

Avant-propos

Que peut-on apprendre de la distribution des mots dans un texte ? Cette question traduit l'une des préoccupations centrales de la statistique textuelle : tirer de l'observation quantitative des mots (observation de leur présence, leur absence, leurs co-occurrences ou leur répartition au fil des textes et des corpus) des enseignements permettant de mieux connaître le ou les texte(s) qu'ils composent ou plus largement les langues auxquelles ils appartiennent.

La statistique des données textuelles intéresse donc des disciplines scientifiques variées qui prennent le texte comme objet d'étude, par exemple dans le cadre de l'analyse littéraire, de la stylométrie, de l'analyse du discours politique, de la linguistique de corpus, de la classification de documents, de la recherche d'information ou encore de la fouille de texte ? Comme l'illustrent les nombreux articles de cet ouvrage, la statistique textuelle, stimulée non seulement par la croissance rapide du volume de textes accessibles sur support numérique, mais aussi par de nouveaux instruments et des méthodes innovantes, comme le recours à l'Intelligence Artificielle, offre également de puissants outils pour faire face aux nouveaux besoins et défis de la société de l'information.

Les Journées internationales d'analyse statistique des données textuelles (JADT) permettent à des chercheurs travaillant dans les différents domaines concernés par les traitements automatiques et statistiques des données textuelles de se rencontrer tous les deux ans afin de confronter leurs expériences. Après 16 éditions, organisées à Barcelone (1990), Montpellier (1993 et 2020), Rome (1995, 2010 et 2018), Nice (1998 et 2016), Lausanne (2000), Saint-Malo (2002), Louvain-la-Neuve (2004), Besançon (2006), Lyon (2008), à Liège (2012), à Paris (2014), Naples (2022), les JADT 2024, organisées à Bruxelles par le SeSLa (Séminaire des Sciences du Langage de l'UCLouvain – Site Saint-Louis), en collaboration avec le LASLA (Laboratoire d'Analyse statistique des Langues anciennes de l'Université de Liège), ont réuni du 25 au 27 juin statisticiens, linguistes, sociologues, historiens, informaticiens, spécialistes d'analyse du discours, de lexicographie et d'exploration de données textuelles. Vingt ans après la publication des JADT 2004 dans cette même collection, ce nouveau volume des *Cahiers du Cental* regroupe en deux tomes les travaux présentés à cette occasion et offre ainsi un panorama représentatif des tendances de la recherche dans un domaine en permanente évolution.

Le premier tome s'ouvre sur une contribution de Damon Mayaffre (Université Côte d'Azur / CNRS) qui questionne l'identité et la nature même de l'analyse des données textuelles en la définissant comme une herméneutique interdisciplinaire. Cette interdisciplinarité propre à

l'ADT se marque ensuite très clairement dans la multiplicité des domaines auxquels sont appliqués les modèles et les outils statistiques au fil de cet ouvrage :

- Textométrie, statistique textuelle
- Analyse exploratoire de données textuelles
- Linguistique de corpus et linguistique quantitative
- Traitement automatique du langage naturel
- Étiquetage, lemmatisation, enrichissement linguistique
- Analyse statistique des données structurées et non structurées
- Classification de textes, cartographie lexicale et textuelle
- Recherche documentaire, recherche d'informations
- Fouille de textes, fouille du web, Web sémantique
- Stylométrie, analyse du discours
- Logiciels pour l'analyse textuelle
- Apprentissage automatique pour l'analyse des données textuelles
- Corpus multilingues et corpus parallèles

La publication de cet ouvrage n'aurait pas été possible sans l'appui de plusieurs institutions, à savoir l'UCLouvain - Saint-Louis - Bruxelles, l'Université de Liège et le FNRS (Fonds de la Recherche Scientifique). Les articles qui composent les deux tomes de l'ouvrage ont été soigneusement relus par les membres du Comité scientifique. La mise en page du volume a été réalisée par Thomas Baruh (SeSLa), Valérie Thon et Pascal Tonnaer (LASLA), la couverture par Cédric Fairon (Cental). Leur contribution a été essentielle à la réalisation de l'ouvrage. Puissent toutes les personnes qui y ont contribué trouver ici l'expression de nos chaleureux remerciements.

Anne Dister – Dominique Longrée