A Hybrid Relational Approach for Word Sense Disambiguation

Lucia Specia[1,2], Maria das Graças Volpe Nunes[1], Mark Stevenson[3]

[1]NILC - Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo
Caixa Postal 668, 13560-970, São Carlos, SP, Brazil
{lspecia, gracan}@icmc.usp.br

[2]Xerox Research Centre Europe
6 Chemin de Maupertuis, Meylan, 38240, France

[3]Department of Computer Science – University of Sheffield
Regent Court, 211 Portobello St., Sheffield, S1 4DP, UK
marks@dcs.shef.ac.uk

Abstract. We propose a novel approach for word sense disambiguation which makes use of corpus-based evidence combined with background knowledge. Using an inductive logic programming technique, it generates expressive models which exploit several knowledge sources and also the relations between them. The approach is evaluated in two tasks: identification of the correct translation for verbs in English-Portuguese and disambiguation of verbs from the Senseval-3 competition. The accuracy obtained in the multilingual task outperforms the alternative learning techniques investigated. The models also yielded significant improvement to the translation quality when integrated into a machine translation system. In the monolingual task, the approach performs as well as the state-of-the-art systems for Senseval verbs.

Keywords: Natural Language Processing, Word Sense Disambiguation, Inductive Logic Programming.

Additional information: student level = PhD, conclusion = 28/09/2007, thesis: http://www.dcs.shef.ac.uk/~lucia/publications/TeseRevisada_LuciaSpecia.pdf.

1 Introduction

Word Sense Disambiguation (WSD) is concerned with the identification of the meaning of ambiguous words in context. For example, among the possible senses of the verb "run" are "to move fast by using one's feet" and "to direct or control". WSD can be useful for many applications, including information retrieval and machine translation. Sense ambiguity has been recognized as one of the most important obstacles to successful language understanding since the early 1960's and many techniques have been proposed to solve the problem. Recent approaches focus on the use of various lexical resources and corpus-based techniques in order to avoid the substantial effort required to codify linguistic knowledge. These approaches have shown good results; particularly those using supervised learning [2]. However, current approaches rely on limited knowledge representation and modeling techniques: traditional machine learning algorithms and attribute-value vectors to represent disambiguation instances. This has made it difficult to exploit deep knowledge sources in the generation of the disambiguation models, that is, knowledge that goes beyond simple features extracted directly from the corpus, like bag-of-words and collocations. For example, with attribute-value vectors it is not possible to utilize relational information, such as semantic relations among the words in the sentence.

In this paper we present a novel approach for WSD that follows a hybrid strategy, i.e. combines knowledge and corpus-based evidence, and employs a first-order formalism to allow the representation of deep knowledge about disambiguation examples together with a powerful modeling technique. This is achieved using Inductive Logic Programming (ILP) [7], which has not yet been applied to WSD.

Our hypothesis is that by using a highly expressive representation formalism, a range of (shallow and deep) knowledge sources and ILP as learning technique, it is possible to generate models that, when compared to models produced by machine learning algorithms conventionally applied to WSD, are both more accurate for fine-grained distinctions, and more "interesting", from a knowledge acquisition point of view (i.e., may convey potentially new knowledge, in a format that can be easily interpreted by humans).

WSD systems have generally been more successful in the disambiguation of nouns than other grammatical categories [6]. Disambiguation of verbs generally benefits from very specific knowledge sources, such as the verb's relation to other items in the sentence. We believe this is a task to which ILP is particularly well-suited. Therefore, we focus on the disambiguation of verbs, as opposed to most of the previous work.

WSD is usually approached as an independent task, however, it has been argued that different applications may have specific requirements. For example, in machine translation, WSD, or *translation disambiguation*, is responsible for identifying the correct *translation* of an ambiguous source word. This paper focuses on the application of our approach to the translation of verbs from English to Portuguese, although experiments with a monolingual task are also described.

In the remainder of this paper, we first present current approaches to WSD and discuss their limitations (Section 2). We then briefly introduce ILP and show how we apply this technique to WSD (Section 3). Finally, we describe our experiments and their results (Section 4).

2 Related Work

WSD approaches can be classified as (a) knowledge-based approaches, which make use of linguistic knowledge, manually codified or extracted from lexical resources [1]; (b) corpus-based approaches, which make use of shallow knowledge automatically acquired from corpus and statistical or machine learning algorithms to induce disambiguation models [14]; and (c) hybrid approaches, which mix characteristics from the two other approaches to automatically acquire disambiguation models from corpus supported by linguistic knowledge [21].

Hybrid approaches can combine advantages from both strategies, potentially yielding accurate and comprehensive systems, particularly when deep knowledge is explored. Linguistic knowledge is nowadays available in electronic resources suitable for practical use, such as WordNet [4], dictionaries and parsers. However, the use of this information has been hampered by the limitations of the modeling techniques that have been explored: using deep sources of knowledge is beyond the capabilities of such techniques, which are in general based on attribute-value vector representations.

Attribute-value vectors consist of a set of attributes intended to represent properties of the examples. Each attribute has a type and a single value for a given example. Therefore, attribute-value vectors have the same expressiveness as propositional formalisms, that is, they only allow the representation of atomic propositions and constants. These are the representations used by the vast majority of the machine learning algorithms conventionally used to perform WSD. More expressive formalisms, such as first-order logic, which is employed by ILP, have not yet been exploited for WSD. They allow the representation of variables and n-ary predicates, i.e., relational knowledge.

In the hybrid approaches that have been explored so far, deep knowledge like selectional preferences is either pre-processed into a vector representation to be accommodated by the machine learning algorithms, or used in previous steps to filter out possible senses [21]. This may cause information to be lost; moreover, it prevents knowledge sources from interacting to each other during the learning process. As a consequence, the models produced reflect only the shallow knowledge that is provided.

Another limitation of attribute-value vectors is the need for a unique representation for all the examples. This usually results in a very sparse representation of the data, given that values for certain features will not be available for many examples. This becomes more and more problematic as more knowledge is exploited and can interfere negatively in the performance of the learning algorithms.

Finally, a disadvantage of attribute-value vectors is that equivalent features may have to be bounded to distinct identifiers. An example of this occurs when the syntactic relations between words in a sentence are represented by attributes for each possible relation, sentences in which there is more than one instantiation for a particular grammatical role cannot be easily represented. First-order formalisms, on the other hand, allow a generic predicate to be created for every possible syntactic role, relating two or more elements.

To sum up, ILP seems to provide a very appropriate general-purpose framework for dealing with the WSD problem: there are explicit provisions for the inclusion of background knowledge of any form, and the representation language is powerful enough to capture contextual relationships.

3 A hybrid relational approach to WSD

In what follows, we provide an introduction to ILP and outline how it is applied in our WSD approach, by showing the knowledge sources used in the experiments.

3.1 Inductive Logic Programming

Inductive Logic Programming [7] employs techniques from Machine Learning and Logic

Programming to build first-order theories from examples and background knowledge, which are also represented by first-order clauses. It allows the efficient representation of substantial knowledge about the problem, which is used during the learning process, and produces disambiguation models reflecting this knowledge. The general approach underlying ILP can be outlined as follows.

Given:

- a set of positive and negative examples $\boldsymbol{E} = \boldsymbol{E}^{+} \cup \boldsymbol{E}^{-}$
- a predicate $\boldsymbol{p}$ specifying the target relation to be learned
- knowledge $\boldsymbol{K}$ of the domain, described according to a language $\boldsymbol{L_k}$, which specifies which predicates $\boldsymbol{q_i}$ can be part of the definition of $\boldsymbol{p}$.

The goal is: to induce a hypothesis (or theory) $\boldsymbol{h}$ for $\boldsymbol{p}$, with relation to $\boldsymbol{E}$ and $\boldsymbol{K}$, which covers most of the $\boldsymbol{E}^{+}$, without covering the $\boldsymbol{E}^{-}$, i.e., $\boldsymbol{K} \wedge \boldsymbol{h} \models \boldsymbol{E}^{+}$ and $\boldsymbol{K} \wedge \boldsymbol{h} \not\models \boldsymbol{E}^{-}$.

We use the Aleph ILP system [20], which provides a complete inference engine and can be customized in various ways.

3.2 Knowledge sources

An important step when designing ILP-based approaches is on identifying, extracting and representing appropriately relevant background knowledge for the problem. This process is not trivial and without it the ILP characteristics that make it different from traditional learning algorithms cannot be truly exploited. The following sources of knowledge were automatically extracted from corpus and lexical resources and used by in our experiments. We used already existing NLP tools whenever possible, and implemented our own tools when necessary, as indicated below. In all cases, we limit the context window to the size of the sentence containing the ambiguous word:

- **KS$_1$**. Bag-of-words consisting of 5 words to the right and left of the verb.
- **KS$_2$**. Frequent bigrams consisting of pairs of adjacent words in a sentence which occur more than 10 times in the corpus.
- **KS$_3$**. Narrow context containing 5 content words to the right and left of the verb, identified by the Mxpost Part-of-Speech (POS) tagger [13].
- **KS$_4$**. POS tags of 5 words to the right and left of the verb, given by Mxpost.
- **KS$_5$**. 11 collocations of the verb: 1st preposition to the right, 1st and 2nd words to the left and right, 1st noun, 1st adjective, and 1st verb to the left and right, also identified using Mxpost.
- **KS$_6$**. Subject and object of the verb, given by the Minipar parser [5].
- **KS$_7$**. Grammatical relations: verb-subject, verb-object, verb-modifier, subject-modifier, and object-modifier, as identified by Minipar, occurring more than 10 times in the corpus.
- **KS$_8$**. The sense with the highest count of overlapping words in its dictionary definition and in the sentence containing the target verb, extracted from the bilingual dictionary *Password* [10], for the multilingual task, and from *Longman Dictionary of Contemporary English* (LDOCE) [11], for the monolingual task.
- **KS$_9$**. Selectional restrictions of the verbs, defined in terms of the features required by its arguments, as extracted from LDOCE, e.g., the verb *come*, in the sense of *move toward*, requires an *animate* subject, and no object. If the

restrictions imposed by the verb are not satisfied by its arguments in the sentence, the features of synonyms and hyperonyms of these arguments – extracted from WordNet – are also verified. A hierarchy of feature types is used to account for restrictions established by the verb that are more general than the features describing its arguments.

The following knowledge sources were designed for multilingual applications only:

- **KS_{10}**. Phrasal verbs potentially occurring in the sentence, identified using a list of phrasal verbs extracted from the same bilingual and monolingual dictionaries and simple heuristics to detect occurrences of separable and inseparable phrasal verbs containing the verb under consideration.
- **KS_{11}**. Bag-of-words consisting of 5 Portuguese words to the right and left of the target verb in the translation of its sentence. This could be obtained using a machine translation system that would first translate the non-ambiguous words in the sentence. We extracted it using a parallel corpus.
- **KS_{12}**. Collocations consisting of 5 Portuguese words to the right and the left of the verb in the translation of its sentence.

Based on the examples, background knowledge and a series of settings specifying the predicate to be learned (i.e., the heads of the rules), the predicates that can be in the conditional part of the rules, how the arguments can be shared among different predicates and several other parameters, the inference engine produces a set of symbolic rules. Fig. 1 shows examples of the rules induced for the verb *come* in the multilingual task**.**

Models learned with ILP are symbolic and can be easily interpreted. Moreover, innovative knowledge about the problem can emerge from the rules learned by the system. For example, **Rule_1** states that the translation of the verb will be "chegar" (*arrive*) if it has a certain subject *B*, which occurs frequently with the word *today* as a bigram, and if the partially translated sentence contains the word "hoje" (the translation of *today*). **Rule_2** states that the translation will be "vir" (*move toward*) if the subject of the verb has the feature *animate* and there is no object, or if the verb has a subject *B* that is also a collocation C, in a position of a proper noun (*nnp*) or personal pronoun (*prp*).

```
Rule_1. sense(A, chegar) :-
  has_rel(A, subj, B), has_bigram(A, today, B),
  has_bag_trans(A, hoje).
Rule_2. sense(A, vir) :-
  satisfy_restriction(A, [animate], nil);
  (has rel(A, subj, B),
  (has collocation(A, C, B),
  (has pos(A, C, nnp); has pos(A, C, prp))).
```

Fig.1. Examples of rules produced for "come" in the multilingual task

4. Experiments and results

The model produced for each verb was tested by applying the rules in a *decision-list* like approach, i.e., retaining the order in which they were produced, using one rule at a time, removing all the examples covered by it from the test set, and backing off to the most frequent sense in the training set to classify cases that were not covered by the rules.

4.1 Multilingual task

For the first scenario, a corpus containing 5,000 sentences for 10 highly frequent and ambiguous verbs (500 for each verb) was extracted from corpora of different domains and genres, e.g., literary fiction and European Parliament proceedings. This corpus was automatically annotated with the translation of the verb using a tagging system based on parallel corpus, statistical information and translation dictionaries [19]. This tagging system outputs the most probable translation for each occurrence of the verb in the parallel corpus. In previous experiments, it showed an average precision of approximately 82%, and thus we decided to manually review the automatic annotation. The sense repository of a verb was defined as the set of all the possible translations of that verb in the corpus. 80% of the corpus was used for training, with the remainder retained for test. The verbs (and their number of senses in the corpus) are: ask (7), come (29), get (41), give (22), go (30), live (8), look (12), make (21), take (32) and tell (8).

Table 1 shows the accuracies (percentage of corpus instances which were correctly disambiguated) obtained by the ILP models. Results are compared against the accuracy that would be obtained by using the most frequent translation in the training set to classify all the examples of the test set (baseline). For comparison, we also ran experiments with three learning algorithms frequently used for WSD, which rely on knowledge represented by attribute-value vectors: C4.5 (decision-trees), Naive Bayes and Support Vector Machine (SVM). In order to represent all knowledge sources in attribute-value vectors, KS_2, KS_7, KS_9 and KS_{10} were pre-processed to be transformed into binary attributes. The accuracy of the ILP approach is significantly better than the most frequent sense baseline and the other learning algorithms (paired t-test; $p < 0.05$). As expected, accuracy is generally higher for verbs with fewer possible translations.

Table 1. Accuracies in the multilingual task

Verb	Majority sense	C4.5	Naïve Bayes	SVM	Aleph
ask	0.68	0.68	0.82	0.88	**0.92**
come	0.46	0.57	0.61	0.68	**0.73**
get	0.03	0.25	0.46	0.47	**0.49**
give	0.72	0.71	**0.74**	**0.74**	**0.74**
go	0.49	0.61	**0.66**	**0.66**	**0.66**
live	0.71	0.72	0.64	0.73	**0.87**
look	0.48	0.69	0.81	0.83	**0.93**
make	0.64	0.62	0.60	0.64	**0.68**
take	0.14	0.41	0.50	0.51	**0.59**
tell	0.65	0.67	0.66	0.68	**0.82**
Average	**0.50**	**0.59**	**0.65**	**0.68**	**0.74**

The models produced by Aleph for all the verbs are very compact, containing 50 to 96 rules each. The various knowledge sources appear in different rules and therefore all of them seem to be useful for the disambiguation of verbs. Details about the experiments are presented in [17].

These results are very positive, particularly if we consider that: (1) the verbs are highly ambiguous; (2) the corpus was automatically tagged, with distinct synonym translations sometimes used to annotate different examples, but only one translation was considered to be correct for a given example; and (3) certain translations were very infrequent. It is likely that a less strict evaluation regime, such as one which takes account of synonym translations, would yield higher accuracies.

4.2 WSD for Machine Translation

Since the "senses" in the multilingual task are actually "translations", the quality of the models produced can be directly evaluated in any application involving translation, particularly Machine Translation (MT) itself. This evaluation would be straightforward in a rule-based MT system: the WSD rules could simply be added to the set of translation rules. However, we chose instead to investigate the contribution of the WSD models to Statistical Machine Translation (SMT), since it yields more promising results for large scale applications.

Although it has been always thought that WSD can be useful for MT, only recently efforts have been made towards integrating both tasks to prove that this assumption is valid, particularly for SMT. While different approaches have been proposed and results started to converge in a positive way, it was not clear yet how these applications should be integrated to allow the strengths of both to be exploited. We propose a new approach to efficiently integrate WSD with SMT, allowing the use of rich contextual WSD features, which is otherwise not done in current SMT systems.

We used a syntactically motivated phrase-based SMT system [12], in which candidate translations are scored according to a linear combination of feature functions. Our approach follows the n-best reranking technique proposed by [8], in which a new feature (in this case, a WSD feature) can be combined to the existing ones at translation time, as opposed to training time, to select the best scoring candidate translation from a list of n-best candidate sentences produced by the SMT system, the so called the *n-best list*. The weights of all the feature functions (including the WSD feature) are optimized against an automatic evaluation metric considering only the candidates in the n-best list. For knowledge-intensive features like the WSD feature, this procedure is computationally much more efficient than estimating the weights during training.

The original SMT system, which we call *baseline SMT system*, has nine features, including the length of the candidate translation, the probability of the source sentence given the target sentence and vice-versa, etc. In our experiments, it was trained on a corpus of 700K English-Portuguese sentences extracted from several sources, mostly from the European Parliament data. The initial weight of these nine features was estimated using 4K sentences.

The value for the WSD feature is “1”, if the translation found in the candidate sentence matches the prediction proposed by the WSD model for that sentence, and “0” otherwise. We look for the translation given by the WSD system in the sentences only in positions that are (at least partially) aligned to the ambiguous source verb. The morphological variations of the possible translations (person, tense, number, etc.) are taken into account through the expansion of the list of possible WSD predictions using a Portuguese lexicon.

The estimation of this new feature weight, as well as the re-estimation of the remaining feature weights, is performed using the n-best list of another 4K-sentence development set, corresponding to the WSD training data, for which the sense annotation was available. The new feature weights are then used to rerank the n-best list of the test data. For example, consider in Fig. 2 the top-2 candidate translations produced by the baseline SMT system for the sentence *s* (its reference translation being *r*) in the experiments with the English-Portuguese translation of the verb *ask*.

s: He returned and *asked* me if I wanted anything else and whether I had enjoyed my meal.
r: Ele voltou, e **perguntou** se eu queria mais alguma coisa, se eu tinha gostado

Ele voltou, e **pediu-me** se eu queria mais alguma coisa e se eu tinha gostado.
Ele voltou, e **perguntou** se eu queria mais alguma coisa, se tinha gostado.

Fig. 2. Top-2 candidate translations for *s* as given by the SMT system

The prediction given for this sentence by the WSD models is "perguntar" (*inquire, enquire*). The top-scored sentence in Fig. 2 uses a different translation: "pedir" ("pediu-me") (*make a request*). The second candidate contains the correct prediction according to the WSD system, inflected as "perguntou". After including the WSD feature, which has the values "0" and "1" for the two sentences, respectively, and optimizing the feature weights for the whole set of features, we obtain new scores for each candidate and re-rank the n-best list of candidates. The second candidate, which is the most appropriate, becomes the top one.

In order to evaluate the contribution of the WSD feature to the overall quality of the SMT system, we evaluate the system using an automatic evaluation metric, BLEU [9], which computes the average of matching n-grams between the system's output and reference translations. The score of the SMT system improved from 0.3248 to 0.34, which is statistically significant according to a paired t-test with $p < 0.05$. In fact, the WSD variations occur mostly in a single word, i.e., the translation of the ambiguous verb, and this usually does not have a significant impact in BLEU scores. This improvement is comparable to that obtained by other approaches integrating WSD and SMT for other language pairs and datasets, e.g. [3]. Details about the integration method and results can be found in [15].

4.3 Monolingual task – Senseval[1] verbs

For the monolingual scenario, we use the sense tagged corpus and sense repositories provided for verbs in Senseval-3. There are 32 verbs with between 40 and 398 examples each. The number of senses varies between 3 and 10. Table 2 shows the average accuracy obtained by Aleph and other systems in the monolingual task for Senseval-3 verbs.

[1] Senseval (www.senseval.org) is a Project which provides benchmarks for WSD and regularly organizes evaluation competitions.

Table 2. Accuracies obtained in the Senseval-3 monolingual task

System	Majority sense	Syntalex-3	CLaC1	MC-WSD	Aleph
% Average accuracy	0.56	0.67	0.67	0.72	0.72

Results are very encouraging for this dataset: our approach was not customized for this monolingual task, instead, we simply removed the multilingual knowledge sources, and still, the ILP approach significantly outperformed most of the other approaches and performed as well as the state-of-the-art system. As with the multilingual task, the models produced contain a small number of rules. An evaluation of the contribution of each knowledge source for the overall performance in this dataset, both individually and in combination with others, can be found in [16]. Experiments with another monolingual dataset, namely the official participation on the last edition of Senseval (SemEval-2007), which included not only verbs but also nouns, can be found in [18]. Our system came in fourth place in that competition (out of 15 systems), which is a very positive result, considering that, again, it was not tuned for the monolingual task and, particularly, for the disambiguation of nouns.

5. Conclusion

We have introduced a new hybrid approach to WSD which uses ILP to combine deep and shallow knowledge sources. ILP induces expressive disambiguation models which include relations between knowledge sources. It is an interesting approach to learning which has not been explored for WSD. Results from both multilingual and monolingual tasks demonstrate that the hypothesis put forward in this paper, that ILP's ability to generate expressive rules which combine and integrate a wide range of knowledge sources is beneficial for WSD systems, is correct. Results for the multilingual task are validated in the experiments with the use of the WSD predictions in a machine translation system, yielding significant improvement in the translation accuracy.

By customizing the sense repository and knowledge sources, the proposed approach could be exploited for any other application requiring lexical disambiguation. We believe that, by accurately identifying the sense of words, where the definition of "sense" is dependent on the application under consideration, this approach could contribute to advance the state-of-the-art of many applications, particularly Information Retrieval, Information Extraction and Question Answering, both in monolingual and multilingual scenarios, where it is crucial to connect the meaning of the words in the query to the meaning of the words in the documents/information to be retrieved. Our goal for future work is to customize and integrate this approach to such applications.

References

1. Agirre, E., Rigau, G.: Word Sense Disambiguation using Conceptual Density. 15th Conference on Computational Linguistics, pp. 16--22, Copenhagen (1996)

2. Agirre, E., Marquez, L., Wicentowski, R.: Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), Prague (2007)
3. Carpuat, M., Wu, D.: Improving Statistical Machine Translation Using Word Sense Disambiguation. Empirical Methods in Natural Language Processing, pp. 61--72, Prague (2007)
4. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Massachusetts (1998)
5. Lin, D.: Principle based parsing without overgeneration. 31st Meeting of the Association for Computational Linguistics, pp. 112--120, Columbus (1993)
6. Mihalcea, R, Chklovski, T. and Kilgarriff, A.: The Senseval-3 English Lexical Sample Task. Senseval-3: 3rd International Workshop on the Evaluation of Systems for Semantic Analysis of Text, pp. 25--28, Barcelona (2004)
7. Muggleton, S.: Inductive Logic Programming. New Generation Computing, 8(4):295--318 (1991)
8. Och, F. J., Gildea, D., Khudanpur, S., Sarkar, A., Yamada, K., Fraser, A., Kumar, S., Shen, L., Smith, D., Eng, K., Jain, V., Jin, Z., Radev, D.: A Smorgasbord of Features for Statistical Machine Translation. Human Language Technology, pp. 161--168, Boston (2004)
9. Papineni, K., Roukos, S., Ward, T. and Zhu, W. J.: BLEU: a method for automatic evaluation of machine translation. 40th Annual Meeting of the Association for Computational Linguistics, pp. 311--318, Philadelphia (2002)
10. Parker, J., Stahel, M.: Password: English Dictionary for Speakers of Portuguese. Martins Fontes, São Paulo (1998)
11. Procter, P. (editor): Longman Dictionary of Contemporary English. Longman Group, Essex (1978)
12. Quirk, C., Menezes, A., Cherry, C.: Dependency Treelet Translation: Syntactically Informed Phrasal SMT. 43rd Annual Meeting of the Association for Computational Linguistics, pp. 271--279, Ann Arbor (2005)
13. Ratnaparkhi, A.: A Maximum Entropy Part-Of-Speech Tagger. Conference on Empirical Methods in Natural Language Processing, pp. 133--142, New Jersey (1996)
14. Schütze, H.: Automatic Word Sense Discrimination. Computational Linguistics, 24(1): 97--123 (1998)
15. Specia, L. Sankaran, B., Nunes, M.G.V.: n-Best Reranking for the Efficient Integration of Word Sense Disambiguation and Statistical Machine Translation. CICLing, pp. 399--410, Haifa (2008a)
16. Specia, L., Nunes, M.G.V., Stevenson, M.: Assessing the contribution of shallow and deep knowledge sources for word sense disambiguation. Journal of Language Resources and Evaluation [under review] (2008b)
17. Specia, L., Nunes, M.G.V., Stevenson, M.: Learning Expressive Models for Word Sense Disambiguation. 45th Annual Meeting of the Association for Computational Linguistics, pp. 41--48, Prague (2007a)
18. Specia, L., Nunes, M.G.V., Srinivasan, A., Ramakrishnan, G.: USP-IBM-1 and USP-IBM-2: The ILP-based Systems for Lexical Sample WSD in SemEval-2007. 4th International Workshop on Semantic Evaluations, pp. 442--445, Prague (2007b)
19. Specia, L., Nunes, M.G.V., Stevenson, M.: Exploiting Parallel Texts to Produce a Multilingual Sense Tagged Corpus for Word Sense Disambiguation. Conference on Recent Advances on Natural Language Processing, pp. 525--531, Borovets (2005)
20. Srinivasan, A.: The Aleph Manual. Technical Report. Computing Laboratory, Oxford University (2000)
21. Stevenson, M., Wilks, Y.: The Interaction of Knowledge Sources for Word Sense Disambiguation. Computational Linguistics, 27(3):321--349 (2001)