

## Installation:

- Installer treetaggger : <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>
- Installation de virtualenv et installation des packages
  - `./install_venv.sh`
- Installer le project en mode editable :
  - `source venv/bin/activate` (Linux, macOS) or `venv/Scripts/activate` (Win)
  - `pip3 install -e .`

## Configuration

- Configurer treetaggerwrapper : dans le fichier `tag_util` mettre le chemin du dossier Treeetagger (celui où `bin`, `lib` and `cm` sont des sous-répertoires) dans la variable `dir_tree_tagger`

## Utilisation

- Préparer le corpus en utilisant le script `pre_process.py` soit en utilisant Treeetagger pour éliminant certaines catégories de mots et lemmatiser ou bien en enlevant les mots de la liste des mots peu informatifs.
- Construire le modèle avec le script `w2v_model.py`
- `Tsne_reduction.py` réduit les dimensionnalités du modèle à 2 dimensions ensuite `tsne_plot.py` permet de visualiser le modèle réduit.
- `Elbow.py` permet de retrouver le nombre de classe optimal
- `K_means.py` et `h_cluster` permettent de générer les classes de mots à partir du modèle w2v.