

Expressions polylexicales et parsing

Proposition d'action pour la fédération ICVL



LIFO Orléans/Bourges
LI Tours/Blois



28 mai 2015

Expressions polylexicales (EP)

Multi-word expressions

The **prime time** speech by **first lady Michelle Obama** **set** the house **on fire**. She made **crystal clear** which issues she **took to heart**, but she was **preaching to the choir**.

Facts

- MWEs are prevalent (40% of text items),
- MWEs are inpredicable and thus challenge grammatical description,
- MWEs are notably semantically non-compositional,
- MWEs are under-represented in language resources and tools,
- MWEs are hard to detect, understand, translate, etc.

Objectifs

- Inscrits dans le réseau européen COST IC1207 **PARSEME**
 - développement de **ressources** d'EP,
 - prise en compte des EP dans le **parsing profond** (symbolique et probabiliste),
 - développement de méthodologie d'**annotation** d'EP dans des **corpus arborés**.
- Rapprochement du domaine de la **reconnaissance d'entités nommées** polylexicales à celui du **parsing** profond.
- Rapprochement du domaine de **Entity Linking** et **Linked Open Data** au **parsing** syntaxique et sémantique.
- Liens avec les autres actions sur le TAL:
 - Techniques d'**apprentissage** pour le TAL,
 - Analyse syntaxique (**parsing**) pour l'**annotation** de la temporalité.

Personnes réellement et potentiellement impliquées

Yannick Parmentier	CA	LIFO
Agata Savary	BdTIn	LI
Jakub Waszczuk	BdTIn	LI
Denis Maurel	BdTIn	LI
Jean-Yves Antoine	BdTIn	LI
Béatrice Bouchou-Markhoff	BdTIn	LI
Denys Duchier	CA	LIFO
Nicolas Labroche	BdTIn	LI
Nathalie Friburger	BdTIn	LI

Historique des collaborations

- Action COST IC1207 **PARSEME** (www.parseme.eu) - en cours depuis 2013 (YP, AS, JW).
- **Thèse** de doctorat de Jakub Waszczuk - co-encadrement YP/AS (initialement aussi DM); comité de pilotage avec DD et NL.
- Dépôts de projets :
 - **PARSEME-FR** - proposition de projet **ANR**, spin-off français de PARSEME, déposée en 2013 et 2014 (en cours d'évaluation),
 - **SHAGRALER** – proposition de projet européen **STREP** (appel FP7-ICT-2011-SME-DCL), soumise en 2011 (non sélectionnée),
 - **ASEF** (Analyse Syntaxique des Expressions Figées) – proposition de projet **régional** de recherche d'initiative académique (APR-IA 2012), soumise en 2012 (non sélectionnée).