

SL0720X - TreeTagger en ligne de commande

Franck Sajous (CLLE-ERSS) - sajous@univ-tlse2.fr
<http://w3.erss.univ-tlse2.fr/membre/fsajous/>

Les lignes de commande données ci-dessous sont des exemples. Adaptez les chemins, noms de fichiers (e.g. french.par) en fonction des répertoires dans lesquels sont installés vos données/programmes.

1 Tokenisation d'un texte non segmenté

```
c:\Cours\Outils\TreeTagger\cmd\tokenize.pl -f texteFrancais.txt > texteFrancaisTOK.txt
ou :
c:\Cours\Outils\TreeTagger\cmd\tokenize.pl -e texteAnglais.txt > texteAnglaisTOK.txt
```

2 Étiquetage d'un texte déjà segmenté

2.1 Obtenir de l'aide

Appel de TreeTagger dans argument :
c:\Cours\Outils\TreeTagger\bin\tree-tagger

2.2 Syntaxe générale

```
CHEMIN\TreeTagger\bin\tree-tagger CHEMIN\TreeTagger\lib\fichierParametreLangue [options]
fichierTokenise > fichierSortie
```

2.3 Exemples

Par défaut, TreeTagger ne produit que l'étiquette en sortie. Pour obtenir la forme du token et son lemme, il faut spécifier les options `-token` et `-lemma`.

```
c:\Cours\Outils\TreeTagger\bin\tree-tagger c:\Cours\Outils\TreeTagger\lib\french.par
-token -lemma texteFrancaisTOK.txt > fichierEtiquete.txt
```

ou :

```
c:\Cours\Outils\TreeTagger\bin\tree-tagger c:\Cours\Outils\TreeTagger\lib\english.par
-token -lemma texteAnglaisTOK.txt > fichierAnglais.txt
```

Utilisation d'un fichier auxiliaire :

```
c:\Cours\Outils\TreeTagger\bin\tree-tagger c:\Cours\Outils\TreeTagger\lib\french.par
-token -lemma -lex lexiqueAuxiliaire.txt texteFrancaisTOK.txt > fichierEtiquete.txt
```

Activation de l'heuristique sur la casse (majuscules/minuscules) : option `-cap-heuristics`

Pour la afficher les différentes étiquettes envisagées et les probabilités associées :

```
c:\Cours\Outils\TreeTagger\bin\tree-tagger c:\Cours\Outils\TreeTagger\lib\french.par
-token -lemma -proba -threshold 0.01 texteFrancaisTOK.txt > fichierEtiquete.txt
```