

Named Entity Recognition for Catalan Using Spanish Resources

Xavier Carreras, Lluís Màrquez, and Lluís Padró

TALP Research Center, LSI Department
Universitat Politècnica de Catalunya
Jordi Girona, 1–3, E-08034, Barcelona
{carreras, lluis, padro}@lsi.upc.es

Abstract

This work studies Named Entity Recognition (NER) for Catalan without making use of annotated resources of this language. The approach presented is based on machine learning techniques and exploits Spanish resources, either by first training models for Spanish and then translating them into Catalan, or by directly training bilingual models. The resulting models are retrained on unlabelled Catalan data using bootstrapping techniques. Exhaustive experimentation has been conducted on real data, showing competitive results for the obtained NER systems.

1 Introduction

A *Named Entity* (NE) is a lexical unit consisting of a sequence of contiguous words which refers to a concrete entity —such as a person, a location, an organization or an artifact. Figure 1 contains an example sentence, extracted from the Spanish corpus referred in section 2 and translated into Catalan, including several entities.

There is a wide consensus about that Named Entity Recognition and Classification (NERC) are Natural Language Processing tasks which may improve the performance of many applications, such as Information Extraction, Machine Translation, Question Answering, Topic Detection and Tracking, etc. Thus, interest on detecting and classify-

ing those units in a text has kept on growing during the last years.

Named Entity processing consists of two steps, which are usually approached sequentially. First, NEs are *detected* in the text, and their boundaries delimited (Named Entity Recognition, NER). Second, entities are classified in a predefined set of classes, which usually contain labels such as *person*, *organization*, *location*, etc. (Named Entity Classification, NEC). In this paper we will focus on the first of these stages, that is, Named Entity boundary detection.

Previous work in this topic is mainly framed in the *Message Understanding Conferences* (MUC), devoted to Information Extraction, which included a NERC task. Some MUC systems rely on data-driven approaches, such as Nymble (Bikel et al., 1997) which uses Hidden Markov Models, or ALEMBIC (Aberdeen et al., 1995), based on Error Driven Transformation Based Learning. Others use only hand-coded knowledge, such as FACILE (Black et al., 1998) which relies on hand written unification context rules with certainty factors, or FASTUS (Appelt et al., 1995), PLUM (Weischedel, 1995) and NetOwl Extractor (Krupka and Hausman, 1998) which are based on cascaded finite state transducers or pattern matching. There are also hybrid systems combining corpus evidence and gazetteer information (Yu et al., 1998; Borthwick et al., 1998), or combining hand-written rules with Maximum Entropy models to solve coreference (Mikheev et al., 1998).

More recent approaches can be found in the proceedings of the shared task at the 2002 edition

“El presidente del [Comité Olímpico Internacional]_{ORG}, [José Antonio Samaranch]_{PER}, se reunió el lunes en [Nueva York]_{LOC} con investigadores del [FBI]_{ORG} y del [Departamento de Justicia]_{ORG}.”

“El president del [Comitè Olímpic Internacional]_{ORG}, [Josep Antoni Samaranch]_{PER}, es va reunir dilluns a [Nova York]_{LOC} amb investigadors del [FBI]_{ORG} i del [Departament de Justícia]_{ORG}.”

Figure 1: Example of a Spanish (top) and Catalan (bottom) sentence including several Named Entities between brackets (PER=person, LOC=location, ORG=organization).

of the *Conference on Natural Language Learning*, CoNLL’02 (Tjong Kim Sang, 2002a), where several machine-learning systems were compared at the NERC task. Usually, machine learning (ML) systems rely on algorithms that take as input a set of labelled examples for the target task and produce as output a model (which may take different forms, depending on the used algorithm) that can be applied to new examples to obtain a prediction. CoNLL’02 participants used different state-of-the-art ML algorithms, such as Support Vector Machines (McNamee and Mayfield, 2002), AdaBoost (Carreras et al., 2002; Tsukamoto et al., 2002), Transformation-Based methods (Black and Vasilakopoulos, 2002), Memory-based techniques (Tjong Kim Sang, 2002b) or Hidden Markov Models (Malouf, 2002), among others.

One remarkable aspect of most widely used ML algorithms is that they are *supervised*, that is, they require a set of labelled data to be trained on. This may cause a severe bottleneck when such data is not available or is expensive to obtain, which is usually the case for minority languages with few pre-existing linguistic resources and/or limited funding possibilities.

Our goal in this paper is to develop a low-cost Named Entity recognition system for Catalan. To achieve this, we take advantage of the facts that Spanish and Catalan are two Romance languages with similar syntactic structure, and that —since Spanish and Catalan social and cultural environments greatly overlap— many Named Entities appear in both languages corpora. Relying on this structural and content similarity, we will build our Catalan NE recognizer on the following assumptions: (a) Named Entities appear in the same contexts in both languages, and (b) Named Entities are composed by similar patterns in both languages.

The work departs from the use of existing annotated Spanish corpora and machine learning tech-

niques to obtain Spanish NER models. We first build low-cost resources (about 10 person-hours each), namely a small Catalan training corpus and translation dictionaries from Spanish to Catalan. We then present and evaluate several strategies to obtain a low-cost Catalan system. Simple naive strategies consist of learning from the large Spanish corpus a model which makes no use of lexical information, or learning a model for Catalan using the small Catalan corpus. More sophisticated strategies are translating a Spanish model into Catalan, or directly learning a bilingual model applicable to both languages. Experimentation shows that the latter strategies, specially the bilingual models, provide very good performance, somewhat better than the former techniques. We also study the evolution of these models within a bootstrapping process, observing no significant improvement.

Next section of the paper describes the used corpora and evaluation measures. Section 3 describes the NER learning system. Section 4 presents the strategies to obtain a low-cost Catalan NER system and provides results. Bootstrapping is studied in section 5, and, finally, section 6 concludes.

2 Data and Evaluation

The experimentation of this work has been carried on two corpora, one for each language. In both cases, the corpora consist of sentences extracted from news articles of same year, namely year 2,000. The Spanish data corresponds to the CoNLL 2002 Shared Task Spanish data, the original source being the EFE Spanish Newswire Agency. It consists of three files: a training set, a development set and a test set. The first two are used respectively to train and tune a system, and the latter is used to evaluate and compare systems. Table 1 shows the number of sentences, words and Named Entities in each set. For Catalan, we had

lang.	set	#sent.	#words	#NEs
es	train.	8,322	264,715	18,797
es	dev.	1,914	52,923	4,351
es	test	1,516	51,533	3,558
ca	train.	817	23,177	1,232
ca	test	844	23,595	1,338
ca	unlab.	83,725	2,201,712	—

Table 1: Sizes of Spanish and Catalan data sets

available a large amount of news articles extracted from the Catalan edition of the daily newspaper *El Periódico de Catalunya* (also from year 2,000). From this corpus, we selected two sets for manual annotation: a training set, to train a system, and a test set, to perform the evaluation. The remaining data was left as unlabelled data.

As evaluation method we use the common measures for recognition tasks: precision, recall and F_1 . Precision is the percentage of NEs predicted by a system which are correct. Recall is the percentage of NEs in the data that a system correctly recognizes. Finally, the F_1 measure computes the harmonic mean of precision (p) and recall (r) as $2 \cdot p \cdot r / (p + r)$.

3 The Spanish NER System

The Spanish NER system is based on the best system at CoNLL’02, which makes use of a set of AdaBoost-based binary classifiers for recognizing the Named Entities in running text. See (Carreras et al., 2002) for details.

The NE recognition task is performed as a sequence tagging problem through the well-known BIO labelling scheme. Here, the input sentence is treated as a word sequence and the output tagging codifies the NEs in the sentence. In particular, each word is tagged as either the beginning of a NE (B tag), a word inside a NE (I tag), or a word outside a NE (O tag). In our case, a NER model is composed by: (a) a representation function, which maps a word and its context into a set of features, and (b) three binary classifiers (one corresponding to each tag) which, operating on the features, are used for tagging each word. When tagging, a sentence is processed from left to right, selecting for each word the tag with maximum confidence that is coherent with the current solution (I-tag

sequences must be preceded by a B-tag). When learning a model, all the words in the training set are used as training examples, applying a *one-vs-all* binarization of the 3-class classification problem.

The representation consists in a shifting window anchored in a word w , which encodes the local context of w with which a classifier will operate. In the window, each word around w is codified with a set of primitive features, together with its relative position to w . Each primitive feature with each relative position and each possible value forms a final binary feature for the classifier (e.g., “the **word form** at **position -2** is **calle**”). Particularly, the set of primitive features applied to each word in the window is the following:

- **Lexical Features** The word forms.
- **Orthographic Features** These are binary and not mutually exclusive features that test whether the following predicates hold in the word: *initial-caps*, *all-caps*, *contains-digits*, *all-digits*, *alphanumeric*, *roman-number*, *contains-dots*, *contains-hyphen*, *acronym*, *lonely-initial*, *punctuation-mark*, *single-char*, *functional-word*, and *URL*. Functional words are determiners and prepositions which typically appear inside NEs.
- **Affixes** Test whether a word beginning (or ending) matches with a common NE prefix (or suffix). The list of affixes has been automatically extracted from the Spanish training set, by taking those NE affixes of up to 4 symbols which occur more than 100 times.
- **Word Type Patterns** The type of a word is either *functional*, *capitalized*, *lowercased*, *punctuation mark*, *quote* or *other*. Each conjunction of types of contiguous words is a word type pattern, but only patterns in the window which include the anchoring word are considered.
- **Left Predictions** The tags being predicted in the current classification. These features only apply to the words in the window to the left of the anchoring word w .

As learning algorithm we use the binary AdaBoost with confidence rated predictions. The

idea of this algorithm is to learn an accurate strong classifier by linearly combining, in a weighted voting scheme, many simple and moderately-accurate base classifiers or rules. Each base rule is learned sequentially by presenting the base learning algorithm a weighting over the examples, which is dynamically adjusted depending on the behavior of the previously learned rules. We refer the reader to (Schapire and Singer, 1999) for details about the general algorithm, and to (Schapire, 2002) for successful applications to many areas, including several NLP tasks.

In our setting, the boosting algorithm combines several small fixed-depth decision trees. Each branch of a tree is, in fact, a conjunction of binary features, allowing the strong boosting classifier to work with complex and expressive rules.

4 Porting to Catalan

In this section we study the portability of a NER system from Spanish to Catalan. Our approach is to port a NER system by porting the model features from Spanish to Catalan. In particular, we concentrate on the features which are language dependent, namely, the lexical features (or word forms) and the functional words. All other features are left unchanged.

Two alternative translation dictionaries from Spanish to Catalan and vice-versa have been built for the task. They contain a one to one correspondence between Spanish and Catalan words. For instance, an entry in a dictionary is “calle ~ carrer”, meaning that the Spanish word “calle” (“street” in English) corresponds to the Catalan word “carrer”.

In order to obtain the relevant vocabulary for NER, we have run several trainings of the Spanish NER system by varying the system parameters, and we have extracted from the learned models all the involved Spanish lexical features. These Spanish words form a set of 5,024 entries.

The first dictionary has been manually completed, with an estimated cost of about 10 person hours of a bilingual speaker (7.2 sec/word). Note that translations are made with no context information, and with no linguistic criteria. The translator’s common sense is blindly assumed to select the best choice among all possible translations.

The second dictionary has been automatically

completed using the InterNOSTRUM Spanish–Catalan machine translation system developed by the Software Department of the University of Alacant¹. In this case, the translations have also been resolved without any context information, and the entries not recognized by InterNOSTRUM (about 17%) have been left unchanged.

4.1 Model Translation

Our first approach to obtain a NER model for Catalan consists in first learning a NER model for Spanish using Spanish annotated data, and then translating its lexical features from Spanish into Catalan using the translation dictionary.

In our particular case, a NER model is composed by the B, I and O classifiers, each of which is a combination of a number of base decision trees. The model translation, therefore, consists in translating every decision tree by translating those nodes in the tree which evaluate lexical features. For instance, considering the translation “calle ~ carrer”, a node for Spanish with feature “word:-2:calle”, testing whether the word form at relative position -2 is “calle”, will be translated into the node for Catalan “word:-2:carrer”, which will test whether the -2 word is “carrer”.

As a result, we obtain models which are trained on Spanish and applied to Catalan text.

4.2 Cross-Linguistic Features

As a more sophisticated alternative, we propose a bilingual model which works for Spanish and Catalan at the same time. We do this by using what we call *cross-linguistic features*, instead of the monolingual word forms specified above. Assume a feature *lang* which takes value *es* or *ca*, depending on the language under consideration. A cross-linguistic feature is just a binary feature corresponding to an entry in the translation dictionary, “es_w ~ ca_w”, which is satisfied as follows:

$$\text{X-Ling}_{\text{es}_w \sim \text{ca}_w}(w) = \begin{cases} 1 & \text{if } w = \text{es}_w \\ & \text{and } \textit{lang} = \textit{es} \\ 1 & \text{if } w = \text{ca}_w \\ & \text{and } \textit{lang} = \textit{ca} \\ 0 & \text{otherwise} \end{cases}$$

¹The InterNOSTRUM system is freely available at the following URL: <http://www.intemostrum.com>.

This representation allows to learn from a corpus consisting of mixed Spanish and Catalan examples. The idea here is to take advantage of the fact that the concept of NE is mostly shared by both languages, but differs in the lexical information, which we exploit through the lexical translations. With this we can learn a bilingual model which is able to recognize NEs both for Spanish and Catalan, but that may be trained with few — or even any— data of one language, in our case Catalan.

4.3 Direct Learning in Catalan

A third approach is the usual learning of a NER system using training data of the same language. Since our interest relies on developing a low-cost NER system for Catalan, we have performed standard learning on a small training set (described in table 1), with an annotation cost comparable to the cost of building the translation dictionary (about 10 person hours).

4.4 Results

Preliminary tuning on Spanish was performed on the Spanish development set, in order to fix learning parameters. The window sizes were set to 3 words around, except for the orthographic window, with size of 1 word around. Concerning classifiers, the depth of the base decision trees was fixed to 4 levels (i.e., tree branches represent conjunctions of up to 4 basic features). When applicable, the number of decision trees per classifier was automatically tuned in the Spanish development set selecting, from up to 2,000 base trees, the number which maximizes the F_1 measure. Otherwise it was fixed to 800.

First, in order to have a baseline for the data sets, two basic models were learned. The first, NO_LEX, makes no use of lexical information at all, that is, focuses only on orthographic features, affixes, type patterns and left predictions. We trained this model on the Spanish training data and we directly applied it to both languages. As a second baseline, a model for Catalan (including lexical information) LEX.ca, was trained using the small Catalan training set.

Following the approach described in Section 4.1, a model was learned on the Spanish training

set, and then translated into Catalan, generating the model LEX.es2ca. Note that this model is also applicable both to Spanish and Catalan, considering, respectively, the learned set of Spanish lexical forms or the translated Catalan ones. In addition, we tested the influence of cross-linguistic features presented in Section 4.2. We trained one model, X-LING_{es}, only with the Spanish training data, and a second model, X-LING_{mix}, using both the Spanish training data and the Catalan training set. In both approaches the experiments were replicated using the two available translation dictionaries.

Table 2 presents the results of all the learned models on the test sets. Clearly, comparing the performance of the NO_LEX model versus the others, it can be stated that lexical information significantly helps on the NER task on both languages.

Looking at the results on the Catalan test (right block), all the models using the manual dictionary achieve a very competitive performance over 90% of F_1 measure. Therefore, the techniques to adapt a NER model to Catalan seem to work considerably well. The LEX.ca model performs somewhat worse (89.18%) than others (probably because of the reduced size of the training set), indicating that, in similar conditions of annotation effort, it is preferable to translate the models than to learn from the small Catalan corpus.

The LEX.es2ca and X-LING_{es} models perform nearly the same. Actually, since they are trained on the same Spanish data, the models are fairly equivalent, and the minor differences may be attributed to the fixed vocabulary of the cross-linguistic model. Besides, the X-LING_{mix} model, trained with mixed corpora, achieves the best results (91.18%), which supports our arguments on learning simultaneously from both languages.

Another positive result shown in table 2 is that the X-LING models using the automatically generated dictionary perform almost as well as using the manual dictionary (a loss of about 0.5 points in F_1 is observed in both cases). After a manual inspection, we explain the bad results of LEX.es2ca with the automatic dictionary (87.53% compared to 90.55%) by the large number of errors coming from the translation of Spanish words, which are directly applied on the Catalan data. X-LING models perform instead a new training step and they

	es train	ca train	dicc.	es test			ca test		
				prec.	rec.	F_1	prec.	rec.	F_1
NO_LEX	yes	no	—	89.31	88.03	88.67	82.80	82.21	82.50
LEX.ca	no	yes	—	—	—	—	90.98	87.44	89.18
LEX.es2ca	yes	no	man. aut.	92.81	92.89	92.85	89.14	92.00	90.55
							83.85	91.55	87.53
X-LING _{es}	yes	no	man. aut.	92.25	92.64	92.44	90.78	89.76	90.27
				92.23	92.69	92.46	89.95	89.61	89.78
X-LING _{mix}	yes	yes	man. aut.	92.27	92.53	92.40	91.95	90.43	91.18
				92.57	92.39	92.48	91.29	90.13	90.71

Table 2: Evaluation of the learned models on the test datasets for Spanish (es) and Catalan (ca). The “es” and “ca train” columns indicate the training material used in each model. The “dicc.” column specifies the dictionary (either manual or automatic) used for translating models. The NO_LEX model learns without making use of lexical information. The LEX.ca model is a baseline standard model developed on Catalan. The LEX.es2ca is a translated model from Spanish to Catalan. The X-LING models are bilingual models using cross-linguistic features.

are capable of discarding useless erroneous cross-linguistic features.

Regarding the performance on Spanish (left block), the original model, LEX.es2ca, working with Spanish lexical information, obtains the best results (92.85%), but cross-linguistic models are still competitive (with a small loss of 0.4 points in F_1). This fact indicates that training with both languages at the same time does not significantly hurt the performance of the individual Spanish model. Additionally, the multilingual models are simpler to use, since they work straightforwardly with both languages, whereas form-based translated models are specific for each language.

We would like to note also that the systems achieve the same order of performance for both languages, which was shown to be very competitive in CoNLL’02. Although the table figures correspond to evaluations in different sets, and thus, can not be directly compared, the two corpora are similar, since both consist of news article from the same dates and geographical area.

As far as the cost concerns, it happens that the better the performance of a model, the more the resources needed to obtain it. Probably, the best tradeoff is observed in the case of X-LING_{mix} with the automatic dictionary, which allows to almost automatically construct an accurate NER system for Catalan (90.71%) at the only cost of 10 person hours of corpus annotation.

5 Bootstrapping the models

This section describes an attempt to improve the NER models via bootstrapping techniques, that is, making use of the available large amount of unlabelled data in Catalan.

We describe a simple, naive strategy for the bootstrapping process. The unlabelled data in Catalan has been randomly divided into a number of equal-sized disjoint subsets $S_1 \dots S_N$, containing 1,000 sentences each. Given an initial NER model M_0 and a base labelled data set T_L , the process is as follows:

1. For $i = 1 \dots N$ do :
 - (a) Identify the Named Entities in $S_1 \dots S_i$ using model M_{i-1} .
 - (b) Learn a new model M_i using as training data $T_L \cup \bigcup_{j=1}^i S_j$.
2. Output Model M_N .

At each iteration, a new unlabelled fold is included in the learning process. First, the folds are labelled by the current model, and then, a new model is learned using the base training data plus the label-predicted folds.

We have run the process for three of the models above, always using the manual dictionary: LEX.ca, with Catalan training set as T_L ; X-LING_{es}, with Spanish training set as T_L ; and X-LING_{mix},

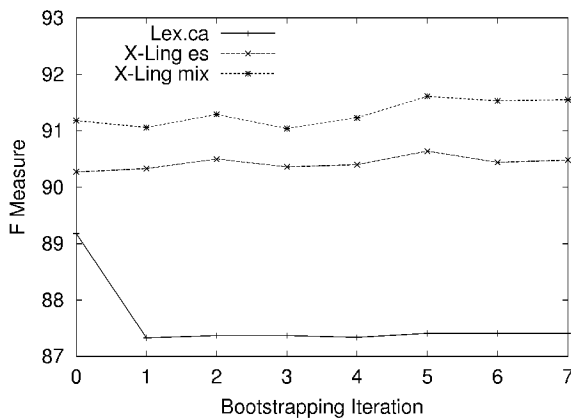


Figure 2: Progress of the F_1 measure through bootstrapping iterations.

with T_L as the union of the Spanish and Catalan training material. Since the LEX.es2ca model can not mix its initial Spanish training with the Catalan folds, we have avoided the model in the experiment. Figure 2 depicts the evolution of the F_1 measure through the bootstrapping process, for 7 iterations.

The model LEX.ca experiments a sharp drop of 2 points in the first iteration, and beyond iteration 5 gets stable at 87.41%. In our opinion, the Catalan training set is not big enough and the errors in the retraining folds degrade the performance of the bootstrapped model. On the other hand, the cross-linguistic models show a slightly better behavior, achieving a maximum increase of about 0.5 points, getting also somewhat stable beyond iteration 5. Again, X-LING_{mix} is slightly better than X-LING_{es}. Bootstrapping, therefore, is not very helpful on improving models. However, these models seem to have learned a robust concept which overcomes the errors produced when relabelling folds. It is also interesting to realize that the inclusion of the Catalan training is crucial in the difference in performance between the cross-linguistic models: the X-LING_{es} model is not able to acquire from the unlabelled data the same behavior than the X-LING_{mix} model, which has access to the manually annotated Catalan set (nearly of the same size than each fold).

More complex variations to the above bootstrapping strategy have been experimented. Basically, our direction has concentrated on select-

ing from the unlabelled material only the “good” sentences for the learning process, by taking those which maximize a mean of the confidences of the predictions on a sentence, or those in which two different models agree on the prediction. In all cases, results lead to conclusions similar to the ones described above.

6 Conclusions and Further Work

We have presented an experimental work on developing low-cost Named Entity recognizers for a language with no available annotated resources, using as a starting point existing resources for a similar language. We have devised and evaluated several strategies to build a Catalan NER system using only annotated Spanish data and unlabelled Catalan text, and compared our approach with a classical bootstrapping setting where a small initial corpus in the target language is hand tagged.

The main conclusions drawn from the presented results are: 1) At same cost, the hand translation of a Spanish model is better than hand annotating a small Catalan training corpus from which directly learn a model. 2) The translation of the Spanish model can be automatically done by using a Spanish-Catalan machine translation system, obtaining also very competitive results. 3) The best strategy turned out to be the use of cross-linguistic features, which enables the training of models using mixed corpora, and results in a system able to work reasonably on both languages.

Results of the experiments with a simple bootstrapping strategy suggest several conclusions. First, LEX.ca is not improved via bootstrapping, probably due to the small size of the Catalan training corpus. Second, bootstrapping slightly improves initial X-LING models, producing robust models which are not degraded by the noise introduced in subsequent iterations of bootstrapping.

Some open issues that should be addressed in the future include an improvement of the quality and coverage of the automatic translation of dictionary entries, and a further development of the idea of cross-linguistic features, extending it either from bilingual to multilingual translations, or including semantic relations, through the use of WordNet or similar ontologies. This could open the door to apply the method to groups of similar

languages (e.g., between Romance languages like Catalan, French, Galician, Italian, Spanish, etc.).

In addition, bootstrapping techniques should be better studied in this domain, in order to take advantage of the large quantities of available unlabelled data. Particularly, we think that it is worth investigating the size and selection of the retraining corpora, and the combination of several algorithms or example views like in the co-training algorithms presented in (Collins and Singer, 1999; Abney, 2002).

Acknowledgements

The authors thank the anonymous reviewers for their valuable comments and suggestions in order to prepare the final version of this paper.

This research has been partially funded by the Spanish Research Department (HERMES TIC2000-0335-C03-02, PETRA TIC2000-1735-C02-02), by the European Commission (MEANING IST-2001-34460), and by the Catalan Research Department (CIRIT's consolidated research group 2001SGR-00254 and research grant 2001FI-00663).

References

- J. Aberdeen, J. Burger, D. Day, L. Hirschman, P. Robinson, and M. Vilain. 1995. MITRE: Description of the ALEMBIC System Used for MUC-6. In *Proceedings of the 6th Message Understanding Conference*, pages 141–155, Columbia, Maryland.
- S. Abney. 2002. Bootstrapping. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Taipei, Taiwan.
- D. Appelt, J. Hobbs, J. Bear, D. Israel, M. Kameyama, A. Kehler, D. Martin, K. Myers, and M. Tyson. 1995. SRI International Fastus System MUC-6 Test Results and Analysis. In *Proceedings of the 6th Message Understanding Conference*, pages 237–248, Columbia, Maryland.
- D. Bikel, S. Miller, R. Schwartz, and R. Weischedel. 1997. Nymble: A High Performance Learning Name-Finder. In *Proceedings of the 5th Conference on Applied Natural Language Processing, ANLP*, Washington DC.
- W. Black and A. Vasilakopoulos. 2002. Language-Independent Named Entity Classification by Modified Transformation-Based Learning and by Decision Tree Induction. In *Proceedings of CoNLL-2002*, pages 159–162. Taipei, Taiwan.
- W. Black, F. Rinaldi, and D. Mowatt. 1998. Facile: Description of the NE System Used for MUC-7. In *Proceedings of the 7th Message Understanding Conference*.
- A. Borthwick, J. Sterling, E. Agichtein, and R. Grishman. 1998. NYU: Description of the MENE Named Entity System as Used in MUC-7. In *Proceedings of the 7th Message Understanding Conference*.
- X. Carreras, L. Màrquez, and L. Padró. 2002. Named Entity Extraction Using AdaBoost. In *Proceedings of CoNLL-2002*, pages 167–170. Taipei, Taiwan.
- M. Collins and Y. Singer. 1999. Unsupervised Models for Named Entity Classification. In *Proceedings of EMNLP/VLC-99*, College Park MD, USA.
- G. Krupka and K. Hausman. 1998. IsoQuest, Inc.: Description of the NetOwlTM Extractor System as Used for MUC-7. In *Proceedings of the 7th Message Understanding Conference*.
- R. Malouf. 2002. Markov Models for Language-Independent Named Entity Recognition. In *Proceedings of CoNLL-2002*, pages 187–190. Taipei, Taiwan.
- P. McNamee and J. Mayfield. 2002. Entity Extraction Without Language-Specific Resources. In *Proceedings of CoNLL-2002*, pages 183–186. Taipei, Taiwan.
- A. Mikheev, C. Grover, and M. Moens. 1998. Description of the LTG System Used for MUC-7. In *Proceedings of the 7th Message Understanding Conference*.
- R. Schapire and Y. Singer. 1999. Improved Boosting Algorithms Using Confidence-rated Predictions. *Machine Learning*, 37(3):297–336.
- R. Schapire. 2002. The Boosting Approach to Machine Learning. An Overview. In *Proceedings of the MSRI Workshop on Nonlinear Estimation and Classification*, Berkeley, CA.
- E. Tjong Kim Sang. 2002a. Introduction to the CoNLL-2002 Shared Task: Language-Independent Named Entity Recognition. In *Proceedings of CoNLL-2002*, pages 155–158. Taipei, Taiwan.
- E. Tjong Kim Sang. 2002b. Memory-Based Named Entity Recognition. In *Proceedings of CoNLL-2002*, pages 203–206. Taipei, Taiwan.
- K. Tsukamoto, Y. Mitsuishi, and M. Sassano. 2002. Learning with Multiple Stacking for Named Entity Recognition. In *Proceedings of CoNLL-2002*, pages 191–194. Taipei, Taiwan.
- R. Weischedel. 1995. BBN: Description of the PLUM System as Used for MUC-6. In *Proceedings of the 6th Message Understanding Conference*, pages 55–69, Columbia, Maryland.
- S. Yu, S. Bai, and P. Wu. 1998. Description of the Kent Ridge Digital Labs System Used for MUC-7. In *Proceedings of the 7th Message Understanding Conference*.