Tuning lexicons to new operational scenarios

Roberto Basili, Maria Teresa Pazienza, Michele Vindigni, Fabio Massimo Zanzotto

University of Rome Tor Vergata
Department of Computer Science, Systems and Production
Via di Tor Vergata 110, 00133 Roma (Italy)
basili,pazienza,vindigni,zanzotto@info.uniroma2.it

Abstract

In this paper the role of the lexicon within typical application tasks based on NLP is analysed. A large scale semantic lexicon is studied within the framework of a NLP application. The coverage of the lexicon with respect the target domain and a (semi)automatic tuning approach have been evaluated. The impact of a corpus-driven inductive architecture aiming to compensate lacks in lexical information are thus measured and discussed.

1. Introduction

Real word NLP applications require background knowledge to support their inferential processes, but also specific information on involved domain, jargon and other typical lexicalised expressions. Availability of lexical information becomes crucial to drive the analysis and design of NLP systems. Moreover adaptivity to changes in operational environments is a crucial issue for an efficient reuse of the architectural components and resources in future applications: similar performances should be guaranteed throughout different domains. It is commonly agreed that appropriate lexical resources determine the linguistic quality of the overall system; domain specific information in lexicons has been proofed to improve syntactic analysis ((Basili et al., 1999)).

General lexicons suffer for several problems. Firstly, they include often overwhelming information. The effect of taking into account too many senses for a particular word is often a useless burden for the linguistic analysis. Secondly, general-purpose lexicons could lack in sublanguage specific (not only jargon) information. Thirdly, the granularity of lexical information required in a domain may strongly differ from the adopted "standards". Specific realisations of domain senses could be highly significant in applications, although they have never been systematically taken into account in dictionaries. These aspects result obviously in weaker system performances.

While specialised lexicons are unquestionably valuable resources, their production from scratch is a very expensive process. It includes, at least, the following activities:

- pruning irrelevant information from the existing (general-purpose) lexicon, L
- adding specific information that is missing from L
- rewriting of information yet represented in L to better express its different specific behaviour into the target domain.

As also described in (Basili et al., 1998a), these are typical activities meant to characterize the notion of *lexical tuning*. Lexical Tuning has been experimented also within the framework of the TREVI system ((R. et al., August 1998)) in order to specialise the adopted lexical server (i.e. the Lexicon Management System, LMS, component). In that project lexical information is used as a support for the following activities:

- Syntactic parsing, as the TREVI language processor, based on CHAOS (see (Basili et al., 1998b)), uses a subcategorization lexicon as a control strategy for robust parsing;
- Event Recognition, based on the robust analysis of CHAOS and on a coarse interpretation, according to high-level semantic categories available from LMS lexical descriptions;

Text Categorisation is the target tasks that relies on part of the lexical content of the LMS as it is based on POS tagged lemmas and typical events of each class (i.e. Topics)[1]

2. The ideal lexicon

LMS ((Weigand and Hoppenbrouwers, 1998)) is a general-purpose lexicon that includes morphosyntactic and semantic information for two languages (English and Spanish). The semantic component is organized around an ontology network of synonymy sets as in Wordnet, where concepts are bilingual. Spanish and English lexical entries refer to the same nodes where possible. These last are organised in three different components:

- A basic level (BLO) ontology, containing those general ontological concepts, commonly shared by different user community users;
- A top level (TLO) ontology, in line with Eurowordnet (Vossen, 1998), that describes senses by means of very general categories or, better, sets of explanatory features (e.g. NATURAL vs. ARTIFICIAL, GROUP vs. UNIQUE). Each BLO nodes points, as in Eurowordnet, to subsets of the TLO catalogue, so that both levels are available for each noun/verb sense.

[1] The current version of the linear profile-based classifier (NL/RDS, see (Basili et al., 2000)) makes only use of lemmatization in a quantitative model. Further extensions of the Text Categorization model foresees the use of more language sensitive information: events recognized in training data sets will be used as trigger rules for categorization. These lasts are further examples of lexicalized knowledge needed for an application (not a purely linguistic) task.

- A user ontology that includes domain-specific concepts like, for example, person names related to industrial sectors and companies.

In the ontology, verbal concepts are also enriched by the corresponding verb thematic structures. To a consistent subsets of concepts of this type is associated a set of thematic descriptions reflecting the underlying meaning of verb entries. LMS thematic structures define arguments in terms of case roles (e.g. *agent*, *patient*, ...) and selectional constraints, usually expressed as sets of TLO. The use of TLO depends on the involved process: disjunction is used in the unification mechanisms applied during parsing in the satisfaction of selectional retrictions, but other inferences may be triggered (e.g. synonimy) only when stricter constraints (e.g. conjunction) are adopted. In the rest of the paper matching at the argument level will be always intended as successful unification between the disjunction of features of verb thematic descriptions and TLO noun sense definitions.

3. ... and the real one

The target lexical information managed by the LMS has been described in the previous section. This clearly reflects most of the widely accepted traditions in lexical semantics (e.g. (Beckwith et al., 1991)) and is of course an ideal design framework. However, when dealing with real NLP applications a significant distance emerges between the potentials of a systematic description offered by the lexicon and the needs coming from the textaual material in the domain. This is what we call the *real* lexicon here.

Our interest, here, is to provide a methodology for evaluating the real lexicon content in terms of the support it is supposed to give to the background application (i.e. not as *it is*, but from a *client* perspective). In our case, lexical information is used as a domain-specific background knowledge for the TREVI event recogniser. In this context lexical knowledge is expected to disambiguate the results of the syntactic analysis and support the interpretation process. Ambiguous material, produced by the robust parser, is matched against the semantic information supplied by the lexicon, mainly selectional restrictions on verbal arguments. Compositionality is applied to build complex event structures, in case the world model provided by the lexical information is coherent enough with corpus sentences. This activity as a whole represents a valid benchmark for evaluating a lexicon, as it involves relevant aspects related to lexical information: definitory knowledge, interpretative (i.e. reasoning) rules and support for the resolution of referential issues in sentence analysis.

Moreover, in order to analyse how our framework can support few tuning abilities, we selected a small and specific collection from a Reuters news corpus [2]. From the news related to advertising category (Reuters subject code: ADV) a set of about 70 sentences for 6 prototypical verbs has been collected as target test set. Complex phenomena for the parser or elliptical phrases have been removed in order to set up test material as significant as possible from a semantic point of view and minimize the influence of parsing noise. All the statistics and measurements discussed below have been carried out over this set.

3.1. Coverage

Due to the rich nature of the lexical information available from the target LMS lexicon, several figures have been considered relevant in order to evaluate its coverage. Consider that the lexicon is accessed on a client-server basis and it is usually queried during the sentence analysis on a paragraph basis[3].

First of all, we would expect that a significant number of lemmas in the corpus, mainly nouns and verbs, be described in the lexicon. We will refer to this aspect as *Lemma Lexical Coverage* and measure this as the percent of lemmas met in the lexicon.

In this framework, in order to evaluate the impact of missing entries, the lack of frequent words is worse than the absence of the rare ones (as it will cause a larger loss in performances) and we will refer to it as *Words Lexical Coverage*, i.e. the percentage of words met in the lexicon. Results over the test set are in tab 1 where data on nouns and verbs are separately discussed.

# Sentences in the Test Set:	70
# Syntactically covered Sentences:	25
# Semantically fully covered Sentences:	0

	Nouns	Verbs
# word	154	68
# lemma	98	12
Lex Coverage (% lemma)	0.63	0.82
Lex Coverage (% word)	0.64	0.83

Table 1: Description of the Text Set

Notice that no significant difference is observed between tokens and lemmas even if no definite results can be claimed over this small test set.

In order to better evaluate the test set, we also investigated the intrinsic word polysemy. The average number of senses per lemma for the words in the 70 sentences AND in the lexicon was about 2.6. Different word senses can be too specific or not very useful within an application, so that we also measured the polisemy at the level of the coarse grain semantics provided by Top Level Ontological information. As a combination of TLO features is meant to describe a concept, we count the average number of different TLO sets for each word in the test set. We obtained a score of 2.09 suggesting that about 2 senses out of 5 share one or more TLO descriptions. In any case it should be noticed that the implicit assumption that each concept can be explained by

[2] Reuters made available these texts within the TREVI project during the operational evaluation phase.

[3] Paragraphs are the basic atomic units of the queries at the different levels, i.e. morphologic, syntactic (i.e. subcategorization patterns) and semantic, so that all word senses for nouns in the same paragraph are made available during just one LMS access

TLO features is sometime false as there is a number of concepts that are not linked to the Basic Level Ontology.

These preliminary statistics on average coverage and ambiguity already suggest the need for some (semi)automatic adjustment of the lexicon. More information can be derived from an analysis of the entire Advertisment corpus.

3.2. Adherence to the source overall corpus

Evaluation becomes more complex when considering argument structures. Within a further set of 180 more significant verbs of the *Sport* domain (Reuters code GSPO)[4], only about the 25% have argument structures postulated by the lexicon, exhibiting an average of 2.1 argument structures for verb.

In our test set, among the six choosen verbs, argument structures were described in the lexicon for only two (used in 28 sentences). Syntactic agreement between the postulated verb projections in the verbal lexicon and the syntactic phenomena is verified for only 25 of them (see Table 1 - Syntactically covered sentences) . Unfortunally, this drops to 0 when semantic constraints are imposed. However, among the above 25 sentences only two had at least a semantic entry in the ontology for each noun in argumental position.

What we have done in this first experiment has been to match nouns in argument position with the semantic constraints on verb arguments, both being described in terms of TLO features. Most of the sentences are lacking one or more information (e.g. one or more possible arguments are unknown proper names, or nouns not in the lexicon or no argument structure was available for the verb). In case we consider the real lexical contribution in sentence interpretation as the percentage of sentences having (at least) all the arguments instantiated in the lexicon we obtain a contribution for only a small 2.8% of the test set. This is partially due to the massive presence of unkonwn Proper Nouns in argumental position.

It is possible to evaluate the $predictivity$ of argument structures in the domain, i.e. at which extent lexical information is reflected in the corpus. This could allow a first re-rank of different senses. We highlight here two different contributions, "syntactic" and "semantic" adherence, representing respectively how well the syntactic and semantic information required by the corpus matches the linguistic models in the lexicon. We first counted the sentences compliant [5] with at least one argument structure, by allowing a weak unification between arguments, i.e. leaving TLO constraints uninstantiated. We also counted sentences according to the stricter match, i.e. by requesting that all TLO features for nouns could be unified with selectional constraints, thus obtaining resultsa previously summarised in table 1.

[4]These verbs have been selected by a relevance test similar to the χ^2.

[5]Adherence here is caclculated via a non lexicalized parser that do not use the grammatical constraints of the thematic structures, and then by comparing the possibly redundant dependencies against the suggested arguments.

4. Filling the gap

The above section suggests that the currently available lexicon is insufficient to adequately model the lexical and ontological principles required in the underlying domain. In particular the following evidences can be summarized:

i) 42 out of the source 70 sentences cannot be described for lack of thematic verb information.

ii) The 28 sentences that are in the scope of available thematic information includes only two sentences for which all nouns in argumental positions have a semantic description in the lexicon. Unfortunately none of them produces a successfull interpretation due to failures in satisfying selectional constraints.

iii) By limiting constraints to syntactic projections postulated by the argument structures, 25 out of the 28 sentences are covered, i.e. they produce syntactic structures/dependencies reflected by the argument structures.

The result of this quite uncomfortable situation is that we are faced with deficiencies at the *completeness* and *soundness* level. First, completeness is weaker as case (*i*) already suggests that verb senses (or just some of their thematic descriptions) are missing. Moreover, noun usually not satisfying the selectional constraints of existing thematic structures (case *ii*) suggest that some of their senses can be missed. It is also the case that mism,atches at the argument level are due to problems of inconsistency. In fact, mismatches may be generated for wrong descriptions of either noun senses (especially their TLO features) or selectional constraints (e.g. too narrow prediction of argument semantics). Often, selectional restrictions are too vague to allow to prune out wrong interpretations, leaving both arguments or verb senses ambiguous. In fact, TLO descriptions related to colliding senses tend to be similar (as evidencies in the previous section suggest). Being this last phenomenon due to wrong sense definitions or missing ones is a matter of deeper analysis.

Basically we could investigate two different solutions:

1. guessing argument semantics either by global (i.e. the corpus) or local (i.e. the sentence) context analysis.

2. suggesting new senses

Evidencies in the first case could be provided by the lexicon itself, by allowing selectional constraints in the thematic description to act as a predictor for the unknown argument(s), or be deduced from the distributional analysis of the different information in the rest of the sentence; otherwise, it is possible to induce a (sort of) thematic description, by observing regularities emerging from the corpus, given various examples of the missing sense. In the next sections, we will explore further these two possibilities, studying the capability each one has to generate interpretations for missing phenomena that could allow a specific customisation of the lexical resource under analysis.

4.1. Guessing missing arguments

As you noticed, a frequently emerging situation in corpus data is that one or more arguments are missing from the lexicon. This happen for instance when one argument is a proper name. To fulfill this problem, (and assuming that lexical information is, when present, complete), we try first to guess missing features, by allowing argument structures to be filled by matching syntax contraints only. This could be viewed as considering the unkonwn argument as having all the features underspecified, and allowing the argument structure to select which features are necessary for the matching .

Matched arguments are then enriched with a partial description, containing the feature(s) the argument structure predicts. This clearly leaves uncertain the final position in the lexicon for the newly identified lexical concept, being undetermined the (sub)hyerarchy dominated by the feature(s) assigned to it. Hopefully, if the head lemma is present more than once in the corpus, with the same sense [6], the cumulated evidence will help the knowledge engineer to fullfill its description, identifying its meaning. When this approach has been applied to our test set (see Table 2 - Lexicon invent Arguments), we have been able to cover 15 of the previously uncomplete sentences, giving raise to 30 possible interpretations, deriving from different selectional constraints in argument structures.

4.2. Suggesting Argument Structures

When no argument structure is available (42 sentences in case *i*), the information potentially derivable from corpus analysis becomes crucial. We use here an architecture based on CHAOS (the above mentioned lexicalised dependency-based parser, able to work at different levels of lexicalisation) and RGL (an incremental conceptual clustering engine, based on conceptual lattice theory). It has been used in several applications over different *real* corpora (see for example (Basili et al., 1999)).

First, CHAOS is run without any lexical evidence on the domain texts. As a result dependency-graphs are obtained, including ambiguities (e.g. multiple attachment-sites for prepositional phrases). The syntactic representation of the sentence is a graph whose node are words and whose edges are inter-chunks dependencies (*iwds*[7]). The graph gathers the set of alternative planar graphs (D.Grinberg et al., 1996) representing the grammatical information of the sentence. The strength of the syntactic processor is the ability to run at different levels of lexicalization. This allows to design an adaptive approach to parsing. The grammatical information gathered from the corpus by the parser Chaos without subcategorization lexical information is used to feed RGL, the learner of verb subcategorization frames. RGL is applied on the derived contexts (i.e. linearized verb sub-graphs, that is a vector representation of the sentence in terms of couples of syntactic relations/lexical handles). RGL derives the set of (potential) syntactic realizations of verb arguments. The derived description expresses:

- the number of arguments
- their syntactic relations (e.g. Subj/Obj relations are guessed by the learner)
- potential argument handlers (like prepositions or relative pronouns)

This representation provides a basic form of argument structure (i.e. the syntactic component, often referred as *subcategorization pattern*).

In order to add semantics and get thematic descriptions, selectional restrictions for detected arguments have to be derived. We can rely here on the senses postulated for nouns occurring (in the corpus) in (syntactic) argument position. Proliferation of interpretation here arise as all senses trigger potentially independent sentence readings. The incremental nature of the RGL method support a refinement of this newly discovered information as whenever more sentences suggest the same senses a ranking of the corresponding thematic interpretations is possible. This controls the growth of the induced lexicon by narrowing the induction according to the corpus material.

This inductive approach has been applied to the subset of sentences headed by verbs missing from the lexicon. The 42 sentences related to the four verbs lacking thematic information have been used (i.e. parsed in a nonlexicalized fashion) to fed the clustering engine. 8 different subcategorisation schemes (2 for each verb) have been proposed by the RGL component (in Table 2 - Covered by synt). They fully represented 16 new sentences, i.e. they provided all the syntactic predictions useful to cover the dependencies proposed by the parser. Thematic interpretations then were expected for fullfilling the argumental (syntactic) slots. Semantic information that the lexicon offered for thouse nouns have been used here. Unfortunately no semantic interpretation could be derived as none of these 16 sentences included all head nouns with lexical descrition in LMS (see Table 2 - RGL + Noun Lex). Again, the completeness problem arises.

4.3. Using Word Sense Disambiguation in Lexical Tuning

As a further inform,ation coming from the corpus, semantic descritpions of words (i.e. their word senses according to an ontological catalogue) can be derived by inductive methods. Corpus-driven word sense disambiguation (WSD) algorithms (Yarowsky, 1992; Basili et al., 1993) here can be adopted. They can be trained on the corpus, and used as sense guessers to support two complementary tasks:

- provide sense hypothesis for unknown words (e.g. Proper Nouns and items missing from the lexicon)
- filter ambiguous senses proposed by the lexicon (i.e. original notion of WSD)

[6] This hypothesis being suggested to hold in most cases by (Yarowsky, 1992)

[7] For the purpose of this paper, inter-chunk dependencies will be treated here as inter-word dependencies, where chunks are mapped into words corresponding to their heads

As combinatorial explosion within the space of possible semantic interpretations leads to hundreds of different possibilities for even simple sentences, corpus data offer evidencies to limit the inherent complexity of this step. Simple frequency-based models are often impractical given the huge amount of required data and the sparsness problems in dealing with NL semantics.

This is where a corpus-driven unsupervised Word Sense Disambiguation and Classification algorithm come in hand. A model proposed in (Basili et al., 1993) has been here adopted. The set of ontological target classes for this classification algorithm is the TLO feature set. The task can be briefly described as:

1. Given an description of word senses in terms of coarse semantic catergories, O, and a corpus C
2. for each class $o_i \in O$ derive a probabilstic model of its typical contexts $ctxt(o_i)$ as observed in C, i.e. contextual information of words known as o_i
3. for each (unknown) word w occurring in the corpus, and given its contexts $ctxt(w_i)$ in C, classify w in all o_i such that similarity between $ctxt(o_i)$ and $ctxt(w_i)$ can be assessed.

Our method, inspired by (Yarowsky, 1992), works as follows:

- Step 1. Select the most typical words in each core category, o_i;
- Step 2. Acquire the collective contexts of these words and use them as a (distributional) description of each category;
- Step 3. Use the distributional descriptions to evaluate the (corpus-dependent) membership of each word to the different categories.

Step 1 is carried out detecting the more significant (and less ambiguous) words in any of the core classes : these sets are called the *kernel* of the corresponding class. Rather than training the classifier on all the nouns in the learning corpus as in (Yarowsky, 1992), we select only a subset of *prototypical* words for each category.

Step 2 uses the kernel words to build (as in (Yarowsky, 1992)) a probabilistic model of a class: this model is based on the distribution of class relevance of the surrounding terms in typical contexts.

In Step 3 a word is assigned to one, or more, classes according to the contexts in which it appears. Many contexts may enforce the selection of a given class, or multiple classifications are possible when different contexts suggest independent classes.

This algorithm has been applied to the underlying advertisement corpus of our test bed. About 1,000 documents have been used as a systematic decsription of the domains. Collective contexts have been derived from all the documents and estimation of simlarity has been projected for all the nouns in our test set.

Notice that if estimation is run on each single context s (i.e. the sentence in which an unknown word $w$is used) an $local$ notion of similarity is derived and can be adopted as Sense hypothesis for the target word w in s. The outcome of this algorithm is usually a subset of TLO features (i.e. the target sense description of the method)) for each unknow word w in each sentence s.

There are basically two ways WSD becomes useful in lexical tuning:

- To enhance lexical coverage, that is, to tag unknown words. This could be used to acquire massive information to increment lexicon content. Guessed features could be verified by hand or by acquiring a statistics about their distribution in the corpus.
- To narrow the proliferation of interpretations for known words, that is, to select only those senses whose TLO information is compatible with WSD guesses. This could be used to increase the precision of the system, cutting out senses that are never realised in the corpus.

We used, when possible, the last approach, by querying the lexicon and using WSD to decide among resulting interpretations. If the lexicon and the algorithm were inconsistent, lexical information is preferred, and the ambiguity is left untouched.

When information is missing from LMS we are forced to the first usage. All possible interpretations are collected, including potential ambiguities. A later analysis based on the frequency of sense occurrence in sentences accomplishes the task of validating the most relevant senses.

In our test set, we thus completed the subcategorization information induced by RGL with selectional restrictions derived from the lexicon (when possible) or otherwise automaically proposed by the WSD algorithm. In this way (see in table 2 - RGL + WSD + Noun Lex), the 16 sentences syntactically covered by the RGL method suggets now 37 thematic structures. The role of WSD here is to suggests *anyhow* a sense for some of the argument nouns.

Given the ambiguity that persists with this approach, we tried to focus on selectional restrictions that occurr more than once (for the same argument). In order to reduce the ambiguity we restrict the choice onyl to the most frequent TLOs: if several interpretations (due to combination of ambiguous senses for argument nouns) are possible, only those generalizations occurring more thanm once for the example sentences of the same verb are left. This brings the number of different interpretations from 37 to 26 (about 2 for each sentence, Best Sense in Table 2), with a 29% of compression.

5. Conclusions

The role of selective lexical information in any application task based on NLP has been discussed and motivated. Due to the lack of specific and domain dependent resources, we have deeply discussed on needs for lexicon tuning to new domains and applications. Several problems emerge when dealing with real sentences due to either lack of information (both words and senses) in the lexicon, or to the

Lexicon		
Covered by synt		
	Sentences	25
	Senses	63
Partially Realised (synt + sem)		
	Sentences	12
	Senses	16
Lexicon Invent Arguments	Sentences	15
	Senses	30
WSD invent Arguments + Verb Lex		
	Sentences	6
	Senses	7
Extracted Lex Info		
RGL induced patterns		8
Covered by synt	Sentences	16
	Senses	na
RGL + Noun Lex	Sentences	0
	Senses	0
RGL + WSD + Noun Lex	Sentences	16
	Senses	37
	Best Senses	26

Table 2: Source vs. Extracted Lexical Information

overwhelming knowledge not enabling the end system to disambiguate among several possible senses and sysntactic attachements. In order to produce specialized lexicons and evaluate their impact on a final application, a dedicated software architecture has been defined based on CHAOS (a lexicalised dependency-based parser) and RGL (an incremental conceptual clustering engine). In this framework several experiments have been run and results have been shown and fully discussed in the paper. The test bed (composed of selected sentences from a Reuters corpus on advertisement news) has proven the soundness of the approach allowing both to enhance the lexical coverage of a general lexicon by domain specific information, and to narrow ambiguities relates to different senses met for words. Further results will be reached when a larger test bed could be run on the same framework in a systematic activity for lexicon tuning.

6. References

Basili, R., R. Catizone, M.T. Pazienza, M. Stevenson, P. Velardi, M. Vindigni, and Y. Wilks, 1998a. An empirical approach to lexical tuning. In *Proc. of the Adapting Lexical and Corpus Resources to Sublanguages and Applications Workshop, held jointly with 1st LREC*. Granada, Spain.

Basili, R., A. Moschitti, and M.T. Pazienza, 2000. A language-sensitive text classification model. In *RIAO'2000 International Conference*. Paris.

Basili, R., M.T. Pazienza, and P. Velardi, 1993. Semiautomatic extraction of linguisitc information for syntactic disambiguation. *Applied Artificial Intelligence*, 4.

Basili, R., M.T. Pazienza, and F.M. Zanzotto, 1998b. Efficient parsing for information extraction. In *Proc. of the ECAI98*. Brighton, UK.

Basili, R., M.T. Pazienza, and F.M. Zanzotto, 1999. Lexicalizing a shallow parser. In *Proceedings of TALN '99, 6e confrence annuelle sur LE TRAITEMENT AUTOMATIQUE DES LANGUES NATURELLES*. Aiaccio, Corse.

Beckwith, R., C. Fellbaum, D. Gross, and G. Miller, 1991. *Wordnet: a lexical database organized on a psycholinguistic principles*. Lawrence-Erlbaum Ass.

D.Grinberg, J.Lafferty, and D.Sleator, 1996. A robust parsing algorithm for link grammar. In *4th International workshop on parsing tecnologies*. Prague.

R., Basili, Mazzucchelli L. Di Nanni M., Marabello M.V., and Pazienza M.T., August 1998. Nlp for text classification: the trevi experience. In *Proceedings of the Second International Conference on Natural Language Processing and Industrial Applications, Universite' de Moncton, New Brunswick (Canada)*.

Vossen, P., 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer Academic Publishers.

Weigand, H. and S. Hoppenbrouwers, 1998. Experiences with a multilingual ontology-based lexicon for news filtering. In *Proc. of the DEXA 98, (R. Wagner, ed.), IEEE Computer Society Press*.

Yarowsky, D., 1992. Word sense disambiguation using statistical models of roget's categories trained on large corpora. In *Proc. of COLING-92, Nantes, France*.