

Journée d'étude

« L'approche textométrique de corpus textuels : pratiques & méthodes »

Université de Neuchâtel (Espace Tilo-Frey, salle R.O 12)

Mardi 30 Novembre 2021

Pour des raisons sanitaires, si vous souhaitez assister à cette journée d'étude, merci d'annoncer votre venue par mail à l'adresse suivante : cyrielle.montrichard@unine.ch

Programme

9h15 : Accueil des participants

9h30 : **Corinne Rossari, Cyrielle Montrichard, Aylin Pamuksaç, Claudia Ricci & Linda Sanvido** (Université de Neuchâtel, chaire de linguistique française). *L'orientation argumentative à travers l'étude de connecteurs et de formes modales : une approche textométrique*

10h30 : Pause café

10h45 : **Margareta Kastberg** (Professeur, université de Franche-Comté). *Le président et son peuple ou le président du peuple : Analyse comparative de discours politiques en Afrique subsaharienne francophone.*

12h : Repas

13h45 : **Annabel Richeton** (doctorante, université Bourgogne Franche-Comté). *Appréhender finement la matérialité d'un corpus de grande ampleur : enjeux et méthodologie de la construction des données. Le cas du discours du ministère des Affaires étrangères français (1997-2019)*

14h15 : **Virginie Lethier** (Maître de conférence, université de Franche-Comté). *Que peut apporter la textométrie à l'analyse du discours institutionnel ? Retour d'expériences.*

15h15 : Pause café

15h30 : **Marion Bandinelli** (Maître de conférence, université de Franche-Comté). *Les expressions modales dans les débats présidentiels états-uniens (1960-2016) : Approche textométrique des auxiliaires modaux et des semi-modaux pour une analyse argumentative*

16h30 : Clôture de la journée d'étude

Programme détaillé de la journée

9h30

Corinne Rossari, Cyrielle Montrichard, Aylin Pamuksaç, Claudia Ricci & Linda Sanvido
(université de Neuchâtel, chaire de linguistique française).

L'orientation argumentative à travers l'étude de connecteurs et de formes modales : une approche textométrique

Les connecteurs constituent des enjeux pour transmettre outre des indications argumentatives, des indications énonciatives. Leur emploi sollicite plus ou moins visiblement la présence du locuteur dans son dire. Leur étude dans cette présentation sera menée dans une perspective outillée alliant étude qualitative et quantitative dans le but d'observer statistiquement les formes avec lesquelles les connecteurs se combinent de façon privilégiée dans de grands corpus représentant deux genres à vocation informative (la presse et l'encyclopédie) et dans deux langues (le français et l'italien). Ces données constitueront des pistes pour différencier les connecteurs selon leurs propriétés énonciatives et argumentatives selon deux pôles d'une même échelle, à savoir : ceux qui mettent en avant les choix discursifs menés par une instance énonciative et ceux qui mettent de tels choix en retrait au bénéfice d'une mise en place des événements se présentant comme se déroulant indépendamment d'une instance énonciative responsable de leur disposition en discours.

10h45

Margareta Kastberg (Professeur, université de Franche-Comté).

*Le président et son peuple **ou** le président du peuple : Analyse comparative de discours politiques en Afrique subsaharienne francophone.*

Mots-clés : discours politiques linguistique textométrie argumentation

Cet exposé consiste dans l'étude des discours de six présidents de l'Afrique de l'Ouest francophone, de la période après l'indépendance et de la décolonisation.

Nous avons constitué un corpus numérisé à partir des discours prononcés par des présidents de plusieurs pays francophones d'Afrique ; la Guinée Conakry, le Burkina Faso, le Sénégal, le Cameroun et le Bénin. Le corpus regroupe des discours prononcés et diffusés pendant des campagnes électorales, durant le mandat d'exercice en tant que président dans des situations différentes ; discours devant l'assemblée générale, discours qui s'adressent au peuple, à la jeunesse, des discours de vœux etc. L'analyse qui s'appuie sur les méthodes textométriques vise une étude comparative de la communication et de l'argumentation du pouvoir politique en Afrique de l'Ouest.

13h45

Annabel Richeton (doctorante, université Bourgogne Franche-Comté).

Appréhender finement la matérialité d'un corpus de grande ampleur : enjeux et méthodologie de la construction des données. Le cas du discours du ministère des Affaires étrangères français (1997-2019)

À travers l'exemple d'un corpus de près de 20 millions de mots, constitué de prises de paroles des représentants du ministère français des Affaires étrangères entre 1997 et 2019, nous proposerons de centrer notre regard sur l'établissement des données. Ainsi, nous aborderons la question de la constitution d'un corpus dont les textes, issus de situations de communication différentes, demandent une attention accrue quant à leur annotation : comment déterminer des critères suffisamment fins pour que l'analyse qui en résulte apporte des résultats probants ? Comment gérer des données dont l'origine peut être écrite ou orale, avec ou sans interaction, et dont l'auditoire varie d'un texte à l'autre ? Nous nous attacherons également à montrer comment le lissage d'un discours institutionnel et diplomatique débute dès la catégorisation des discours mis à disposition du grand public pour, enfin, présenter une première analyse de l'emploi des catégories grammaticales selon les locuteurs et leur statut.

14h15

Virginie Lethier (Maître de conférence, université de Franche-Comté).

Que peut apporter la textométrie à l'analyse du discours institutionnel ? Retour d'expériences.

Notre communication vise à illustrer les apports de la textométrie à l'analyse du discours institutionnel (Oger & Ollivier-Yaniv, 2003 ; Krieg-Planque, 2012). En vue de permettre une meilleure compréhension des singularités de la textométrie (Heiden *et al.*, 2010) et des possibilités offertes par cette pratique d'analyse de données textuelles faisant suite à la lexicométrie (Tournier, 1980 ; Lebart et Salem, 1994), nous présenterons différentes classes d'exploration de vastes corpus relevant du discours institutionnel.

Dans un premier mouvement, nous insisterons sur une première singularité de la démarche d'analyse textométrique : celle-ci permet en effet la construction de parcours de lecture nourris par l'articulation de prises de vue globales et de prises de vue locales sur la matérialité textuelle (Pincemin, 2010). Afin d'illustrer les apports de cette démarche fondamentalement contrastive, mettant à mal l'opposition traditionnellement admise entre analyse quantitative et analyse qualitative (Duchastel & Laberge, 2014; Mayaffre, 2014), nous montrerons comment la textométrie permet de mettre au jour des contrastes difficilement décelables par une lecture humaine (c'est-à-dire linéaire et non-outillée). Nous insisterons sur la façon dont de telles analyses différentielles, respectueuses de la matérialité, permettent d'éclairer des documents de planification urbaine, réputés être aussi neutres que lisses et « techniques » (Buhler & Lethier, 2019).

Dans un deuxième mouvement, nous mettrons la focale sur les unités de segmentation et d'indexation mobilisées en textométrie pour l'analyse du discours institutionnel. Afin d'illustrer les potentialités de l'exploration des dimensions plurielles de la textualité, nous rendrons compte de classes d'analyse menées sur les discours (« speeches ») délivrés par les différents représentants de la Banque Centrale Européenne (BCE) lors d'événements publics entre 1998 et 2020. Alors que des analyses thématiques pointent une permanence du stock lexical de la BCE en dépit des crises que traverse l'Europe à partir de 2007 (Lebaron, 2015), nous montrerons comment l'analyse de la consistance morphosyntaxique de ces discours livre, pour sa part, un faisceau d'indices attestant d'un tournant effectif des stratégies communicationnelles de la BCE dès 2009 (Lethier, 2019).

Bibliographie

Buhler T., Lethier V. (2019) "Analysing urban policy discourses using textometry: An application to French urban transport plans (2000-2015) ". *Urban Studies*, SAGE Publications, [10.1177/0042098019873824](https://doi.org/10.1177/0042098019873824). [hal-02316662](https://hal.archives-ouvertes.fr/hal-02316662)

Dacos M., Mounier P. (2014). *Humanités numériques. État des lieux et positionnement de la recherche française dans le contexte international*. Paris : Institut Français, <http://www.institutfrancais.com/sites/default/files/if_humanites-numeriques.pdf>.

Duchastel J., Laberge D. (2014) « Au-delà de l'opposition quantitatif/qualitatif. Convergence des opérations de la recherche en analyse du discours ». *Corela* [En ligne], HS-15 | 2014. URL : <http://journals.openedition.org/corela/3524> ; DOI : <https://doi.org/10.4000/corela.3524>

Guilbert T. (2014) « Introduction : articuler les approches qualitatives et quantitatives dans l'analyse de discours. ». *Corela* [En ligne], HS-15 | 2014. URL : <http://journals.openedition.org/corela/3545> ; DOI : [10.4000/corela.3545](https://doi.org/10.4000/corela.3545)

Heiden S., Magué J.-P., Pincemin B. (2010). « TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement », in I. C. Bolasco S. (éd.), *Proceedings of the 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, 2, Rome, Edizioni Universitarie di Lettere Economia Diritto, pp.1021-1032.

Jenny J. (1997) « Méthodes et pratiques formalisées d'analyse de contenu et de discours dans la recherche sociologique française contemporaine : état des lieux et essai de classification ». *Bulletin de méthodologie sociologique (BMS)* N° 54, pp.64-112.

Krieg-Planque A. (2012) *Analyser les discours institutionnels*. Paris : Armand Colin.

Lebaron F. (2015) « La BCE : flexible dans la stabilité ? ». *Économie et institutions* [En ligne], 22 | 2015. URL : <http://journals.openedition.org/ei/986> ; DOI : <https://doi.org/10.4000/ei.986>

Lebart L., Salem A. (1994) *Statistique textuelle*. Paris : Dunod.

Lethier V. (2019) « Sonder le discours de la Banque Centrale Européenne (1998-2018) », in Desmedt L., Gautier L., Llorca M. (Eds) (2019). *Les discours de conjoncture : approches croisées*. Paris : L'Harmattan.

Oger C., Ollivier-Yaniv C. (2003) « Du discours de l'institution aux discours institutionnels : vers la constitution de corpus hétérogènes », X° Colloque bilatéral franco-roumain, CIFSIC Université de Bucarest, 28 juin – 3 juillet 2003, Oct 2003. [ffsic_00000717](https://doi.org/10.1007/978-2-7099-0000-7_17)

Mayaffre D. (2014) « « Ça suffit comme ça ! ». La fausse opposition quantitatif / qualitatif à l'épreuve du discours sarkozyste ». *Corela* [En ligne], HS-15 | 2014. URL : <http://journals.openedition.org/corela/3543> ; DOI : <https://doi.org/10.4000/corela.3543>

Offner J.-M. (2006) *Les Plans de déplacements urbains*, Paris : La Documentation française.

Pincemin B. (2011). « Sémantique interprétative et textométrie– Version abrégée ». *Corpus*, 10, pp.259-269.

Tournier M. (1980) « D'où viennent les fréquences de vocabulaire ? La lexicométrie et ses modèles ». *Mots*, n°1, octobre 1980. Saussure, Zipf, Lagado, des méthodes, des calculs, des doutes et le vocabulaire de quelques textes politiques. pp. 189-209. DOI : <https://doi.org/10.3406/mots.1980.1010>

15h30

Marion Bendinelli (Maître de conférence, université de Franche-Comté)

Les expressions modales dans les débats présidentiels états-uniens (1960-2016) : Approche textométrique des auxiliaires modaux et des semi-modaux pour une analyse argumentative.

Les débats présidentiels sont l'occasion pour les candidat·es de faire connaître leur programme, de le défendre, d'asseoir ainsi leur légitimité et leur crédibilité par une maîtrise (souhaitée parfaite) des dossiers et de l'art de la rhétorique (Martel, 1983 ; Hinck, 1993). Ces débats, bien que dénigrés pour leur faible effet sur les intentions de vote, sont attendus, largement suivis par les spectateurs (radio, tv puis web) et commentés dans les médias. Pour cette raison, nous en faisons notre corpus privilégié pour étudier le discours politique états-unien.

Nous envisageons ces débats comme le produit d'une triple structuration textuelle (Adam, 2015 ; Halliday & Webster, 2010) : à un niveau macro-textuel, ils s'organisent autour d'un plan textuel à diverses facettes scénographiques (*press conference style*, *town-hall meeting style* et *conversation-like style*) et conversationnelles (séquences d'ouverture et de clôture ritualisées, types de questionnement différents – question avec ou droit de réfutation et/ou droit de suite, échange libre) ; à un niveau méso-textuel, la cohésion est assurée par les relations inter-propositionnelles définissant les séquences textuelles articulées (narration, argumentation, explication...) ; à un niveau micro-textuel, la cohésion est garantie par les relations intra-propositionnelles. Egalement, les débats sont le lieu de l'élaboration de l'identité discursive des candidats, une identité hybride et mouvante, modulée en fonction des contraintes génériques et discursives définissant les débats (Fairclough, 1998), de l'appartenance partisane des candidats qu'ils embrassent à des degrés distincts, de leur statut durant les débats (président sortant vs challenger avec ou sans mandat électoral), et des thématiques abordées durant les débats exigeant par exemple de paraître plus ou moins ferme et autoritaire.

Notre travail repose sur l'étude de marqueurs de modalité dont l'intérêt se justifie au regard de leur capacité à participer à la mise en œuvre de stratégies d'influence (légitimation, crédibilisation, captation, coercition – Charaudeau, 2005 ; Chilton et Schäffner, 2002) en fournissant divers moyens linguistiques pour valoriser le locuteur et ses connaissances,

orienter l'interprétation de l'interlocuteur à travers la manipulation et la hiérarchisation des informations (Winter et Gärdenfors, 1995 ; Lillian, 2008). Notre attention se portera en particulier sur les marqueurs *can*, *must*, *need to* et *will*.

Sur le plan méthodologique, nous avons constitué un corpus de près de 500.000 formes graphiques, à partir des débats à deux candidats (les années 1980 et 1992 sont donc omises) organisés en 1960 et 2016 ; les transcriptions des débats, disponibles sur le site debates.org ont été vérifiées grâce aux vidéos, corrigées quand nécessaire, puis établies au langage XML-TEI. Nous adoptons une démarche textométrique, permettant de dépasser la lecture linéaire des débats par une exploration heuristique des données réunies en corpus conduite à travers l'élaboration de diverses tables et représentations graphiques (listes, tables, analyses factorielles de correspondances, carte des sections). Utilisant les fonctionnalités du logiciel TXM (Heiden, Magué, Pincemin, 2010), nous décrirons dans un premier temps la ventilation des auxiliaires modaux et des semi-modaux ainsi que leur profil d'emploi de manière à dégager les ethos élaborés par les candidats ; dans un second temps, nous procéderons à une annotation fine d'un sous-corpus réduit aux seules introductions et conclusions des débats de 1960 afin de dégager le rôle joué par les expressions modales dans ces séquences, et *in fine* d'en proposer une modélisation au moyen du logiciel Lexico3 (Lamalle, Martinez, Fleury, Salem [en ligne]).