Towards a Bootstrapping Framework for Corpus Semantic Tagging

Roberto Basili, Michelangelo Della Rocca, Maria Teresa Pazienza
Dipartimento di Informatica, Sistemi e Produzione,
Universita' di Roma, Tor Vergata (ITALY)
{basili,dellaroc,pazienza}@info.utovrm.it

Abstract

Availability of source information for semantic tagging (or disambiguating) words in corpora is problematic. A framework to produce a semantically tagged corpus in a domain specific perspective using as source a general purpose taxonomy (i.e. WordNet) is here proposed. The tag set is derived from higher level Wordnet synsets. A methodology aiming to support semantic bootstrapping in a NLP application is defined. Results from large scale experiments are reported[1].

1 Introduction

Lexical Acquisition (LA) processes strongly rely on basic assumptions embodied by the source information and training examples. Several approaches to LA rely on some forms of declarative descriptions of source data: bracketed or POS tagged corpora are just examples. Many authors claim that class-based methods are more robust against data sparseness problems (Dagan,1994), (Pereira, 1993), (Brown et al.,1992). Other works (Basili et al.,1993a, 1996) demonstrated that a variety of lexical acquisition methods over small size corpora are viable whenever a domain specific semantic bias is available: using high level semantic classes (rather than simple words) increase the robustness of the probability driven methods, usually affected by coverage and data sparseness problems. Furthermore, domain specific semantic classes add expressivity to the underlying statistical acquisition model (Basili et al., 1996): this saves the knowledge engineer from having to deal with mysterious scores with no linguistic flavor. Semantic data possess an explanatory power that is truly required in specific knowledge domains.

Modeling semantic information is much more corpus and domain dependent than POS or syntactic tagging. Bracketed corpora are core components of an underlying grammatical knowledge to which results of different inductive methods equivalently refer. Such equivalence is no longer valid for semantic tagging when corpora (as well as underlying domains) change. In order to design tagging capabilities at a semantic level, it is more important to design adaptation capabilities to process a given corpus in a domain driven fashion. Tagging is a dynamic process that aims to produce a core semantic information to support several induction processes over the same domain.

Availability of source information to support any tagging activity is problematic: general purpose sources (e.g. MRDs and static Lexical Knowledge Bases) may be too generic and worsen the induction quality, while specific domain sources are usually absent. Although semantic information is crucial to the induction of most lexical knowledge, accessing it is often impossible. As gold standards are fairly questionable, it is necessary to rely on sources that are as much systematic as possible and adapting their description to the underlying corpus. The widespread diffusion of WordNet, and its large scale as well, have motivated in several recent studies to start using it as a common source and adapt it for the purpose of the target LA task.

In this framework, we consider tagging as a process carried out in two phases: (1) selection of the semantic tag system specific to the domain (tuning Wordnet); (2) use of the specific classification to tag the corpus *in vivo*.

2 Semantically-driven Induction of Lexical Information

Several phenomena have been (more or less successfully) modeled in the LA literature:

[1] This work has been partially supported by the Esprit LRE project n. 2110 - ECRAN

- Acquisition of word taxonomies from a corpus by means of syntactic (Hindle,1990) (Pereira et al.,1993) as well semantic (Basili et al.,1993a,1996) evidence
- Probability driven PP-disambiguation (Hindle and Rooths,1993), (Basili et al.,1993c), (Brill and Resnik,1994) ,(Resnik,1995), (Frank,1995), (Collins and Brooks,1995). Some of these methods rely on semantic classes in order to improve robustness.
- Verb Argument Structure derivation. Many selectional constraints in argumental information have a semantic nature (e.g. ± animate), like in (Resnik and Brill,1994) (Resnik,1995) or (Basili et al.,1996)

Semantic tagging is thus crucial to all the above activities. We propose the following strategy:

1. *Tune* a predefined (general) classificatory framework using as source an untagged corpus
2. *Tag* the corpus within the defined model eventually adjusting some of the tuning choices;
3. Use tagged (i.e. semantically typed) contexts to derive a variety of lexical information (e.g. verb argument structures, PP-disambiguation rules, ...)

The design of the overall process requires a set of modeling principles:

1. ***to focus*** on the suitable tag system
2. ***to customize*** the classification to a corpus
3. ***to tag*** the corpus correspondingly.

3 Tuning a Classification Framework to a Domain

The wide-spectrum classification adopted within WordNet is very useful on a purely linguistic ground, but creates unacceptable noise in NLP applications. In a corpus on Remote Sensing (RSD) [2], for example, we computed an average ambiguity of 4,76 senses (i.e. Wordnet ***synsets***). Table 1 counts the WN synsets of some of the most ambiguous verbs found in our RSD corpus.

Table 1: RSD verbs with the highest initial polysemy

verb	#WN senses
give	34
run	34
break	29
cut	25
take	23
make	21
go	20
pass	20
set	20
draw	19
raise	18

Several problems are tackled when a domain driven approach is used. First, ambiguity of words is reduced in a specific domain, and enumeration of all their senses is unnecessary. Second, some words function as sense primers for others. Third, raw contexts of words provide a significant bundle of information able to guide disambiguation. Applying semantic disambiguation as soon as possible is useful to improve later LA and other linguistic tasks. Our aim is thus to provide a systematic bootstrapping framework in order to:

- Assign sense tags to words
- Induce class-based models from the source corpus
- Use the class-based modesl (that have a semantic nature) within a NLP application.

The implemented system, called *GODoT* (General purpose Ontology Disambiguation and Tuning), has two main components: a classifier *C-GODoT*, that tunes WordNet to a given domain, and *WSD-GODoT* that locally disambiguates and tags the source corpus contexts. The Lexical Knowledge base (i.e. WordNet) and the (POS tagged) source corpus are used to select relevant words in each semantic class. The resulting classification is more specific to the sublanguage as the exhaustive enumeration of general-purpose word senses has been tackled, and potential new senses have been introduced. The tuned hierarchy is then used to guide the disambiguation task over local contexts thus producing a final sense tagged corpus from the source data. Class-based models can be derived according to the tags appropriate in the corpus and used to derive lexical information according to generalized collocations.

3.1 A semantic tag system for nouns and verbs

Experimentally it has been observed that sense definitions in dictionaries might not capture the domain specific use of a verb (Basili et al, 1995). This strongly motivated our approach mainly based on the assumption that the corpus itself, rather than dictionary definitions, could be used to derive disambiguation hints. One such approach is undertaken in

[2] The tuning phase has been evaluated over different corpora but results will be discussed over a collection of publications on Remote Sensing, sized about 350.000 words.

(Yarowsky 1992), which inspired our tuning method, although objectives and methods of our classifier (C-GODoT) are slightly different.
First, the aim is *to tune an existing word hierarchy to an application domain*, rather than selecting the best category for a word occurring in a context.
Second, since the training is performed on an unbalanced corpus (and also for verbs, that notoriously exhibit more fuzzy contexts), we introduced local techniques to reduce spurious contexts and improve reliability.
Third, since we expect also domain-specific senses for a word, during the classification phase *we do not make any initial hypothesis on the subset of consistent categories of a word.*
Finally, we consider globally all the contexts in which a given word is encountered in a corpus, and compute a (domain-specific) *probability distribution* over its expected senses (i.e. hierarchy nodes)

A domain specific semantics is obtained through the selection of the suitable high level *synsets* in the Wordnet hierarchy. A different methodological choice is required for verbs and nouns. As, WordNet hyperonimy hierarchy is rather bushy and disomogeneous, we considered inappropriate, as initial classification, the WordNet *lowest level synsets.* A more efficient choice is selecting the topmost synsets, called *unique beginners*, thus eliminating branches of the hierarchy, rather than leaves. This is reasonable for nouns, (only 25 *unique beginners*), but it seems still inappropriate for verbs, that have hundreds of *unique beginners* (about 208). We hence decided to adopt as initial classification for verbs the 15 semantically distinct categories (verb semantic fields) in WordNet. The average ambiguity of verbs among these categories is 3.5 for our sample in the RSD. A similar value is the ambiguity of nouns in the set of their unique beginners. The first columns in Tables 2 and 3 report the semantic classes for nouns and verbs.

3.2 Tuning verbs and nouns

Given the above reference tag system, our method works as follows:

- Step 1. Select the most typical words in each category;
- 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 class (semantic fields of verbs and unique beginners for nouns): any of these sets is called kernel of the corresponding class. Rather than training the classifier on all the verbs or noun in the learning corpus, we select only a subset of *prototypical* words for each category. We call these words w the *salient words* of a category C. We define the *typicality* $T_w(C)$ of w in C, as:

$$T_w(C) = \frac{N_{w,C}}{N_w} \qquad (1)$$

where:
N_w is the total number of synsets of a word w, i.e. all the WordNet synonymy sets including w.
$N_{w,C}$ is the number of synsets of w that belong to the semantic category C, i.e. synsets indexed with C in WordNet.

The *typicality* depends only on WordNet. A *typical verb* for a category C is one that is either non ambiguously assigned to C in WordNet, or that has most of its senses (synsets) in C.

The *synonymy* S_w of w in C, i.e. the degree of synonymy showed by words other than w in the synsets of the class C in which w appears, is modeled by the following ratio:

$$S_w(C) = \frac{O_{w,C}}{O_w} \qquad (2)$$

where:
O_w is the number of words in the corpus that appear in at least one of the synsets of w.
$O_{w,C}$ is the number of words in the corpus appearing in at least one of the synsets of w, that belong to C.

The *synonymy* depends both on WordNet and on the corpus. A verb with a high degree of synonymy in C is one with a high number of synonyms in the corpus, with reference to a specific sense (synset) belonging to C. *Salient verbs* for C are frequent, typical, and with a high synonymy in C. The *salient words* w, for a semantic category C, are thus identified maximizing the following function, that we call *Score*:

$$Score_w(C) = OA_w \times T_w(C) \times S_w(C) \qquad (3)$$

where OA_w are the absolute occurrences of w in the corpus. The value of *Score* depends both on the corpus and on WordNet. OA_w depends obviously on the corpus.
The *kernel* of a category *kernel*(C), is the set of salient verbs w with a "high" $Score_w(C)$. In Table 2 and 3 the kernel words for both noun and verb classes are reported. The typicality of the words in the Remote Sensing domain is captured (in the

tables some highest relevance words in the classes are reported). This is exactly what is needed as a semantic domain bias of the later classification process.

Step 2 uses the kernel words to build (as in (Yarowsky,1992)) a probabilistic model of a class: distributions of class relevance of the surrounding terms in typical contexts for each class are built.

In Step 3 a words (verb or noun) is assigned to a class according to the contexts in which it appears: collective contexts are used contemporarily, as what matters here is domain specific class membership and not contextual sense disambiguation. Many contexts may cooperate to trigger a given class and several classifications may arise when different contexts suggest independent classes. For a given verb or noun w, and for each category C, we evaluate the following function, that we call *Domain Sense* ($DSense(w, C)$):

$$DSense(w, C) = \frac{1}{N} \sum_{k} Y(k, C) \qquad (4)$$

where

$$Y(k, C) = \sum_{w' \in k} Pr(w', C) \times Pr(C) \qquad (5)$$

where k's are the contexts of w, and w' is a generic word in k.
In (5), $Pr(C)$ is the (not uniform) probability of a class C, given by the ratio between the number of collective contexts for C [3] and the total number of collective contexts.

The tuning phase has been evaluated over the RSD corpus, and the resulting average ambiguity of a representative sample of 826 RSD verbs is 2.2, while the corresponding initial WordNet ambiguity was 3.5. For the intrinsic difficulty of deciding the proper domain classes for verbs we designed two tests. In the first ambiguous verbs in WordNet have been evaluated: the automatic classification is compared with the WordNet initial description. A recall (shared classes) of 41% denotes a very high compression (i.e. reduction in the number of senses) with a corresponding precision of 82% that indicate a good agreement between WordNet and the system classifications: many classes are pruned out (lower recall) but most of the remaining ones are among the initial ones. A second test has been carried out on WordNet unambiguous verbs (e.g. *flex, convoy, ...*). For such verbs a recall of 91% is obtained over their unique (and confirmed) senses. These results show that tuning a classification using word contexts is enough precise to be used in a semantic bootstrapping perspective and by its nature it can be used on a large scale.

3.3 Tagging verbs and nouns in a corpus

After the tuning phase local tagging is obtained in a similar fashion: given a context k for a word w and the set of the proposed classes $\{C_1, C_2, \ldots C_n\}$ for w, a tag $C \in \{C_1, C_2, \ldots C_n\}$ is assigned to w in k iff adherence of k to the probabilistic model of C is over a given threshold and it is maximal.

The WSD algorithm (WSD-GODoT) can be sketched as follows:

1. Let k be a context of a noun/verb w in the source corpus and $\{C_1, C_2, ..., C_n\}$ be the set of domain specific classifications of w, as they have been pre-selected by C-GODoT;

2. For each class C_i, the normalized contextual sense, NCS, is given by:

$$NCS(k, w, C_i) = \frac{Y(k, C_i) - \mu_{C_i}}{\sigma_{C_i}} \qquad (6)$$

where $Y(k, C_i)$ is defined as in (5), and μ_{C_i}, σ_{C_i} are the *mean* and *standard deviation* of the $Dsense(w, C_i)$ over the set of kernel words w in C_i.
3. The *sense* C that w assumes in the context k is expressed by:

$$\begin{array}{l} \text{Sense(w,k)=C} \\ \text{iff} \ \ \text{NCS(k,w,C)}= \max_{C_i}(NCS(k, w, C_i)); \end{array} \qquad (7)$$

Experimentation has been carried out over set of 1,000 disambiguated contexts of about 97 verbs randomly extracted from RSD. All these 97 verbs where ambiguous, with an average of 2.3 semantic classes per verb persisting ambiguity, even after the semantic tuning phase. Recall and Precision have been measured against a manual classification carried out by three human judges (about 70% cases received the same tag by all the judges, this suggesting a certain complexity ot the task). In 98.74% of cases the tagging system selected one tag. A recall of 85.97% has been obtained. Precision is of about 62.19%.

Comparing these figures with related works is very difficult, due to the differences in the underlying semantic type systems and mainly to the variety of information used by the different methods. (McRoy,1992) (and recently (Wilks and Stevenson,1997) described a word sense disambiguation methods based on multiple models, acting over different linguistic levels (e.g. MRD senses, POS tags,

[3] those collected around the kernel verbs of C

corpus contexts). Our methodology is less demanding from the point of view of the required source information and possibly should be compared against one only of the levels mentioned in these works. (Resnik,1995) reports a *human precision* of about 67% but on a noun disambiguation task carried out at the level of true WordNet senses (i.e. synsets): this task seems fairly more complex than ours as we estimated an average of 2.9 synsets per noun on a set of 100 nouns of the RSD. However one of the results of our method is also to eliminate most of these senses from the hierarchy, during the tuning phase, so that precision of the two method cannot be directly compared. Exhaustive experimental data on nouns are not yet available. However the significant results obtained for verbs are important, as several authors (e.g. (Yarowsky,1992)) report verb as a category that is more problematic than noun for context driven classification tasks.

4 Discussion

The relevance of word classes for a variety of lexical acquisition tasks has been described in several works. In (Brown et al.,1993) class-based language models for text processing are described. Classes are derived by pure collocational analysis of corpora. Approaches of this type aim to improve the statistical significance of probability estimations, tackle the data sparseness problems and reduce the number of the model parameters. The derived clusters are very interesting but are not amenable for a direct linguistic analysis. Difficulties in interpreting data derived from numerical cluster analysis emerge also in other studies (e.g. (Pereira et al.,1993)) where additional work is required to assign a suitable meaning to groups of words. The essential difficulty in separating word senses, when conflating data are derived from distinct senses, is due to the fact that simple collocations are often the surface results of independent linguistic phenomena. Collocationally derived lexical constraints (as in the *strong tea* vs. *powerful tea* example given in (Smadja,1989)) may be very different from other types of relations, like verb-argument relations. In this case, in fact a statistical significant relationship is not to be detected betwen verb and its lexical arguments, but between the verb and a whole class of words that play, in fact, the role of such arguments. For example, in the RSD corpus the verb *catalogue* appears 33 times. It takes as a direct object the word *information* only once, that is an evidence too small to support any probabilistic induction. *Information* indeed is a typical *abstraction* that can be *catalogued.* There is no hope for any inductive method making use of simple lexical collocations instead of class based collocations (e.g. *abstraction*) to acquire enogh evidence of most of the phenomena.

Class methods based on taxonomic information may provide more comprehensive information for a larger number of lexical acquisition tasks. In PP-disambiguation tasks several works based on bi-gram statistics collected over syntactic data (e.g. Hindle and Rooths,1993) show evident limitations in coverage and efficacy to deal with complex forms. In (Franz,1995) weak performances are reported for ambiguities with more that two attachment sites. These last are very frequent in a language like Italian where prepositional phrases play a role similar to English compounds. Class-based approaches (e.g. (Basili et al.,1993) and (Brill and Resnik, 1994) are more promising: the implied clustering also tackles the data sparseness difficulties, but mainly they produce selectional constraints that have a direct semantic interpretation. Smaller training data set can be used and also unknown collocates are deal with, if they are able to trigger the proper semantic generalizations.

The method proposed in this paper suggests and provides evidences that processing a corpus, first, to tune a general purpose taxonomy to the underlying domain and, then, sense disambiguating word occurrences according to the derived semantic classification is feasible. The reference information (i.e. the Wordnet taxonomy) is a well-known sharable resource with an explicit semantics (i.e. the hyperonimy/hyponimy hierarchy): this has a beneficial effect on the possibility to extract further lexical phenomenon (e.g. PP disambiguation rules) with a direct semantic interpretation. Let for example:

Future Earth observation satellite systems for worldwide high resolution observation purposes require satellites in low Earth orbits, supplemented by geostationary relay satellites to ensure intermediate data transmission from LEO to ground.

be a potential source document, taken form our RSD domain. Given a preliminary customization of the Wordnet hirerachy, according to the set of kernel verbs and nouns exemplified in Tables 2 and 3, the described methods allow to apply local semantic taggin to the set of verbs and nouns in the document. Some vrbs/nouns are no longer ambiguous in the domain: their unique tag is retained. For the remaining ambiguous words the local disambiguation model is applied (by (7). The tagged version of the source document results as follows:

Future Earth/LO observation/AC satellite/OB systems/CO for worldwide high resolution/AT observation/AC purposes/MT require/CG satellites/OB in low

Earth/LO orbits/SA, supplemented/CT by geostationary relay/OB satellites/OB to ensure/CG intermediate data/GR transmission/AC from LEO/AR to ground/LOC. [4]

The data now available for any lexical acquisition techniques are not only bigrams or trigrams, or syntactic collocations (like those derived by a robust parser (as in (Grishman and Sterling,1994) or (Basili et al, 1994)) but also disambiguated semantic tags for co-occurring words. For example, for the verb *require* , we extract the following syntactic collocations from the source document:

```
V_N(systems/CO, require/CG)
N_V(require/CG, satellites/OB)
```

These data support several inductions. First, semantic tags allow to cluster togheter source syntactic collocations according to similar classifications. Other occurrence of the verb require, as they have been found in the RSD corpus are:

```
V_N(model/CO, require/CG)
N_V(require/CG, satellites/OB)

V_N(process/CO, require/CG)
N_V(require/CG, sensors/OB)

V_N(satellite/OB, require/CG)
N_V(require/CG, beam-antenna/OB)   ...
```

When arguments are assigned with the same tags (e.g. OB for the direct objets) basic instances can be generalized into selectional rules: a typical structure induced from the reported instances is thus

$$\begin{array}{c}(require(Subj/[+COor+OB]) \\ (Obj/[+OB]))\end{array} \tag{8}$$

where explicit semantic selectional restrictions ($+OB$) for syntactic arguments (e.g. *Obj*) are expressed. A method for deriving a verb subcategorization lexicon from a corpus, according to an example based learning technique applied to robust parsing data is described in (Basili et al,forthcoming). Availability of explicit semantic tags (like OB) allows to derive semantic selectional constraints as in (8). Further induction would allow to assign thematic descriptions to arguments in order to extend (8) in:

$$\begin{array}{c}(require(Subj/Theme/[+CO]) \\ (Obj/Instrument/[+OB]))\end{array} \tag{9}$$

Previous work on the acquisition of high level semantic relations is described in (Basili et al.,1993b), where the feasibility of the derivation of lexical semantic relations from several corpora and domains has been studied. Interesting results on applicability of semantic filtering to syntactic data, for the purpose of acquiring verb argument information is reported in (Dorr and Jones,1996). Semantic information greatly improve the precision of a verb syntactic classification.

The proposed tag system (e.g. Wordnet high level classes) has several advantages. First it puts some limit to enumeration of word senses, thus keeping limited the search space of any generalization process. Learning methods are usually search algorithms through concept spaces. The larger is the set of basic classes, the larger is the size of the search space. It is questionalble how expressive is the resulting tag system. Previous research in ARIOSTO (Basili et al,1996a) demonstrated the feasibility of complex corpus driven acquisition based on high level semantic classes for a variety of lexical phenomena. A naive semantic type system allows a number of lexical phenomena to be captured with a minimal human intervention. As an example acquisition of verb hierarchies according to verb thematic description is described in (Basili et al.,1996b). Whenever an early tuning of the potetial semantic classes of a given verb in a corpus has been applied and local disambiguation has been carried out as corpus semantic annotation, more precise verb clustering can be applied:

- first, local ambiguities have been removed during corpus tagging,
- second, clustering is applied with an intra-classes strategy and not over the whole set of verbs. First, a set of thematic verb instances from source sentences are collected for each given semantic class, so that *social* verbs are taken separate from *change* or *cognition* verbs. Then, separate hirarchies can be generated for each semantic class, in order to have a fully domain driven taxonomic description within general classes , e.g. *social*, for which a general agreement exists. Later reasoning processes could thus exploit general primitives augmented with domain specific lexico-semantic phenomena.

5 Conclusions

In this paper a framework to bootstrapping lexical acqusition in a given domain has been outlined. A source semantic tag system is proposed able to guarantee an explicit semantic description of lexical phenomena, with a minimal size in order to minimize the complexity of the inductive algorithms. A

[4] The reported tags are as in Tables 2 and 3.

methodology to customize the general purpose tag system (i.e. the high level classes of the Wordnet hierarchy) to a domain is described and a semantic disambiguation model to semantically tag source raw texts is defined. The result is a semantically annotated corpus where lexical phenomena can be studied with a reduced ambiguity. Experimental evidences for verb and nouns tagging in different domains have been outlined and extensive data froma remote sensing (medium size) corpus have been reported. Implications of the proposed semantic tagging for typical lexical acquisition tasks (e.g. derivation of PP-disambiguation rules or verb argument structures) have been discussed. Further research in assessing the semantic tagging evaluation, customizing the lexical acquisition models to the proposed semantic type system and evaluating extensively the acquisition results are on going.

REFERENCES

Basili et al. 1993a. Basili R., M.T. Pazienza, P. Velardi. "What can be learned from raw text?". Vol. 8: Machine Translation.

Basili et al. 1993b. Basili R., M.T. Pazienza, P. Velardi. "Acquisition of selectional patterns in sublanguages". Vol. 8: Machine Translation.

Basili et al. 1993c. Basili R., M.T. Pazienza, P. Velardi. "Semi-automatic Extraction of Linguistic Information for Syntactic Disambiguation". Vol.7:339-364: Applied Artificial Intelligence.

Basili et al. 1995a. Basili R., M.T. Pazienza, P. Velardi. "A context driven conceptual clustering method for verb classification". in Corpus Linguistics for Lexical Acquisition, B. Boguraev & J Pustejovsky Eds., MIT press, 1996.

Basili et al. 1995b. Basili R., M. Della Rocca, M.T. Pazienza, P. Velardi. "Contexts and categories: tuning a general purpose verb classification to sublanguages". Proceeding of RANLP95, Tzigov Chark, Bulgaria, 1995.

Basili et al 1996a. Basili R., M.T. Pazienza, P. Velardi. "An Empirical Symbolic Approach to Natural Language Processing". To appear in Artificial Intelligence, Vol.85, August 1996.

Basili et al 1996b. Basili R., M.T. Pazienza, P. Velardi. "A Context Driven Conceptual Clustering Method for Verb Classification", in Corpus Processing for Lexical Acquisition, J. Pustejovsky and B. Boguraev Eds., MIT Press 1996.

Beckwith et al 1991. Beckwith R., C. Fellbaum, D. Gross, G. Miller. "WordNet: A Lexical Database Organized on Psycholinguistic Principles, in Lexical Acquisition: Exploting On-Line Resources to Build a Lexicon". U. Zernik Ed. & Lawrence-Erlbaum Ass.

Brill and Resnik, 1994. E. Brill, P. Resnik, "A Rule-Based Approach to Prepositional Phrase Attachment Disambiguation", in Proc. of COLING 1994, Kyoto, Japan.

Brown et al.1992. P.F. Brown, P. deSouza, R.L. Mercer, V.J. Della Pietra, J. C. Lai, "Class-Based *n*-gram Models of Natural Language", in Computational Linguistics, Vol. 18, n. 4., 1992.

Collins and Brooks,1995. Collins M. and Brooks J., *Prepositional Phrase Attachment trough a Backed-off Model*, 3rd. Workshop on Very Large Corpora, MT, 1995

Dagan et al.1994. Dagan I., Pereira F., Lee L. "Similarly-based Estimation of Word Cooccurrences Probabilities ". Proc. of ACL, Las Cruces, New Mexico, USA, 1994.

Dorr and Jones,1996. B.J. Dorr, D. Jones, "Acquisition of Semantic Lexicons: Using Word Sense Disambiguation to Improve Precision", in Proc. of ACL SIGLEX Workshop "Breadth and Depth of Semantic Lexicons", 28 June 1996, University of California, Santa Cruz, USA.

Franz,1995. Franz A., *A statistical approach to learning prepositional phrase attachment disambiguation*, in Proc. of IJCAI Workshop on New Approaches to Le arning for Natural Language Processing, Montreal 1995.

Grishman and Sterling,1994. R. Grishman, J. Sterling, Generalizing Automatically Generated Selectional Patterns, in Proc. of COLING 1994, Kyoto, Japan.

Hindle and Rooth,1993. Hindle D. and Rooth M., *Structural Ambiguity and Lexical Relations*, Computational Linguistics, 19(1): 103-120.

McRoy,1992. McRoy S., "Using Multiple Knowledge Sources for Word Sense Discrimination" Computational Linguistics, vol. 18, n. 1, March 1992, 1-30.

Pereira et al. 1993. Pereira F., N. Tishby, L. Lee. "Distributional Clustering of English Verbs". Proc. of ACL, Columbus, Ohio, USA, 1994.

Resnik,1995, P. Resnik "Disambiguating noun groupings with respect to WordNet senses" in Proceeding of the ACL Third Workshop on Very Large Corpora, June 30, 1995, MIT, Cambridge, MA.

Yarowsky, D. 1992. "Word-Sense disambiguation using statistical models of Roget's categories trained on large corpora". Nantes: Proceedings of COLING 92.

Table 2: Semantic Classes and (some) kernel elements for RSD verbs

Class (Tag)	Kernel verbs
body (BD)	*produce, acquire, emit, generate, cover*
change (CH)	*calibrate, reduce, increase, measure, coordinate*
cognition (CG)	*estimate, study, select, compare, plot, identify*
communication (CM)	*record, count, indicate, investigate, determine*
competition (CP)	*base, point, level, protect, encounter, deploy*
consumption (CS)	*sample, provide, supply, base, host, utilize*
contact (CT)	*function, operate, filter, segment, line, describe*
creation (CR)	*design, plot, create, generate, program, simulate*
emotion (EM)	*like, desire, heat, burst, shock, control*
motion (MO)	*well, flow, track, pulse, assess, rotate,*
perception (PC)	*sense, monitor, display, detect, observe, show*
possession (PS)	*provide, account, assess, obtain, contribute, derive*
social (SO)	*experiment, include, manage, implement, test*
stative (ST)	*consist, correlate, depend, include, involve, exist*
weather (WE)	*scintillate, radiate, flare*

Table 3: Semantic Classes and (some) kernel elements for RSD nouns

Class (Tag)	Kernel Nouns
act (AC)	*experiment, record, calibration, measurement, sensing, services, use*
animal (AN)	*whale, chlorophyll, fur, beluga, gosling*
artifact (AR)	*tape, spacecraft, radar, file, network, radio, pixel, filter, camera,*
attribute (AT)	*density, energy, accuracy, temperature, measurement, magnitude, intensity,*
body (BO)	*limb, water, plasma, region, lineament*
cognition (CO)	*parameter, method, coordinate, study, imagery, temperature*
communication (CMM)	*information, catalog, channels, number, description, summary, signal*
event (EV)	*waves, spin, earthquake, noise, orbit, result, pulse, rotation*
feeling (FE)	*shock, gravity, identification, sensitivity, concern*
food (FO)	*ice, table, potato, fish, board*
group (GR)	*data, field, galaxy, cluster, set, subset, forest, masses, vegetation*
location (LO)	*longitude, field, range, plot, latitude, profile, zenith, terrain, ionosphere*
motive (MT)	*purpose*
object (OB)	*electron, ion, sea, ocean, cloud, sky, planet, satellite, comet*
person (PE)	*user, author, experimenter, investigator, computer, researcher, pioneer*
phenomenon (PH)	*ultraviolet, rainfall, x-ray, energy, result, microwave*
plant (PL)	*galaxy, axis, tree, pollen, composite, crop*
possession (PO)	*coverage, residual, rate, list, resource, fee, record, cost*
process (PR)	*processing, flow, period, evaporation, cooling, absorption, emission*
quantity (QU)	*x, mm, km, p, reflectance, coefficent, inch*
relation (RE)	*altitude, mapping, ratio, map, spectrum, average, correlation*
shape (SH)	*azimuth, vector, surface, region, zone, angle, star*
state (SA)	*humidity, moisture, potential, orbit, climate, atmosphere, polarization, dependence*
substance (SB)	*particle, plasma, ozone, al, proton, ice, helium, gas*
time (TM)	*day, time, hour, khz, year, month, phase, period*