

Module Textométrie avec R

Durée : 10h, lieu : ENS de Lyon

Ce module montre comment **réaliser l'analyse quantitative d'un corpus de textes sous R**. Le cours décomposera le processus en **plusieurs étapes méthodologiques**, notamment :

- **lecture / mise en forme d'un corpus** (lecture d'un texte et automatisation du processus pour création d'un objet R)
- **nettoyage d'un corpus via l'usage d'expressions régulières** (pour repérer certains patterns : par exemple repérer des noms propres, des nombres, des urls, etc. et éventuellement les retirer ou les corriger)
- **tokénisation du corpus** (pour décomposer les textes en unités pertinentes pour l'analyse - mot par exemple-).
- **filtres, tris et jointures** (pour lemmatiser, et retirer les mots outils n'ayant pas de réelle fonction sémantique pour la suite de l'analyse)
- **statistiques descriptives** (catégorisations, comptages)
- **analyses de contrastes** (test du Chi-2, spécificités)
- **analyses factorielles** (AFC / ACM)
- **réalisation de graphiques** (valorisation des résultats en termes de statistiques descriptives, contrastes et analyses factorielles)

Les étudiants mettront en œuvre ces méthodes sur un **corpus de textes correspondant à l'ensemble des communiqué de presse mis en ligne par le Ministère de la Transition Ecologique** (<https://www.ecologie.gouv.fr/presse>), à travers un ensemble d'**exercices interactifs sous R**.