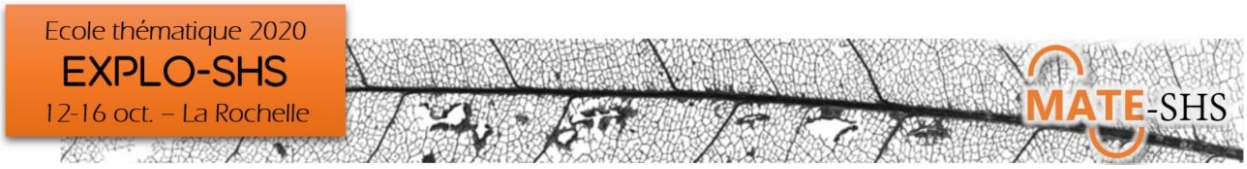

Comment explorer des corpus de textes

Bénédicte Garnier, Lucie Loubère, Gaëlle Delétraz

Atelier, mardi 13 octobre 9h30–12h15 et 13h45-16h30 (atelier répété 2 fois)

Présentation générale

L'exploration de textes avec un logiciel de statistique textuelle permet d'en déceler des structures, de produire des graphiques synthétiques ou d'extraire des verbatims.

Cette « fouille » de textes quasi automatique et très rapide nécessite de structurer les données disponibles en amont et de comprendre les méthodologies sous-jacentes issues de la linguistique (comme la lemmatisation) ou de la statistique multidimensionnelle (comme la classification automatique).

Lors de cet atelier nous illustrerons les méthodes et les appliquerons sur un corpus de réponses à une question ouverte dans une interface gratuite simple d'utilisation et qui fait référence : IraMuTeQ.

Environnement informatique

Utilisation du logiciel libre IRaMuTeQ sous Environnement PC ou MAC : http://www.iramuteq.org/

Il sera demandé aux stagiaires d'installer R et IRaMuTeQ avant l'ET. Un tutoriel d'installation et un fichier d'essai seront fournis en amont.

Type de données traitées - droit d'accès

Réponses aux questions ouvertes du questionnaire sur les attentes des participants de l'ET. Données anonymes et associées aux caractéristiques des répondants (sexe, statut, discipline, …).

Niveau requis

Aucun – Niveau initiation- pas besoin de savoir programmer.

Objectifs

- S'approprier le protocole scientifique d'exploration de données textuelles dans un outil dédié : IraMuTeQ
- Utiliser les méthodes et techniques permettant de détecter des structures dans de grands volumes de textes et les restituer
- Expliquer la méthodologie embarquée des procédures exploratoires intégrées dans l'outil : lemmatisation, classifications automatiques, extraction de verbatims (concordances)
- Interpréter et présenter les résultats.

Formule pédagogique

Atelier de 2h45 structuré en 3 temps :

- Exposé sur la méthodologie embarquée
- Présentation du logiciel et d'exemples de résultats
- Manipulation des participants sur un corpus fourni (en groupe de 2 ou 3 si les conditions le permettent) et retour d'expérience, trucs et astuce, …

Références bibliographiques

- Baril, Élodie, et France Guérin-Pace. 2016. Compétences à l'écrit des adultes et événements marquants de l'enfance : le traitement de l'enquête Information et vie quotidienne à l'aide des méthodes de la statistique textuelle. Économie et Statistique 490 (1): 17-36. https://doi.org/10.3406/estat.2016.10719.
- Baudelot, Christian, et Gollac, Michel. 1997. Faut-il travailler pour être heureux ? Insee Première, no 560. https://www.epsilon.insee.fr/jspui/bitstream/1/759/1/ip560.pdf.
- Benzécri, Jean-Paul. 1984. Description des textes et analyse documentaire. Cahiers de l'analyse des données Tome 9 (2): 205-11.
- Bonvalet, Catherine, Maison, Dominique, et Ortalda, Laurent. s. d. La place des univers familiaux, résidentiels et professionnels dans la structure du discours - Analyse textuelle des entretiens de “Proches et parents”. In La famille et ses proches, Ined-Puf diffusion. Travaux et documents 143. Ined.
- Demazière, Didier, Claire Brossaud, Patrick Trabal, et Karl Van Meter. s. d. Analyses textuelles en sociologie. Logiciels, méthodes, usages Presses universitaires de Rennes, Rennes (2006). 224 p.
- Garnier, Bénédicte, et France Guérin-Pace. 2010. Appliquer les méthodes de la statistique textuelle. Paris: CEPED. https://www.ceped.org/fr/publications-ressources/editions-du-ceped-1988-2011/les-clefs-pour/article/appliquer-les-methodes-de-la. Guérin-Pace, France. 1997. La statistique textuelle : un outil exploratoire en sciences sociales. In Population, revue de l'INED, 865-87. 4. https://www.persee.fr/doc/pop*0032-4663*1997*num*5*246*471.
- Guérin-Pace, France, Thérèse Saint Julien, et Anita W. Lau-Bignon. 2012. Une analyse lexicale des titres et mots-clés de 1972 à 2010. Espace géographique 41 (1): 4. https://doi.org/10.3917/eg.411.0004. http://lexicometrica.univ-paris3.fr/. s. d.
- Lebart, Ludovic, Bénédicte Pincemin, et Céline Poudat. 2019. Analyse des données textuelles. Mesure et évaluation 11. Québec: Presses de l'Université du Québec.
- Lebart, Ludovic, et André Salem. 1994. Statistique textuelle. Paris: Dunod. http://lexicometrica.univ-paris3.fr/livre/st94/st94-tdm.html.
- Reinert, Max. 1983. Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte. Cahiers de l'analyse des données 8 (2): 187-98.
- Reinert, Max. 1993. Les "mondes lexicaux" et leur "logique" à travers l'analyse statistique d'un corpus de récits de cauchemars. Langage & société 66 (1): 5-39. https://doi.org/10.3406/lsoc.1993.2632.
- Reinert, Max. 2001. Alceste, une méthode statistique et sémiotique d'analyse de discours ; application aux ‘Rêveries du promeneur solitaire'. La revue française de psychiatrie et de psychologie médicale 49: 32-36.
- Reinert, Max. 2008. Mondes lexicaux stabilisés et analyse statistique de discours.