



Lattice

(Laboratoire Langues, Textes, Traitements
Informatiques et Cognition)

Isabelle Tellier

Professor of computational linguistics

Paris 3 - Sorbonne Nouvelle

Research themes

(15 permanents, 3 engineers)

➤ **Linguistic studies**

- Expression of space and time in languages
- Diachronic studies (Old French)
- Discourse structure

➤ **NLP-oriented studies**

- Supervised machine learning for NLP
- POS tagging
- Name Entity Recognition, structured NE, entity linking...
- Model adaptation for multilingual/multimodal/multidomain texts
- Recognition of co-reference chains
- Syntax learning, modelisation of natural learning/evolution

Researchers on NLP

- Thierry Poibeau, DR CNRS, director of the lab : Information Extraction, subcategorisation frames, summarisation, texts and social networks...
- Isabelle Tellier, Pr : supervised (symbolic/statistical) machine learning for NLP, resource integration
- Marco Dinarelli, CR CNRS : machine learning for NLP (CRF, neural networks), NE
- Others : Frederic Landragin (co-reference chains), Jeanne-Marie Debaisieux (syntax of oral), Benjamin Fagard (modélisation of the evolution of languages)...
- 6 PhD students

On-going research projects on NLP

- ANR Orfeo (J-M. Debaisieux) : collect, annotation and interrogation of a large multimodal French corpus
- CNRS project Mastodons (Big Data) : efficient annotation (POS, NE) of large collections of texts
- AIM-West !
- Data Mining and Pattern extractions, Relations between entities (post-doc with Paris 13, labex EFL)
- Modelisation of the evolution of languages (Cifre with Sony)
- Structured Name Entities (Cifre with Temis)
- Model adaptation for POS tagging/NE (Cifre with Synthesio)
- Entity linking for digital humanities (PhD student)
- Semantic Indexing (PhD student with BNF)

What we don't do

- Speech recognition
- Machine translation

Our concern with MWE

- Development of a POS tagger for French combined with MWE recognition (Constant & Tellier 2012) :

<http://www.lattice.cnrs.fr/sites/itellier/SEM.html>

- Learned with CRFs with POS+ BIO annotations
- Combined with MWE lexicons

Quant	prep+b	PREP-B
à	prep_prep+l	PREP-I
la	cn_det_pro	DET
technique	cn	CN
...		
(concerning the technique....)		

- Recognition of indexing terms in social networks texts (thesis of Elisa Omodei with T. Poibeau)

Conclusion

- MWE: interesting for tagging, NE recognition, indexing...
- Machine learning orientation but compatible with external resources
- We are users of manual resources, producers of (machine-learned) models
- Multilingual interest (even if not for translation)
- internships, post-docs, PhDs... are welcome in Lattice!