

Projet M2 TAL-IL
Extraction automatique des règles d'accord à partir de treebanks

Sylvain Kahane (Modyco, Université Paris Nanterre)
Bruno Guillaume (INRIA Grand Est)

L'objectif du projet est d'extraire automatiquement des règles d'accord, c'est-à-dire des motifs du type pattern { GOV -[relation]-> DEP ; GOV.feature = DEP.feature } (où relation est une relation syntaxique donnée et feature est un trait morphosyntaxique donné) des treebanks SUD. La difficulté est de décider quand on peut considérer la proportion de (GOV,DEP) pour lesquels GOV.feature = DEP.feature est suffisante pour considérer qu'il y a effectivement accord et éventuellement de déterminer des conditions qui déclenchent l'accord (par exemple, en français, l'accord avec du participe passé avec son objet n'a lieu que si DEP est un pronom antéposé à GOV).

On se basera sur le travail de Chaudhary et al. (2020), sur les treebanks SUD (<https://surfacesyntacticud.github.io/>), sur les outils Grew (<https://grew.fr/>) et Grew-match qui permettent de rechercher des motifs dans graphes et donc des treebanks.

Ce travail s'inscrit dans un projet plus large d'extraction de règles de grammaires à partir de treebanks.

Le livrable pourra s'inspirer du site développé par Chaudhary et al. (2020) : <https://neulab.github.io/lase/>

Aditi Chaudhary, Antonios Anastasopoulos, Adithya Pratapa, David R. Mortensen, Zaid Sheikh, Yulia Tsvetkov, Graham Neubig (2020) Automatic Extraction of Rules Governing Morphological Agreement, <https://arxiv.org/abs/2010.01160>