

TD 9 - NLTK, écriture de grammaires

NLTK est une bibliothèque de fonctions dédiées au traitement automatique du langage naturel au sens large, le but étant de couvrir toutes les manipulations informatiques sur les langues naturelles : calcul de fréquences de mots sur des corpus, apprentissage, construction de représentations syntaxiques et sémantiques. La suite fournit des modules et des corpus (corpus libres ou parties libres de certains corpus). Cette suite se base sur le langage de programmation Python. Parmi ses fonctionnalités, `nltk` permet d'écrire des grammaires de réécriture manipulant des arbres décorés. Le but de ce TP est d'écrire une grammaire minimale pour le français.

Ce TP se base sur le chapitre 10 du livre *Natural Language Processing with Python*, disponible en ligne à l'adresse suivante : <http://www.nltk.org/book>.

Mise en place

- À partir du site du cours
- Téléchargez la grammaire `gram0.fcfg`
- Téléchargez le script python `script0.py`
- Vérifiez que le script tourne et traite bien les phrases de bases qui s'y trouvent.

Extension

En appliquant les principes vus en cours, étendez la grammaire ci-dessus pour qu'elle couvre les phénomènes suivants (pour le moment les ambiguïtés ne doivent pas être traitées) :

- accord au sein du GN
- accord entre sujet et verbe
- verbes transitifs
- verbes obliques (*parle à Marie*)
- verbes ditransitifs (*donner un paquet à Marie*)
- négation (en *ne...pas*)
- adjectifs épithètes
- adjectifs attributs (du sujet)
- coordination de phrase
- adjectifs relationnels (*fier de jean*)