

Outils d'analyse automatique de corpus



Présentation

DESCRIPTION

SL4BY020

Ce bloc concerne l'analyse automatique de textes: tokenisation, lemmatisation, identification des parties du discours (tagging) et des relations syntaxiques (parsing). On abordera les problématiques liées à l'annotation (choix de tokenisation, de lemmes, d'étiquettes...), l'utilisation de ressources existantes (modèles de langue) pour l'analyse automatique de texte, l'évaluation de l'analyse obtenue, et l'adaptation des ressources en fonction du type de texte.

Pour en savoir plus, rendez-vous sur > u-paris.fr/choisir-sa-formation