From efficiency to portability: acquisition of semantic relations by semi-supervised machine learning

Vincent CLAVEAU
OLST - University of Montréal
C.P. 6128, succ. Centre-Ville
Montréal, QC, H3C 3J7
Canada,
Vincent.Claveau@umontreal.ca

Pascale SÉBILLOT
IRISA - University of Rennes 1
Campus de Beaulieu
35042 Rennes Cedex
France,
Pascale.Sebillot@irisa.fr

Abstract

Numeric approaches to the corpus-based acquisition of lexical semantic relations offer robust and portable techniques, but poor explanations of their results. On the other hand, symbolic machine learning approaches can infer patterns of a target relation from examples of elements that verify this relation; the produced patterns are efficient and expressive, but such techniques are often supervised, *i.e.* require to be (manually) fed by examples. This paper presents two original algorithms to combine one technique from each of these approaches, and keep advantages of both (meaningful patterns, efficient extraction, portability). Moreover the extraction results of these two semi-supervised hybrid systems, when applied in an illustrative purpose to the acquisition of semantic noun-verb relations defined in the Generative Lexicon framework (Pustejovsky, 1995), rival those of supervised methods.

1 Introduction

Natural language processing applications often require semantic knowledge, *e.g.* semantic relations between words that vary from one domain/application to another; automatic acquisition of such a knowledge from corpora is thus necessary. Much of the work developed within this area of automatic acquisition of lexical semantic relations—mainly considered as collocations—is dedicated to numeric approaches (Pearce, 2002; Manning and Schütze, 1999); those methods, even if they often obtain satisfactory results, cannot however provide any explanation about those results but a statistical score. On the other hand, other much rarer studies focus on a symbolic approach of the extraction, using predefined morpho-syntactic patterns of the relations if they are known, or using symbolic machine learning (ML) techniques to infer those patterns from examples (and counter-examples) of elements verifying the target relation, if they are unknown. The advantages of these symbolic approaches lie in the expressiveness of their produced patterns, that can be verbalized in order to explain the relation. One drawback of the majority of ML methods comes however from the supervision phase, that is, the fact that they must be (manually) fed by sets of examples and counter-examples, built for each new corpus/studied relation by a human expert.

In this paper, we show that it is possible to take advantage of key-points from both numeric and symbolic approaches, by combining in a so-called *semi-supervised* acquisition technique one method from each of these families. We thus present semi-supervised versions of ASARES, a symbolic ML method that allows us to infer morpho-syntactic and semantic patterns of semantic relations from the descriptions of some pairs of elements linked (or not) by the target relations. ASARES, based on inductive logic programming (ILP, (Muggleton and De-Raedt, 1994)), is deeply presented in (Claveau et al., 2003). This supervised system produces efficient and linguistically motivated patterns, useful for the study of the corpus-specific structures conveying a target relation, but its portability is limited by its supervised nature. This paper focuses on the proposition of two different and original algorithms to combine a statistic approach of acquisition to ASARES's supervised version in order to solve this drawback, while keeping the previous good results and corpus-specific expressive patterns. The resulting semi-supervised versions of ASARES thus fulfill our three objectives: first, they produce efficient patterns, able to extract, once applied to a corpus, pairs of elements actually bound by a relation; secondly, these patterns are expressive, that is, linguistically relevant; last, these methods are generic and easily portable from one corpus/relation to another.

Though not dedicated to any particular semantic link, ASARES and its semi-supervised versions, in order to be evaluated, are presented here applied to the acquisition of one type of semantic relations: the *qualia* relations as defined in the Generative Lexicon (GL) formalism (Pustejovsky, 1995). GL is a lexicon model in which lexical entries consist of structured sets of predicates that define a word. In one of the components of this model, called the *qualia structure*, words are described with semantic roles such as the purpose or function (*e.g.* *cut* for *knife*), or the creation mode (*build* for *house*)... For a given word, each role can get numerous realizations, and the qualia structure of each word, especially for common nouns (N), is mainly made up of verbal (V) associations, encoding relational information. Such N-V pairs, in which V plays one of the qualia role of N, are called *qualia pairs* hereafter. Positioning our work within GL linguistic theory provides a framework to evaluate the performances of the various versions of ASARES by clearly defining the semantic relations they have to focus on. Indeed, our choice of qualia relations has been essentially guided by 3 reasons:

1- qualia extraction patterns are not known; using a symbolic method has therefore a linguistic interest since we learn patterns that characterize globally qualia relations (without currently trying to distinguish each role);

2- only a few projects have been undertaken to construct qualia structures. Among them, Lapata and Lascarides (2003) and others make strong assumptions on the structures conveying the qualia relations and relies on the good results of syntactic parsers, not available for most of the languages. This lack of good syntactic parsers also prevents us from adapting for the qualia acquisition task other acquisition methods producing patterns developed within different frameworks (*e.g.* (Briscoe and Carroll, 1997));

3- moreover some authors have pointed out the relevance of N-V links for index expansion in information retrieval systems (Fabre and Sébillot, 1999) without being able to fully test their hypothesis because of the lack of such lexical resources. ASARES will be used to solve this problem[1].

This paper will be divided into 4 parts. Section 2 is dedicated to the presentation of ASARES's supervised version when applied to a technical corpus in order to extract qualia relation patterns. The main principles of ILP are given, and the supervised, problematic aspect is pointed out. Section 3 explains the kind of statistic approaches that has to be associated with ASARES in order to avoid the necessity of feeding it with examples, and the 2 semi-supervised, hybrid, versions are described. Section 4 compares the results of the 4 systems (supervised ASARES, statistic approach alone, and 2 semi-supervised versions), and clearly shows that semi-supervised methods reach the same scores for the extraction of N-V qualia pairs when applying the inferred patterns to the corpus. A brief discussion of the linguistic relevance of these patterns and a positioning of our work within the domain dedicated to combination of ML methods is also provided in this section.

2 Supervised version of ASARES

After the description of the corpus used for our experiments, this section presents the supervised version of ASARES, based on ILP, on which our hybrid systems rely.

2.1 Corpus and tagging

The corpus used during our different experiments is a collection of helicopter maintenance handbooks, provided to us by MATRA-CCR Aérospatiale. This French 700 KBytes technical corpus contains more than 104,000 word occurrences. It has some special characteristics that are especially well-suited for our task: its vocabulary and syntactic structures are homogeneous; it contains many concrete terms that are frequently used in sentences together with verbs indicating their functions or modes of creation. The corpus has been segmented into sentences and words and then lemmatized, Part-of-Speech (POS) tagged and disambiguated with the MULTEXT tools (Armstrong, 1996). The quality of this morpho-syntactic tagging is very good: less than 2% of errors are detected when compared to a 4,000 word manually tagged test-sample of the corpus. Following the work described in (Bouillon et al., 2000), a semantic tagging has also been carried out on the already POS-tagged corpus. This tagging relies on a tagset composed by the most generic WordNet classes and by other classes with a granularity more adapted to our corpus. Here again, the error rate, estimated from a 6,000 word manually tagged sample, is very low: 98.82% of the words are correctly tagged.

[1]This is why we do not try to learn each qualia role separately: knowing that N and V are semantically linked is sufficient for a use in information retrieval.

2.2 Supervised extraction rule learning by ILP

The use of symbolic ML methods for NLP tasks is becoming more common. Among these methods, ILP, thanks to its expressiveness and flexibility, has been applied to numerous problems (see (Cussens and Džeroski, 2000) for an overview of this domain). We briefly present the use of ILP we propose in the qualia extraction framework (Claveau et al., 2003).

ILP aims at producing general rules (more precisely Horn clauses) explaining a concept from examples and counter-examples of the concept and from a background knowledge. Here, the concept to be learned is the qualia nature of a N-V pair occurring within a sentence. A GL expert has manually built up a set of examples E^+ and counter-examples E^-; that is, she has extracted from our corpus qualia and non-qualia N-V pairs with their contexts (all the words and their tags occurring with the pair within a sentence). The ILP system is given this way about 3,000 examples and 3,000 counter-examples.

A hypothesis language is also provided to the ILP system; it is used to precisely define the expected form of the generated rules (or hypotheses). In the qualia extraction case, this language makes the most of the POS and semantic tags of words occurring in the examples (N-V pairs and their contexts) and distance information between N and V. For example, the rules produced, which are then used as patterns to extract new qualia pairs, look like[2]:
is_qualia(N,V) :- precedes(V,N), near_verb(N,V), infinitive(V), action_verb(V), artifact(N).
This rule means that a pair composed by a noun N and a verb V will be considered as qualia if N appears in a sentence after V, V is an action verb in the infinitive and N is an artifact.

According to this language, the ILP algorithm infers rules that cover (that is, explain) a maximum of examples and no counter-examples (or only a few, some *noise* can be allowed in order to produce more general patterns), by generalizing the examples. More precisely, the inference process follows the following steps:

1 - select one example $e \in E^+$ to be generalized. If none exists, stop.

2 - define a hypothesis search space $\mathcal{H}$ according to e and the hypothesis language;

3 - search $\mathcal{H}$ for the rule h that maximizes a score function Sc;

4 - remove the examples that are covered by the chosen rule. Return to step 1.

The score function Sc depends on the number of examples and counter-examples covered by a hypothesis h. These two sets are respectively called E_h^+ and E_h^- and their cardinals $|E_h^+|$ and $|E_h^-|$. The inferred rules obtained at the end of these steps are then used as qualia N-V pair extraction patterns to retrieve new qualia pairs from the corpus.

As explained in (Bouillon et al., 2002), this method (hereafter supervised ASARES) gives good results for the qualia pair extraction task; moreover, it gives access, through the produced rules, to a linguistically interpretable support to the concept of qualia role. However, the cost of this method, essentially lying in the construction by an expert of the example and counter-example sets, makes this technique too time-consuming, and thus difficult to apply to any new corpus or relation.

3 Semi-supervised versions of ASARES

In order to fulfill our goal of making ASARES fully automatic, we must suppress the expert supervision step. In this respect, we propose to combine the ILP algorithm with standard statistical extraction techniques. After describing these statistical techniques, we successively present in this section two original implementations of such a combination. The first one is a sequential combination of the two approaches; the second one integrates the statistical results inside the ILP algorithm. The two resulting systems, relying on a supervised ML technique but without the need of supervision, are then called semi-supervised.

3.1 Statistical extraction of qualia pairs

A lot of work has been conducted about cooccurrence extraction by statistical approaches (Manning and Schütze, 1999). In this framework, the qualia N-V pairs are then seen as a special kind of collocations. The experiments reported in (Bouillon et al., 2002) present results obtained for this task with some of the most common statistical association criteria (Kulczinsky, Ochiai, Yule, Loglike, Simple Matching, Mutual Information, cubed Mutual Information, Φ^2). Among them, the cubed Mutual Information (Daille, 1994) (MI^3 hereafter) gives the best results. With the notations given in the contingency table 1 (the cooccurrences indicated are

[2] A discussion on the produced patterns is proposed in (Bouillon et al., 2002) and in section 4.2.

computed in the scope of a sentence with the lemmas of the words), the MI^3 coefficient is defined by: $log_2 \frac{a^3}{(a+b)(a+c)}$.

	V_j	V_k, $k \neq j$
N_i	a	b
N_l, $l \neq i$	c	d

Table 1: Contingency table for the pair N_i-V_j

However qualia extraction conducted using this technique (Bouillon et al., 2002) leads to much less satisfactory results than those obtained with supervised ASARES. In particular, interesting pairs occurring rarely in the corpus cannot be retrieved; such pairs are said to be "*under the noise level*". Moreover, this kind of methods does not provide any comprehension element to interpret the results and thus does not meet our requirements. Nevertheless, their complete autonomy (no human intervention is required) and utilization ease are qualities we want to preserve in the approaches we present below.

3.2 Sequential extraction system

The hybrid system proposed here relies on a sequential combination of the statistical and symbolic systems described above (given in algorithm 1). Each system iteratively uses as input the output data of the other one. More precisely, the N-V pair list generated by a system (L_{ILP} for ILP, L_{MI^3} for the statistical system) is used by the other one to construct its own N-V pair list. The only constraint is to begin this iteration with the statistical system since it does not need any data but the corpus. At the initialization step, every N-V pair occurring within a sentence is considered as potentially qualia; this is indicated through the rule is_qualia(N,V). provided in the extraction pattern list L_R.
The loop terminates when the same set of rules is obtained during two successive iterations. During our experiments, n_1 was chosen (at each iteration) such that the n_1 first N-V pairs of L_{MI^3} were the pairs with a positive association score; and n_2 was chosen such that $n_2 = n_1$. The resulting extraction technique is called hereafter sequential hybrid system.

3.3 Integrated extraction system

Unlike the system presented above in which the statistical and ILP-based systems are used without major modifications, the second hybrid extraction technique we propose combines them

Algorithm 1 Sequential hybrid system

Initialization

- $L_R = \{\texttt{is\_qualia(N,V).}\}$
- application of the rules in L_R to the corpus; the retrieved N-V pairs and their number of occurrences are inserted in L_{ILP}

Iteration

1. for every $N_i - V_j$ pair in L_{ILP}
   - setting up of the $N_i - V_j$ contingency table with number of occurrences given in L_{ILP}
   - MI^3 score computing of $N_i - V_j$
   - according to its score, insertion of the pair in the list L_{MI^3} in descending order
2. setting up of E^+ (respectively E^-) from every occurrences in the corpus of the n_1 first (resp. n_2 last) pairs of L_{MI^3}
3. ILP learning with E^+ and E^-; the produced rules are put into L_R
4. application of the rules in L_R to the corpus; the retrieved N-V pairs and their number of occurrences are put into L_{ILP}

more finely and implies some changes in the ILP algorithm.

As mentioned in subsection 2.2, during the third learning step, a rule h is chosen from a hypothesis space $\mathcal{H}$ if it maximizes a score function Sc that depends on the number of examples and counter-examples it covers, that is, $h = \underset{h \in \mathcal{H}}{\text{argmax}} \ Sc\left(|E_h^+|, |E_h^-|\right)$.

The principle of our second hybrid system is to weight the examples according to their statistical scores so that the hypotheses are now evaluated with the help of these weighted examples. The sets of weighted examples and counter-examples are thus built with the MI^3 extraction system: the highest MI^3 scored pairs are put in E^+, and conversely, the lowest in E^-; their weights w are computed from their MI^3 scores. Therefore, the more the example is considered interesting (that is, highly scored) by the statistical technique, the more it influences the choice of rules. Finally, the rules that are kept are those maximizing $Sc(h)$ redefined as:

$$h = \underset{h \in \mathcal{H}}{\text{argmax}} \ Sc\left(\sum_{e^+ \in E_h^+} w(e^+), \sum_{e^- \in E_h^-} w(e^-)\right)$$

With this setting, and the weighted examples and counter-examples, and according to the hypothesis language defined, our modified ILP algorithm produces rules that can be used as extraction patterns for our qualia acquisition task; this extraction technique is called integrated hybrid system hereafter.

4 Evaluation and comparison of performances

This section describes the results of the 2 semi-supervised versions of ASARES compared with the supervised one and the statistical extraction system alone. It also presents some of the patterns obtained with the 3 versions of ASARES and proposes some further comments on the semi-supervised approaches.

4.1 Extraction results

To clearly evaluate the two hybrid extraction systems in real-world conditions, an empirical test-set has been constructed by 4 GL experts. The test corpus on which the qualia-pair extraction is performed is a 32,000 word subset of the MATRA-CCR corpus. In spite of its relatively small size, it is impossible to manually examine every N-V pair to class it as qualia or non-qualia. We have thus focused our attention on 7 domain relevant common nouns: *vis, écrou, porte, voyant, prise, capot, bouchon* (screw, nut, door, indicator signal, plug, cowl, cap)[3]. Every N-V pair such that N is one of the 7 nouns occurring within a sentence in the sub-corpus is retrieved. Then, the 4 experts manually tag each one as relevant (that is, qualia) or not[4]. Divergences are discussed until complete agreement is reached. Finally, among the 286 examined pairs, 66 are classified qualia (each N has between 4 and 17 V in qualia relations). This test-set is therefore used to compare the extraction results of each system with the human expert ones.

In order to draw this comparison, and since extraction systems based on statistical measure (such as MI^3) assign a coefficient value to each N-V pair, a coefficient threshold (s hereafter) has to be chosen: all the pairs whose statistical scores reach s are considered as qualia, the others are considered as non-qualia. In the same way, a threshold must be set up for our supervised and semi-supervised versions of ASARES since a pair can be considered as qualia only when at least s of its occurrences are retrieved by the extraction patterns. For statistical as well as for symbolic systems, the higher s is, the higher the precision rate is, and conversely, the lower s is, the higher the recall rate is. The recall and precision rates, measured on our test set, are thus defined (TP means True Positives, FP False Positives and FN False Negatives) according to s: $R(s) = \frac{TP(s)}{TP(s)+FN(s)}$, $P(s) = \frac{TP(s)}{TP(s)+FP(s)}$.

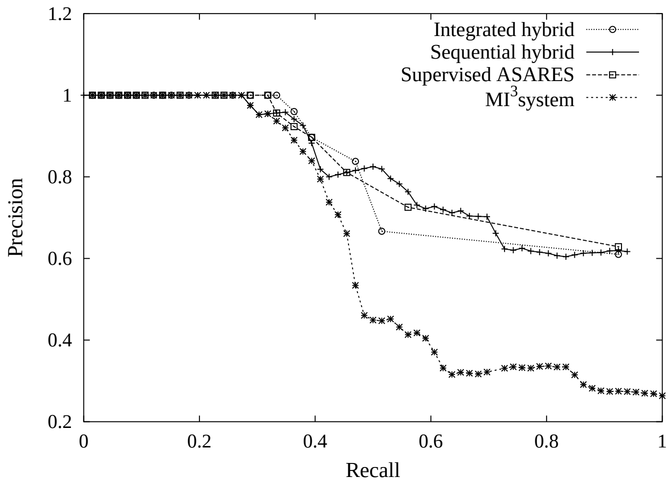

Figure 1: Recall-Precision graph for the 4 systems

To represent performances of such systems for every possible values of s we use a recall-precision graph; thus, each point represents the precision of the system according to its recall for a given s. Figure 1 presents these graphs for the 4 systems. The similarity between the curves of the two hybrid systems and the supervised ASARES one shows that these 3 systems react very closely. Indeed, their precision rates are quite similar whatever the recall rate, even if the sequential hybrid system performs slightly better than the integrated one. This is confirmed in table 2 where we indicate the recall and precision rates and the F-measure (the harmonic mean of R and P: $F(s) = \frac{2P(s)R(s)}{P(s)+R(s)}$) obtained when s is such that F is maximal. Moreover, the 3 symbolic systems always perform better than the statistical one, especially when recall is high. The aim of extraction efficiency is thus fulfilled by our semi-supervised versions of ASARES.

When examining the results, it appears that several causes are responsible for the errors, that is for non-retrieved valid pairs and non-valid retrieved ones. These causes, peculiar to our sym-

[3] To prevent distortion of results, none of these common nouns were used as examples or counter-examples for the pattern induction in the 2 hybrid and the supervised systems.

[4] The potential hyperonymic links between verbs given by our semantic tagging are not taken into account, each N-V pair is examined separately.

	recall	precision	F-meas.
Supervised ASARES	92.4%	62.2%	0.744
MI^3	36.4%	92.3%	0.522
Sequential hybrid	93.9%	62.0%	0.747
Integrated hybrid	89.4%	60.2%	0.720

Table 2: Optimal performances of the 4 systems

bolic approach, are examined in detail in (Bouillon et al., 2002). They mainly rely on tagging errors and on a lack of semantic and syntactic information on the corpus.

4.2 Discussion on the inferred patterns

As previously mentioned, one of the interests of symbolic learning techniques like ILP is the production of meaningful patterns. A qualitative evaluation of our two hybrid methods has thus been performed by examining the linguistic relevance of the generated extraction patterns. In that respect, the rules obtained by the two semi-supervised systems are very similar to the ones using the supervised version of ASARES (see (Bouillon et al., 2002) for details). More precisely, these rules are very general linguistic patterns making the most of morpho-syntactic information (POS tags) of the N or V (infinitive verb for example) and distance information between the N and the V. However, little semantic information (semantic tags) is used in these patterns, except for the V (action verbs are favored), like in the pattern seen in section 2.2 equivalent to *infinitive action V + (anything but a verb)* + N of artifact*. Moreover, these rules point out surface clues, very specific to the MATRA-CCR corpus like punctuation marks, that are generally neglected by manual analysis. For example, ASARES in its 3 versions produces a pattern similar to: is_qualia(N,V) :- precedes(V,N), singular_common_noun(N), suc(V,C), colon(C), pred(N,D), punctuation(D). This pattern, equivalent to *V + : + (any token)* + [:,;] + singular N*, covers the very frequent enumerations of our corpus. All the patterns produced are thus very interesting for a linguistic analysis, highlighting corpus-specific structures. In this respect, our two semi-supervised versions of ASARES thus fulfill our objective of producing interpretable results.

4.3 Further comments and related work

Several studies aim at improving performances and costs of supervised ML methods, not by working on new algorithms, but by using existing ones (and the classifiers they produce) in a particular way. Some work, sharing our concerns, tries to construct semi-supervised learning methods from supervised ones. Most of these last techniques rely on *bootstrapping* variants (Jones et al., 1999): a few hand-annotated examples are first used to produce a primary classifier; this one then serves to annotate more examples which are used to generate a second classifier and so on. High-level bootstrapping versions such as *co-training* (Blum and Mitchell, 1998) or Yarowsky's algorithm (Yarowsky, 1995) ensure interesting theoretical properties on the PAC learnability, but are granted to strong assumptions about the data. In the case of our two semi-supervised versions of ASARES, such learnability results cannot be found, due to the inadequacy of the PAC model to describe ILP. However, the use of statistics (seen as a sort of probability distribution over E^+ and E^-) in ILP, as it is done in our integrated hybrid system, raises interesting issues in the ML domain (Muggleton, 1994).

5 Concluding remarks and future work

We have presented two semi-supervised symbolic hybrid systems that extract from corpus semantic relations. They rely on a combination of a supervised symbolic pattern learner, ASARES, and a statistical extraction technique. They preserve advantages of each of these two different extraction approaches: unsupervised aspect of the statistical acquisition, linguistically meaningful contextual pattern generation of the supervised symbolic one. These two semi-supervised versions of ASARES have been applied and evaluated on the extraction of N-V qualia relations. The extraction results of the two hybrid systems both rival the supervised ASARES. Moreover, their symbolic nature gives access to corpus-specific structures conveying the target relation. Furthermore, thanks to the unsupervised aspect of the MI^3 extraction they make the most of, our two hybrid systems are fully-automatic and easily portable from a corpus to another. Indeed, the semi-supervised systems only need about 15 minutes to perform the acquisition while several hours are required to annotate manually examples for the supervised system. These semi-supervised systems thus fulfill our 3 objectives of producing good results and interpretable patterns in a fully automatic way.

Further studies are currently undertaken or planed on several aspects of this work. First, these semi-supervised versions of ASARES are currently being applied to the extraction of semantic relations defined in a different framework (Lexical Functions (Mel'čuk, 1998)), and thus confirm their genericity. Secondly, the qualia pairs acquired by ASARES are to be used in information retrieval experiments to extend queries; first results confirm the interest of qualia relation in this context (Claveau and Sébillot, 2004). Moreover, the newspaper corpus used in these experiments confirms the interest of our symbolic approach in a more heterogeneous context. Finally, more efficient ways to combine statistical and symbolic approaches should be studied in order to improve the extraction results and thus to exceed those of the supervised technique.

References

Susan Armstrong. 1996. Multext: Multilingual Text Tools and Corpora. In H. Feldweg and W. Hinrichs, editors, *Lexikon und Text*. Tübingen: Niemeyer.

Avrim Blum and Tom Mitchell. 1998. Combining Labeled and Unlabeled Data with Cotraining. In *COLT: Workshop on Computational Learning Theory*, Madisson, WI, USA.

Pierrette Bouillon, Robert H. Baud, Gilbert Robert, and Patrick Ruch. 2000. Indexing by Statistical Tagging. In *Journées d'Analyse statistique des Données Textuelles, JADT2000*, Lausanne, Switzerland.

Pierrette Bouillon, Vincent Claveau, Cécile Fabre, and Pascale Sébillot. 2002. Acquisition of Qualia Elements from Corpora - Evaluation of a Symbolic Learning Method. In *3rd International Conference on Language Resources and Evaluation, LREC 02*, Las Palmas de Gran Canaria, Spain.

Ted Briscoe and John Carroll. 1997. Automatic Extraction of Subcategorisation from Corpora. In *5th ACL Conference on Applied Natural Language Processing, ANLP97*, Washington, WA, USA.

Vincent Claveau and Pascale Sébillot. 2004. Extension de requêtes par lien sémantique nom-verbe acquis sur corpus. In *11ème conférence de Traitement automatique des langues naturelles, TALN'04*, Fès, Morocco.

Vincent Claveau, Pascale Sébillot, Cécile Fabre, and Pierrette Bouillon. 2003. Learning Semantic Lexicons from a Part-of-Speech and Semantically Tagged Corpus using Inductive Logic Programming. *Journal of Machine Learning Research, special issue on ILP*, 4:493–525.

James Cussens and Sašo Džeroski, editors. 2000. *Learning Language in Logic*. LNAI. Springer Verlag.

Béatrice Daille. 1994. *Approche mixte pour l'extraction de terminologie : statistique lexicale et filtres linguistiques*. Ph.D. thesis, Université Paris 7.

Cécile Fabre and Pascale Sébillot. 1999. Semantic Interpretation of Binominal Sequences and Information Retrieval. In *International ICSC Congress CIMA/AIDA*, Rochester, NY, USA.

Rosie Jones, Andrew McCallum, Kamal Nigam, and Ellen Riloff. 1999. Bootstrapping for Text Learning Tasks. In *IJCAI-99 Workshop on Text Mining: Foundations, Techniques and Applications*, Stockholm, Sweden.

Maria Lapata and Alex Lascarides. 2003. A Probabilistic Account of Logical Metonymy. *Computational Linguistics*, 29(2):263–317.

Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press.

Igor Mel'čuk. 1998. Collocations and Lexical Functions. In *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press.

Stephen Muggleton and Luc De-Raedt. 1994. Inductive Logic Programming: Theory and Methods. *Journal of Logic Programming*, 19-20.

Stephen Muggleton. 1994. Bayesian Inductive Logic Programming. In *7th Annual ACM Conference on Computational Learning Theory*, New Brunswick, NJ, USA.

Darren Pearce. 2002. A Comparative Evaluation of Collocation Extraction Techniques. In *3rd International Conference on Language Resources and Evaluation, LREC 02*, Las Palmas de Gran Canaria, Spain.

James Pustejovsky. 1995. *The Generative Lexicon*. The MIT Press.

David Yarowsky. 1995. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. In *33rd Annual Meeting of the Association for Computational Linguistics, ACL*, Cambridge, MA, USA.