

TD – Langue et Informatique - Analyse morphologique automatique

P. Mercuriali – L2 – Centre Tesnière – CRIT – UMLP

Année 2025-2026

Objectifs pédagogiques

- Comprendre le principe de l'analyse morphologique
- Savoir utiliser quelques outils en ligne (gratuits)
- Savoir comparer les résultats de différents outils
- Savoir critiquer les outils de TAL

1 Introduction et rappels théoriques

Qu'est-ce que l'analyse morphologique ?

- Notion de lemme, lexème, forme fléchie
- Identification du lemme
- Identification des traits morphologiques (genre, nombre, temps, personne...)
- Catégories grammaticales (POS)
- Vocabulaire : affixes, préfixe, suffixe, inflection, racine, etc.
- Exemple d'analyse morphologique :
 - "analysaient" = analyser, verbe, imparfait, 3ème personne du pluriel
 - les DET, pl
 - étudiantes NOM, fém, pl
 - travaillaient VERBE, imparfait, 3pl

2 Outils disponibles (en ligne)

Choisissez au moins deux outils dans la liste suivante. (Attention aux langues disponibles)

- TreeTagger <http://corpora.lancs.ac.uk/tree-tagger/>
- UDPipe <https://lindat.mff.cuni.cz/services/udpipe/>
- Morpho http://corpora.lancs.ac.uk/vocab/analyse_morph.php
- Etymonline <https://www.etymonline.com/>
- GATE <https://cloud.gate.ac.uk/shopfront/displayItem/pos-tagging-and-morphological-analysis>

3 Activité pratique : tests comparés

Corpus (texte)

Les étudiantes et les étudiants analysaient attentivement les résultats obtenus.

Expérimentations avec les outils

1. Analysez le texte à la main.
 - Quels sont les lemmes ?
 - Les catégories grammaticales ?
 - Les informations morphologiques ?
2. Pour chaque outil (au moins deux) :
 - Analysez le texte
 - Observez les lemmes, catégories grammaticales, informations morphologiques.
 - L'analyse est-elle correcte ?

Vous pouvez remplir le tableau suivant pour mettre en forme vos résultats.

Exemple :

Mot	Outil	Lemme	Catégorie	Informations morphologiques	Exactitude
Students	GATE	student	NNS	affixe : s	correct

Comparaison des résultats

Mot	Outil 1	Outil 2	Problème repéré

Analyse critique des outils

1. Quel outil donne le plus d'informations morphologiques ?
2. Quel outil est le plus simple à utiliser ?
3. Quel outil est le plus correct ?
4. Y a-t-il des différences entre les lemmes ?

Analyse des problèmes

Les erreurs sont-elles graves pour :

1. la traduction ?
2. l'extraction terminologique (extraction de termes) ?
3. la génération de résumés automatiques ?

4 Synthèse

- Ces outils ne fonctionnent pas par LLMs
- Aucun outil n'est parfait
- Les programmeurs ont fait des choix qui reflètent des modèles linguistiques

5 Autre exemples plus extrêmes

Les poules du couvent couvent. (ambiguïté lexicale)

J'avais le cafard ce dimanche.

Ils ont forwardé des mails non-replyables. (emprunts, hybridations)

L'anti-anti-européen n'est pas anti-anti-anti-européen. (segmentation)

Les moyens moyens moyennent moyennement leurs moyens. (!)

Le mannequin enceinte défile. (incompatibilité ontologique)

Questions

- Problèmes linguistiques ou techniques ?
- Un traducteur automatique pourrait-il s'en sortir ?
- Un humain non natif comprendrait-il mieux qu'un outil ?
- Ces phrases sont-elles traduisibles mot à mot ?

Sources de confusion possibles

- nom propre/verbe
- adjectif/verbe
- absence de traits (temps, genre)
- dépendance excessive au contexte statistique
- difficultés avec jeux de mots, homographes, capitalisation