Etat de l’art

<http://blog.fouadhamdi.com/introduction-a-nltk/>

tokenisation permet de découper le texte en groupes, puis en mots et les mots inutiles sont supprimés (principalement les articles et mots de liaison).

Classification : permet de labéliser les tokens

Chunking : créer des groupes dans la phrase (avant tokenisation)

Clustering : regroupe des items

Corpus : lis les données texte

Featuring : feature structure est un dictionnaire. Le featuring permet de travailler sur des groupes de mots.

Grammaring : permet de trouver les possibles structures syntaxiques d’une phrase.

Parsing : permet de trouver les structures syntaxiques du texte.

Sementic :