

## RESPONSABLES D'ACTION

### Marion Bendinelli

Maître de conférences en sciences du langage, laboratoire Édition, langages, littératures, informatique, arts, didactiques, discours (EA 4161), UFC

### Margareta Kastberg Sjöblom

Professeure en sciences du langage, laboratoire Édition, langages, littératures, informatique, arts, didactiques, discours (EA 4161), UFC

## FINANCEMENT

Projet ANR Textométrie

**DURÉE DE L'ACTION** 2017-2020 (prolongement de l'action 2012-2016)



## RÉSUMÉ

L'action a pour objet de maintenir et développer, sous l'égide de la MSHE, le carrefour méthodologique et épistémologique de la textométrie et de diverses branches de la linguistique générale et comparée : linguistique textuelle, analyse de discours, linguistique de corpus, linguistique informatique. Du dialogue de ces communautés, qui s'ignorent encore partiellement, dépendent en effet la mise à l'épreuve et l'actualisation des concepts et des outils méthodologiques et conceptuels de chacune et la qualité de la contribution des analyses assistées sur corpus aux sciences humaines.

Elle pilote l'École CNRS MISAT (Méthodes informatiques et statistiques en analyse de textes) et organise ses propres événements scientifiques internationaux (e.g. colloque Cooccurrence en 2012). Elle constitue le support institutionnel des partenariats conclus au titre de son programme et garantit notre expertise face aux appels d'offre pour les contrats de recherche.

## OBJECTIFS DE L'ACTION

A travers le pôle « Archive bases corpus » de la MSHE Ledoux CITELA demeure un lieu identifié internationalement comme un carrefour et un creuset importants de la statistique linguistique et lexicale, rassemblée aujourd'hui sous la bannière de la textométrie, et de diverses branches de la linguistique générale et comparée : linguistique textuelle, analyse de discours, linguistique de corpus, linguistique informatique. La rencontre des linguistes, des

spécialistes de sciences sociales, et des statisticiens sur des objets et problèmes communs a permis et permettra la mise à l'épreuve et le perfectionnement des concepts et des outils de chacun. Le but de l'École MISAT est de permettre une réflexion sur les corpus textuels numériques et la maîtrise d'outils informatiques et statistiques pour leur traitement. Le public concerné doit assimiler les protocoles méthodologiques pour objectiver les parcours de lecture dans les textes ; épistémologiquement, il doit réfléchir à l'objet « corpus textuel numérique » ; techniquement, il doit apprendre à manier les logiciels de lexicométrie / textométrie / logométrie qui sont présentés en détail.

L'action a pour objet de maintenir et développer cette activité de carrefour aussi bien méthodologique qu'épistémologique. Elle pilote pour le compte du CNRS l'École thématique MISAT (et sa 4<sup>e</sup> session en 2016 à Besançon) – école ouverte à tous les chercheurs en sciences humaines soucieux de se perfectionner en analyse assistée de corpus, et bénéficiant du concours formateur des spécialistes reconnus. Elle est le support institutionnel des partenariats conclus au titre de son programme.

L'action couvre l'ensemble des participations de ses membres (spécialistes en textométrie et informatique textuelle) aux événements scientifiques connexes (Journées d'analyse des données textuelles JADT, Journées de linguistique de corpus JLC, ATALA, colloques de linguistique textuelle et discursive, etc.). Elle organise ses propres événements scientifiques (colloque Cooccurrence en 2012 et différentes journées d'étude). C'est elle aussi qui garantit notre expertise en matière d'analyse de données lors de la réponse à des appels d'offre pour contrats de recherche.

## **PROGRAMME DE TRAVAIL**

---

Publication d'un Vocabulaire de statistique appliquée aux textes et à l'analyse de discours en coopération avec le CEDITEC – Upec.

Poursuite des coopérations entreprises dans le cadre du projet ANR Textométrie 2007-2010 (partenaire de 1<sup>er</sup> rang).

Création du cadre textométrique pour l'édition électronique de plusieurs corpus de FANUM.

Exploitation textométrique du corpus PRESNAT.

Participations aux Journées internationales d'analyse statistique des données textuelles (2018 Rome).

Participation aux Journées de la linguistique de corpus (2017 Université de Grenoble).

## **RÉSULTATS ATTENDUS**

---

Visibilité des laboratoires franc-comtois en matière d'expertise statistique et de coopération interdisciplinaire en matière d'exploration statistique de corpus de sciences humaines.

Modernisation des méthodes et concepts de l'analyse statistique de corpus.

Perfectionnement de l'adéquation des environnements d'analyse et d'exploration.

Renforcement et élargissement des partenariats autour de ces thématiques.

Capacité de réponse à de nombreux besoins de secteurs de recherche avec lesquels les diverses composantes du pôle et de la recherche UFC entretiennent des rapports interdisciplinaires.

## **CRITÈRES DE L'ACTION**

---

- Inter- et pluridisciplinarité : Sciences du langage, informatique, statistique, applications en sciences humaines et sociales.

- International : réseaux internationaux JADT, JLC, ATALA.

- Interinstitutionnalité : université - CNRS, UQAM (centre ATO), Éditions scientifiques (Champion), Édition électronique (Champion-Electronique)

- Éléments d'innovation : Visualisations et cartographie statistique, approfondissement de concepts récents, ergonomie.

## **PARTENAIRES**

---

- Institutionnels : Éditions scientifiques Champion-Slatkine et Champion-Electronique.

- Scientifiques :

- ENS de Lyon, IHRIM, UMR5317
- Centre de Linguistique et des Sciences du Langage (CLSL, Lausanne)
- CEDITEC, Centre d'étude des discours, images, textes, écrits, communication, Université Paris XII - Upec
- Centre ATO de l'UQAM
- TGE ADONIS (CRN MEET et CRN TL)
- UMR 6039 Bases Corpus Langages, université Nice-Sophia Antipolis

• SYLED, SYstèmes Linguistiques, Énonciation et Discours (EA2290), université de la Sorbonne nouvelle -Paris 3

• LASLA, Laboratoire d'Analyse Statistique des Langues Anciennes, Université de Liège