

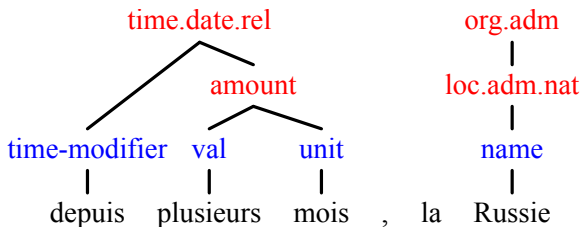
Entity Detection Task in ETAPE: the LIMSI participation

Marco Dinarelli, Eric Bilinski, Sophie Rosset

LIMSI/CNRS Orsay

Observations

Entities structure and approach



- Two annotations' levels : components and types
- Components are simple sequence labels on words
- Types are semantic trees

→ Use of CRF for labeling the components

→ then PCFG for semantic tree reconstruction

Observations

Entities structure and approach

ASR outputs

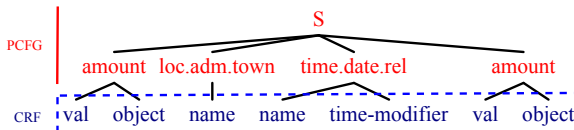
- Spoken data, ASR outputs
- Sometimes lack of cases
- Sometimes lack of punctuation marks

→ Normalization of the ASR outputs to do tokenization and capitalization

Our approach

Models cascade

- Conditional Random Fields (CRF) for tagging entity components (Lavergne et al., 2010)
- Probabilistic Context-Free Grammar (PCFG) and chart-parsing (Johnson, 1999) for entity trees



90 personnes *toujours présentes à* **Atambua** *c'est là qu'hier*
matin ont été tués **3 employés du haut commissariat des**
Nations unies aux réfugiés¹

1. 90 people still present in Atambua is where yesterday morning killed three employees of the United Nations High Commissioner for Refugees

CRF+PCFG Model

- CRF model (Lavergne et al., 2010) :

$$P(E_1^N | W_1^N) = \frac{1}{Z} \prod_{n=1}^N \exp \left(\sum_{m=1}^M \lambda_m \cdot h_m(e_{n-1}, e_n, w_{n-2}^{n+2}) \right) \quad (1)$$

- PCFG model (Johnson, 1999) :

$$P(\tau) = \prod_{X \rightarrow \alpha} P(X \rightarrow \alpha)^{C_{\tau}(X \rightarrow \alpha)} \quad (2)$$

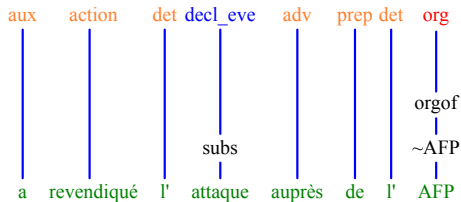
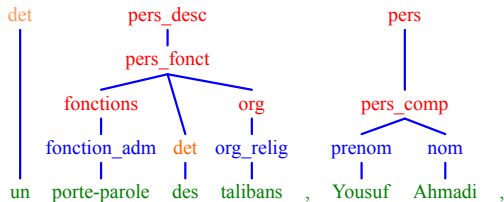
- CRF + PCFG : $P(E_1^N | W_1^N) \cdot P(\tau)$
 - E_1^N are entity components, i.e. leaves of τ
 - $P(E_1^N | W_1^N)$ encodes many lexical features : prefixes, suffixes, capitalization, morpho-syntactic

Features for CRF models

Standard system (Dinarelli and Rosset, IJCNLP 2011)

- word prefixes and suffixes of length from 1 to 5
- yes/no features : does the word start with a capital letter ? does the word contain a non-alphanumeric characters ?...
- 4 POS features extracted from the POs tagger presented in (Allauzen and Bonneau-Maynard, 2008)
- 4 features extracted from the Ritel analyzer

Ritel analysis example



Improving component model

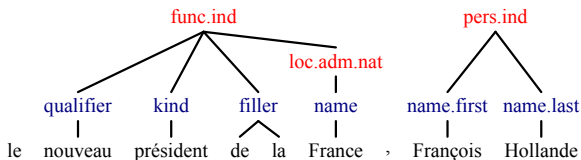
Other experiments

- The better is the component tagging, the better is the overall result
- Can we improve the component CRF model ? We tried
 - More simple features : results worsened
 - More complex features : results worsened
 - Incremental Features using a CRF types model as features, hypothesised component as features etc : result worsened
- Conclusion : we stayed with our tree contextualization model (see next section)

Tree representations (Dinarelli and Rosset, EACL 2012)

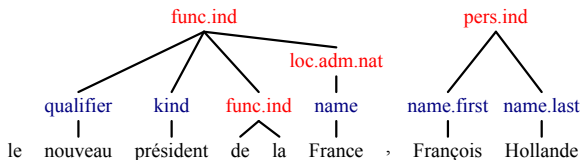
Default representation (**baseline**)

- English : *The new president of France , François Hollande*

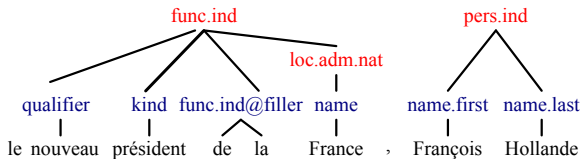


Tree contextualization (1/3)

Filler contextualization (1) (**filler-parent**)

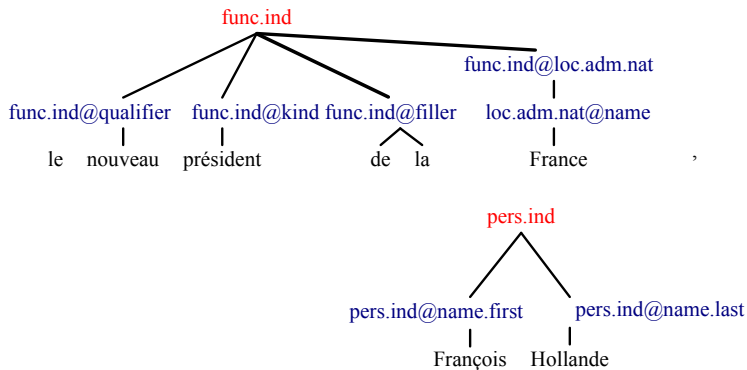


Filler contextualization (2) (**parent-context**)



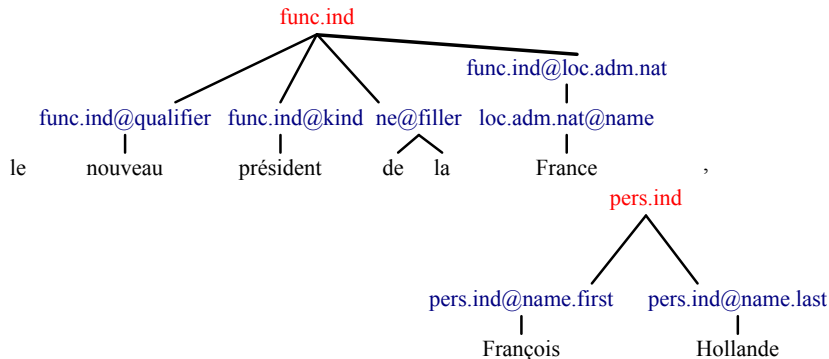
Tree contextualization (2/3)

Node contextualization (1) (parent-node)



Tree contextualization (3/3)

Node contextualization (2) (parent-node-filler)



Evaluation of different tree representations

- SER : Slot Error Rate (Makhoul et al., 1999)
- Evaluation on the Broadcast data Quaero corpus

	DEV		TEST	
Model	SER	F1	SER	F1
baseline	33.5%	0.725	33.4%	0.728
filler-parent	31.3%	0.744	33.4%	0.727
parent-context	30.9%	0.746	33.3%	0.728
parent-node	31.2%	0.778	31.4%	0.795
parent-node-filler	28.7%	0.789	30.2%	0.803

Table: Comparison of different tree representations

Handling ASR outputs

Normalization steps (Rosset et al., ASRU 2007)

- Tokenization (words separation)
- Punctuation separation
- Case reconstruction and punctuation addition : using a fully-cased, punctuated language model.
 - a word graph is built covering all the possible variants
 - a 4-gram language model is used to select the most probable hypothesis.
- Sentence splitting

De-normalization

- Annotation done on normalized data
- Alignment between normalized and non-normalized data
- Projection of annotation from normalized data to non-normalized data

Normalization example

Before

le rendez vous de bfm tv première chaîne info de
france c' est l' heure de bfm story nous sommes
...

After

le rendez vous de BFM TV première chaîne info de France c' est l' heure
de BFM story nous sommes ensemble ...

Final Thoughts

On our participation

- CRF models
 - Parent-node-filler model is too constrained, too rigid
 - CRF model does almost everything (every thing falls into components!)
- Time management :
 - Too much time trying to improve our models without any success. Same with Lexicalization. No time to do optimization (2-3% of SER).
- Results
 - Normalization : 12-14% gain of SER
 - Baseline system on BC Quaero data : 37.0% → ETAPE harder

On the ETAPE Entities Task

- Overlap speech is interesting. Nothing done. We should.
- Data more complex than for the Quaero program. Interesting.
- Data useful to study WER vs SER