

L'Analyse Morphologique

T. Gennuso, 2018

1. Qu'est ce que l'analyse morphologique ?

L'analyse morphologique constitue une partie primordiale du traitement automatique des langues dans des domaines aussi divers que la reconnaissance de la parole, la traduction automatique ou la recherche d'informations. La présente synthèse s'intéresse aux problématiques qui y sont inhérentes. Sur un plan purement étymologique, on définit l'analyse morphologique comme l'étude de la forme des mots. On s'intéresse notamment à la manière dont ceux-ci sont construits à partir d'unités linguistiques minimales nommées « morphèmes ».

Dans l'histoire de la recherche dans ce domaine du traitement automatique des langues, on distingue deux grandes directions qui ont été prises :

- les recherches se basant sur des règles de traitements puis sur un système de règles automatiquement acquises.
- les recherches se basant sur des modèles probabilistes. Cette approche probabiliste est très souvent considérée comme plus robuste, plus facile à mettre en place et plus facile à optimiser. Ces méthodes utilisèrent dans un premier temps des modèles de Markov avant de s'intéresser à d'autres types de structures comme les arbres de décision. Plus récemment, d'autres structures furent utilisés dans les recherches sur modèles probabilistes, on peut notamment citer les divers travaux sur les réseaux neuronaux comme ou plus récemment des travaux sur les réseaux de neurones convolutifs.

L'analyse morphologique est un terme ultra-généraliste. Il regroupe en effet un certain nombre de méthodes, modèles, traitements, analyses qui permettent d'analyser et de décomposer les mots pour en extraire leur utilité dans la phrase ou bien encore leur sens dans cette dernière.

2. Définitions majeures de l'analyse morphologique :

L'étiquetage Morphologique : L'étiquetage Morphologique ou morphological tagging permet d'apporter plus d'informations que l'étiquetage Morphosyntaxique sur un mot en particulier. Il peut par exemple apporter le genre, le cas, la personne, ou encore le temps. Cet étiquetage est très important pour les langues morphologiquement riches.

L'étiquetage Morphosyntaxique : En linguistique, l'étiquetage morpho-syntaxique ou POS tagging (part-of-speech tagging en anglais) est le processus qui consiste à associer aux mots d'un texte les informations grammaticales correspondantes comme le fait qu'un mot soit un verbe, un nom commun, une entité nommée, etc.

Exemple d'étiquetage Morphosyntaxique :



Morphème : Un Morphème est le plus petit élément significatif, isolé par segmentation d'un mot, le plus souvent dépourvu d'autonomie linguistique.

Exemple : le mot « lecteurs » est composé de trois morphèmes :

- lect : « lire »
- eur : « celui qui »
- s : marque du pluriel

Lemmatisation : Principe qui vise à analyser les mots en les ramenant à leur forme infinitive dans le cas d'entités verbales ou au masculin singulier pour les autres mots.

Exemple de lemmatisation :

- les mots "joue", "jouerait", "jouions" sont reliés à la forme "jouer"
- les mots "joueuses", "joueurs", "joueuse" sont ramenés au mot "joueur"

Racinisation : C'est un procédé de transformation des flexions en leur radical ou racine. Il ne faut pas confondre ce terme avec celui de lemmatisation. En effet, La lemmatisation a pour objectif de retrouver le lemme d'un mot, par exemple l'infinitif pour les verbes. La racinisation consiste à supprimer la fin des mots, ce qui peut résulter en un mot qui n'existe pas dans la langue.

Exemple de racinisation : "joueurs" aura pour racine "jou"