

Exploration de corpus

Infos pratiques

- > ECTS : 3.0
- > Nombre d'heures : 24.0
- > Période de l'année : Enseignement quatrième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Travaux dirigés
- > Ouvert aux étudiants en échange : Oui
- > Composante : Philo, Info-Comm, Langues, Littératures & Arts du spectacle
- > Code ELP : 4L4SC01P

Présentation

Le cours s'attachera à présenter deux méthodes d'exploration de corpus. La première s'appuiera sur la notion d'automate telle qu'elle est mobilisée dans le domaine de la linguistique informatique. Nous utiliserons le logiciel Unitex, dont l'interface graphique permet de constituer et de maintenir des automates complexes. Trois objectifs seront alors visés : 1) apprendre à construire des ressources linguistiques sous la forme d'automates qui permettent à leur tour l'annotation de textes (en particulier, une annotation sémantique) ; 2) se confronter à la complexité de la modélisation linguistique que cette démarche implique ; 3) tester et étudier sur des corpus diversifiés ces ressources et en tirer des conclusions quant à leurs performances.

La seconde méthode d'exploration de corpus s'appuiera sur l'utilisation du logiciel, TXM, fondé sur des méthodes statistiques. Quatre objectifs seront alors visés : 1) produire des concordances à partir de recherches de motifs lexicaux complexes construits à partir des propriétés des mots ; 2) construire différents tableaux de contingence croisant les mots, les textes et leurs structures ; 3) calculer la liste des mots apparaissant de façon préférentielle dans les mêmes contextes qu'un motif lexical complexe (cooccurents statistiques).

Objectifs

Il s'agit de donner à manipuler deux outils d'exploration de corpus (l'un symbolique, l'autre statistique) en offrant des repères méthodologiques.

Évaluation

M3C en deux sessions

REGIME STANDARD Session 1: CC 100%

Contrôle Continu : 2 notes au moins dont une évaluation en temps limité (50%)

REGIME DEROGATOIRE : CT 100%

1 note, évaluation écrite en temps limité - 1H

SESSION 2 (dite de 2nd chance) : CT 100%

1 note, évaluation écrite en temps limité - 1H

Pré-requis nécessaires

Aucun

Compétences visées

Savoir manipuler les outils d'exploration de corpus

Bibliographie

Habert, Benoît. Portrait de linguiste(s) à l'instrument. *Texto!* [en ligne], décembre 2005, vol. X, n°4

Marandin, Jean-Marie, Cori, Marcel. La linguistique au contact de l'informatique : de la construction des grammaires aux grammaires de construction. *Histoire Épistémologie Langage*, 23-1, 2001, pp. 49-79

Poudat, Céline, Landragin Frédéric. Explorer un corpus textuel. De Boeck 2017