

LES STRATÉGIES DISCURSIVES À L'ŒUVRE DANS LE DISCOURS INSTITUTIONNEL SUR LES RÉSEAUX SOCIAUX : LE CAS DU PROFIL TWITTER DU MINISTÈRE DE L'IMMIGRATION, LA FRANCISATION ET L'INTÉGRATION AU QUÉBEC

ELEONORA MARZI
UNIVERSITÀ DI BOLOGNA

Abstract - Institutional discourse, defined as any discourse issued by an entity with an official position within the state apparatus, has specific discursive characteristics. The presence of official institutional profiles on social networks stimulates the analysis of forms of enunciation in search of discursive strategies influenced by the medium. Our contribution examines the Twitter profile of the Ministère de l'Immigration, de la Francisation et de l'Intégration du Québec (MIFI), from which we extracted a corpus of around 3,000 statements that we analysed using the methodology of corpus-assisted discourse analysis, in search of syntactic and lexical-semantic structures. The results of our analysis, which is mainly qualitative, highlight the discursive role of the enunciator, which varies according to the typology of the utterances: it fulfils an assistance function in the sub-corpus of comments and a socio-cultural amplification function in the sub-corpus of retweets. More specifically, for the tweets sub-corpus (considered as the first production utterances), we examined the use of fictitious proper names to identify the discourse strategies employing them. We have identified a discursive strategy of training narrative and a discursive strategy of testimony: in the latter case, there are differences when the fictitious proper noun designates a member of the host community as opposed to a member of the host community.

Keywords: corpus-assisted discourse analysis; immigration; Twitter; textometry; press discourse analysis; Québec.

Introduction

Dans le cadre théorique de l'analyse du discours, le discours institutionnel est défini comme tout discours produit par un énonciateur singulier ou collectif occupant une position juridiquement inscrite dans l'appareil d'Etat (Oger, Ollivier-Yaniv 2008) et qui présente certains traits stylistiques comme l'emploi de formes stéréotypées, l'usage des performatifs, et une rigidité en termes de dialogisme interactionnel (Moirand 1990). Si, parmi les contraintes qui définissent les genres situationnels, il y a le médium et le mode de diffusion (Maingueneau 2006), il est indéniable que l'entrée dans l'usage courant des réseaux sociaux pour diffuser les discours institutionnels ou les discours politiques implique des changements dans les stratégies discursives. Cela n'équivaut pas à affirmer que le discours institutionnel change en raison de sa diffusion sur les réseaux sociaux - car cela présupposerait que la totalité des discours institutionnels se fasse sur les réseaux sociaux - mais que les réseaux sociaux façonnent en contaminant les discours institutionnels avec certains traits typiques de la nature même du support.

Le discours institutionnel est traversé par plusieurs thématiques, et celui de la migration occupe une place importante en raison de sa portée sociale et politique et qui se concrétise aussi par l'émission de discours à travers des agents spécialisés tels que les Ministères de l'Immigration. Nous voulons donc croiser le discours institutionnel présent sur les réseaux et la thématique de l'immigration en nous concentrant en particulier sur le cas du MIFI – Ministère de l'Immigration, de la Francisation et de l'Intégration¹ du Québec, province francophone du Canada caractérisée par son

¹ <https://www.quebec.ca/immigration>

approche multiculturelle à la société et par sa profonde action d'intégration envers les personnes immigrantes (Lamy, Mathieu 2020).

Notre contribution veut montrer certaines stratégies discursives mises en œuvre au sein du discours institutionnel du Profil Twitter du Ministère de l'Immigration, de la Francisation et de l'Intégration² du Québec en employant la méthodologie de l'analyse du discours assistée par corpus, qui s'insère dans la plus ample mais très récente tradition d'aire anglophone des *Corpus-Assisted Discourse Studies* (Partington *et al.* 2013) et qui est le résultat d'une fusion entre la linguistique de corpus et l'analyse du discours.

Le corpus est composé des messages publiés sur le profil Twitter du MIFI et il a été interrogé à travers le logiciel en accès ouvert TXM³, qui permet d'importer le corpus préalablement construit, de l'annoter par parties du discours, et de l'analyser grâce aux calculs de la textométrie (Pincemin, Heiden 2008). En particulier nous nous sommes concentrés sur l'aspect lexico-sémantiques de trois sous-corpus définis à partir des différentes typologies d'énoncé (*tweets*, *retweets* et *commentaires*) dont nous avons analysé la composition lexicale pour identifier la fonction discursive que détient l'énonciateur et dans le cas particulier du sous-corpus des *tweets* nous nous sommes focalisés sur les stratégies discursives qui se développent autour des noms propres fictifs.

Le présent travail est ainsi structuré : le premier paragraphe §1 nous permettra de fournir une introduction à certains traits typiques du discours institutionnel à travers le cadre méthodologique de l'analyse du discours ; en nous focalisant, en particulier, sur le réseau social Twitter, ce qui nous donnera également l'occasion de dresser un bref état de des lieux des analyses sur ces genres de support. Dans le deuxième paragraphe §2, nous illustrerons la méthodologie de l'analyse du discours assistée par corpus, tout en traitant les principes qui viennent de ses deux composantes, qui sont la linguistique du corpus avec la textométrie et l'analyse du discours. Avec le paragraphe §3, nous entrons dans le vif de l'analyse, en illustrant quelques aspects techniques des outils employés pour l'interrogation du corpus, et en abordant la composition du corpus en terme de variété lexicale selon les sous-corpus créés à partir de la variété des messages (*tweet*, *commentaires*, *retweet*), pour nous concentrer dans le paragraphe §4 sur l'usage rhétorique des noms propres dans les stratégies discursives mises en place par l'énonciateur et qui impliquent à la fois la personne immigrante et le membre de la culture d'arrivée.

1. Le discours institutionnel et le réseau social Twitter

Parmi les genres situationnels repérés par l'analyse du discours, le discours institutionnel se définit comme le discours qui se produit officiellement par un énonciateur singulier ou collectif occupant une position juridiquement inscrite dans l'organisation de l'État et qui se caractérise par un fort lien entre les contenus véhiculés et leur concrétisation en actions politiques ou législatives. (Maingueneau 1979). Définir de manière stricte le discours institutionnel n'est pas l'objectif de notre communication, cependant nous partageons la théorie proposée par Claire Oger et Caroline Ollivier-Yaniv : elles considèrent le discours institutionnel comme un ensemble de cercles concentriques qui se développent autour d'une définition juridique restrictive comme celle que nous venons de présenter, et où la différence majeure réside dans le caractère officiel de la parole, ce qui se traduit dans le degré de performativité. Le deuxième cercle représente le discours qui peut être tenu par le même énonciateur, par exemple une personnalité politique en dehors de son cadre institutionnel – qui ne possède donc pas le droit d'agir au niveau juridique, et qui pourrait indiquer une différence substantielle entre discours institutionnel et discours politique :

² https://twitter.com/mifi_qc

³ <https://txm.gitpages.huma-num.fr/textometrie/Telecharger/>

Il est important de distinguer ces deux cercles, car à notre sens, le statut de ces deux discours n'est pas le même : seul le premier a véritablement une valeur performative au sens où il peut comporter des conséquences juridiques, y compris des formes de coercition. La distinction n'est donc nullement de pure forme. (Oger Ollivier-Yaniv 2003)

Si l'on suit cette idée du discours institutionnel, on peut parvenir à des définitions plus « amples » :

Enfin, au sens large, le discours institutionnel (institutionnalisé et institutionnalisant) peut comprendre l'ensemble des discours que l'on peut considérer à des degrés divers comme des discours « autorisés » dans un milieu donné, sans référence nécessaire à l'État (productions des syndicats, des états-majors des partis politiques, chartes des associations professionnelles, programmes et règlements d'écoles privées). (Oger Ollivier-Yaniv 2003)

À cette définition, l'on peut faire correspondre des traits discursifs spécifiques : l'utilisation des performatifs est un pivot important du discours institutionnel. Très présentes également sont les expressions figées qui ont la fonction de réduire la distance entre l'énonciateur et le co-énonciateur, et la stratégie de l'opposition est gérée grâce au principe de stabilisation des énoncés et au principe d'effacement de la conflictualité. (Krieg-Planque et Oger 2010)

Considérant que parmi les contraintes définitoires des genres situationnels du discours, il y a le médium et les modes de diffusion, mais aussi l'organisation et la longueur d'un texte (Maingueneau 2006), nous ne pouvons nier que la diffusion des réseaux sociaux a modifié l'état de choses lorsqu'ils imposent une longueur de texte maximale et qu'ils permettent, par leur matérialité, des typologies de messages spécifiques. À ce propos Maingueneau avait annoncé :

avec les mutations provoquées par les nouvelles technologies, on est de plus en plus conscient que le médium joue un rôle fondamental dans l'émergence et la stabilisation d'un genre. (Maingueneau 2009, p. 69)

Sans vouloir affirmer que nous nous trouvons face à un nouveau genre de discours, nous estimons que le médium peut influencer et modifier certaines pratiques discursives propres de la partie du discours institutionnel publié sur les réseaux sociaux. Le dispositif en fait ne se limite pas à être un conteneur ; il est, au contraire, un élément qui intervient dans la création du discours car le mode d'existence matérielle du discours et de sa diffusion intervient dans la constitution même du discours.

La structure interne des messages est modifiée dans les réseaux sociaux par sa longueur prédéfinie ou au moins « °limitée° » – si l'on considère les messages de maximum 140 caractères de Twitter, mais aussi par une série d'éléments typographiques qui permettent certaines interactions comme on va voir par la suite. L'interaction change donc et avec elle la forme des énoncés. En général, les réseaux sociaux interviennent dans le discours institutionnel en faisant percevoir une distance réduite entre énonciateur (l'institution) et le co-énonciateur (tout public qui a un rapport plus ou moins étroit avec l'institution) qui influence à son tour les contenus véhiculés.

En particulier, parmi la multitude des réseaux sociaux, nous observons comment la plateforme de micro-blogging Twitter a rapidement intégré des stratégies de communication des institutions et des personnalités politiques, orientation qui a été très claire dès son ouverture au public en 2006 (Lotan *et al.* 2011). Il suffit de penser au rôle que le réseau a joué par exemple dans la campagne Présidentielle en France en 2012 (Eyriès 2015) ou aux mouvements sociaux comme le Printemps arabe (Faris, 2012).

Twitter permet, comme tout réseau social, la création d'un profil personnel (qui peut être une personne physique, ou bien un groupe ou une institution), que nous considérons comme énonciateur et qui peut publier divers messages (énoncés) d'une longueur de maximum 280 caractères, accompagnés éventuellement de photos ou vidéos. L'énonciateur peut accomplir diverses réactions :

tweet, commentaires, mentions, *retweet* et *likes* (Williams et al 2013 ; Broniatowski et al. 2014), de plus il a à sa disposition des symboles typographiques, notamment l'arobase @ et le *hashtag* # qui peuvent se trouver à l'intérieur des trois typologies de réactions écrites.

Le *tweet* est le message dans sa forme simple, la source primaire écrite par l'énonciateur, à laquelle il est possible de répondre à travers le commentaire, créant ainsi l'unité minimale de conversation. Nous parlons d'unité minimale car le mécanisme de réponse peut se répéter potentiellement à l'infini aussi bien sur le plan horizontal, lorsque de nombreux utilisateurs répondent ou commentent le même message, que sur le plan vertical au moment où différents utilisateurs, y compris l'auteur du premier message, se répondent les uns aux autres. Dans les deux cas, des flux de dialogue polyphonique sont créés et peuvent être regroupés avec des #. Ainsi, le *hashtag* définit un topic, un sujet mais de manière fluide et dynamique. Selon Segerberg et Bennett (2011), en dépit de sa simplicité apparente, le *hashtag* permet de générer des *streams*, c'est-à-dire des flots de messages dont le fonctionnement est complexe. Ils peuvent servir à rallier une communauté, à créer un champ de débats pour des points de vue contradictoires, à attirer l'attention du public de manière ponctuelle ou à effectuer de la veille autour d'une thématique spécifique.

La mention se matérialise avec le symbole typographique de l'arobase @ et se trouve dans les *tweets*, les commentaires et les *retweets*. Il équivaut à une forme de citation d'un autre compte twitter et se traduit concrètement dans l'action d'appeler un tiers dans la conversation (Honeycutt, Herring 2009).

Enfin, le *retweet* est la copie et la rediffusion du message d'un autre participant, souvent enrichi par un commentaire, et il a pour fonction d'amplifier la diffusion du message, en prenant en même temps une position identitaire (Boyd 2010). Pour conclure, nous soulignons que le *likes* (j'aime) sont des réactions sans texte.

La quantité d'études sur la communication politique sur Twitter est très prolifique par plusieurs raisons, dont celle de la standardisation de taille et de structure des données et leur facilité de repérage : Twitter met à disposition des API⁴ pour recueillir les données selon plusieurs filtres : par Profils, par *stream*, par date, par zone géographique. L'analyse des données extraites par Twitter a été prêté à diverses finalités et les études sur ce sujet se trouvent au carrefour entre la linguistique, la sociologie, l'analyse du discours, la théorie de la communication, comme le démontrent la variété des études que nous allons présenter pour mieux situer notre contribution. Certaines études visent des approches lexicales en cherchant à définir le lexique d'un phénomène politique tel le populisme à travers l'analyse des *tweets* des politiciens (Combei 2020), ou à chercher les mots-clés à l'intérieur d'un *stream* particulier comme Black Lives Matters (Giorgi et al. 2022), ou encore à étudier la composition lexicale des expressions émotionnellement polarisées (Ana Zwitter Vitez 2020). D'autres approches d'études sont orientées vers l'analyse du discours qui porte sur les relations énonciatives et discursives s'établissant entre la partie textuelle du *tweet* et l'aperçu du lien URL qui y est intégré (Cagninelli 2020). D'autres approches adoptent un point de vue sociologique pour étudier le rôle que Twitter a joué dans le débat sur le droit des femmes à conduire en Arabie Saoudite (Altoaimy 2018).

Notre perspective d'étude se positionne à l'intérieur de l'analyse du discours assistée par corpus et que nous allons traiter en détail dans le paragraphe suivant.

2. La méthodologie : l'analyse du discours assistée par corpus (ADAC)

D'époque assez récente est l'apparition dans l'aire anglophone des *Corpus-assisted critical discourse Studies analysis* (Ancarno 2006, Baker, Gabrielatos, McEnery 2006 2008 2013, Partington 2008, Partington *et al.* 2013) qui se traduit en français avec le calque *analyse du discours assistée par*

⁴ En informatique, API est l'acronyme d'Application Programming Interface, que l'on traduit en français par interface de programmation applicative ou interface de programmation d'application.

corpus. Il s'agit de l'union de deux perspectives méthodologiques : l'analyse du discours – comme cadre interprétatif, la linguistique du discours pour collecter les données et qui se sert de la textométrie pour les interrogations des corpora. Bien que la locution *analyse du discours assistée par ordinateur* puisse paraître plus familière, elle pourrait occulter le fait que ce n'est pas uniquement l'ordinateur (ce qui se limiterait à la textométrie) mais aussi la linguistique du corpus qui sont convoquées ici.

L'analyse du discours assistée par corpus, dorénavant ADAC, adopte de la linguistique du corpus la perspective corpus-based, qui s'oppose à celle corpus-driven (Tognini-Bonelli 2001) qui vise à étudier les traits d'un corpus spécifique, construit selon une nécessité de recherche et devant donc être représentatif du phénomène à étudier. La construction du corpus reflète les nécessités pour répondre à la problématique de recherche, qui à son tour permet le partitionnement du corpus et des interrogations contrastives.

De l'analyse du discours, qui considère la langue comme pratique sociale, l'ADAC adopte les concepts d'énonciateur et de co-énonciateur, medium de diffusion, scène d'énonciation, dans un cadre général visant à étudier la manière dont les stratégies discursives contribuent à définir les identités et leurs relations. Les représentations que l'énonciateur décide de construire sont à interpréter dans un contexte, celui-ci étant doté de règles qui influencent les productions discursives et leur réception : il s'agit d'envisager le contrat que les participants à une énonciation acceptent, ce qui entraîne un certain nombre de stratégies discursives qui peuvent être analysées d'un point de vue syntaxique et lexical.

Le dernier aspect est l'interrogation du corpus : pour rendre opératoires les questions sur les stratégies discursives, une transformation des interrogations en ancrages linguistiques est nécessaire. La question de recherche, théorique (dans notre cas les stratégies discursives des noms propres) doivent trouver leur traduction en unités linguistiques à chercher ou à interpréter (des calculs textométriques appliqués aux résultats de requêtes CQL, dont nous donnerons le détail par la suite). Du point de vue de la sémantique lexicale, l'unité prend du sens du moment qu'elle s'intègre à un réseau d'autres mots avec des relations de similitude. L'interrogation du corpus se fait par le biais de la textométrie, qui applique la statistique aux textes, permettant de décrire les caractéristiques lexicales, syntaxiques, morphologiques, d'un texte. Cette discipline s'est développée en France à partir des années 1970, dans la lignée des recherches pionnières de Charles Muller (Muller 1974) en statistique lexicale et qui permettent d'évaluer la richesse en vocabulaire d'un texte, ou d'identifier le vocabulaire spécifique d'un texte, et, dans un autre filon de recherches, celle de l'analyse multidimensionnelle de Benzécri qui permet des analyses lexicales contrastives parmi plusieurs textes collectés en un corpus (Benzécri 1977).

Les outils de la textométrie permettent d'aborder le texte non seulement grâce à une lecture linéaire, mais aussi grâce à l'observation ciblée de certains de ses éléments et certaines de ses caractéristiques que l'on sélectionne en fonction des questions que l'on se pose (Mayaffre 2019).

Pour résumer, l'approche ADAC s'applique dans notre contribution de la manière suivante : la question de recherche des stratégies discursives du MIFI nous guide pour la composition du corpus qui est dessiné et construit selon les principes de linguistique du corpus alors que le cadre des questions posées et de l'interprétation des données est offert par l'analyse du discours. Il se produit ainsi l'union d'un aspect quantitatif, basé sur le corpus, qui permet de calculer la distribution et le comportement des éléments lexicaux au sein d'un corpus de manière massive, et l'approche qualitative qui permet l'analyse ponctuelle des éléments lexicaux selon les catégories opérationnelles de l'analyse du discours.

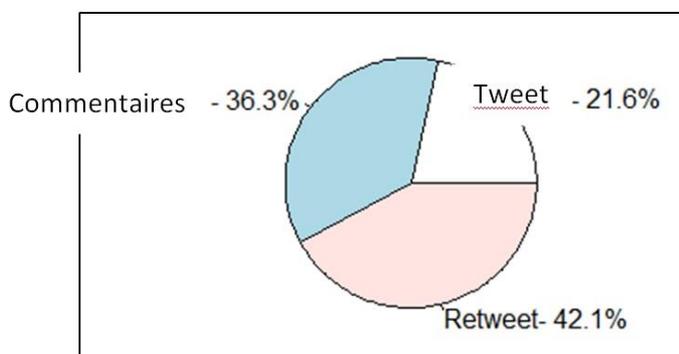
3. Composition et caractéristiques du corpus MIFI

Comme nous l'avons déjà annoncé, notre question de recherche se focalise sur le cas spécifique du discours institutionnel du Ministère de l'Immigration, la Francisation et l'Intégration du Québec, dorénavant MIFI, en raison de son orientation interculturelle.

Le corpus est composé par les messages publiés sur le Profil Twitter du *Ministère de l'Immigration, de la Francisation et de l'Intégration* du Québec (MIFI)⁵ que nous avons pu extraire grâce à un *script* écrit avec le langage de programmation R⁶ et que nous avons interrogé par le biais du logiciel de textométrie en distribution libre TXM⁷. Lors de son importation en TXM, grâce au logiciel Treetagger⁸ (Schmid, 1995) le corpus est annoté par parties du discours selon des algorithmes d'intelligence artificielle⁹. Pour l'interrogation des structures lexicales et syntaxiques, nous utiliserons des expressions CQL – *corpus query language*¹⁰, un langage d'interrogation pour rechercher des modèles grammaticaux ou lexicaux complexes.

Le corpus est composé de 3165 messages considérant un période qui va de mai 2014 (début de l'ouverture du compte) jusqu'à novembre 2022. Le corpus en détail est composé par 95568 occurrences et 8503 formes, et repartis en trois typologies : *tweet*, *retweet*, *commentaire*. Ces informations intégrées dans notre importation sous forme de méta-données nous permettent de diviser le corpus en trois sous-corpus, chacun correspondant à une typologie de message, ce qui nous servira pour mener des analyses contrastives et obtenir une vue d'ensemble plus complète.

Après une analyse quantitative sur la distribution des messages (Graphique 1), nous observons comment la majeure concentration se trouve sur les typologies *retweet* et *commentaires*.



Graphique 1. Distribution dans le corpus par typologie de message

Nous considérons tous les messages, tels que les *tweets*, les *commentaires* et les *retweets*, comme des énoncés. Dans le contexte de la scène d'énonciation, qui est le réseau social, l'énonciateur est toujours le MIFI, tandis que le co-énonciateur représente le public. Ce public est envisagé à la fois dans sa globalité, c'est-à-dire l'ensemble des utilisateurs, ainsi que dans sa singularité, notamment lorsqu'il s'agit d'un individu en particulier.

Une première exploration nous conduit vers une analyse contrastive des trois sous-corpus à la recherche de leurs traits caractéristiques au niveau discursif à travers l'analyse lexicale. En particulier, nous nous sommes concentrés sur les prédicats et après les avoir extraits¹¹, nous nous sommes concentrés sur la partie haute du classement par fréquence (tableau 1) qui est composée de mots tels

⁵ Le profil possède 221 following et 31.044 followers (dernière consultation 08/12/2022).

⁶ <https://www.r-project.org/>

⁷ <https://txm.gitpages.huma-num.fr/textometrie/Documentation/>

⁸ <https://www.ims.uni-stuttgart.de/en/research/resources/tools/treetagger/>

⁹ Ces systèmes fonctionnent par apprentissage automatique : un algorithme est entraîné avec de bons et de mauvais exemples d'annotation par parties du discours, ; il est ensuite capable d'annoter un nouveau texte avec une pourcentage d'erreur qui va de 10 à 20%.

¹⁰ Les CQL furent développées initialement par le Corpora and Lexicons group, IMS, University of Stuttgart au début des années '90 puis ont été enrichies toujours dans l'éprit d'un langage universel. <https://www.ims.uni-stuttgart.de/forschung/projekte/corpus-workbench/>

¹¹ CQL= Index of <[frpos = "VER.*"]> (442 item for 4.082 occurrences).

« appelez » (305), « consultez » (214), « connaître » (116). Nous avons ensuite extrait¹² les noms qui s'attachent aux prédicats dans une fenêtre de trois mots (tableau 2) : « dossier » (261), « questions » (341), « demande » (74), « programmes » (79), « déclaration » (73).

Forme du prédicat	Fréquence
Appelez	305
Consultez	214
connaître	116
savoir	64
peuvent	63
.....

Tableau 1. Fréquence des parties du discours du type prédicats du sous-corpus « Commentaires »

Forme des noms	Fréquence
dossier	261
questions	341
programmes	79
demande	74
déclaration	73
.....

Tableau 2. Fréquence des parties du discours du type noms du sous-corpus « Commentaires »

La composition lexicale utilisée ici indique clairement une prédominance du contenu assistantiel. En examinant plus en détail le contexte, nous pouvons observer que, lorsque nous regroupons les structures syntaxiques de manière similaire, les messages semblent suivre un schéma relativement uniforme, comme cela est illustré dans le tableau 3.

Structures syntaxiques	Fréquence
Appelez -nous pour des questions/vos questions concernant votre dossier	305
Consultez la grille	4
consultez le communiqué	4
Consultez notre/le site	109
Consultez notre/cette/la page	39
Consultez la section	5
compléter votre déclaration	21
recevrez sous peu un courriel	17
recevrez un message	15
obtenir de l'information	10
connaître les délais	9
connaître les plus récentes informations	22
connaître les détails	8
connaître les prochaines missions	10
connaître vos chances	6

Tableau 3. Concordances des prédicats les plus fréquents dans le sous-corpus "Commentaires"

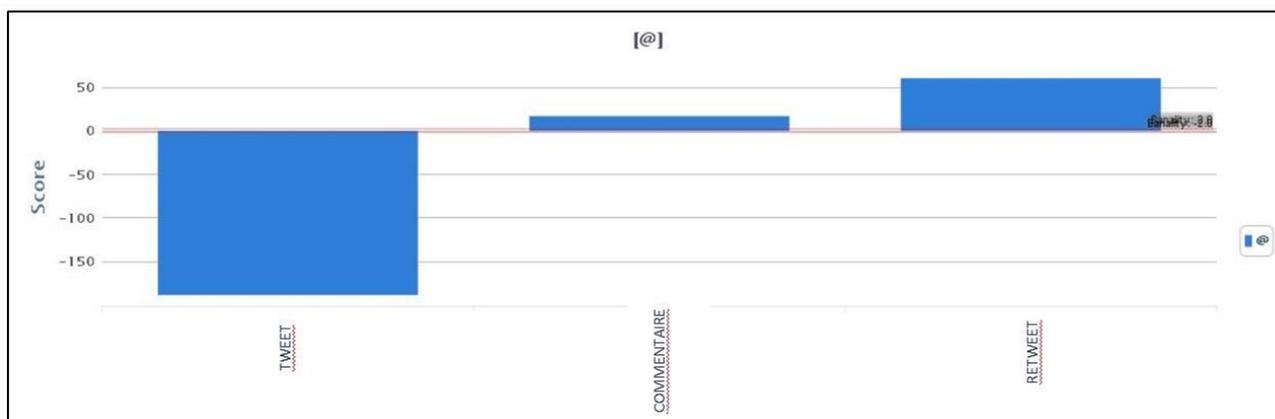
Les commentaires présentent un degré élevé de similitude : ils comportent des instructions, des indications ou fournissent les liens des documents administratifs. L'énonciateur agit comme un

¹² CQL= Index of <[frpos = "VER.*"][]]{0,3}@[frpos="NOM"]>, property @word, in MIFITWEET/Commentaries corpus (254 item for 2.979 occurrences).

guichet d'information, sa fonction déclarative se réduit au profit d'une fonction performative, très similaire à celle de son co-énonciateur.

Cela suggère une similitude potentielle entre les premiers *tweets*, ceux émis par le co-énonciateur, qui peuvent être de nature informative ou de recherche d'informations. Ainsi, si l'énonciateur est défini au niveau du discours comme une source d'information, le co-énonciateur est plutôt dans une position de demande d'information.

Le sous-corpus des *retweets* remplit une fonction discursive qui tient à l'amplification du message. En effet, le *retweet* se distingue par l'utilisation du symbole « @ » (arobase), qui indique un tiers destinataire de l'énoncé, lequel est invoqué dans le contexte de l'énonciation. Ce tiers représente à son tour le premier énonciateur dans une autre scène d'énonciation. Pour évaluer la fréquence d'apparition de l'arobase « @ » dans chacun des sous-corpus, nous examinons la proportion des occurrences en utilisant le calcul des spécificités, comme illustré dans le graphique 2. Cela nous permet de déterminer la probabilité d'apparition de ce symbole dans chacun des sous-corpus. Dans le sous-corpus des *tweet* la valeur est de signe négatif alors que dans le sous-corpus des *commentaires* s'affiche un faible pourcentage due au fait que certains commentaires sont des réponses exactes à des questions d'utilisateurs individuels. Les *retweet* montrent ces énonciateurs tiers, qui dérivent d'une autre scène d'énonciation et qui ont des dénominations qui sont signe d'une appartenance au monde culturel et au monde de l'emploi : @Culturepourtous (47), @Acces_travail (4), @chambremontréal (14), ConcertationMTL (12), @diversitebsl (17), Emploi_Quebec (29), FinancesQuebec (8), GouvQc (32), lemoisesnoirs (9), poleemploi (40), talentmontreal (11).



Graphique 2. Calcul des spécificités pour le symbole arobase @ sur les trois sous-corpus

L'hypothèse d'une stratégie discursive caractérisant le sous-corpus des *retweets*, à savoir son rôle d'amplificateur, notamment pour des initiatives à caractère culturel, social et liées au travail, est renforcée par l'utilisation la présence d'un lexique tel que « emploi », « entreprise », « recrutement », « salon », « événements », « prix », « journée », « finalistes », ainsi que par l'abondance des mentions (arobase @) par rapport aux autres parties du corpus. Cette utilisation accrue des mentions contribue à renforcer la fonction amplificatrice de l'énonciateur, qui vise à promouvoir l'intégration entre les membres de la communauté d'arrivée (personnes immigrantes) et les membres de la communauté d'accueil (Marzi 2020).

Le troisième sous-corpus est celui des *tweets*, que nous considérons comme la communication de première source et qui, pour cette raison, couvre plusieurs sujets ayant trait aux lois et aux actualités en général. Considérant l'ampleur des sujets traités, notre choix s'est posé sur les stratégies discursives qui se développent autour des noms propres fictifs à qui nous consacrerons le paragraphe suivant.

4. L'analyse des stratégies discursives autour des noms propres

En considérant l'ouverture que Twitter active dans le mouvement dialogique, notre intérêt se porte sur les formes rhétoriques que l'énonciateur emploie autour des noms propres fictifs. Il ne s'agit pas de noms correspondant à des personnes réelles, car ceux-ci sont indiqués avec des mentions (dont typographiquement avec des arobases @), mais de noms propres qui font référence à des personnes fictionnelles avec une fonction rhétorique que nous mettrons en évidence grâce à des analyses lexicales et syntaxiques.

Nous suivons la démarche est la suivante : grâce à des CQL nous extrayons¹³ les phrases où se trouvent les noms propres (terme pivot) qui sont automatiquement étiqueté avec le sigle NAM¹⁴, pour les distinguer de NOM (nom communs). Après un filtrage manuel pour éliminer les entités géographiques, nous regroupons les données selon une structure syntaxique et que nous allons par la suite détailler et analyser en appliquant une perspective lexico-sémantique.

Il est essentiel de préciser que notre analyse est principalement qualitative. En d'autres termes, le corpus représentatif du discours institutionnel que nous avons étudié ne se focalise pas spécifiquement sur l'utilisation des noms propres, ce qui explique leur présence en quantité limitée. Néanmoins, en considérant l'ensemble des noms propres fictifs comme une valeur relative, nous sommes en mesure de tirer des résultats dignes d'intérêt.

Le résultat des regroupements au niveau syntaxique fait ressortir deux typologies (tableau 4) : la première [1] veut exprimer un témoignage : « Selon Andrea... », la deuxième [2] présente un récit : « Voici l'histoire de ».

STRUCTURE SYNTAXIQUE	PATRON LEXICAL (EXEMPLE)	FREQUENCE
[1] PREP + NAM	<i>Pour Andrea [...]</i>	(8)
[2] ADV + NOM + PREP + NAM	<i>Voici l'histoire de Chloe [...]</i>	(4)

Tableau 4. Regroupement des structures syntaxiques autour des noms propres (NAM)

Les deux structures syntaxiques, présentent une partie immuable (l'incipit) et puis une structure similaire qui se remplit d'un lexique qui varie selon les énoncés et qui contribue ainsi à définir et à délimiter un domaine sémantique spécifique qui est l'objet de l'analyse qui suit.

La première structure que nous avons défini de témoignage [1] PREP + NAM est systématiquement remplie par la proposition « pour » suivie d'un nom propre. Il s'agit d'une forme indirecte du discours rapporté, où l'énonciateur second (l'entité désignée par le nom propre fictif) devient le vecteur de transmission d'un message, qui, en réalité, représente l'expérience de l'immigration. L'énonciateur premier de ce message (en l'occurrence, le Ministère) incarne une garantie de fiabilité de ce contenu et dans ce cadre l'entité fictive désignée par le nom propre renforce la fonction de l'énonciateur. Le tableau ci-dessus (tableau 3) montre un échantillon d'exemples en contexte que nous allons détailler et expliquer dans les tables 4 et 5.

	[WORD = "POUR"] [PART OF SPEECH = "NAM"]
1	<i>Pour Andrea, le Québec est l'endroit rêvé pour transformer ses idées en réalité. Avec son laboratoire de cosmétiques qui emploie une trentaine de personnes et son engagement envers de nombreuses causes, elle embellit la vie des gens.</i>
2	<i>Pour Hicham Louridi, opérateur de machine numérique originaire du Maroc, c'est la simplicité de la [ville de Saint-Georges, en Chaudière-Appalaches, qui fait tout son charme.</i>
3	<i>Pour Geneviève Paris, [directrice générale de Paber Aluminium] à Cap-Saint-Ignace, l'expérience des travailleurs issus de l'immigration bénéficie à son entreprise et à sa région.</i>

¹³ CQL = Concordance [frpos="NAM"] @ [word !="@.*.*"] [frpos="NAM"] <whitin \s.

¹⁴ L'étiquetage est fait de manière automatique par le biais du logiciel Tretagger et les abbreviations pour l'analyse du discours sont standard.

4	<i>Pour Laura Hernandez, professeure de yoga [originaire du Mexique], [la ville de Knowlton, en Estrie], lui a offert de multiples opportunités.</i>
5	<i>Pour Mario Séguin, président de Medwave Optique à Belœil, l'expertise des travailleurs issus de l'immigration est un incontournable pour son entreprise et sa région.</i>
6	<i>Pour Martin Lafrance, DRH chez Supermétal à Sherbrooke, les employés issus de l'immigration sont indispensables pour l'atteinte du plein potentiel de son entreprise.</i>
7	<i>Pour Nicolas Clusiault, DRH chez Momentum Technologies à Québec, la contribution et l'expertise des employés issus de l'immigration sont essentielles à son entreprise.</i>
8	<i>Pour Rockefeller Faustin, analyste-programmeur originaire d'Haïti, la région de la Capitale-Nationale représente une multitude d'occasions d'emploi.</i>

Tableau 3 – Echantillon d'exemples en contexte de la structure syntaxique de témoignage [1]

Nous pouvons identifier deux sous-groupes distincts dans notre analyse discursive : ceux où le nom fictif désigne une personne immigrante et ceux où le nom fictif désigne un membre de la société d'accueil. Notre analyse révèle comment chacune de ces deux catégories présente des caractéristiques lexico-sémantiques spécifiques. Dans le cas où le témoin est un membre de la société d'accueil (comme indiqué dans le tableau 4), c'est-à-dire la personne immigrante, l'identité est caractérisée par l'origine (« originaire du Mexique », « originaire du Maroc », « originaire d'Haïti ») et sa profession (« opérateur de machine numérique », « professeur de yoga », « analyste-programmeur »). Le contenu de son témoignage met l'accent sur des aspects « géographiques » qui sont perçus comme positifs, tels que « la ville de Knowlton en Estrie », « la ville de Saint-Georges en Chaudière-Appalaches », « le Québec », « la région de la Capitale-Nationale ». Ici c'est la qualité de vie québécoise qui est en question, avec la mention d'une variété de lieux géographiques aux allures didactiques. Les lieux géographiques sont en fait le sujet principal du témoignage, et les verbes associés à ces lieux sont principalement de nature étatique et inchoative (par exemple, « offert », « fait », « est », « représente »). Les caractéristiques mentionnées, qui décrivent la qualité de l'expérience migratoire, se manifestent à travers des noms à connotation positive tels que « occasions », « opportunité », « charme », ainsi que des adjectifs numéraux vagues comme « multiples », « multitude ».

IDENTITE (PROVENANCE)	IDENTITE (PROFESSION)	
*originaire du Mexique *originaire du Maroc *originaire d'Haïti	*opérateur de machine numérique *professeure de yoga *analyste-programmeur	
LIEUX GEOGRAPHIQUES	(PREDICATS) D'ETAT ET COACTIFS	QUALITES
* la ville de Knowlton, en Estrie *la simplicité de la ville de Saint-Georges, en Chaudière-Appalaches *Le Québec *La région de la Capitale-Nationale	*a offert *fait *est *représente	*tout son charme *multiples opportunités *L'endroit rêvé *multitude d'occasions d'emploi

Tableau 4 – Fonctions discursives où le nome propre fictif dénote la personne immigrante

Pour ce qui concerne le cas des noms propres fictifs représentant des membres de la société d'accueil les éléments lexicaux informent sur une stratégie rhétorique différente par rapport au cas du témoin en tant que personne immigrante. Notre analyse (comme illustré dans le tableau 5) révèle que l'identité du témoin est définie par son poste de travail, qui est généralement de haut niveau (directrice générale, président, DRH - Directeur des Ressources Humaines), sans mention d'une provenance géographique. Le sujet sur lequel le témoin est invité à donner son avis concerne la contribution des personnes immigrantes au tissu professionnel de la société québécoise, mettant en avant des termes

tels que « expérience », « expertise », qui ajoutent de la valeur aux entreprises et aux régions. Les prédicats utilisés qui sont fortement polarisés de manière emphatique, comme « est un incontournable », « sont indispensables », « sont essentielles », « bénéficie », semblent viser à rassurer les membres de la société d'accueil.

Le seul point commun entre les deux sous-groupes est la polarisation positive évidente, exprimée par des termes tels que « multiples opportunités », « multitude d'occasions », ainsi que des structures nominales comme « est un incontournable », « sont indispensables », « sont essentielles ».

IDENTITE (PROFESSION)		
*directrice générale *président *DRH – Directeur Ressources Humaines		
OBJET DU TEMOIGNAGE	PREDICAT	BENEFICIAIRE
*l'expérience des travailleurs issus de l'immigration *l'expertise des travailleurs issus de l'immigration *les employés issus de l'immigration *la contribution et l'expertise des employés issus de l'immigration	*bénéficie *est un incontournable *sont indispensables pour l'atteinte du plein potentiel *sont essentielles	*entreprise et à sa région *son entreprise

Tableau5 - Fonctions discursives où le nome propre fictif dénote le membre de la société d'accueil

Nous remarquons que malgré des structures syntaxiques similaires et une fonction discursive constante en tant que témoin, les représentations diffèrent. Le membre de la société d'accueil est principalement caractérisé par son rôle professionnel, souvent de haut niveau de responsabilité. Son message semble s'adresser aux autres membres de la société d'accueil en montrant comment les personnes immigrantes peuvent apporter une valeur ajoutée au tissu socio-économique du Québec : la stratégie discursive remplit une fonction rassurante par rapport à la diversité de l'Autre. D'autre part, le membre de la société de départ (la personne immigrante) est principalement caractérisé par son lieu d'origine, avec une mention accessoire de son rôle professionnel (qui n'est généralement pas au même niveau hiérarchique que ceux présentés par les membres de la société d'arrivée). Le sujet sur lequel il est appelé à s'exprimer concerne le tissu social et géographique du Québec, mettant en avant de multiples caractéristiques positives qui rendent la vie attrayante dans la région. Il semble s'adresser aux autres membres de la société de départ, qui est cependant multiple, et l'indication de l'origine géographique semble témoigner de cette variété : ce témoignage est valable pour toutes les personnes immigrantes : même si le lieu de départ est différent pour chacun, la qualité de vie au Québec est valable pour tous.

En résumé, le contenu implicite est un message promotionnel en faveur du Québec, qui est décrit sous deux perspectives différentes : celle de ceux qui arrivent et celle de ceux qui vivent sur place. Il met en avant leurs avantages respectifs pour souligner les atouts de la région.

Dans le cas de la deuxième structure syntaxique [2] que nous avons défini celle du récit, composée de « ADV + NOM + PREP + NAM », le modèle se répète : le syntagme « Voici l'histoire de » demeure constant et précède toujours le nom propre, suivi du reste du message où l'énonciateur nous présente une histoire, il bâti une structure narrative autour de l'entité fictive représentée par le nom propre. Dans le tableau ci-dessus (tableau 6) nous proposons un échantillon d'exemple.

[2]	
1	Voici l'histoire d'Atiquallah, originaire d'Afghanistan. Arrivé au Québec en 2019, il continue de progresser en français afin de travailler dans le domaine de la photo et de la vidéo
2	Voici l'histoire de Chloé Yao Shi, originaire de Chine, arrivée au Québec en 2020. Aujourd'hui, elle travaille en santé et en français.

Voici l'histoire de Claudia Lora, originaire de Colombie, venue s'installer au Québec avec son conjoint et ses enfants. Apprendre le français a été un incontournable pour elle.

Tableau 6. Exemples des *tweets* pour la structure discursive [2] du récit

L'analyse des stratégies discursives met en évidence que le contenu du récit tourne autour de l'expérience de l'immigration. Comme toute narration, elle comporte un point de départ et un point d'arrivée, un objectif pour le protagoniste et un élément déclencheur. Le point de départ, à la fois spatio-temporel et identitaire, est marqué par la provenance du protagoniste – dénotée par le nom propre fictif – ce qui le définit comme une personne immigrante (« originaire de [...] »). Le point d'arrivée est le Québec et le but ultime du protagoniste est de s'intégrer dans le tissu social économique et culturel, ce qui n'est possible que grâce à l'élément déclencheur : l'apprentissage de la langue française (tableau 7). Cette narration incarne le message central de l'énonciateur et qui est clairement reflété par la dénomination « Ministère de l'Immigration, la Francisation et l'Intégration » : où l'on voit bien que le mot « francisation » est la condition d'accès, le pacte implicite que la personne immigrante doit honorer pour accomplir l'« intégration » (Marzi 2022). La valeur pédagogique de la narration, qui devient un modèle réconfortant, un exemple à suivre pour toute personne immigrante, est la manière implicite pour l'énonciateur d'accomplir sa fonction qui est celle de pousser à l'intégration à travers la participation au tissu productif de la société québécoise.

DEBUT DE L'ACTION	IDENTITE (PROVENANCE)	ELEMENT DECLENCHEUR	OBJECTIF
*arrivée au Québec en 2020	* originaire de Colombie	*le français	*travailler
*Arrivé au Québec en 2019	*originaire d'Afghanistan	*il continue de progresser en français	*travaille en santé et en français
*venue s'installer au Québec	*originaire de Chine	*apprendre le français	*a été incontournable

Tableau 7. Fonctions discursives de la structure [2]

Conclusions

Suite à une analyse du discours assistée par corpus collectée à partir du Profil Twitter du Ministère de l'immigration, la Francisation et l'Intégration du Québec, nous sommes en mesure de tirer les conclusions suivantes. En général, le discours institutionnel sur Twitter du MIFI se caractérise par une réduction de l'espace entre énonciateur et co-énonciateur, et, selon la typologie, des messages nous avons pu observer certaines tendances discursives.

Dans les messages du sous-corpus *commentaires* les stratégies discursives transforment l'énonciateur en une source d'information, un dispositif d'assistance matérielle, dessinent un discours institutionnel qui dirige et indique le chemin à suivre, se faisant près de son co-énonciateur. Dans les messages du sous-corpus *retweet* l'énonciateur joue également le rôle d'amplificateur culturel, social et économique en renforçant ses fonctions d'intégration entre la personne immigrante et le membre de la société d'accueil, qui revient à s'adresser aux deux composantes du public et leur proposer des narrations adaptées.

En particulier, au sein du sous-corpus des *tweets*, considérés comme les énoncés premiers car ils ne naissent pas pour répondre à des stimuli, nous observons un usage discursif des noms propres fictifs qui se résume dans la mise en place d'une stratégie discursive de deux typologies : celle du témoignage et celle du récit de formation.

Dans le cas de la stratégie discursive du témoignage, nous pouvons distinguer deux scénarios distincts. D'abord, lorsque le témoin est la personne immigrante qui partage son expérience de l'endroit, la valeur de cette expérience (qui symbolise implicitement l'expérience de l'immigration) est présentée de manière positive mais significative en termes quantitatifs, bien que vague, avec une

forte emphase. Ensuite, lorsque le témoin est un membre de la société d'accueil, le contenu du témoignage met en avant la valeur de l'immigré en termes d'expertise et d'expérience professionnelle, soulignant son rôle fondamental dans le développement du pays et de ses infrastructures professionnelles, auxquelles appartiennent les entités fictives.

Quant à la stratégie discursive du récit de formation, elle dépeint la migration comme un moment formateur pour le sujet, dont l'objectif principal est l'intégration au sein du tissu social, culturel et économique du Québec. Cette intégration est perçue comme étant conditionnée par l'apprentissage du français, qui est considéré comme un élément déterminant. Ces deux stratégies discursives, en général, servent d'exemples pour soutenir la mission du MIFI visant une intégration pacifique basée sur une convenance mutuelle.

Bionota: Eleonora Marzi est chercheuse en Langue et Linguistique françaises auprès du Département de Langues, Littératures et Cultures Modernes de l'Université de Bologne. Elle s'occupe d'analyse du discours avec une attention particulière au discours politique et institutionnel dans le cadre de la linguistique du corpus et à travers l'application des technologies d'extraction automatique des connaissances (TAL, textométrie). Elle s'occupe également de terminologie numérique, en se focalisant en particulier sur la terminologie liée à la culture et à la politique. Parmi ses dernières publications la monographie *Langue, Littérature et Informatique : interculturalité en perspective* (2021) et parmi ses derniers articles : « La traduction automatique neuronale et les biais de genre : le cas des noms de métiers entre l'italien et le français » (2021) et « Les stratégies linguistiques à l'œuvre dans le discours de presse » (2023).

Recapito dell'autore: eleonora.marzi@unibo.it

Références bibliographiques:

- Altoaimy L. 2018, "Driving Change on Twitter: A Corpus-Assisted Discourse Analysis of the Twitter Debates on the Saudi Ban on Women Driving", in *Social Sciences* 7 [5], DOI: <https://doi.org/10.3390/socsci7050081> (15.11.2023).
- Ancarno C. 2020, "Corpus-Assisted Discourse Studies", in De Fina A., Georgakopoulou A. (dir.) *The Cambridge Handbook of Discourse Studies*, Cambridge University Press, Cambridge, pp. 165-185, DOI: <https://doi.org/10.1017/9781108348195.009> (15.11.2023).
- Baker P., Erez L. 2015, "Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity", in *Discourse & Communication* 9 [2] pp. 221-236. DOI <https://doi.org/10.1177/1750481314568542> (15.11.2023).
- Baker P., Gabrielatos C., Khosravinik M., McEnery T., Wodak R. 2008, "A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press", in *Discourse and Society* 19 [3], pp. 273-306.
- Baker P. 2006, *Using Corpora in Discourse Analysis*, A&C Black, London.
- Benzécri J.-P. 1977, « Analyse discriminante et analyse factorielle », in *Cahiers de l'analyse des données* 4 [2], pp. 369-406.
- Biber D. 1993, "Representativeness in Corpus Design", in *Literary and Linguistic Computing* 4 [8], pp. 243-257, DOI: <https://doi.org/10.1093/lc/8.4.243> (15.11.2023).
- Boyd D. 2010, "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter", in *Hawaii International Conference on System Sciences*, DOI: 10.1109/HICSS.2010.412 (15.11.2023).
- Broniatowski D. A., Paul M., Dredze M. 2014, "Twitter: Big data opportunities", in *Science* 345 [6193], pp. 148-148, DOI: 10.1126/science.345.6193.148-a (15.11.2023).

- Cagninelli C. 2020, « Le débat public sur Twitter : construction énonciative et discursive de tweets intégrant des liens URL », in *Repères DoRiF* 22, <https://www.dorif.it/reperes/le-debat-public-sur-twitter-construction-enonciative-et-discursive-de-tweets-integrant-des-liens-url/> (15.11.2023).
- Combei C. R. 2020, “‘Parlo come mangio’. Il lessico populista su Twitter”, in *CLUB Working Papers in Linguistics* 4, pp. 103-122.
- Eyriès A. 2015, « Deux campagnes électorales dans la twittosphère. L’élection présidentielle française et l’élection générale du Québec en 2012 », in *Les Cahiers du numérique* 11 [4], pp. 75-90.
- Faris D. M. 2012, « La révolte en réseau : le ‘printemps arabe’ et les médias sociaux », in *Politique étrangère* 1, pp. 99-109.
- Giorgi S. 2022, “Correcting Sociodemographic Selection Biases for Population Prediction from Social Media”, in *Proceedings of the Sixteenth International AAAI Conference on Web and Social Media* 16, pp. 228-240.
- Heiden S., Magué J.-P., Pincemin B. 2010, « TXM: Une plateforme logicielle open-source pour la textométrie - conception et développement », in *JADT 2010: 10th International Conference on the Statistical Analysis of Textual Data* (Rome).
- Honeycutt C., Herring S. 2009, “Beyond Microblogging: Conversation and Collaboration via Twitter”, *Proceedings of the 42nd Hawaii International Conference on System Sciences*, pp. 1 - 10. DOI : 10.1109/HICSS.2009.89 (15.11.2023).
- Krieg-Planque A., Oger C. 2010, « Discours institutionnels : perspectives pour les sciences de la communication », in *Mots. Les langages du politique* 94, pp. 91-96, DOI : <https://doi.org/10.4000/mots.19870> (15.11.2023).
- Krieg-Planque A. 2012, *Analyser les discours institutionnels*, Armand Colin, Paris.
- Lamy G., Mathieu F. 2020, « Les quatre temps de l’interculturalisme au Québec », in *Canadian Journal of Political Science*, 53 [4], pp. 777-799, DOI :10.1017/S0008423920001080 (15.11.2023).
- Maingueneau D. 2004, « Hyperénonciateur et participation », in *Langages* Effacement énonciatif et discours rapportés 156, pp. 111-126.
- Maingueneau D. 2006, « Le dialogue comme hypergenre », in Guérin P. (dir.) *Le dialogue : ou les enjeux d’un choix d’écriture*, Presses universitaires de Rennes, Rennes. DOI : <https://doi.org/10.4000/books.pur.30110> (15.11.2023).
- Maingueneau D. 2009, *Les termes clés de l’analyse du discours*, Seuil, Paris.
- Marzi E. 2021, *Langue, Littérature et informatique : l’interculturalité en perspective. Etudes de traitement automatique des langues et analyse du discours*, I libri di Emil, Bologna.
- Mayaffre D. Pincemin B., Poudat C. 2019, « Explorer, mesurer, contextualiser. Quelques apports de la textométrie à l’analyse des discours », in *Langue française* 203 [3], pp. 101-115.
- Moirand S. 1990, *Une grammaire des textes et des dialogues*, Hachette, Paris.
- Muller C. 1974, *La statistique linguistique*, Hachette, Paris.
- Oger C., Ollivier-Yaniv C. 2003, « Analyse du discours institutionnel et sociologie compréhensive : vers une anthropologie des discours institutionnels », in *Mots. Les langages du politique* 71, p. 125-145, DOI : <https://doi.org/10.4000/mots.8423> (15.11.2023).
- Partington A., Duguid A., Taylor C. 2013, *Patterns and Meanings in Discourse. Theory and practice in corpus-assisted discourse studies (CADS)*, John Benjamin, Philadelphia.
- Partington A. 2008, “The armchair and the machine: Corpus-assisted discourse research”, in Torsello C. T. (dir.), *Corpora for university language teachers*, Peter Lang, Berna, pp. 95-118.
- Pincemin B., Heiden S. 2008, « Qu’est-ce que la textométrie ? Présentation », Site du projet Textométrie, <https://pages.textometrie.org/textometrie/Introduction> (15.11.2023).
- Schmid H. 1994, “Probabilistic Part-of-Speech Tagging Using Decision Trees”, in *Proceedings of International Conference on New Methods in Language Processing* (Manchester).

- Segeberg A., W. L. Bennet 2011, "Social Media and the Organization of Collective Action: Using Twitter to Explore the Ecologies of Two Climate Change Protests", in *The Communication Review*, 14 [3], pp. 197-215, DOI: 10.1080/10714421.2011.597250 (15.11.2023).
- Tognini-Bonelli E. 2001, *Corpus linguistics at work*, John Benjamins Publishing, Amsterdam.
- Williams S., Terras M., Warwick C. 2013, "What do people study when they study Twitter? Classifying Twitter related academic papers", in *Journal of Documentation* 69 [3], pp. 384-410.
- Zwitter Vitez A. 2020, « Le discours politique et l'expression de l'opinion sur Twitter : analyse syntaxique, lexicale et orthographique », in *Ars & Humanitas* 14 [1], pp. 157-170 DOI:10.4312/ars.14.1.157-170 (15.11.2023).