Automatic Semantic Tagging of Unknown Proper Names

Alessandro CUCCHIARELLI
Università di Ancona
Istituto di Informatica
Via Brecce Bianche
60131 Ancona, Italia
alex@inform.unian.it

Danilo LUZI
Università di Ancona
Istituto di Informatica
Via Brecce Bianche
60131 Ancona, Italia
luzi@inform.unian.it

Paola VELARDI
Università di Roma 'La Sapienza'
Dip. di Scienze dell'Informazione
Via Salaria 113
00198 Roma, Italia
velardi@dsi.uniroma1.it

Abstract

Implemented methods for proper names recognition rely on large gazetteers of common proper nouns and a set of heuristic rules (e.g. *Mr.* as an indicator of a PERSON entity type). Though the performance of current PN recognizers is very high (over 90%), it is important to note that this problem is by no means a "solved problem". Existing systems perform extremely well on newswire corpora by virtue of the availability of large gazetteers and rule bases designed for specific tasks (e.g. recognition of Organization and Person entity types as specified in recent Message Understanding Conferences MUC).

However, large gazetteers are not available for most languages and applications other than newswire texts and, in any case, proper nouns are an open class.

In this paper we describe a context-based method to assign an entity type to unknown proper names (PNs). Like many others, our system relies on a gazetteer and a set of context-dependent heuristics to classify proper nouns. However, due to the unavailability of large gazetteers in Italian, over 20% detected PNs cannot be semantically tagged.

The algorithm that we propose assigns an entity type to an unknown PN based on the analysis of *syntactically and semantically* similar contexts already seen in the application corpus.

The performance of the algorithm is evaluated not only in terms of precision, following the tradition of MUC conferences, but also in terms of Information Gain, an information theoretic measure that takes into account the complexity of the classification task.

Introduction

In terms of syntactic categories, proper nouns are lexical NPs that can be formed by primitive proper names (*Adolfo_Battaglia*), groups of proper nouns of different semantic categories (*San_Paolo di Brescia*), and also of non-proper nouns (*Banca dei regolamenti internazionali*). 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 the use of a *context-sensitive grammar* (McDonald, 1996). McDonald points out that the context sensitivity requirement involves two complementary types of evidence: *internal* and *external*.

The *internal evidence*, can be derived from the sequence of words in a text (proper nouns and trigger words, such as Inc., &, Ltd., Company, etc.), and is gained in almost all state-of-art PNs recognisers 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, if any, or supplies some classificatory evidence. In fact, proper names form an open class, making the incompleteness of gazetteers an obvious problem.

The methods for recognition of proper nouns (PNs) described in literature closely reflects this view of the problem. PN identification typically includes:

- a gazetteer lookup, which locates simple and complex nominals identifying common PNs, such as companies, person names, locations, etc.
- a set of patterns or rules, stated in terms of part-of-speech, syntactic or lexical features (e.g. *Mr.* as an indicator of a PERSON entity type), orthographic features (e.g. capitalization), etc.

286

Proper nouns recognition has recently attracted much attention especially in the area of Information Extraction, where this problem is known as the Named Entity recognition task. The highest performing systems include large numbers of hand-coded rules, or patterns, such as VIE (Humphreys et al. 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 tagging manually around half a million words. Since PNs are mostly domain-specific, presumably a comparable effort is needed when shifting to different domains.

High performances of the existing systems are by no means the result of many years of studies and research in the area of IE from newswire English texts, promoted and funded by the Message Understanding Conferences (MUC) organizers. Yet, there is no evidence that a similar performance could be obtained in other languages and domains, if not at the price of a similar effort for rule writing (or manual training), and for the compilation of a high-coverage gazetteer. A recent study (Palmer and Day, 1997) established that the baseline performances of the PN 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.

The method we propose in this paper combines symbolic and statistical approaches to classify unknown PNs using context evidence previously extracted from the application corpus. *The method can be used to overcome the limitation of small gazetteers and poorly encoded rule bases.*

Our method is untrained: what is needed is a learning (raw) corpus, a surface syntactic analyzer, a dictionary of synonyms, a list of category names for classifying PNs (we used the categories proposed in the forthcoming MUC-7), and a "start-up" gazetteer and rule base, used to acquire an initial model of typical PNs contexts.

In the next section, we describe the method in detail. Section 3 is dedicated to a discussion of experimental results.

2 The Method

The problem of PN recognition has been considered in our group in the context of the European project ECRAN, aimed at improving domain adaptability of IE systems through the integrated use of corpora and MRDs.

A first version of the Named Entity (NE) recognizer, in Italian, closely reproduced the architecture of the VIE recognizer, developed at the University of Sheffield (Humphreys et al. 1996).

Proper noun recognition is initially performed in two steps:

1) common proper nouns are identified using a gazetteer, structured in files and related lists of trigger words for each proper nouns category (e.g. "Gulf" for LOCATIONs, or "Association" for ORGANIZATIONs);

2) a context-sensitive grammar of about 250 rules is used to parse proper nouns in contexts. The majority of rules uses internal evidence to identify and classify proper nouns made of complex NPs. For example the following rule is used to recognize street names:

rule(tagged_location_np(s_form:[via,",F2 ,' ',F3],sem:A^B),
  [nome(s_form:via,sem:_^_),
  organ_names_np(s_form:F2,sem:_^_),
  num(s_form:F3)])

Ex: " via Giorgio Marini 34 ".

When running these first two modules on a one million word corpus of economic news (extracted from the newspaper *Il Sole 24 Ore*), we obtained the following performances: 84% precision, 85% recall, about 20% proper nouns correctly identified as such, but NOT classified. Unknown proper nouns are identified initially by the Brill part-of-speech tagger (Brill, 1995). Complex unknown nominals (e.g. *Quick Take 200*) are partly detected by simple heuristics.

One of the motivations for such a high percentage of unknowns and relatively low performance (as compared with state-of-art PN recognizers) is that at the present state of implementation the gazetteer has a

limited coverage[1]; yet, the problem of unknowns is generally recognized as crucial in real-world applications, because proper nouns are an open class.
We have therefore devised a method to reinforce external evidence, using a corpus-driven algorithm to incrementally update the gazetteer and classification of unknown PNs in running texts.

The algorithm to classify unknown proper nouns uses the following linguistic resources: a (raw text) learning corpus in the same domain as the application, a shallow corpus parser, a "seed" gazetteer, and a dictionary of synonyms.
The shallow parser (Basili et al. 1994), extracts from the learning corpus *elementary syntactic relations* such as subject-object, noun-preposition-noun, etc. A 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. PP(of), PP(for), SUBJ-Verb, Verb-DirectObject, etc.).
The learning corpus is previously morphologically and syntactically processed. Step 1 and 2 described at the beginning of this section are used to detect PNs. A database of esls including known PNs[2] is then created and used by the algorithm to assign a category to unknown PNs. The algorithm works as follows:
let **PN_U** 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 *elementary syntactic links* (esl) extracted from the

learning corpus that include **PN_U** as one of its arguments.

For each esl_i in **ESL** let:

$$esl_i(w_j, mod(type_i, w_k)) = esl_i(x, PN\_U)$$

where $x = w_j$ or w_k and PN_U $= w_k$ or w_j, $type_i$ is the syntactic type of esl (e.g. N-di-N, N_N, V-per-N ecc), and further let:

$$pl(esl_i (x, PN\_U)$$

be the *plausibility* of a detected esl. The plausibility is a measure of the statistical evidence of a detected syntactic link (Basili et al, 1994b), 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 defined as follows: for each $esl_i(x,PN\_U)$ in **ESL** put in **ESL_A** the set of $esl_j(x,PN_j)$, in the corpus, with type=$type_i$, x in the same position of esl_i, and PN_j a known proper noun, in the same position as PN_U in esl_i,

- **ESL_B** be the set of esl_k defined as follows: for each $esl_i(x,PN\_U)$ in **ESL** put in **ESL_B** the set of $esl_j(w,PN_j)$, in the corpus, with type=$type_i$, w in the same position of x in esl_i, Sim(w,x)> δ, and PN_j a known proper noun, in the same position as PN_U in esl_i. Sim(w,x) is a similarity measure between x and w. In our first experiments, Sim(w,x)> δ iff w is a **synonym** of x.

For each semantic category C_{pnj} compute *evidence(C_{pnj})* as shown in Figure 1, where:
- *amb(esl(x, PN_i))* is a measure of the ambiguity of x and PN_j in esl_i;
- α and β are experimentally determined weights (currently, $\alpha=0.7$ and $\beta=0.3$).
The selected category for PN_U is:

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

The underlying hypothesis is that, in a given application corpus, a PN has a unique sense. This is a reasonable restriction supported by empirical evidence (see also (Gale et al. 1992)). An alternative solution would be to select the "best performing" tags, and then apply

---

[1] The context sensitive grammar closely reflects, with extension, that developed for a similar application in the English VIE system. Therefore, low performance is likely due to the low-coverage gazzetteer. The absence of available linguistic resources in languages other than English is a well known problem.

[2] Note that the database is not manually inspected for correctness (POS tagging and parsing errors). However, the parser assigns to each detected esl a statistical measure of confidence, called *plausibility* (Basili et al. 1994b).

$$(1) evidence(C_{pnj}) = \alpha \frac{\displaystyle\sum_{esl_i \in ESL_A, C(PN_j)=C_{pnj}} (pl(esl_i(x,PN_j)) * amb(esl_i(x,PN_j)))}{\displaystyle\sum_{esl_i \in ESL_A, anyPN} pl(esl_i(x,PN_j))} +$$

$$\beta \frac{\displaystyle\sum_{eslj \in ESL_B, C(PN_j)=C_{pnj}} (pl(esl_i(w,PN_j)) * amb(esl_i(x,PN_j)))}{\displaystyle\sum_{eslj \in ESL_B, anyPN} pl(esl_i(w,PN_j))}$$

Figure 1 - The evidence(C_{pnj}) computation formula

some WSD algorithm to predict the precise sense in running texts.

3   Discussion of the Experiment

In our experiment, we used a corpus of one million words extracted from articles in the *Il Sole 24 Ore* economic newspaper. A database of 76055 esls including proper nouns was obtained.

Table 1 shows the distribution of esls by category, and the *prior probability* (i.e. relative distribution) of each category.

Category	N° ESL$_i$	Prior Prob.
ORGANIZ	26418	0.347
LOCATION	25087	0.330
PERSON	20558	0.270
DATE	544	0.007
TIME	879	0.011
MONEY	1076	0.014
PERCENT	520	0.007
PRODUCT	2671	0.035
OTHERS	1112	0.015
Tot. ESL	**76055**	

Table 1 - PN distribution by category

The semantic categories in Table 1, with the addition of Product, are those that will be used for Named Entity task evaluation in the forthcoming MUC-7 contest.

In Figure 2, a complete experiment is reported. In the figure, an esl is represented as a list, for example (*0.5 G_N_P_N Quick_Take_200 0 1 in documento*). The detected esl is '*Quick_Take_200 in documento*' (*Quick_Take_200 in document*), the syntactic type is G_N_P_N (noun-preposition-noun), the plausibility is 0.5,

the initial category of Quick_Take_200 is 0 (= unknown) and its ambiguity is initially set to 1.

It is seen in the figure that some detected esls do not contribute to the computation of (1) (e.g. acquisire con Quick_Take_200 *to acquire with Quick_Take_200*) while some other esl turns out to be particularly informative (e.g. qualita' di Quick_Take_200 *quality of Quick_Take_200*)

For the name Quick_Take_200 (a software product), the category 8 is finally selected (*PRODUCT*, as shown in the figure).

An extended experiment was designed as follows:

We selected from the corpus 35 PNs for each of the following categories: Organization, Person, Location and Product[3]. The PNs are selected by ranges of frequency in the corpus, except for Producs, that are very rare in our excerpt of the *Il Sole 24 Ore*: here we selected the 35 top frequency PNs.

We then removed each of the 140 PNs from the gazetteer, one at the time, and attempted a re-classification using our algorithm.

To evaluate the performances we used, in addition to the classical Precision measure, the *Information Gain* (Kononenko and Bratko, 1991).

The Information Gain is an information-theoretic measure that takes into account the *complexity* of the classification task.

---

[3]The other categories are less interesting in our view. Numbers, dates etc. are recursive and regular phenomena that can be detected in a more general way by the use of specific grammars or pattern matchers.

289

PROPER NAME: Quick_Take_200

0.5 G_N_P_N Quick_Take_200 0 1 in documento

1.0 G_N_V Quick_Take_200 0 1 nil dotare

ESLB= 1.0 G_N_V Apple 1 1 nil fornire
ESLB= 1.0 G_N_V Power_Pc 1 1 nil fornire
ESLB= 1.0 G_N_V Tank_Francaise_Chronoreflex 8 1
                nil equipaggiare

0.1 G_Agg_P_N acquisito con Quick_Take_200 0 1

0.1 G_pp_P_N acquisire con Quick_Take_200 0 1

0.333000 G_N_P_N Forza di Quick_Take_200 0 1

ESLA= 0.333000 G_N_P_N Forza di Linea_Pret 2 1

0.333000 G_N_P_N Punti di Quick_Take_200 0 1

0.333000 G_N_P_N acquisizione di Quick_Take_200 0 1

0.333000 G_N_P_N capacita' di Quick_Take_200 0 1

ESLB= 0.167000 G_N_P_N portata di 280_Kg 9 1
ESLB= 0.2 G_N_P_N portata di 300_Kg 9 1
ESLB= 0.333000 G_N_P_N mezzo di Cartier 3 1
ESLB= 0.333000 G_N_P_N facilita' di Apple_Share 8 1

0.333000 G_N_P_N immagine di Quick_Take_200 0 1

0.333000 G_N_P_N importante di Quick_Take_200 0 1

ESLB= 0.333000 G_N_P_N grande di Weil 3 1
ESLB= 0.250000 G_N_P_N grande di Europa 2 1
ESLB= 0.2 G_N_P_N grande di Casa 1 1

0.333000 G_N_P_N qualita' di Quick_Take_200 0 1

ESLA= 1.0 G_N_P_N qualita' di Epm 8 1
ESLB= 0.333000 G_N_P_N sorta di Iri 1 1
ESLB= 1.0 G_N_P_N generazione di G 3 1
ESLB= 0.125000 G_N_P_N caratteristica di Casa 1 1
ESLB= 0.250000 G_N_P_N caratteristica di
                Macintosh_Performa 8 1
ESLB= 0.250000 G_N_P_N caratteristica di Vs 8 1
ESLB= 0.5 G_N_P_N marca di Arese 2 1

0.1 G_V_P_N acquisire con Quick_Take_200 0 1

0.2 G_V_P_N utilizzare con Quick_Take_200 0 1

Coefficient α: 0.7
Coefficient β: 0.3

NUM	CLASS	SUM_ESLA	SUM_ESLB	EVIDENCE
1	ORG	0.000	2.658	0.109
2	LOC	0.333	0.750	0.205
3	PERSON	0.000	1.666	0.068
4	DATE	0.000	0.000	0.000
5	TIME	0.000	0.000	0.000
6	MONEY	0.000	0.000	0.000
7	PERCENT	0.000	0.000	0.000
8	PRODUCT	1.000	1.833	0.600
9	OTHERS	0.000	0.367	0.015

SUM_ESLA= 1.333   SUM_ESLB= 7.274

Max evidence category is: PRODUCT

Selected category: PRODUCT

Figure 2 - A complete example

If $P(C)$ is the prior (a-priori) probability[4] that an instance c is a member of class C, and $P'(C)$ is the probability of $c \in C$, as computed by the classifier in a given test t_i, the Information Gain $I(t_i)$ is defined as:

$$I(t_i) = \log(1-P(C)) - \log(1-P'(C))$$
$$if\ P(C) > P'(C)$$

or

$$I(t_i) = \log(P'(C)) - \log(P(C))$$
$$if\ P'(C) > P(C)$$

That is, if the classification is wrong, $I(t_i)$ is a penalty as high as the classification task

was an easy one (i.e. the prior probability of C was high). If the classification is correct, $I(t_i)$ is a price as high as the classification task was complex (i.e. the prior probability of C was low).

Over a test set of T cases, I is given by:

$$I = \frac{1}{T} \sum_{i=1}^{T} I(t_i)$$

Table 2 illustrates the results. It is seen that unknown PNs in the three major categories (those for which there is evidence in the corpus and in the gazetteer) have a very high probability of being correctly classified (up to 100% for Organizations). On the contrary, we obtain poor performances with Products.

However, Product is interesting because:

- there are no more than 50-60 product names in the gazetteer (which we

---

[4]The prior probability can be easily computed in a learning set as the ratio between the number of training instances belonging to a class C and the total number of training instances. In our experiment, the prior probabilities are listed in Table 1.

290

manually added for the purpose of this experiment)
- there are no contextual rules for Products in the context-sensitive grammar.

Thus, both prior probability and prior knowledge on Products are close to zero. This is numerically evidenced by the Information Gain: though we are not learning much about Products, the Information Gain is higher than for the other categories, and also as an absolute value (in (Kononenko and Bratko, 1991) a 0,5 bit improvement is among the highest measured values in a comparative experiment). In addition, the relative precision of classifying PNs as Product is 100%. This means that most products are misclassified, but, if something is classified as Product, this information can be reliably used to enrich the gazetteer.

Category	Precision	Inf. Gain
ORGANIZ.	100.00%	0.11
LOCATION.	91.43%	0.14
PERSON	80.00%	0.23
PRODUCT	22.86%	0.65

Table 2 - Precision and Information Gain of the method

Table 3 reports an experiment on a small corpus extracted from another portion of *Il Sole 24 Ore*, indexed as "New Products".

Category	N° ESL$_i$	Prior Prob.
ORGANIZ	735	0.160
LOCATION	583	0.126
PERSON	902	0.196
DATE	7	0.001
TIME	8	0.001
MONEY	31	0.007
PERCENT	114	0.025
PRODUCT	2184	0.473
OTHERS	262	0.057
Tot. ESL	4615	
	Precision	Inf. Gain
PRODUCT	88.57%	0.12

Table 3 - Experiment with a small "New Product" Corpus

Here, the prior probability of Products is obviously higher, though -due to the poor gazetteer- there is an elevated number of unrecognized products.
In this corpus we selected and then removed 35 product names, and now the

system correctly classifies 31. Notice that in this experiment the gazetteer and the PN grammar are the same as before, The only difference is that the corpus provides more evidence (contexts) concerning those products that have been recognized as such. Notice on the other side, that the Information Gain now is very low.

4 Conclusions and Future Work

Our current implementation of a PN analyzer still has a limited performance, caused by a variety of problems that range from unsatisfactory performance of state-of-art POS taggers in inflected languages, to limited availability of linguistic resources,in Italian, such as PN gazetteers.

The algorithm that we propose has indeed the purpose of overcoming limitations of gazetteers and manually defined contextual rules for PN recognition. In (Cucchiarelli et al. 1998) we also show how to extend our method to incrementally update the initial gazzeteer.

The performance of the proposed algorithm is more than satisfactory. A comparison with existing systems is difficult because in the literature global PN recognition performances are reported, without considering the semantic classification of unknowns as a subtask. The only exception is in (Wacholder et al, 1997) where the reported performance for the sole semantic disambiguation task of PNs is 79%. In that paper, however, semantic disambiguation is performed among a lower number of classes[5].

The performance of our system is clearly affected by the dimension of the initial seed gazetteer and contextual rules. If the sets ESL_A and ESL_B are large enough, obviously more examples of similar contexts are found, even for unknown PNs with a single occurrence.

In our test experiment, we always managed to find at least one or two similar contexts of an unknown PN, but in some cases they were misleading and caused a wrong classification, especially for Products. However, it may be possible to increase the evidence provided by the set ESL_B by including contexts in which the words are

---

[5]One of the advantages of Information Gain is that, if widely adopted, this measure facilitates the comparison among learning methods with different complexity of the classification task.

not strictly synonyms, but belong to the same semantic category.

One such experiment requires a word taxonomy, like for example WordNet. WordNet is currently unavailable in Italian (the first known results of the EuroWordNet project are too preliminary), therefore we plan to reproduce our experiment in English.

Another strategy to improve performances in absence of a substantial evidence is the definition of general (not contextual) rules to capture unknown complex nominals. For example, looking at the Product experiment in more detail, we found that product names are often formed by very complex nominals, e.g. *Fiat- Marea Weekend 2000* (the name of a car model). Capturing complex nominals in absence of anchors and specific contextual rules (here the only anchor is *Fiat*, which appears in the gazetteer as an Organization name) may be difficult, and if a complex nominal is not captured as a unit, the resulting syntactic context may be misleading (e.g. N_ADJ(Fiat_Marea_Weekend, 2000)).

We believe that finding class-independent heuristics for capturing complex nominals is a more "general" way of improving the performance of the method, rather than adding specific rules for specific entity types and enriching the gazetteer.

Acknowledgments

References

Basili, R., Pazienza M.T., Velardi P. (1994) *A (not-so) shallow parser for collocational analysis*. Proc. of Coling '94, Kyoto, Japan, 1994.

Basili, R., Marziali A., Pazienza M.T. (1994b) *Modelling syntax uncertainty in lexical acquisition from texts*. Journal of Quantitative Linguistics, vol.1, n.1, 1994.

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

Brill, E (1995). *Transformation-based Error-Driven Learning and Natural Language Processing: A case study of Part of Speech Tag-*

ging. Computational Linguistics, vol. 21, n. 24, 1995

Cucchiarelli A., Luzi D., Velardi P. Using Corpus evidence for Automatic Gazetteer Extension in Proc. of first Language Resources and Evaluation, Granada, May 1988

ECRAN: *Extraction of Content: Research at Near Market.* http://www2.echo.lu/langeng/en/le1/ecran/ecran.html

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

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

Grishman R., Macleod C. and Meyers A. (1992) *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.

Kononenko I. and Bratko I. (1991) *Information-based Evaluation Criterion for Classifier's Performance.* Machine Learning 6, pp. 67-80, 1991

Mani I., McMillian R., Luperfoy S., Lusher E., Laskowski S. (1996) *Identifying Unknown Proper Names in Newswire Text.* in Corpus Processing for Lexical Acquisition, J. Pustejovsky and B. Boguraev Eds., MIT Press 1996.

McDonald D. (1996) *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.

Paik W., Liddy E., Yu E. and McKenna M. (1996) *Categorizing and standardizing proper nouns for effcient Information Retrieval.* in Corpus Processing for Lexical Acquisition, J. Pustejovsky and B. Boguraev Eds., MIT Press 1996.

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

Wacholder N., Ravin Y. and Choi M. (1997) *Disambiguation of Proper Names in Text.* in Proc. of 5th Conference on Applied natural Language Processing, Washington, 1997