Using Part-of-Speech and Semantic Tagging for the Corpus-Based Learning of Qualia Structure Elements

Pierrette Bouillon
TIM/ISSCO, Université de Genève, 40 Bvd du Pont-d'Arve, CH-1211 Geneva 4, Switzerland
Pierrette.Bouillon@issco.unige.ch

Vincent Claveau
IRISA, Campus de Beaulieu, 35042 Rennes cedex, France
vclaveau@irisa.fr

Cécile Fabre
ERSS, Université de Toulouse II, 5 allées A. Machado, 31058 Toulouse cedex, France
cfabre@univ-tlse2.fr

Pascale Sébillot
IRISA, Campus de Beaulieu, 35042 Rennes cedex, France
sebillot@irisa.fr

Abstract

This paper describes the implementation and results of a machine learning method, developed within the inductive logic programming (ILP) framework (Muggleton and De-Raedt, 1994), to automatically extract, from a corpus tagged with parts of speech (POS) and semantic classes, noun-verb pairs whose components are bound by one of the relations defined in the qualia structure in the Generative Lexicon (Pustejovsky, 1995). We demonstrate that the semantic tagging of the corpus improves the quality of the learning, both on a theoretical and an empirical point of view. We also show that a set of the rules learnt by our ILP method have a linguistic significance regarding the detection of the clues that distinguish, in terms of POS and semantic surrounding context, noun-verb pairs that are linked by one qualia role from others that are not semantically related.

1 Introduction

In the generative lexicon framework (Pustejovsky, 1995), the qualia structure gives access to relational information that prove to be crucial both for linguistic analysis and for NLP applications. In particular, the qualia roles express, in terms of predicative formulae, the basic features (telic, agentive, constitutive, formal) of the semantics of nouns. In this model, the noun is linked not only to other nouns via traditional lexical relations (such as meronymy and hyperonymy) but also to verbs that correspond to typical actions in which the noun is involved. Previous works ((Pustejovsky et al., 1997), (Fabre and Sébillot, 1999)) have focused on such Noun-Verb relations to provide lexical resources in the context of terminology and information retrieval applications.

Given the lack of such resources and the fact that the notion of typical activities associated with a noun considerably varies from one register to another, methods for corpus-based acquisition of Noun-Verb relations are needed. Previous experiments related in (Pustejovsky et al., 1993) are based upon the assumption that the extraction of the qualia structure of a noun can be performed by spotting a set of syntactic structures related to qualia roles. We propose to go one step further as we have no *a priori* concerning the structures that are likely to convey these roles in a given corpus. We therefore develop and apply a symbolic learning method which automatically produces general rules explaining what, in terms of surrounding context, characterizes examples of Noun-Verb (N-V) pairs related by one qualia role from pairs that are not[1]. Learning is first achieved on a part-of-speech(POS)-tagged version of a corpus and the theoretical and empirical evaluations of the method are already good; it is then done on the version of the same corpus that has been this time annotated both with POS and semantic tags and we show that semantic tagging improves the quality of the results, even if we do not up to now exploit all the richness of the semantic information (further research on this aspect is pointed out).

In this paper, we first describe the learning method we have elaborated within the Inductive Logic Programming (ILP) framework and initially tested on a POS-tagged corpus (section 2). We then show how a probabilistic Hidden Markov Model (HMM) tagger can be used to perform the semantic annotation of the corpus (section 3). We finally present the results of the application of our ILP learning method on the POS- and semantically tagged corpus. We conclude this paper by a description of the linguistic significance of some of the general rules obtained with the help of this double level of tagging (section 4).

2 Learning elements of GL from a POS-tagged corpus

This section describes the ILP learning method that we have developed in order to automatically acquire from a POS-tagged corpus N-V pairs whose components are linked by one of the qualia roles. We firstly present our corpus and its POS-tagging; we then explain the machine learning method, and finally explicit its results and the remaining problems.

2.1 The Matra CCR corpus and the POS tagging

The French corpus used in this project is a 700 kBytes handbook of helicopter maintenance, given to us by MATRA CCR Aérospatiale, which contains more than 104000 word occurrences. It has some specific characteristics that are especially well suited for our task: it is coherent; it contains many concrete terms (*screw, door, etc.*) that are frequently used in sentences together with verbs indicating their telic (*screws must be tightened*, etc.) or agentive roles. This corpus has been POS-tagged with the help of annotation tools developed in the MULTEXT project (Armstrong, 1996); sentences and words are first segmented with *MtSeg*; words are analyzed and lemmatized with *Mmorph* (Petitpierre and Russell, 1998; Bouillon et al., 1998), and finally disambiguated by the *Tatoo* tool, a Hidden Markov Model tagger (Armstrong et al., 1995). Besides the tagger itself, the tool includes modules to prepare the text for training and tagging, define new tagsets or modify existing sets, declare linguistic preferences, train or retrain with hand annotated data, and facilitate the comparison between re-

[1]This research is funded by the Agence universitaire de la Francophonie (AUF) (Action de recherche partagée *"Acquisition automatique d'éléments du Lexique Génératif pour améliorer les performances de systèmes de recherche d'information"*, réseau FRANCIL).

sults and hand corrected data. For the experiments described here, the language model is calculated in three stages. First the syntactic model is automatically built on the basis of the ambiguous text (following the Baum-Welch algorithm). The transition probabilities between two POS tags are then refined with a set of linguistic rules (bigrams), very useful to bias the model when training data are not sufficient to make generalizations. For example, the following rule says that a singular determiner (DET-SG) cannot precede a plural verb (VERB-PL): `DET-SG VERB-PL -10`. Finally a small part of the text (4000 words), manually disambiguated, is used to automatically reestimate the model. With this method, the detected error rate for POS-tagging is less than 2 %.

2.2 The machine learning method

Machine learning aims at automatically building programs from examples that are known to be positive (E^+) or negative (E^-) examples of their runnings (Mitchell, 1997). Among different machine learning techniques[2], we have chosen the Inductive Logic Programming framework (Muggleton and De-Raedt, 1994) to learn from the corpus N-V pairs whose components are related in terms of one of the relations defined in the qualia structure in GL.

In the ILP framework, the aim is to infer logic programs, that is, sets of Horn clauses, from a set of facts (the examples) and a background knowledge. The set of generalized clauses that is produced must be as specific as possible, while explaining the positive examples; more precisely, it must be sufficiently generic to cover the majority of the E^+ and sufficiently specific to rightly correspond to the concept we want to learn and to cover no (or a few - some noise can be allowed) E^-.

Here, given a set of E^+, corresponding to N-V pairs related by one of the qualia roles within their POS contexts in the MATRA-CCR corpus, and a set of E^-, corresponding to N-V pairs that are not semantically linked, the ILP method must infer general rules (clauses) that explain these E^+, that is, that explain what, in the POS context of the N-V pairs, distinguishes the relevant pairs from the others. This particular explanatory characteristic of ILP has motivated our choice: ILP does not just provide a predictor (this N-V pair is relevant, this one is not) but also a data-based theory. Contrary to some statistical methods, it does not just give raw results but explains the concept that is learnt.

We use Progol (Muggleton, 1995) for our experiment, Muggleton's ILP implementation that has already been proven well suited to deal with a big amount of data in multiple domains, and to lead to results comparable to other ILP implementations (Roberts et al., 1998).

The first task therefore consists in building up E^+ and E^- that will be furnished to Progol and to determine the contextual elements of the N-V pairs that must be put inside these examples. Our choice of context has been led by the quality of the learning that is obtained with the chosen parameters (see further for a brief description of the learning method evaluation, or (Sébillot et al., 2000; Bouillon et al., 2000) for a complete explanation). For each of the 1489 different N (29633 noun occurrences) of the MATRA-CCR corpus, we have firstly selected the 10 most strongly associated V, in terms of Chi-square, that co-occur with it within a sentence. This step both produces pairs that are really bound by one qualia role (*(écrou, serrer)* (nut, tighten)) and pairs that are fully irrelevant (*(roue, prescrire)* (wheel, prescribe)). Each pair

[2] See, for example, (Wermter et al., 1996) for a survey of machine learning techniques used in natural language processing.

is manually annotated as relevant or irrelevant according to the qualia structure principles. Then for each occurrence of each relevant pair, a Perl program permits a manual control to verify if the N and the V really are in the expected relation within the sentence. After this control, a second Perl program automatically produces the E^+ of the following form: *positive(POS-tag_word_before_N, POS-tag_word_after_N, POS-tag_word_before_V, V_type, distance, position).*, where *V_type* indicates if the V is an infinitive or a conjugated form, *distance* corresponds to the number of verbs between the N and the V, and *position* is POS (for positive) if the V appears before the N in the sentence, NEG if the N appears before the V. E^- are automatically built the same way from the previous highly correlated N-V pairs that have been manually annotated as irrelevant, i.e. whose components are not linked by a qualia role.

The 4000 E^+ and 7000 E^- produced from the corpus are then supplied to Progol in order for it to generalize the E^+ and infer a set G of generalized clauses that explain what distinguishes them from E^-. Here is one example of the 66 elements of G that we have obtained: *positive(A,B,C,D,E,F) :- aux_etre(C), prepositionpar(A), pres(E).*, which means that N-V pairs (i) in which the category before the N is the French preposition *par* ("by") (*prepositionpar(A)*), (ii) in which the word before the V is the auxiliary *être* ("be") (*aux_etre(C)*), (iii) in which there is no verb between the N and the V (proximity denoted by *pres(E)* (close(E)), are relevant; no constraint is set on the category of the word after the N, on the type of V and on N/V order in the sentences.

2.3 Results

The theoretical evaluation of the learning method consists in measuring the percentage of E^+ that are covered by the generalized clauses of G, and the percentage of E^- that are rejected by them. These results can be summarized in a Pearson coefficient. With the chosen contextual elements, 88% of the initial E^+ are covered and only 5% of E^-; Pearson value is 0.84. The method is also empirically evaluated by applying the G clauses on the corpus and by comparing the decisions with a manual expert tagging for 7 significant nouns (*vis, écrou, porte, voyant, prise, capot, bouchon)* (screw, nut, door, indicator signal, plug, cowl, cap). If we consider a N-V pair tagged as *"relevant"* by the generalized clauses if at least one of them covers it once, 49 pairs are correctly found, 54 incorrectly and 10 not found (Pearson value: 0.5138); if we consider that to be said correctly detected by the generalized clauses one pair must be at least covered six times (by different clauses of G or by a same clause in different phrases), values are respectively 23, 4 and 36 (Pearson value: 0.5209). These numbers can be also compared with the initial Chi-square pair selection for which the values are respectively 38, 124 and 21 (Pearson value: 0.1206).

Our learning method on a POS-tagged corpus therefore leads to very promising results: 83.05% of relevant pairs are detected (64% for Chi-square). Moreover ILP learning, contrary to statistical methods, has an explanatory characteristic; our ILP method does not just extract statistically correlated pairs but it permits to automatically learn rules that distinguish relevant pairs from others. The fact that E^- have to be covered by some elements of G however means that something is missing in our E^+ to fully define the concept "qualia pair" *versus* "not qualia pair":

some E^- have to be covered to define it better. A piece of information, maybe syntactic and/or semantic, is missing in our E^+ to fully characterize it. This fact can be easily illustrated by the following example: in structures like 'Verbinf det N1 prep N2', the Verbinf and the N2 are not related in some cases (*vérifier l'absence de corrosion*[3]), but they are related in others, for example when the N1 indicates a part of an object (*vider le fond du réservoir*[4]). A simple POS-tagging of the sentences offers no difference between them. We have therefore semantically tagged our corpus and conducted the same learning experiment with this new information. Here are the descriptions of this tagging and of the first learning results.

3 Learning elements of GL from a POS- and semantically tagged corpus

In this section, we firstly present the semantic classification of the main POS categories in the MATRA-CCR corpus. We then describe how we use these classes as a tagset for the semantic tagging of the corpus and give the tagging results[5]. We finally detail the learning of the relevant N-V pairs from the corpus.

3.1 Noun and verb semantic classification

In order to systematically classify the nouns, the most generic classes of WordNet (Fellbaum, 1998) have been used, and modified and refined in two ways: irrelevant classes (for the corpus) have been withdrawn; for large classes, a more precise granularity has been chosen. This has led to 33 classes. Here is a part of their hierarchical organization as defined in WordNet (classes not used for tagging are in italics):

event
- **hap** (happening, natural event)
- **act** (act)
    - **acy** (human activity)
- **phm** (phenomenon)
    - **pro** (process)

sta (state)

Only 8.7% of the entries are ambiguous. Most of them correspond to complementary polysemy, in particular classical semantic alternations (for example, *enfoncement* (hollow) can both indicate a process or its result; it is therefore both classified as **pro** and **sta**) or contextual variants (for example, *bout* (end) can be temporal or locative).

Concerning verbs, WordNet classification was judged too specific and divided into too many classes. A minimal partition into 5 classes was chosen: **acc**: cognitive activity, **acp**: physical activity, **eta**: state, **mod**: modality and **tem**: temporality. Very few entries are ambiguous. Adjectives and prepositions have also been categorized.

To sum up, here is a few numerical information about the lexicon that has finally been created and used for the tagging. It contains 1489 nouns, 129 of them being ambiguous; the most frequent ambiguity (about one sixth) is between **art** (artefact) and **pro** (process). The lexicon also contains 8 acronyms, one of them being ambiguous; 567 verbs, 6 of them being ambiguous; 68 adjectives, 4 of them being ambiguous; 53 prepositions, 9 of them being ambiguous; about 15 determiners, none of them being ambiguous and around 30 pronouns and relative pronouns.

[3]Check the corrosion absence.

[4]Empty the tank bottom.

[5]We do not deal here with a Word Sense Disambiguation task consisting in the assignment of the right meanings to some word occurrences with the help of a predefined list of meanings or definitions found in a dictionary or a thesaurus (see for example (Kilgarriff and Palmer, 2000)); we rather perform the tagging of each N, V, etc. occurrence in the corpus with its right semantic tag among those associated with its lexical entry in a semantic lexicon created for this purpose.

3.2 Semantic annotation

The main hypotheses guiding the method of semantic tagging are that: (i) syntax can help to distinguish meanings of words that are polyfunctional[6] (see also (Wilks and Stevenson, 1996; Yarowsky, 1992; Ceusters et al., 1996)), (ii) syntactic analysis can be done by a probabilistic tagger (HMM, Hidden Markov Model; (Rabiner, 1989; Kupiec, 1992); etc.) and, more daringly, (iii) remaining semantic ambiguity can also be solved (mutatis mutandis) by an HMM tagger. These hypotheses are not new, but we describe here the way we have implemented them, and we evaluate our method with the MATRA-CCR corpus.

After the POS-tagging and disambiguation of the corpus previously explained, the lexicon which contains the semantic tags is used to associate one or more semantic tags with words in the corpus. The HMM tagger, applying a model trained on the ambiguous semantic tags, resolves the remaining semantic ambiguities. As we are in a restricted domain after syntactic analysis, homonyms are very rare; as we have already said, what need to be disambiguated here are mostly polysemes whose senses are related in a systematic way (Pustejovsky, 1995), as *enfoncement* that denotes both the process of hollowing something and the result of this process. These polysemes are particulary suitable for this kind of method as by definition the correct sense can be identified by the context around the word and their disambiguation does not require pragmatic disambiguation. The training is done in the same way as for POS-tagging, but with two differences: firstly, the model has to be simplified as much as possible. In this experiment for example, adverbs and attributive adjectives are not taken into consideration. These categories seem to be irrelevant for the disambiguation of nouns or verbs and for the identification of links based on qualia roles between them. Moreover, since the ambiguities are very limited, the training has been done with a set of interesting sentences. A subset of about 5850 words of the MATRA-CCR corpus has been manually tagged and compared with the output of the tagger. In this subset, 455 words were ambiguous (7.78%). The application of the semantic tagging method, with training on the ambiguous corpus and one retrain with the hand-tagged part, has led to the correct resolution of about 70% of the ambiguities. As for the POS-tagging, some biases written in bigram form can be added and allow us to improve these results; in our case, only a few simple linguistic preferences has led to a score of 1.18% of semantic tagging errors, that is (when compared with 7.78%) 85% of good disambiguation. More than one third of the remaining errors are due to prepositions. The errors concerning nouns and verbs – which are the key elements of the qualia structures we are willing to extract – are therefore relatively rare in the disambiguated corpus.

[6]i.e have different syntactic categories, as *règle* in French which can be the indicative of the verb *to regulate* and the noun *rule*.

3.3 Learning

In order to be able to compare results obtained with the POS-tagged version of the corpus and the semantically-tagged one, we have first deliberately chosen a similar general form for the E^+ and E^-: *positive(semantic-tag_before_N, semantic-tag_after_N, semantic-tag_before_V, V_POS-tag, distance).*, where *distance* corresponds to the number of verbs between the N and the V and is positive if the V precedes the N, and negative otherwise (this last argument summarizes the two previous *distance* and *position* arguments for the POS-tagged corpus examples); we do not therefore fully exploit the available semantic information, as, for example, we do not deal with the semantic tags of the two components of the pairs (this task will constitute one of our future works). The same 4000 E^+ and 7000 E^-, but with this new form, are automatically built from the corpus and are furnished to Progol in order for it to make generalizations of the E^+ and infer generalized clauses that explain what, in terms of this new kind of contextual information, distinguishes N-V pairs whose components are linked by one of the roles defined in the qualia structure from others.

The theoretical evaluation of the learning method on the semantically-tagged version of the MATRA-CCR corpus is better than on the POS-tagged corpus: the 346 generalized clauses produced by Progol cover (i.e. explain) 89.9% of the initial E^+ and only cover 0.7% of the E^-; the value of the Pearson coefficient is therefore closer to its highest possible value (1) and is now 0.91. Here is one example of an infered clause: *positive(A,B,C,D,E) :- vide(A), etat_verbe(C), modalite(B).*, which means that N-V pairs (i) in which there is nothing before the N (*vide(A)*[7]), (ii) in which the word after N is a modal verb (*modalite(B)*), and (iii) in which the word before V is a stative verb (*etat_verbe(C)*), are relevant; no constraint is set either on N/V order and distance, or on the POS-tag of V in the sentences. This rule, for example, recognizes the pairs *(platine, déposer)* and *(train, sortir)* as relevant in the two following sentences extracted from our corpus: *"Les **platines** doivent être **déposées** s'il y a échange du réservoir."* and "Le **train** peut être **sorti** à l'aide de l'électropompe secours en cas de perte soit : de pression des pompes, de l'alimentation 28 V du circuit de commande.".

[7](Empty). Or one of the three categories that we do not consider for example elaboration: determiners, adverbs and some adjectives.

The same technique as before is also used to empirically evaluate the learning method by applying the generalized clauses on the corpus and by comparing the decisions with the manual expert tagging for the same 7 significant nouns. If we consider a N-V pair tagged as *"relevant"* by the generalized clauses if at least one of them covers it once, 57 pairs are correctly found, 61 incorrectly and 13 not found (Pearson value: 0.4678); if we consider that to be said correctly detected by the generalized clauses one pair must be covered at least twice (by different clauses of G or by a same clause in different phrases), values are respectively 47, 21 and 23 (Pearson value: 0.5817). These new numbers lead to two kinds of comments: first, the highest Pearson value here is better (0.58) than for the POS-tagging (0.52); moreover, the number of "cross-validations" needed to obtain this highest value is lower (2 against 6), which means that each generalized clause is here a better characterization of the concept we want to learn: "qualia-pair" versus "non qualia-pair". We now discuss the linguistic interest of the clauses that we have infered from the two versions of the corpus, with a special focus on what semantic tagging brings out.

4 Linguistic evaluation of generalized clauses

The two previous sections show that the results of the learning method are interesting from a theoretical (and empirical) point of view, especially in the case of the semantic tagging. We conclude this paper by a short evaluation of the linguistic significance of the generalized clauses. In other words, what do we learn?

The first four points concern linguistic information grasped at the level of POS tagging, showing that superficial clues about word order, punctuations and strings of grammatical categories are useful to spot the textual areas that are likely to convey telic or agentive relations:

- As expected, the proximity is a very important criterion. 14 clauses express the fact that the related N-V must be separated by a single element at the most (preposition, proper noun, comma, etc.), for example: "*tâches* **à** *effectuer*"; "*écrous*, *serrer*"; "*voyants* **joker** *allumés*". 35 clauses contain the information that the verb and the noun cannot be separated by another verb;

- 20 clauses show the significance of punctuation marks. In our corpus, they are indeed sufficient for identifying related N-V, for example if the N and the V are separated by a colon (or a comma) and if the N and the V are directly preceded/followed by a comma: "**...,** *écrou* **:** *serrer* au couple**,**";

- the position of the N-V in the sentence is often made explicit: 16 clauses generalize the fact that a verb (infinitive) in initial position is a strong candidate. Most of them express a general and impersonal order that is a typical action that should be done with the object, for example: "*procéder* à un *essai*"; "*éliminer* toute trace de *calcaire*", etc. This is quite typical of our corpus, which is full of instructions of use;

- finally the clauses show the relevance of some syntactic constructions. Unsurprisingly, one clause characterizes the passive construction: the N and V are related if the V is preceded by the auxiliary *être* ("be") and if it is followed by the preposition *par* ("by"). More surprising at first glance, two clauses specify that the N and V that follow a subordinating conjunction are relevant. This constraint generalizes the fact that a lot of verbs in our corpus subcategorize for a complement that indicates typical action like "s'assurer que" or "vérifier" ("s'assurer **que** l'*alimentation* est *coupée*"; "vérifier **que** le *feu* anti-collision *clignote*"), etc.

The generalized clauses learnt from the semantically tagged corpus make the same kind of generalizations, but also specify some interesting semantic properties of the words that follow or precede the semantically related N-V. Two further rules are mainly using information related to verb and preposition semantic categorization:

- Modal verbs like *permettre*, *devoir* or *pouvoir* are strong indicators of relevant N-V, i.e. "le *tableau* **doit** être *éclairé*", "l'*adhésion* **peut** être *atteinte*", etc.;

- the semantic type of the preposition can help identifying relevant N-V, especially if the preposition indicates the manner or the purpose: "*fixer* **avec** leurs *vis* sans serrer" or "**pour** *emmancher* l'arbre d'entraînement **dans** la *prise* de mouvement".

These first results have been compared with explorations made manually to discover structures that convey the telic relation in the same corpus (Galy, 2000). This comparison shows that much of the structures are correctly automatically learnt by our method, but also that some reliable structures are not detected by it, in particular when the markers are polylexical expressions such as "avoir pour but de", "être utilisé pour", "avoir pour fonction", etc. However our learning method suggests the importance of positional criteria (such as N or V must occur at the beginning of a phrase) that have been left unnoticed by observations made in (Galy, 2000).

To sum up our research, we can say that we have demonstrate that symbolic machine learning methods can be used on a POS-tagged corpus in order to automatically acquire N-V pairs whose components are linked by one of the qualia structure roles. Moreover, this acquisition can be improved with the help of a semantic tagging of the corpus. Finally we have shown that a set of the rules (clauses) learnt by our ILP method have a linguistic significance. We have up to now not fully exploited all the possibilities offered by the semantic tagging of the corpus, and are therefore not fully able to solve the problems mentioned at the end of section 2; our next task will consequently consists in taking into account the semantic types of the noun and the verb that are in a semantic relation.

References

Susan Armstrong, Pierrette Bouillon, and Gilbert Robert. 1995. Tagger Overview. Technical report, ISSCO, (http://www.issco.unige.ch/staff/robert/tatoo/tagger.html).

Susan Armstrong. 1996. Multext: Multilingual Text Tools and Corpora. In H. Feldweg and W. Hinrichs, editors, *Lexikon und Text*, pages 107–119. Tübingen: Niemeyer.

Pierrette Bouillon, Sabine Lehmann, Sandra Manzi, and Dominique Petitpierre. 1998. Développement de lexiques à grande échelle. In *Colloque de Tunis 1997 "La mémoire des mots"*, Tunis, Tunisia.

Pierrette Bouillon, Cécile Fabre, Pascale Sébillot, and Laurence Jacqmin. 2000. Apprentissage de ressources lexicales pour l'extension de requêtes. *TAL (traitement automatique des langues), special volume on Natural Language Processing and Information Retrieval*, 41(2):367–393.

Werner Ceusters, Peter Spyns, Georges DeMoor, and W. Martin. 1996. *Tagging of Medical Texts: The Multi-TALE Project.* Amsterdam:IOS Press.

Cécile Fabre and Pascale Sébillot. 1999. Semantic Interpretation of Binominal Sequences and Information Retrieval. In *International ICSC Congress on Computational Intelligence: Methods and Applications, CIMA'99, Symposium on Advances in Intelligent Data Analysis AIDA'99*, Rochester, N.Y., USA.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database.* MIT Press, Cambridge, MA.

Edith Galy. 2000. Repérer en corpus les associations sémantiques privilégiées entre le nom et le verbe : le cas de la fonction dénotée par le nom. Master's thesis, Université de Toulouse - Le Mirail.

Adam Kilgarriff and Martha Palmer. 2000. Special Issue on Senseval. *Computers and the Humanities*, 34(1/2).

Julian Kupiec. 1992. Robust Part-of-Speech Tagging using a Hidden Markov Model. *Computer Speech and Language.*

Tom M. Mitchell. 1997. *Machine Learning.* McGraw-Hill.

Stephen Muggleton and Luc De-Raedt. 1994. Inductive Logic Programming: Theory and Methods. *Journal of Logic Programming*, 19-20:629–679.

Stephen Muggleton. 1995. Inverse Entailment and Progol. *New Generation Computing*, 13(3-4):245–286.

Dominique Petitpierre and Graham Russell. 1998. Mmorph - the Multext Morphology Program. Technical report, ISSCO.

James Pustejovsky, Sabine Bergler, and Peter Anick. 1993. Lexical Semantic Techniques for Corpus Analysis. *Computational Linguistics.*

James Pustejovsky, Branimir Boguraev, Marc Verhagen, Paul Buitelaar, and Michael Johnston. 1997. Semantic Indexing and Typed Hyperlinking. In *AAAI Spring 1997 Workshop on Natural Language Processing for the World Wide Web*, Stanford, CA, USA.

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

Lawrence R. Rabiner. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. In A. Waibel and K.F. Lee, editors, *Readings in Speech Recognition*. San Mateo:Morgan Kaufmann.

Sam Roberts, Wim Van-Laer, Nico Jacobs, Stephen Muggleton, and Jeremy Broughton. 1998. A Comparison of ILP and Propositional Systems on Propositional Data. In Springer-Verlag, editor, *Proceedings of 8th International Workshop on Inductive Logic Programming, ILP-98*, Berlin, Germany. LNAI 1446.

Pascale Sébillot, Pierrette Bouillon, and Cécile Fabre. 2000. Inductive logic programming for corpus-based acquisition of semantic lexicons. In *LLL-2000 (Learning Language in Logic)*, Lisbon, Portugal, september.

Stefan Wermter, Ellen Riloff, and Gabriele Scheler, editors. 1996. *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*. Lecture Notes in Computer Science, Vol. 1040, Springer Verlag.

Yorick Wilks and Mark Stevenson. 1996. The Grammar of Sense: is Word-Sense Tagging much more than Part-of-Speech Tagging? Technical report, University of Sheffield, UK.

David Yarowsky. 1992. Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora. In *COLING'92*, Nantes, France.