Réunion PARSEME-FR

4 octobre 2016

Réunion PARSEME-FR

- Round table
- Administrative point
- Scientific point on WPs
- WP1
- Misc.

Round table

Administrative point

- Consortium
- Recruitment

Initial consortium

- Université Paris-Est Marne-la-Vallée, LIGM (coordinator)
- Aix Marseille Université, LIF
- INRIA, Alpage
- Université François Rabelais, LI
- Université d'Orléans, LIFO

Initial consortium

- Université Paris-Est Marne-la-Vallée, LIGM (coordinator)
- Aix Marseille Université, LIF
- INRIA, Alpage
- Université François Rabelais, LI
- Université d'Orléans, LIFO

Change

- Université Paris-Est Marne-la-Vallée $\rightarrow$ out of the consortium
- CNRS, ATILF $\rightarrow$ in consortium + new coordinator
- LIGM could stay under the supervision of CNRS (TBC)
- procedure in progress (but very advanced)

Consortium as of October 1st, 2016

TBC by ANR

- CNRS, ATILF (coordinateur) + LIGM (?)
- Aix Marseille Université, LIF
- INRIA, Alpage
- Université François Rabelais, LI
- Université d'Orléans, LIFO

Recruitments

Past recruitments

- LIGM : Manolo Iborra, 5-month internship $\rightarrow$ WP2
- LIF : Manon Scholivet, 4-month internship

Recruitments in progress

- ATILF : Hazem Al Saied $\rightarrow$ WP3+WP2 (statistical syntactic and semantic parsing), co-supervised by Alpage
- LI : Caroline Pasquer $\rightarrow$ WP2+WP5 (named entity), co-supervised by LIF

Coming recruitments

- LIF : Silvio Cordeiro (in January 2017), 6-month engineer $\rightarrow$ mwetoolkit

Scientific point

- WP1 : Marie and Mathieu (cf. specific point later)
- WP2 : Agata and Mathieu
- WP3 : Marie
- WP4 : Yannick and Éric
- WP5 : Carlos and Agata

WP1

- Annotation guide (quasi-operative)
- Pilot annotation (4th phase)
- Adjudication (today)
- Final annotation
- Lexicon of annotated MWEs (??)

WP2

Recall of objectives

- Extraction of unified lexicon, with "holes" (first version)
- "Filtering" of entries to be compliant with WP1 criteria (coming soon, starting with most frequent ones)
- Enrichment of lexicon (fill in the holes)
- Reference to Linked Open Data for Named Entities
- Lexicon projection on treebanks (also linked to WP3)
- Standard format

WP2

Manolo Iborra's internship

What has been done

- tool to automatically construct a lexicon from existing resources (lexicon-grammar tables, dictionaries DELAC and PROLEX)
- a start of harmonization (e.g. POS tags)
- First version of lexicon (XML and HTML)

A lexical entry

- lemma (raw and processed) + POS tag
- inflected entries with morphological information (DELAC)
- morphosyntactic structure of the lemma (raw, processed by rules, predicted by POS tagger)
- syntactic structures (constraints on constituents, syntactic trees) for LG tables

WP2

Exemple 1 : DELA (not exactly what is produced)

ID : 94008_carte_grise

Lemma : carte grise

Inflected form : carte grise,.N+NA :fs

- POS : Noun
- Structure : Noun Adj
- gender : feminine
- number : singular

Inflected form : cartes grises,carte grise.N+NA :fs

- POS : Noun
- Structure : Noun Adj
- gender : feminine
- number : plural

WP2

Exemple 2 : PROLEX

WP2

Exemple 3 : Lexicon-grammar

Court et moyen terme

Procédure d'extraction

- quelques bugs sur des structures peu fréquentes
- certaines informations à mieux formaliser
- ajout d'autres ressources existantes (ressources de l'ATILF ?)

Ajout d'informations distributionnelles (stats)

- projection (approximative) du lexique existant sur gros corpus
- calculer stats (fréquences, vecteurs distributionnels, degré de compositionalité)
- cf. travaux de Carlos and co (ACL 2016)
- base pour prioriser le travail de nettoyage

Court et moyen terme

Travail linguistique sur ressources existantes

- Stage prévu au LIGM sur les tables (nettoyage des tables et documentation sur propriétés avec règles de constitution des structures syntaxiques à partir des propriétés)
- Nettoyage du DELA (L. Danlos ?)

Few words on other work packages

- WP3 : statistical parsing (Marie)
- WP4 : symbolic parsing (Yannick and Eric)
- WP5 : MWE linking (Agata and Carlos)

WP1

Annotation

- Short update on NE annotation guide (Agata)
- Recall on last agreement scores (70-75% for MWEs, 80-85 on non typed NE)
- Adjudication (start then dispatch)
- Agenda

Misc.

- Publications
- Next PARSEME-FR meeting