmann, pp. 240-243.

Miller, George A. (ed.), 1990. WordNet: An on-line lexical database. Special issue of *International Journal of Lexicography*, 3 (4).

Resnik, Philip, 1998. *WordNet and Class-Based Probabilities.* In: Christiane Fellbaum (ed.), *WordNet: an electronic lexical database.* Cambridge: MIT Press, p. 239-263.

Schmid, Helmut, 1994, Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, September 1994.

Véronis, J., 1998. A study of polysemy judgements and inter-annotator agreement. Programme and advanced papers of the Senseval workshop, 2-4 September 1998. Herstmonceux Castle, England.

Vossen, Piek (ed.), 1999. WordNet1.5 in EuroWordNet format. Deliverable D032D033/2D014 Part B1.