Adaptability of Linguistic Resources to New Domains: an Experiment with Proper Noun Dictionaries

ALESSANDRO CUCCHIARELLI
ISTITUTO DI INFORMATICA
UNIVERSITÀ DI ANCONA
VIA BRECCE BIANCHE, 60131 ANCONA, ITALY
ALEX@INFORM.UNIAN.IT

PAOLA VELARDI
DIP. DI SCIENZE DELL'INFORMAZIONE
UNIVERSITÀ DI ROMA 'LA SAPIENZA'
VIA SALARIA 113, 00198 ROMA, ITALIA
VELARDI@DSI.UNIROMA1.IT

Abstract

One of the major problems of state-of-art NLP systems is the difficulty to adapt the linguistic knowledge bases to new application and languages, without substantial manual work. In this paper we present a method, and related experiments, to build a *self-adaptive linguistic knowledge base*. The method is applied to the problem of automatic Proper Noun Dictionaries extension.

1 Introduction

Information Extraction is an emerging technology [Pazienza 1997] aiming at retrieving facts from textual information available electronically. Information Extraction significantly differs from the more mature field of Information Retrieval, in that IE aims to extract specific *facts* from documents. This task is rather more complex, since there are many ways to express the same fact in natural language, and information may be spread across different sentences.

Broadly speaking we can say that the field of IE grew very rapidly when ARPA funded competing research groups through the Message Understanding Conferences (MUC).

MUC conferences favoured the design of high-performing IE systems by exploiting the recent trend of Natural Language Processing towards more empirical computational linguistics, that is, by putting less emphasis on linguistic theory, and more on the use of language engineering techniques, to ensure higher coverage of the language processing tasks being considered.

In our view, the meritorious "make it work" imperative, while favouring the design of very high-performing systems, somehow lowered the attention on generality and domain adaptability of the devised techniques.

Often, linguistic knowledge and rule bases used for fact extraction are specifically defined for the task at hand, e.g.

identification of management successions in economic newswires or terrorist events.

To facilitate the adaptability of IE to applications, state-of-art systems put more emphasis on the design of engineering tools (e.g. GATE [Humphreys et al. 1996]) and knowledge acquisition interfaces (e.g. Proteus, [Yangarber and Grishman, 1997]) than on the definition of self-adaptive and learning capabilities.

We believe that *adaptive capabilities of NLP systems require the extensive use of language samples as a source evidence for extracting and organizing lexical information*. Data are extracted from real texts, as embodiment of language *in use*, to capture lexical regularities and to code them into operational rules.

In this paper, we present the results of research, carried out within a European project, concerning the design of self-adaptive Named Entities Gazetteers in Information Extraction.

Gazetteers are dictionaries of common proper nouns grouped by category (e.g. *airline*, *person first name*, *organization*, *location*, etc.). It is well known that, in many sublanguages, proper nouns (PNs) represent a significant percentage (20-30% and more) of the words in a corpus.

Dictionaries of PNs are therefore essential to effective understanding of language, at least so that they can be recognized within their context as locations, products, persons, etc. They take on a special significance in many applications, especially where the name is central to the application, such as in information systems through automated telephone call handling, and in information extraction for financial operators. The importance of this task has been recognized within the European Community, which recently funded projects entirely dedicated to the task of Proper Nouns handling (e.g. ONOMASTICA http://www2. echo.lu /search97cgi/).

1.1 The Named Entity recognition task

The Named Entity recognition task is the problem of identifying and semantically tagging proper nouns in running texts. In terms of syntactic categories, PNs are lexical noun phrases, that can be formed of primitive proper names (*Clinton*) groups of proper names of different semantic categories (*Vice Chairman James T Sherwin)*, and also of non-proper names (*Jamaica tourist board)*. In the latter case, capital letters are

optional, making the problem of PN items identification even more complex.
In the literature, it is accepted that an adequate treatment of proper nouns requires two complementary types of evidence [McDonald, 1993]:

- the *internal evidence*, that can be derived from the sequence of words (proper names and trigger words, such as Inc., &, Ltd., Company, etc.). Internal evidence is gained in almost all state-of-art PNs recognizers by the use of large *gazetteers* and lists of trigger words;
- the *external evidence* is the context of a proper noun, that provides classificatory criteria to reinforce internal evidence, or supply some classificatory evidence when there is no internal evidence. In fact, PNs form an *open class*, making the incompleteness of gazetteers an obvious problem.

1.2 State of Art Named Entity recognizers

In Information Extraction, the Named Entity recognition task has been one of the main test cases for evaluating competing systems in MUC contests. The highest performing systems include large numbers of hand-coded rules, or patterns, such as VIE [Humphreys 1996], the UMass system [Fisher et al. 1997] and Proteus [Grishman et al. 1992], but lately a high performance has been obtained by the use of statistical methods. For example, Nymble [Bikel et al. 1997] learns names using a trained approach based on a variant of Hidden Markov Models. However, a 90% success rate is reached at the price of manually tagging around half a million words.
Yet, since PNs are mostly domain-specific, there is no evidence that a similar performance may be obtained in other languages and domains ohter than those considered in the mentioned papers, if not at the price of a similar effort for rule writing (or manual training), and for the compilation of a high-coverage gazetteer. *In any case, since PNs form an open class, adaptable NLP systems should provide automatic means to increase system robustness against unknown items*. A recent study [Palmer and Day, 1997] established that the baseline performances of the PNs recognition task for several languages and application domains vary between 34% and 71%. The lower bound is calculated by considering a simple algorithm that recognizes PNs on the basis of a list of frequent proper nouns seen in a training set.
In this paper, we present a corpus-based incremental method to update a Proper Nouns Gazetteer with minimal manual effort.
The gazetteer extension is performed in 3 phases:
In **phase one**, PNs are detected from a Learning Corpus using an initial Gazetteer and a kernel of PNs detection rules, as in most Named Entity recognizers.
In **phase two**, unrecognized PNs are submitted to a corpus-based module for automatic unknown PNs classification (**U_PN tagger**). The U_PN tagger computes a statistical measure of syntactic and semantic similarity of a given U_PN in contexts, with respect to previously recognized PNs. A semantic category is assigned to the unknown, based on this measure.
The correctly recognized PNs are then added to the gazetteer (**phase 3**), and the three phases are repeated, until the global performance of the system becomes stable (i.e. the U_PN tagger is unable to classify any new item).
Section 2 describes the overall architecture of the system and the U_PN tagger algorithm.
Section 3 provides a detailed discussion of an experiment performed "in the large" on a corpus of economic news (the Italian language journal *Sole24Ore*). In this experiment, the Gazetteer and rule-base used in phase 1 were imported from an existing English system (VIE) after a rather "quick" manual adaptation to the Italian language. The experiment demonstrates that, at the end of the incremental automatic adaptation process, the system reaches the performance of state-of-the-art IE systems (around 90% precision), with a 20% increase in performance with respect to those obtained using only the initial (non-adapted) system.
We then replicated the experiment on the English language *Wall Street Journal*.
In the English experiment, we used an on-line linguistic resource, the *WordNet* word sense taxonomy [Beckwith et al. 1993] (not available in Italian), to investigate the effect of gradually generalizing the notion of contextual similarity.
The objective of this second experiment was to increase the evidence of rarely occurring patterns by "cautious" context expansion.

2 A method for automatic gazetteer extension

This section describes the method for automatic unknown Proper Nouns classification (hereafter U_PN tagging) and incremental extension of a PNs gazetteer.

2.1 Learning from corpus typical contexts

The objective of the initial learning phase is to acquire from the application corpus examples of typical PNs contexts, ordered by category type. Context learning requires some standard NLP tool and a Named Entity recognizer. The NE recognizer needs not be tuned to the application corpus, nor high coverage, or this would beg the question of self-adaptability. However, it should be able to capture at least "some" examples of each PN category. In our system, standard PNs are recognized using the VIE Named Entity recognizer.
In VIE, PNs are identified using a gazetteer, structured in files and related lists of trigger words

for each proper noun category (e.g. "Gulf" for Location, or "Association" for Organization).

A context-sensitive grammar is used to parse proper nouns in contexts. At the end of this phase, “some” PNs are recognized and classified, depending upon the size of the gazetteer, and the coverage of the context-sensitive grammar.

After NE tagging, syntactic processing is applied over the corpus.

The objective is to learn typical PNs contexts, ordered by category. A shallow parser extracts from the learning corpus *elementary syntactic relations* such as subject-object, noun-preposition-noun, etc. An *elementary syntactic link* (hereafter *esl*) is represented as:

$$esl_i(w_j, mod(type_i, w_k))$$

where w_j is the head word, w_k is the modifier, and $type_i$ is the type of syntactic relation (e.g. N_PP: noun-preposition-noun, N_V: subject-verb, V_N: verb-direct object, etc.).

In our study, the *context* of a word *w* in a sentence S is represented by the esls including *w* as one of its arguments. This notion is rather restricted (the large majority of works in this area use the set of words in a more or less narrow window around *w*), however, as proved in our previous work, syntactic contexts are far more reliable indicators.

After corpus parsing, the esls, including semantically classified PNs as one of its arguments, are grouped in a database, called **PN_esl**. This database provides contextual evidence to assign a category to unknown PNs.

Another database, **U_PN_esl**, includes all the esls with an unknown proper noun.

2.2 Classification of unknown PNs

The algorithm for unknown proper nouns classification works as follows:

let **U_PN** be an unknown proper noun, i.e. a single word or a complex nominal. Let $C_{pn} = (C_{pn1}, C_{pn2}, ..., C_{pnN})$ be the set of semantic categories for proper nouns (e.g. Person, Organization, Product etc.). Finally, let **ESL** be the set of *esls* in **U_PN_esl** that include **U_PN** as one of its arguments.

For each esl_i in **ESL** let:

$$esl_i(w_j, mod(type_i, w_k)) = esl_i(x, U\_PN)$$

where $x=w_j$ or w_k and $U\_PN=w_k$ or w_j, $type_i$ is the syntactic type of esl, and further let:

$$pl(esl_i\ (x, U\_PN))$$

be the *plausibility* of a detected esl. The plausibility is a measure of the statistical evidence of a detected syntactic relation [Basili et al, 1994], [Grishman and Sterling, 1992], that depends upon *local* (i.e. at the sentence level) syntactic ambiguity and *global* corpus evidence.

Finally, let:

- **- ESL_A** be a set of esls in **PN_esl** defined as follows: for each $esl_i(x, U\_PN)$ in **ESL** put in **ESL_A** the set of $esl_j(x, PNj)$ with $type_j=type_i$, x in the same position of esl_i, and PN_j a <u>known</u> proper noun, in the same position as U_PN in esl_i,
- **- ESL_B** be the set of esls in **PN_esl** defined as follows: for each $esl_i(x, U\_PN)$ in **ESL** put in **ESL_B** the set of $esl_j(w, PN_j)$, with $type_j=type_i$, w in the same position of x in esl_i, $Sim(w,x) > \varepsilon$, and PN_j a <u>known</u> proper noun, in the same position as U_PN in esl_i. Sim(w,x) is a similarity measure between x and w.

In our experiments, $Sim(w,x) > \varepsilon$ iff w and x have a common hyperonym H in WordNet. The generality of H (i.e. the number of levels L from x to H) is made parametric, to analyze the effect of generalization.

For each semantic category C_{pnj} compute *evidence(C_{pnj})* as:

$$evidence(C_{pnj}) = \alpha \frac{\sum_{esl_i\ ESL_A, C(PNj)=C_{pnj}} weight_{ij}(x)\ \ D(x, C(PN_j))}{\sum_{esl_i\ ESL_A} weight_{ij}(x)\ \ D(x, C(PN_j))} + \beta \frac{\sum_{esl_i\ ESL_B, C(PNj)=C_{pnj}} weight_{ij}(w)\ \ D(w, C(PN_j))}{\sum_{esl_i\ ESLB} weight_{ij}(w)\ \ D(w, C(PN_j))}$$

where:

α and β are parametric, and can be used to study the evidence provided by **ESL_A** and **ESL_B**, and $D(x, C_{pnj})$ is a smoothing factor used to determine the *saliency* [Yarowsky 1992] of a context esl_i (x, _) for a category C_{pnj}, i.e. how good a context is at discriminating between C_{pnj} and the other categories[1].

Furthermore:

$$weight_{ij}(x) = weight_{ij}(esl_i(x, PNj)) = pl(esl_i(x, PNj))\ \ (1 - \frac{amb(x) - 1}{2k - 1})$$

$$weight_{ij}(w) = weight_{ij}(esl_i(w, PNj)) = pl(esl_i(w, PNj))\ \ (1 - \frac{amb(w) - 1}{k - 1})$$

where *amb(esl(x, PN_j))* is the ambiguity of x in esl_i. The intended effect of the constant factor k (experimentally determined) is to gradually reduce the influence of ambiguous words. The smoothing is tuned to be higher in **ESL_B**. The selected category for U_PN is:

$$C = argmax(evidence(C_{pnk})) = max_j(evidence(C_{pnj}))$$

[1] For example, a subject_verb phrase with the verb *make* (e.g. Ace made a contract..) is found almost with equal probability with Person and Organization names

When grouping all evidence of a U_PN in the corpus, the underlying hypothesis is that, in a given application, a PN has a *unique sense*. This is a reasonable restriction for Proper Nouns, supported by empirical evidence, though we would be more sceptical about the applicability of the one-sense-per-discourse paradigm [Gale et al. 1992] to generic words.
Notice that the formula of the evidence has several smoothing factors that work together to reduce the influence of unreliable contexts and increase that of the most reliable ones.

2.3 Incremental updating of the Named Entity Gazetteer

The incremental process for gazetteer adaptation requires, besides the application corpus, a manually tagged Test corpus, used to halt the process when performance becomes stable.

- Phase one: some PNs are detected using the gazetteer and contextual rules. The performance of this phase is measured on a Test Corpus.
- Phase two: unrecognized PNs are submitted to the U_PN tagger for automatic unknown classification. The local performance of the UP_N tagger is measured on the Test Corpus
- Phase 3: the PNs classified in Phase1 and Phase 2 are added to the gazetteer

The three phases are repeated, until the U_PN tagger is unable to classify new items. Notice that the Test corpus is used as a predictor of system performance over the application corpus. If in Phase 3 errors are not tolerated, a manual supervision is necessary.

3. Experimental discussion

In this section we describe two experiments. The first is performed on a corpus of economic news in Italian (the Sole24Ore corpus). This experiment closely follows the three phases described in section 2.3, but the notion of word similarity used to compute the set $\mathbf{ESL_B}$ is limited to strict synonymy.
In the second experiment we explore the effect of gradually generalizing the notion of word similarity, using the English Wall Street Journal corpus and the WordNet word sense hierarchy. Figure 1 illustrates an example (from the English corpus, for sake of readability) of U_PN tagging, for the PN *Angola*.
In the Sole24Ore experiment, we used a Learning Corpus of 500,000 words of economic and financial articles and a Test Corpus of 11,488 words, of which 355 are proper nouns (619 when counted with their frequency). The proper nouns belong to one of three categories: Person, Organization or Location, though 9 categories are considered in total[2].

```
PROPER NOUN:   Angola

1.000 G_N_P_N base in Angola 0
--------------------------------------------
ESLA= 2.000 G_N_P_N base in Philippines 2
ESLA= 1.000 G_N_P_N base in U_._S_. 2
ESLA= 1.000 G_N_P_N base in Tennessee 2
ESLB= 1.000 G_N_P_N home in Colorado 2
ESLB= 1.000 G_N_P_N home in Seattle 2
ESLB= 0.500 G_N_P_N root in England 2

1.000000 G_V_P_N remove from Angola 0
---------------------------------------------
ESLB= 1.000 G_V_P_N take from Canada 2
ESLB= 0.333 G_V_P_N take from William 3
ESLB= 1.000 G_V_P_N take from Mr_._Berry 3
ESLB= 1.000 G_V_P_N transfer from Chicago 2
ESLB= 1.000 G_V_P_N transfer from Japan 2

α: 0.70 β: 0.30

NUM CLASS     SUM_ESLA  SUM_ESLB  EVIDENCE
--- --------  --------  --------  --------
1   ORGAN     0.000000  0.981282  0.049176
2   LOCATION  5.978947  9.919980  0.704194
3   PERSON    0.000000  4.921385  0.246630

Max ev. category: LOCATION
```

Figure 1 - An example of U_PN tagging

In Table 1 the frequency distribution of PNs is shown. It is seen that the large majority (265) has only 1 occurrence.

Frequency	# PNs
1	265
2-4	69
5-9	14
>9	4

Table 1 - Frequency distribution of U_PN

When running the shallow parser over the Learning Corpus 43,847 PNs are recognized, possibly with errors due to over productivity of the grammar rules (linguistic processing over the Learning Corpus is not checked for correctness, but some over productivity phenomenon was observed in the Test Corpus). The set **PN_esl** included 70,070 *different* esls. We then used the Test Corpus to verify the performance of the system before and after U_PN_tagging in an incremental way.

[2] The categories used are those of MUC, with the addition of Product and Other. In practice, however, categories like Date, Percent, etc. are recognized rather well through the use of specific context free grammars, since they are productive phenomena. The category Product is almost absent in the fragment of Sole24Ore we used.

(PNs are not frequency weighted)

	A	B	C	D	E	F	G	H	I	J	K	L
Run #1	239	355	67.32%	339	70.50%	60	83	72.29%	75	80.00%	84.23%	88.20%
Run #2	298	355	83.94%	337	88.43%	1	19	5.26%	13	7.69%	84.23%	88.72%
Run #3	299	355	84.23%	337	88.72%	0	18	0.00%	12	0.00%	84.23%	88.72%

Legenda

A: PN correctly tagged at the end of Phase 1 in the Test Corpus
B: Total PNs in the Test Corpus
C: Recall after Phase 1 (A/B)
D: Total PNs detected at the end of Phase 1 (D = A+A1 (errors)+G (unknown))
E: Precision after Phase 1 (A/D)
F: U_PN correctly tagged after phase 2 in the Test Corpus
G: Total U_PN at the end of Phase 1
H: Local recall of U_PN tagger (Phase2) (F/G)
I: Total U_PN for which a decision was possible by the U_PN tagger
J: Local precision of the U_PN tagger
K: Recall after Phase 2 (A+F)/B

Figure 2 - Global Performance of the Gazetteer Adaptation method (Sole24Ore)

Figure 2 illustrates the results obtained with 3 runs of the incremental method, over the Test Corpus.
After Phase 1, the system classifies PNs in three sets: **A**, the set of correctly classified PNs, **A1**, the set of incorrectly classified PNs and **G**, the set of unknown PNs. In Phase 2, the U_PN tagger does not assign a tag only if no esls are found for the U_PN: in fact, certain esls without any information content are removed (for example auxiliaries).
It can be immediately seen that the performance of the method is saturated at the second run (88.7%, 91% when PNs are weighted with frequency). In Run #2, Phase 1, the performance (**C** and **E**) has a sharp increase with respect to the same values in Run #1, by virtue of the gazetteers update (60 new correctly tagged PNs are added to the gazetteers). Only 19 PNs remain unknown after Phase 1 of Run #2, and are again fed to the U_PN tagger. Of these, 6 cannot be assigned to any class, because there is no evidence in the learning corpus of similar patterns. Of the remaining 13, only one additional PN is correctly classified. Looking at the data, the 12 misclassified PNs appear true "hard cases". The contextual evidence is represented by a limited number of "weak" examples (often just one), for example, locatives, as in *Aarhus in Danimarca (Aarhus in Denmark*)
In general, locative phrases in **PN_esl** have a low discriminating factor for the unknown PN.
Furthermore, the databases **PN_esl** and **U_PN_esl** include syntactic noise, since they are not manually inspected. This means that, especially when a PN occurs once and there is a limited evidence of similar patterns in **PN-esl**, noisy esls cause misclassification.
While wisdom would suggest avoiding a choice when only one example is found, one such policy would dramatically lower the recall of the system, since most unknowns, as show in Table 1, have only one occurrence.
An alternative choice is to increase the evidence of rare patterns by generalizing the notion of similarity in $\mathbf{ESL_B}$.

To evaluate the effect of generalization on performance, we performed a second experiment using WordNet and an English corpus, the Wall Street Journal (WordNet is not available in Italian).
In the Wall Street Journal corpus, we detected (using VIE) 8,712 PNs instances belonging to one of the three main semantic categories: Organization, Location or Person. The frequency distribution is the following:

Freq.	NPs	%
1	3779	43
2	2269	26
3-9	2082	24
≥10	582	7

While the category distribution is:

Cat.	NPs	%
1(Org)	2890	33
2(Loc)	514	6
3(Per)	5308	61

We selected a test set of 250 PNs, that reproduce the frequency and category distribution of the corpus. We removed these 250 PNs from the gazetteer, and then we measured the performance of the U_PN tagger at reclassifying them correctly.
Table 2 summarizes the results. Each row shows the result of an experiment performed with a certain level of generalization L, indicated in the first column of the Table. L=0 means "no generalization", e.g. only $\mathbf{ESL_A}$ is used in the evidence formula. L=1 means that in $\mathbf{ESL_B}$ only the strict synonyms are used (as for the Sole24Ore experiment), and the subsequent numbers represent increasing generalization levels L.
The table shows that increasing the evidence of similar patterns through generalization produces an increase of performance (especially recall) up to the first two levels of generalization (0 and 1).
Further generalization causes a drop in performances, clearly due to the noise produced by semantic ambiguity (WordNet is highly ambiguous).

Generality (L)	Tot. Prec	Tot. Rec.	Prec. Org	Rec. Org	Prec. Loc	Rec. Loc.	Prec. Pers.	Rec. Pers
0	80.60	75.20	80.70	75.90	92.30	85.70	79.50	73.80
1	81.10	77.60	85.00	81.90	92.30	85.70	78.00	74.50
2	82.50	79.20	88.70	85.50	92.30	85.70	78.20	75.10
3	81.40	78.80	87.50	84.30	92.30	85.70	77.10	75.10
4	80.60	78.40	87.50	84.30	92.80	92.80	75.80	73.80

Table 2 - Performance of the U_PN tagger when increasing the level of generalization (Wall Street Journal)

The effect of context generalization on the performance of unknown PN tagging (and on other NLP tasks) is an issue that deserves much more discussion, omitted here for sake of space. For the purpose of this paper, it is interesting to note that the performance of our method is almost the same in two different languages and domains (80% and 81.1% precision of U_PN tagging, in the comparable case of L=1), and furthermore can be improved through a "cautious" use of context generalization.

4. Concluding Remarks

We believe that self-adaptability to languages and domains is a challenge for next-generation NLP systems.

In this paper we presented a method for automatic adaptation of Proper Noun Dictionaries.

The proposed tagging method shows very good performances on two languages and domains, and is easily reproducible. Of course, the one-sense per domain hypothesis, which is realistic for PNs, contributes to the good performance of the method, in that we can reliably group all the contexts in which the U_PN appears in the test corpus, and use these collective contexts to determine similarity with already classified PNs. Noise in data is originated by the ambiguity of words in U_PN contexts, but not (or rarely so) by the U_PN itself.

In the future, we plan to extend the method to the task of unknown common noun tagging. Unknown common nouns are often rare or technical words, therefore the one-sense hypothesis could still be applied.

References

[Basili et al. 1994]Basili, R., Marziali A., Pazienza M.T. *Modelling syntax uncertainty in lexical acquisition from texts*. Journal of Quantitative Linguistics, vol.1, n.1, 1994.

[Beckwith et al. 1993], R. Beckwith, C. Fellbaum, D. Gross, G. Miller *WordNet: A Lexical Database Organized on Psycholinguistic Principles*, in Lexical Acquisition: Exploting On-Line Resources to Build a Lexicon, U. Zernik Ed., Lawrence-Erlbaum Ass., 1991.

[Bikel et 1997] Bikel D.,Miller S., Schwartz R. and Weischedel R. *Nymble: a High-Performance Learning Name-finder*. in Proc. of 5th Conference on Applied natural Language Processing, Washington, 1997

[Fisher et al. 1996] Fisher D., Soderland S., McCarthy J., Feng F. and Lenhart W *Description of the UMass system as used for MUC-6*. http://ciir.cs.umass.edu/info/psfiles/ tepubs/tepubs.htm

[Gale et al. 1992] Gale K., Church W. K. and Yarowsky D. *One sense per discourse.* in Proc. of the DARPA speech and and Natural Language workshop, Harriman, NY, February 1992

[Grishman and Sterling, 1992] Grishman R., J. Sterling, (1992) *Acquisition of selectional patterns*, Proc. of COLING '92, Nantes, July 1992.

[Grishman et al. 1992] Grishman R., Macleod C. and Meyers A *NYU: description of the Proteus System as used for MUC-4.* in Proc. of Fourth Message Understanding Conference (MUC-4) June 1992

[Humphreys 1996] VIE Technical Specifications, 1996/10/1815. ILASH, University of Sheffield.

[McDonald. 1996] McDonald D. *Internal and External Evidence in the Identification and Semantic Categorization of Proper Names.* in Corpus Processing for Lexical Acquisition, J. Pustejovsky and B. Boguraev Eds., MIT Press 1996.

[Palmer and Day 1997] Palmer D. and Day D. *A Statistical Profile of the Named Enity Task.* in Proc. of 5th Conference on Applied natural Language Processing, Washington, 1997

[Pazienza 1997] M.T.Pazienza (Ed.) *Information Extraction: a Multidisciplinary Approach to an Emerging Technology.* Lecture Notes in Artificial Intelligence 1299, Springer 1997

[Yangarber and Grishman, 1997] Yangarber R. and Grishman, R. *Customization of Information Extraction Systems* in Proc. Of Workshop on Lexically Driven Information Extraction, Frascati, Italy, July 16th, 1997

[Yarowski, 1992]Yarowsky D., *Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora* Proc. of COLING-92, Nantes, Aug. 23-28.