

Analyser des questions ouvertes au moyen de la statistique textuelle

Bénédicte Garnier (Ined)

(garnier@ined.fr)

La statistique textuelle se situe à la croisée de plusieurs disciplines : la statistique, la linguistique, l'analyse du discours, l'informatique et le traitement des enquêtes (Lebart, Salem, 1994).

La méthode s'applique à des corpus de textes de nature variée : entretiens, réponses à des questions ouvertes, écrits historiques, littéraires, corpus de presse, titres d'articles, etc... La taille des textes peut varier : de quelques mots à quelques phrases. L'intérêt majeur de ces méthodes est de traiter les corpus tels qu'ils ont été écrits ou recueillis.

L'objectif de cet atelier de 3h est d'acquérir les clés pour explorer des réponses à une question ouverte, repérer des univers lexicaux et des spécificités au moyen de la statistique textuelle. Les participant.e.s pourront mettre en relation les textes avec des métadonnées comme les caractéristiques sociodémographiques sur les répondants (âge, sexe, niveau d'études, etc.).

Il est nécessaire de posséder un ordinateur et d'installer R et Rstudio (disponibles gratuitement) ainsi que les package dédiés : R.temis, Quanteda et Rainette.

Une connaissance de R et de l'environnement Rstudio n'est pas indispensable.

Programme

- Place de la statistique textuelle dans les data science
- Traiter des données textuelles (plus ou moins structurées) : tableaux lexicaux et lemmatisation
- Explorer : lexicométrie, analyses factorielles ou classifications
- Interpréter les résultats : concordances, spécificités, contributions
- Visualiser les résultats : nuages ou graphes de mots, plans factoriels, dendrogrammes

Bénédicte Garnier travaille au service méthodes statistiques de l'Ined.

Elle pratique et enseigne la méthode à l'Ined mais aussi en master à l'Université.

Elle contribue au [package R.temis](#) (R Text Mining Solution).

Garnier B., Guérin-Pace F., 2010 - [Appliquer les méthodes de la statistique textuelle](#), Ceped, les clefs pour, Paris