

PROPOSITION ACTION POUR LA FEDERATION LI-LIFO

(VERSION 1)

1- Titre : Expressions polylexicales et parsing

2- Description succincte (thématiques et objectifs)

Les expressions polylexicales (e.g. *merle blanc*, *Banque centrale européenne*, *chérir l'espoir*, *prendre le taureau par les cornes*) sont l'un des défis majeur dans le traitement automatique des langues, et plus particulièrement pour le parsing syntaxique et sémantique, car elles possèdent des propriétés imprévisibles à différents niveau de traitement linguistique (lexique, morphologie, syntaxe, sémantique, pragmatique). Les objectifs de cette action de la fédération ICVL est inscrite pleinement dans le réseau européen COST IC1207 PARSEME, et vise la prise en compte des expressions polylexicales dans le parsing profond, dans un contexte multilingue. D'autre part, nous souhaitons établir des liens entre le parsing des expressions polylexicales et d'autres compétences présentes au sein des équipes de la fédération. Ainsi, deux sous-objectifs (exprimés notamment dans la proposition de projet ANR PARSEME-FR) concernent :

- le rapprochement du domaine de la reconnaissance d'entités nommées polylexicales à celui du parsing profond,
- le rapprochement du domaine de entity linking et Linked Open Data (incluant beaucoup d'entités polylexicales) au parsing syntaxique et sémantique.

3- Participants (personnes impliquées avec précision sur appartenance)

Yannick Parmentier (CA, LIFO)

Agata Savary (BdTln, LI)

Jakub Waszczuk (BdTln, LI)

Denis Maurel (BdTln, LI)

Jean-Yves Antoine (BdTln, LI)

Béatrice Bouchou-Markhoff (BdTln, LI)

Denys Duchier (CA, LIFO)

Nicolas Labroche (BdTln, LI)

Nathalie Friburger (BdTln, LI)

4- Historique des collaborations (s'il existe des collaborations passées et des résultats déjà obtenus)

- Action COST IC1207 PARSEME (www.parseme.eu) - en cours depuis 2013 (jusqu'à 2017), réseau de 180 membres de 30 pays ; Yannick Parmentier et Agata Savary sont à l'origine de cette action. Yannick, Jakub et Agata sont membres de son Steering Committee.
- Thèse de doctorat de Jakub Waszczuk - co-encadrée par Yannick Parmentier et Agata Savary (et initialement aussi par Denis Maurel). Participation de Denys Duchier et Nicolas Labroche au comité de pilotage de cette thèse.
- Trois dépôts de projets :
 - PARSEME-FR - proposition de projet ANR, le spin-off français de l'action

COST IC1207 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).

5- Proposition de travail (description plus détaillée des collaborations envisagées et des résultats attendus)

- Poursuite du co-encadrement de la thèse de Jakub, dédiée au parsing symbolique des expressions polylexicales en français (et polonais).
- Poursuite de la coordination du réseau COST IC1207 PARSEME (avec 3-4 réunions du réseau par an).
- Collaboration étroite avec d'autres chercheurs français, experts du parsing probabiliste et symbolique (laboratoires ALPAGE Paris, LIF Aix-Marseille, LIGM Paris-Est), si le projet ANR PARSEME-FR est sélectionné pour un financement.

6- Prospectives (Interactions possibles avec autre action ? Auriez vous des idées sur les thématiques d'un axe qui pourraient englober cette action ?)

- Interactions possibles avec d'autres actions : action TAL « parsing » et « apprentissage automatique »
- Axe : traitement des données hétérogènes (?)