Machine Learning for Information Extraction

Claire Nédellec - Inference and Learning Group, LRI, Bât. 490, CNRS UMR 8623
Université de Paris-Sud, F–91405 Orsay cedex, email: cn@lri.fr

1. Introduction and discussion

As an increasing amount of information becomes available in the form of electronic documents, the need to intelligently process such texts makes shallow text understanding methods such as Information Extraction (IE) particularly useful. IE has been restrictedly defined by DARPA's MUC program [MUC Proceedings] as the task of extracting specific, well-defined types of information from text in restricted domains and filling pre-defined template slots. MUC has inspired a huge amount of work in IE and has become the major reference in the field. A typical IE tasks is illustrated by the example in Figure 1 from the MUC-4 corpus that describes terrorist incidents. Even as such it is still a challenging task to build an efficient IE system with good recall (coverage) and precision (correctness) rates.

DOCUMENT: Lima, 16 jan 90 (Television Peruana) - Ten terrorists hurled dynamite stick at U.S. embassy facilities in the Miraflores district, causing serious damage but fortunately no casualties. The attack took place at 2100 on 15 January [0100 GMT on 16 January]. Inside the faculty, which was guarded by 3 security officers, a group of embassy officials were holding a work meeting. According to the first police reports, the attack was staged by 10 terrorists who used 2 Toyota cars which were later abandoned.
FILLED PATTERN:
Weapon: 2 cars
Physical-target: U.S. embassy facilities
Date: 15 January
Perpetrator: 10 terrorists
Victims: /

Fig. 1. A MUC-4 example.

Building IE systems is time-consuming because they rely on manually encoded dictionaries of vocabulary and on extraction rules or patterns which are specific to the domains and the tasks at hand and not easily portable. Therefore, automatically learning extraction rules from examples of pairs of filled patterns and annotated documents has appeared as very attractive since the early nineties [Riloff, 93]. At the end of the decade, the opinion about the relative merits of the trainable approach and the knowledge engineering approach is more contrasted as discussed by Appelt and Israel at IJCAI-99 tutorial on IE. According to them, trainable approaches (statistics and ML-based) should be preferably applied when the training data is cheap and plentiful, the extraction specifications stable and the highest possible performance is not critical (the best recall obtained by the ML-based systems is quite low compared to hand-coded IE systems). Israel's and Appelt's analysis is based on the current state of the art where existing ML-based systems are exploiting few background knowledge for guiding learning, if they do, they usually use quite shallow representation of the training texts and most of them are based on general purpose ML algorithms. They are mainly KNN, grammatical inference, naïve Bayes methods and top-down or bottom-up relational learning based on exhaustive search or information gain measure. The lack of variety of the approaches with respect to the richness of the state-of-the art in ML can be explained by two related facts.

First, on the usual and quite simple IE tasks (MUC tasks, IE on job and seminar announcements), approaches based on linguistic analysis, lexical semantics, and informative representation of the training data do not perform so much better, when they do, than more shallow approaches (see for instance the experimental results in [Freitag, 98] and [Ciravegna, 2000]). This does not encourage the design and the application of novel symbolic and relational ML methods which would be suitable for richer text analysis although no systematic comparison but just limited experiments have been performed.

Second, the main stream in text processing until recently was mostly linguistic and statistic but not ML-based apart some noticeable exceptions of for instance, Soderland's work, Mitchell's group

([Freitag, 97, 98], [Craven & Kumlien, 99]), and Mooney's group research ([Califf & Mooney, 98], [Nahm & Mooney, 2000]). A large part of the effort in learning for IE has also been devoted to lower level tasks such as entity named recognition [Bikel et al., 97].

Things are evolving with the growing interest of ML to text processing and IE in particular. Thus, ML-based pioneer systems such as Crystal [Soderland et al., 95], Liep [Huffman, 96], AutoSlog [Riloff, 93, 96, 98, 99], Alergia [Freitag, 97], have been followed by Rapier [Califf & Mooney, 98], [Thomson et al., 99], SRV [Freitag, 98], FOIL [Craven & Kumlien, 99], Whisk [Soderland, 99], Wawe [Aseltine, 99], RHB+ [Sasaki & Matsuo, 2000], DiscoTEX [Nahm & Mooney, 2000], Inthelex [Esposito et al., 2000], Pinocchio [Ciravegna, 2000] among others. Moreover, the growing application pressure provides many new IE tasks which require deeper understanding and then push towards more sophisticated linguistic and ML approaches. Additionally in real-world applications, training is more viewed as closely complementary to knowledge engineering rather than opposed to it, as illustrated by the interactive approaches [Soderland, 99], [Thomson et al., 99]. At the same time there are tentatives to reduce the tedious annotation tasks but using more training data [Yangarber et al., 2000], multi-strategy learning [Freitag, 98] or existing background knowledge [Craven & Kumlien, 1999].

Additionally, intermediate learning steps of knowledge acquisition from texts towards IE are required and will receive more attention in the future, including for instance learning semantic classes, predicate-argument structures, learning for co-reference resolution, (see for instance [Faure & Poibeau, 2000] and [Maedche & Stabb, 2000]).

The rest of the chapter will be organized as follows, Section 2 will be devoted to classical IE as defined above while Section 3 will present some of the methods for learning knowledge from text which appear as promising for IE.

2. Classical IE

In the classical framework, the ML system is fed with pairs of filled templates and annotated texts where the substrings in the text are associated to the filled slots in the template. Learning can be then viewed as a classification task [Freitag, 97] where the extraction rules to be learned represent the conditions for filling a given slot or as pattern learning where the patterns are regular expression to be matched to text substrings. The methods then differ in

- The type of text: free, semi-structured, structured text, more or less domain restricted, (physician discharges, gene interactions, newswires about company joint ventures and terrorist attacks, job or seminar announcements).
- The type of the slots to fill, (symbolic / numeric, text substring or more abstract).
- The type of the features for describing the documents, which are relational (relative position of two words, word neighborhood, syntactic relation, thematic role) or not (exact word, lemma, word position, part-of-speech tag, semantic category, case information).
- The role of the context of the relevant fragment in the text (taken into account or not, size of the context)
- The use of additional lexicon (semantic categories, hyperonym links, thematic roles, case frames)
- The role of the user for annotating the examples and validating the result, (the whole document is classified as relevant or not, the text fragment is labeled with the slot, the sentence is labeled with a central concept, tags are inserted, seed semantic categories or seed patterns are provided, intermediate learned patterns are validated).
- The type of learning algorithm (case-based, naïve Bayes, grammatical inference, relational learning, ILP) and the learning steps (building a pool of good rules and then specializing them, refining the boundaries).

Let us take some short examples to illustrate this typology. E. Riloff was the pioneer in the domain with the system **AutoSlog** [Riloff, 93]. Auto-Slog-TS as described in [Riloff, 96] differs from AutoSlog in that it does not require annotated documents but takes a set of relevant and irrelevant documents as input. It extracts as potential extraction patterns syntactic dependencies in the corpus from a list of pre-defined dependencies. For example, `subject: ten terrorists active verb: hurled direct object: dynamite stick` gives the pattern `Subject: < > verb: hurled Dobj: < >` where the words in the noun phrases are generalized into wild cards. The relevance rate of the candidate patterns is then computed according to the number of relevant and irrelevant texts where they can be activated. The ranked patterns are then validated and labeled by a

user, for instance `Subject:<Perpretator> hurled Dobj: <explosive>`. Later versions of Auto-Slog-TS include case frames learning (semantic representation of the patterns) [Riloff & Schmelzenbach, 98] and semantic categories learning [Riloff & Jones, 99].
Dayne Freitag in 1998 proposed with the **SRV** system more sophisticated generalization steps viewing the problem of pattern learning as a relational classification problem like [Califf & Mooney, 97] with the Rapier system. On semi-structured texts, SRV performs better than Rapier, Whisk and similarly to Pinocchio [Ciravegna, 2000]. SRV takes as input a set of tagged documents and a set of features richer than Rapier's ones for describing the tokens drawn from the documents such as length, type, part-of-speech, semantic category and synsets (from WordNet [Miller, 90]), adjacency of tokens and syntactic dependencies. SRV combines a naïve Bayes classifier and a relational rule learner which proceeds top-down like FOIL inducing sets of constraints [Freitag, 98a]. The role of the naïve Bayes classifier is to compute an estimated probability that a token is found in correct slot filler. Tokens with the best probabilities are added as constraints by the top-down algorithm. Experiments show that linguistic information (parsing and semantics) on the data yields better precision but lower recall pointing out that the choice of the suitable features for describing the data is a crucial part of the IE problem.
Whisk is a general rule extraction system which learns regular expressions as extraction patterns [Soderland, 99]. It is able to learn sentence-based multi-slot rules. Whisk algorithm induces rules in a top-down and covering manner as opposed to Soderland's previous system Crystal [Soderland, 95]. As Progol, it uses a positive seed as lower bound to constrain the search. Active learning is used to reduce the size of the training annotated corpus. The multi-slot approach seems to augment the precision but to yield a lack of generality, thus badly affecting the recall.
Among the most recent systems, Pinocchio outperforms previous systems on a semi-structured corpus [Ciravegna, 2000]. Pinocchio learns separate extraction rules for the left and right boundaries of the slot fillers in the texts. The tokens in the training data are labeled with their POS tag, case type, and user pre-defined categories (such as *Company*). Learning applies in three steps, (1) bottom-up learning of a Best Rule Pool, (2) completing the best pool by learning additional rules including conditions on a previous or a next tag, (3) adjustment of the boundaries.

2. Knowledge extraction from text

As illustrated by most of the systems described above, learning for IE requires external resources for building more abstract and richer representations of the training text, such as subcategorization frames, restrictions of selection, semantic lexicon, case frames or predicate-argument structures. Automatically learning such resources from training corpora has received much attention in the past ten years. The reasons for building such resources concern many other tasks than IE, namely Information Retrieval, Question Answering (QA), translation, and enriching existing lexicon.
Many different tools have been developed for the unsupervised automatic or semi-automatic acquisition of semantic classes from “near” terms or verbs. The notion of semantic proximity is based upon the distance among terms, defined as a function of the degree of similarity of their contexts following Harris' assumption [Harris *et al.* 89]. The descriptions of the term contexts (the learning examples) and of the regularities to be sought vary in different approaches. Contexts can be purely graphic—words co-occurring within a window—as in the case of the work described in [Sparck Jones & Barber, 71], [Church & Hanks, 89], [Brown et al., 92]; in some cases, the window can cover the whole document (see e.g. [Quiu & Frei, 93]). Contexts can also be syntactic. The learning results can be of different types, depending on the method employed. They can be distances that reflect the degree of similarity among terms [Hirshman et al., 75], [Grishman et al., 86], [Grishman & Sterling, 94], [Sekine et al., 92], [Dagan et al., 94], [Dagan et al., 96], [Resnik, 95], distance-based term classes elaborated with the help of nearest-neighbor methods [Grefenstette, 92], [Hindle, 90], [Bisson et al., 2000], degrees of membership in term classes [Ribas, 94], class hierarchies built by hierarchical conceptual clustering [Pereira et al., 93], [Hogenhout & Matsumoto, 97], [Bouaud et al., 1997], subcategorization frames [Briscoe & Caroll, 97], [Faure & Nédellec, 98], predicative schemata that use concepts to constrain selection [Basili et al., 96], [Thompson, 95], [Gomez, 97], and semantic roles [Sébillot et al., 2000]. Some of these works exploit additional resources for enriching the data, guiding learning or validating the learning results such as terminology, [Grefenstette, 94], dictionaries [Krovetz & Croft, 91], nomenclature such as SNOMED international [Bouaud et al., 1997] specific

ontologies [Soderland, 95] or general ontologies such as WordNet, [Yarowsky, 92], [Resnik & Hearst, 93], [Resnik, 95], [Ribas, 94], [Ribas, 95], [Li & Abe, 96].

Other tools learn semantic relations for enriching thesauri or ontologies which are useful for IE, by learning general extraction patterns from corpora (e.g hyperonymy [Morin & Jacquemin, 99]) or from multiple observations at the syntactic level [Hahn & Schnattinger].

On the one hand more and more complex and abstract semantic knowledge such as semantic classes, thematic roles and case frames [Gomez, 98], [Sasaki & Matsuo, 2000] are used in the extraction patterns of IE systems applied to understanding tasks in free texts.

On the other hand, different kinds of textual sources including highly structured text are explored which do not require external knowledge except pre-determined patterns. The web pages including hyperlinks and neighbor pages receives more and more attention [Craven et al., 99], for example, wrappers which identify regular expression in structured texts such as tables in html pages give rise a growing interest from Machine Learning researchers [Goan et al., 97], [Kushmerick et al. 97], [Kushmerick, 99], [Knoblock et al., 98], [Cohen, 99].

Moreover as the research topics in other neighbor fields, i.e. IR and QA become closer and closer to IE, one may expect that IE will also benefit from the advances of the application of Machine Learning in these fields (e.g. [Harabagiu et al. 2000]).

References

Appelt D. E. and Israel D., "A tutorial Prepared for IJCAI-99", http//: www.ai.sri.com/~appelt/ie-tutorial/

Aseltine J.H., "An Incremental Algorithm for information Extraction", In Proceedings of the *AAAI-99 Workshop on Machine Learning for Information Extraction*, 1999.

Basili R., Pazienza M. T. and Velardi P., "An empirical symbolic approach to natural language processing", in Artificial Intelligence Journal 85, pp. 59-99, 1996.

Bikel D. M., Miller S., Schwartz R. and Weischedel R., "Nymble: a High-Performance Learning Name-finder", *Conference on Applied Natural Language Processing*, 1997.

Bisson G., Nédellec C. and Canamero D. "Designing clustering methods for ontology building - The Mo'K workbench" in Proceedings of the workshop on *Ontology Learning*, workshop of the European Conference on Artificial Intelligence (ECAI-2000), S. Stabb, et al (Eds.). pp. 13-19, Berlin, August 2000.

Briscoe E. and Carroll J., "Automatic extraction of subcategorization from corpora". In Proceedings of the *5th ACL Conference on Applied Natural Language Processing*, pp. 356-363, Washington, DC, 1997.

Brown P. F., Della Pietra V. J., deSouza P. V., Lai J. C. and Mercer R. L. "Class-based n-gram models of natural language." In *Computational Linguistic* 18(4), pp.283-298.

Califf M. E. and Mooney R. J., "Relational Learning of Pattern-Match Rules for Information Extraction." In Proceedings of *AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, Stanford, CA, March, 1998.

Church K. W. and Hanks P., "Word Association Norms, Mutual Information, and Lexicography", in Proceedings of the *27th Annual Meeting of the Association for Computational Linguistics*, pp. 76-83, 1989.

Cohen W., "What Can We Learn From the Web?", in Proceedings of the *Sixteenth International Machine Learning Conference (ICML'99)*, Bratko I. and Dzeroski S. (Eds.), pp. 515-521, Bled, Slovenia, June 1999.

Craven M., DiPasquo D., Freitag D., McCallum A., Mitchell T., Nigam K. and Slattery S., "Learning to Construct Knowledge Bases from the World Wide Web." In *Artificial Intelligence*, 1999.

Ciravegna F., "Learning to Tag for Information Extraction from Text". In Proceedings of the *ECAI-2000 Workshop on Machine Learning for Information Extraction*, F. Ciravegna et al. (Eds.), Berlin, August 2000.

Craven M. and Kumlien J., "Constructing Biological Knowledge Bases by Extracting Information from Text Sources." In Proceedings of the *7th International Conference on Intelligent Systems for Molecular Biology (ISMB-99)*, 1999.

Faure D. and Nédellec C., "A Corpus-based Conceptual Clustering Method for Verb Frames and Ontology Acquisition ", in proceedings of the workshop *Adapting lexical and corpus resources to sublanguages and applications* of the 1st *International Conference on Language Resources and Evaluation (LREC)*, pp. 1-8, P. Velardi (Ed.), Granada, Spain, May 1998.

Faure D. and Nédellec C., "Knowledge Acquisition of Predicate-Argument Structures from technical Texts using Machine Learning" in Proceedings of *Current Developments in Knowledge Acquisition: EKAW-99*, pp. 329-334, Fensel D. and Studer R. (Eds.), Karlsruhe, Germany, April 1999.

Esposito F., Ferilli S., Fanizzi N. and Semeraro G., "Learning from Parsed Sentences with INTHELEX", in Proceedings of the *4th Conference on Computational Natural Language Learning and of the Second Learning Language in Logic Workshop*, Cardie C. et al (Eds), pp. 194-198, Lisbon, Portugal, September 2000.

Dagan I., Pereira F., and Lee L., "Similarity-Based Estimation of Word Cooccurrence Probabilities", in Proceedings of the *32nd Annual Meeting of the Association for Computational Linguistics, ACL'94,* New Mexico State University, June 1994.

Dagan I., Lee L., and Pereira F., "Similarity-Based Methods For Word-Sense Disambiguation", in Proceedings of the *Annual Meeting of the Association for Computational Linguistics, ACL'96*, 1996.

Faure D. and Poibeau T., "First experiments of using semantic knowledge learned by Asium for Information Extraction task using Intex.3, in the Proceedings of *the ECAI-2000 Workshop on Ontology Learning*, Staab et al. (Eds), Berlin, Germany, August 2000.

Freitag D., "Using Grammatical Inference to Improve Precision in Information Extraction", in Working Papers of the *ICML-97 Workshop on Automata Induction, Grammatical Inference and Language Acquisition*, P. Dupont (Ed.), 1997.

Freitag D., "Information Extraction From HTML: Application of a General Learning Approach.", In the Proceedings of the *15th National Conference on Artificial Intelligence (AAAI-98)*, 1998.

Freitag D., "Toward General-Purpose Learning for Information Extraction.", in *Proceedings of the Seventeenth International Conference on Computational Linguistics (COLING-ACL-98)*, 1998.

Freitag D., "Multistrategy learning for information Extraction." In the Proceedings of the Fifteenth *Machine Learning Conference (ML-98)*, J. Shavlik (Ed.), Madison, USA, 1998.

Goan T., Belson N., and Etzioni O., "A Grammar Inference Algorithm for the World Wide Web", In the Proceedings of the *AAAI Spring Symposium on Machine Learning in Information Access*, M. Hearst and H. Hirsh, (Eds.), Stanford, March 25-27, 1996.

Gomez F., Segami C. and Hull R., "Determining Prepositional Attachment, Prepositional Meaning, Verb Meaning and Thematic Roles". In *Computational Intelligence*, Vol 13, Number1, 1997.

Gomez F., "Linking WordNet Classes to Semantic Interpretation", in Proceedings of the *COLING-ACL Workshop on the usage of WordNet in NLP Systems*, 1998.

Grefenstette G., "Use of syntactic Context to Produce Term Association Lists for Text Retrieval", in Proceedings of the *15th International SIGIR'92*, Denmark, 1992.

Grefenstette G., *Explorations in Automatic Thesaurus Discovery*, Klüwer Academic (Pub.), 1994.

Grishman R., Hirshman L., Nhan N. T., "Discovery Procedures for Sublanguages Selectional Patterns: Initial Experiments", in *Computational Linguistics*, Vol 12, (3) pp. 205-215, 1986.

Grishman R., Sterling J., "Generalizing Automaticaly Generated Selectional Patterns", in Proceedings of the *16th Int'l Conf. Computational Linguistics (COLING'94)*, 1994.

Hahn U. and Schnattinger K., "Towards Knowledge Engineering" in Proceedings of the *National Conference on Artificial Intelligence (AAAI-98)*, Madison, USA, 1998.

Harabagiu S., Pasca M., Maiorano, S., “Experiments with open-domain textual questions”, in Proceedings of the *18th International Conference on Computaional Linguistics, COLING-2000*, Saarbrücken, Germany, 2000.

Harris Z., Gottfried M., Ryckman T., Mattick Jr P., Daladier A., Harris T. and Harris S., "The form of Information in Science, Analysis of Immunology Sublanguages", vol. 104 of *Boston Studies in the Philosophy of Science*. Dordrecht, the Netherlands, Kluwer Academic (Pub.), 1989.

Hirshman L., Grishman R. and Sager N., "Grammatically-based automatic word class formation", in *Information Processing and Management*, Vol 11, pp. 39-47, Pergamon Press, 1975.

Hogenhout W. R. and Matsumoto Y., "A preliminary Study of Word Clustering Based on Syntactic Behavior", In Proceedings of *CNLP*, *35th annual meeting of the ACL* and *EACL'97*, 1997.

Huffman S., "Learning Information Extraction Patterns from Examples", in *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*, Springer-Verlag, 1996.

Knoblock C.A., Minton S., Ambite J.L., Ashish N., Modi P.J., Muslea I., Philpot A.G., and Tejada S., "Modeling Web Sources for Information Integration", In Proceedings of the *National Conference on Artificial Intelligence*, 1998.

Krovetz R., "Homonymy and Polysemy in Information Retrieval", In Proceedings of the *35th annual meeting of the ACL (ACL'97)*, 1997.

Kushmerick N., "Wrapper Induction: Efficiency and Expressiveness", in Artificial Intelligence Journal, 1999.

Kushmerick N., Weld D., and Doorenbos B., "Wrapper Induction for Information Extraction.", In Proceedings of the *International Joint Conference on Artificial intelligence (IJCAI-97)*, Nagoya, 1997.

Li H. and Abe N., "Word clustering and desambiguation Based on Co-occurrence Data", in Proceedings of *COLING - ACL'98*, 1998.

Maedche and Stabb, "Discovering Conceptual Relations from Text" in Proceedings of *ECAI-2000*, pp. 321-325, Horn W. (Ed.), Berlin, Germany, August 2000.

Proceedings of the *Message Understanding Conference* (MUC-4-7), Morgan Kaufman, San Mateo, USA, 1992-98.

Miller G., "WordNet: an on-line lexical database", in *International Journal of Lexicography*, 3(4), Special Issue, 1990.

Morin E. and Jacquemin C., "Projecting Corpus-Based Semantic Links on a Thesaurus", in Proccedings of the *37th Annual Meeting of the Association for Computational Linguistics (ACL'99*), pp. 389-396, Maryland, USA, 1999.

Nahm U. Y. and Mooney R. J., "A Mutually Beneficial Integration of Data Mining and Information Extraction". In Proceedings of the *Seventeenth National Conference on Artificial Intelligence (AAAI-2000)*, Austin, TX, pp. 627-632, July 2000.

Pereira F., Tishby N. and Lee L., "Distributional clustering of English words" in *Proceedings of ACL'93*, p. 183-190, 1993.

Resnik P., "Using Information Content to evaluate Semantic Similarity in a Taxonomy." In *Cognitive Modeling*, 1995.

Resnik P. and Hearst M. A. "Structural Ambiguity and Conceptual Relations", in Proceedings of the *Workshop on Very Large Corpora: Academic and Industrial Perspectives*, pp. 58-64, Ohio State University, 1993.

Ribas F., "An experiment on Learning Appropriate Selectional Restrictions from a Parsed Corpus.", in Proceedings of the *16th Int'l Conf. Computational Linguistics (COLING'94)*, 1994.

Ribas F., "On Learning More Appropriate Selectional Restrictions.", in Proceedings of *EACL'95*, 1995.

Riloff E., "Automatically constructing a Dictionary for Information Extraction Tasks". In Proceedings of the *Eleventh National Conference on Artificial Intelligence (AAAI-93)*, pp. 811-816, AAAI Press/The MIT Press, 1993.

Riloff E., "Automatically Generating Extraction Patterns form Untagged Text." in Proceedings of the *Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, pp. 1044-1049, 1996.

Riloff E. and Schmelzenbach M., "An Empirical Approach to Conceptual Case Frame Acquisition", in Proceedings of the *Sixth Workshop on Very Large Corpora*, 1998.

Riloff E. and Jones R., "Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping", in Proceedings of the *Sixth National Conference on Artificial Intelligence (AAAI-99)*, 1999.

Roots M., Riezler S., Prescher D., Carroll G. anf Beil F., "Inducing a semantically Annotated Lexicon with EM-Based Clustering", in Proceedings of the *18th International Conference on Computational Linguistics, COLING-2000*, Saarbrücken, Germany, 2000.

Sasaki Y. and Matsuo Y., "Learning Semantic-Level Information Extraction Rules by Type-Oriented ILP", in Proceedings of the *18th International Conference on Computational Linguistics, COLING-2000*, Kay M. (Ed.), Saarbrücken, 2000.

Sébillot P., Bouillon P. and Fabre C., "Inductive Logic Programming for Corpus-Based Acquisition of Semantic Lexicon" in Proceedings *of the Fourth Conference on Computational Natural Language Learning and of the Second Learning Language in Logic Workshop.* , Cardie C., Daelemans W., Nédellec C. and Tjong Kim Sang E. (Eds.), pp. 199-208, Omni Press (Pub.), Lisbon, September 2000.

Sekine S., Caroll J. J., Ananiadou S. and Tsujii J., "Automatic Learning for Semantic Collocation" in Proceedings of the *third Conference on Applied Natural Language Processing*, pp. 104-109, 1992.

Soderland S., Fisher D., Aseltine J. and Lehnert W., "CRYSTAL: Inducing a Conceptual Dictionary.", Proceedings of the *Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, 1995.

Soderland S., "Learning Information Extraction Rules for Semi-Structured and Free Text" in *Machine Learning Journal*, vol 34, 1999.

Sparck Jones K. and Barber E. B., "What makes an automatic keywords classification effective?" in *Journal of the ASIS*, 18, pp. 166-175, 1971.

Thompson C. A., "Acquisition of a Lexicon from Semantic Representations of Sentences", in the Proceedings of the *33rd Annual Meeting of the Association of Computational Linguistics, (ACL'95)*, pp. 335-337, Boston, M A, July, 1995.

Yangarber R., Grishman R., Tapanainen P. and Huttunen S., "Unsupervised Discovery of Scenario-Level Patterns for Information Extraction.", In Proceedings of *18th International Conference on Computational Linguistics, COLING-2000*, Saarbrücken, Germany, 2000.

Yarowsky D., "Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora", in Proceedings of *COLING'92*, p. 454-460, Nantes, 1992.

Yarowsky D., "Unsupervised Word Sense Disambiguation Rivaling Supervised Methods", In Proceedings of the *33rd Annual Meeting of the Association for Computational Linguistics (ACL'95).* Cambridge, MA, pp. 189-196, 1995.