Recognizing Named Entities using Automatically Extracted Transduction Rules

D. Nouvel, J.Y. Antoine, N. Friburger, A. Soulet

Université François Rabelais Tours
Laboratoire d'Informatique
Equipe BDTLN

Named Entity Recognition

- Named Entity Recognition (NER) task :
  - **Proper Nouns** : person, location, organization (movie, brand. . .)
  - **Definite Descriptions** : time expression, amount, function (. . .)

- Named Entities Recognition (NER) by :
  - **Detecting** / delimiting NEs (determining **frontiers**, **boundaries**)
  - **Categorizing** / classifying / assigning a type to detected NEs

  $\Rightarrow$ Finding **markers** as NEs boundaries

Example

The *<prod>* iPhone 4 *</prod>* was announced during the *<time>* 7th of june, 2010 *</time>* keynote by *<pers>* Steve Jobs *</pers>*, *<fonc>* chief executive officer *</fonc>* of the *<org>* Apple *</org>* company.

Outline

Context of work

- Main approaches of NER :
  - **Knowledge-based** systems (difficult to attain good recall)
  - **Machine learning** systems (generally not easy to customize)

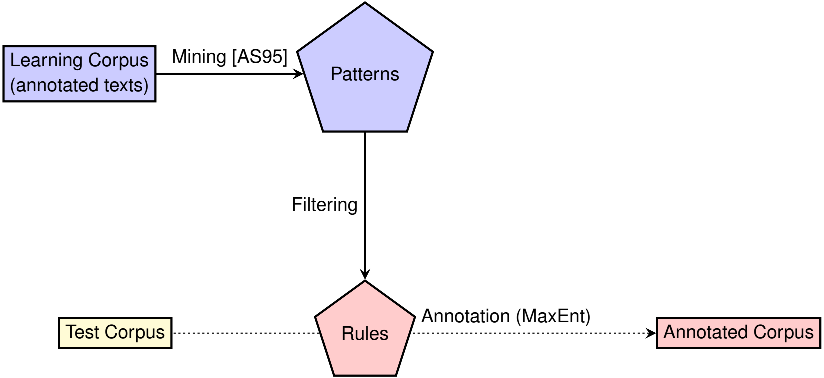
  ⇒ We try to find a common ground for combining / hybriding systems
- Existing system : **CasEN** [Fri06] (transducer / rule-based system)
- Available corpus : **Ester2** [GGC09], corpus of transcription of French radio broadcasts annotated in NEs :

Corpus	Tokens	Sentences	NEs
Ester2-corr	40 167	1 300	2 798
Ester2-held	48 143	1 683	3 074

TABLE: Characteristics of Ester2 corpora

⇒ **Our objective** : from Ester2 corpus (as train), **mine pattern** and find **informative rules** that may **enhance CasEN** for NER

Data Flow for NER Learning and Evaluating

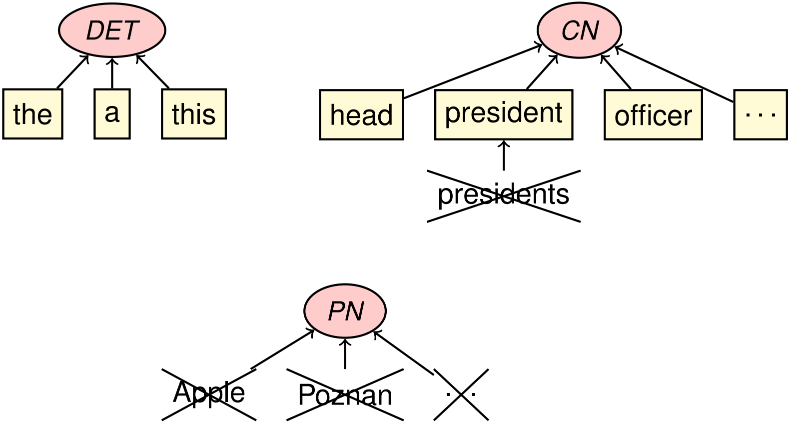


Outline

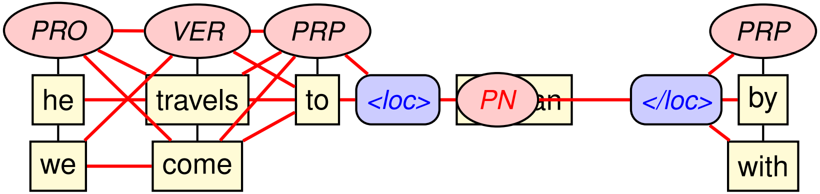
Extracting Patterns

- Finding **rules** that help **detecting** and **categorizing** simultaneously by determining **markers** of NEs
  - he *flies to* Poznan → he flies to *<loc>* Poznan *</loc>*
  - *president* Obama → president *<pers>* Obama *</pers>*
  - the *benefits* of Apple → the benefits of *<org>* Apple *</org>*
- **Preprocessings** : tokens, lemmas, POS-tagging (TreeTagger)
  - ⇒ Regular tokens : we only keep the lemma (generalized patterns)
  - ⇒ Proper Nouns (PN), we only keep POS (avoids overfitting)
- **Pattern Mining** considerations :
  - Exhaustively looking for patterns on pre-annotated corpus
  - Extracting and filtering patterns correlated to NEs markers
  - Apply patterns on unseen (test) corpus

Building hierarchy of items



From Corpus to Patterns : concrete example



Corpus pre-annotated sentence

- (. . .) As he *travels to* Poznan by plane, he thought (. . .)
- (. . .) , this time, we *come to* Barcelona with (. . .)

Extracted Patterns

-
-
-
-

Filtering Patterns as Informative Rules

Transduction Rule

- A **Transduction Rule** is a **morpho-syntactic pattern** (relying on the POS-tagging hierarchy) containing NEs markers for which are defined the standard parameters in pattern mining :
  - **Support** : number of occurrences in corpus
  - **Confidence** : in what proportion pattern appears with its markers

Informative Transduction Rule

- By exhaustively mining corpus, we obtain a very large set of rules

$\Rightarrow$ We need to **filter out** rules

$\Rightarrow$ For two rules which are generalization one of each other, we keep :

- The most **specific** one in terms of POS-tagging hierarchy
- The most **informative** according to markers

Outline

Probability model

- **Many rules** are triggered at a given position
- Define a random variable to define **probability of markers**

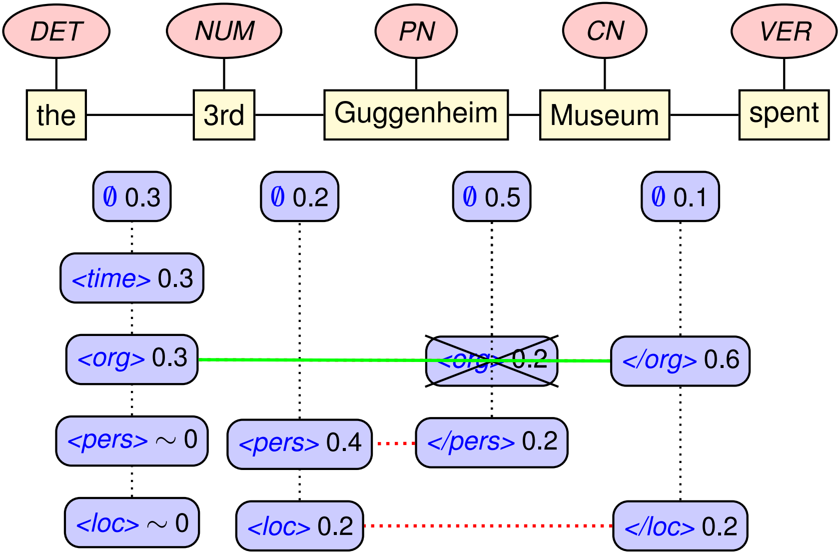
$$P(Mi = m_{j_i})$$

- Annotation probability for a sentence (assumption : markers are independant) :

$$P(M_1 = m_{j_1}, M_2 = m_{j_2}, \ldots, M_n = m_{j_n})$$
$$\approx \prod_{i=1 \ldots n} P(M_i = m_{j_i})$$

- Probability learned by **Maximum Entropy modeling**
- Use dynamic programming to search annotation (XML-like / flat)

Dynamic programming



Outline

Ester2 Corpus

Pattern extraction results over Ester2-Corr (40K tokens, 3K NEs)

Corpus	Sup.	Conf.	Rules	Inf. Rules	Gain
Ester2-corr	10	.5	2 270	1 119	**2.03**
	5	.5	28 047	3 673	**7.63**
	3	.3	458 875	12 653	**36.27**

TABLE: Extraction over Ester2 corpus at support and confidence thresholds

Interpretation

- Number of patterns is **very large** when support / confidence thresholds are lowered
- Filtering pattern is effective and alllows to keep a **reasonnable number of rules**

Predicting Markers

Actual markers	Predicted markers										
	tot	∅	<pers>	</pers>	<loc>	</loc>	<org>	</org>	<fonc>	</fonc>	rec.
∅	**27803**	**27168**	46	5	114	68	91	75	28	28	0.98
<pers>	**583**	86	**430**		20	1	26	1		18	0.74
</pers>	**592**	48		**470**		45		27			0.79
<loc>	**700**	162	20	2	**394**		114	1		2	0.56
</loc>	**698**	137	2	16	2	**407**		127			0.58
<org>	**448**	203	30		45		**157**		2	6	0.35
</org>	**443**	176		59		69		**122**		2	0.27
<fonc>	**225**	84	1	2	3		2		**129**		0.57
</fonc>	**219**	112	27	6		10		14		**48**	0.22
prec.		0.94	0.77	0.83	0.68	0.66	0.40	0.33	0.81	0.46	

TABLE: Confusion matrix between rule markers using a MaxEnt classifier

Interpretation

- Great ambiguity **org/pers** and **org/loc** (known problem)
- Beginning of a NE is not necessarily easier to find (cf pers, loc)

Predictions NEs

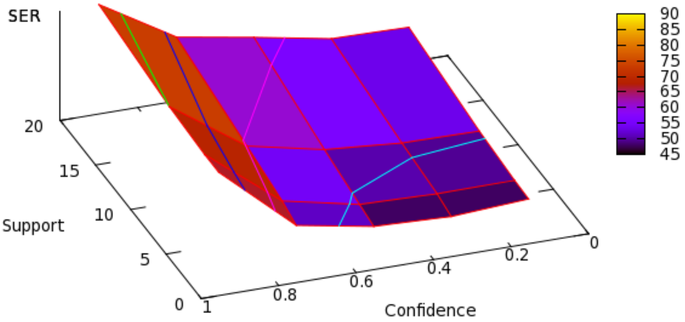


FIGURE: Evaluating (SER, to be minimized) NER annotations

Interpretation

- MaxEnt **accurately weights rules** (even less frequent/confident)

Hybriding Symbolic and Mining Systems

	Ins.	Del.	Typ.	Ext.	SER
Symbolic	43	348	171	257	29.0
fonc	0	-1	+1	0	28.8
loc	+4	-15	+3	+1	16.8
org	0	-13	+11	0	52.8
pers	+1	-20	0	+8	15.3
time	0	-2	0	0	24.6
total	+5	-51	+19	+8	-1.3
Coupled	48	297	190	265	27.7

TABLE: Using informative rules to enhance a symbolic system

Interpretation

- Coupling systems improves system with generic rules
  - *from <pers> PN PN*
  - *to <loc> PN*
  - *for <time> / years </time>* ("for a few years")

Outline

Conclusion

Contributions

- **Extracting** rules using a **morpho-syntactic** hierarchy
- Filtering **specific** and **informative** patterns as **rules**
- Using patterns to **annotate** a texte (Named Entities)
- Hybriding systems

Further investigations

- Better **filtering** patterns to be integrated in the **knowledge base** ?
- How to **enrich** patterns (syntax, semantics, anaphora)
- Assess performance with other models to **predict markers**
- Involved in NER task of project Etape (French National Research Agency, ANR)

Thank you

Rakesh Agrawal and Ramakrishnan Srikant.
Mining sequential patterns.
In *International Conference on Data Engineering (ICDE'95)*, pages 3–14, 1995.

Nathalie Friburger.
Linguistique et reconnaissance automatique des noms propres.
Meta : Translators' Journal, 51-4 :637–650, 2006.

Sylvain Galliano, Guillaume Gravier, and Laura Chaubard.
The ester 2 evaluation campaign for the rich transcription of french radio broadcasts.
In *10th Conference of the International Speech Communication Association (INTERSPEECH'2009)*, 2009.