

## Analyse morphosyntaxique

### Exercice 1. Les catégories grammaticales.

La liste des catégories grammaticales n'est pas fixe. Leur nombre, leurs noms et ce qu'elles englobent varie notamment selon les langues, mais aussi, en ce qui concerne le traitement automatique, selon les systèmes qui effectuent l'étiquetage. Certains standards existent, mais sans faire consensus.

1. Comparez les listes d'étiquettes (*tag sets*) suivantes :
  - Le jeu d'étiquettes utilisé par TreeTagger pour le français :  
<http://www.ims.uni-stuttgart.de/~schmid/french-tagset.html>
  - Le jeu d'étiquettes anglaises du Brown Corpus :  
<http://www.comp.leeds.ac.uk/amalgam/tagsets/brown.html>
  - Le jeu d'étiquettes anglaises de la Penn TreeBank :  
<http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQP-HTMLDemo/PennTreebankTS.html>
2. Quels peuvent être les conséquences d'utiliser tel ou tel jeu d'étiquettes, que ce soit sur les performances de l'étiqueteur ou sur les traitements qui suivent l'étiquetage ?

### Exercice 2. Utilisation de Flemm.

Flemm est installé dans `/net/public/tal/Flemm`.

1. Lancez-le sur un texte préalablement étiqueté par TreeTagger :  
`perl flemm.pl --entree <fichier tree-taggé> --sortie <fichier sortie> --tagger treetagger`
2. Essayez également l'option `--format xml`.
3. Comparez la sortie de TreeTagger et celle de Flemm. Quels peuvent être les avantages d'utiliser les informations supplémentaires apportées par l'analyseur flexionnel (en recherche d'information, pour l'analyse syntaxique, ...)?

### Exercice 3. Désambiguïsation morphosyntaxique.

Écrire des règles de désambiguïsation sous GATE permettant d'analyser correctement les mots soulignés ou de construire des locutions dans les phrases suivantes :

- Le gouvernement donne le la de la réforme.
- C'est bien le moins que chaque nation puisse espérer.
- Le tableau qu'elles tracent est d'ailleurs sans complaisance.
- Le jugement a été rendu en dépit de la pression médiatique.
- Une partie de la population vit sous le seuil de pauvreté.
- Une bonne partie de la population vit sous le seuil de pauvreté.