

## Fiche de projet tutoré / Project form

### Etude de l'impact des indices syntaxiques sur l'identification automatique des connecteurs discursifs

#### Encadrement / Supervisors

1. équipe, laboratoire / team, lab

SYNALP / LORIA

1. encadrant·e principal·e (nom, email) / main supervisor (name, email)

Chloé Braud, chloe.braud@loria.fr

2. autres encadrant·es / other supervisors

#### Description / Description

1. projet global/global project

L'analyse discursive est un défi actuel majeur en TAL : il s'agit globalement de fournir une analyse d'ordre sémantique d'un document, en liant les phrases qui le composent par des relations comme Contraste, Explication, Comparaison etc. Ces relations dites discursives peuvent être lexicalisées : une expression, un connecteur discursif, marque la relation ("mais, parce que, alors que" etc). L'identification automatique de ces connecteurs est une première étape cruciale dans le développement de systèmes complets d'analyse discursive.

Le but du projet est de déterminer l'utilité de l'information syntaxique pour cette tâche, en répondant notamment aux questions suivantes : est-ce qu'une analyse morpho-syntaxique est suffisante ? Quel est l'impact de l'utilisation d'informations (morpho-)syntaxiques prédites vs gold ? Quel effet a l'utilisation d'une analyse (morpho-)syntaxique de grain plus ou moins fin (ex. Penn Treebank vs Universal Dependencies) ? Au final, on veut savoir si la détection de connecteurs est possible pour des langues et domaines où l'on ne dispose pas d'outils ou de ressources aussi riches que pour l'anglais. Des tests seront effectués sur l'anglais et le français.

2. biblio. UE 705 (semestre 7)

L'étude bibliographique portera sur les travaux existants sur la tâche pour l'anglais et le français.

3. réalisation. UE 805 (semestre 8)

Le projet nécessite le développement d'un système d'identification des connecteurs (classification binaire ou de séquence) pour l'anglais et le français, et l'utilisation d'outils de TAL (analyseurs syntaxiques).

#### Informations diverses : matériel nécessaire, contexte de réalisation /

Various information: material, context of realization

**Livrables et échéancier / Deliverable and schedule**

**Bibliographie /References (max. 4-5)**

*[il ne s'agit pas de la bibliographie complète qui sera fournie aux étudiants au début du projet mais d'une bibliographie indicative pour aider à cerner le sujet]*

Pitler, E., & Nenkova, A. (2009, August). Using syntax to disambiguate explicit discourse connectives in text. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*(pp. 13-16). Association for Computational Linguistics.

Laali, M., & Kosseim, L. (2017). Automatic Disambiguation of French Discourse Connectives. *arXiv preprint arXiv:1704.05162*.

Pour se faire une idée de la tâche complète :

Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Attapol T. Rutherford , Bonnie Webber, Chuan Wang, and Hongmin Wang (2016). The CoNLL-2016 shared task on multilingual shallow discourse parsing. *Proceedings of the 20th Conference on Computational Natural Language Learning: Shared Task (CoNLL Shared Task 2016)*