

Assumer ou diluer les responsabilités énonciatives dans la diffusion du savoir encyclopédique

Corinne Rossari¹, Laura Aubry², Chloé Tahar³

¹Université de Neuchâtel– corinne.rossari@unine.ch

²Université de Neuchâtel– laura.aubry@unine.ch

³Université de Neuchâtel– chloe.tahar@unine.ch

Abstract

The aim of this paper is to quantify how knowledge is mediated across the different thematic domains of two encyclopedic works, representing the beginnings of the genre with the *Encyclopédie Diderot et d'Alembert* as well as its most current state with Wikipedia. Adopting a textometric approach based on various statistical methods (specificity scores and CA), we measure the share of constructions of the type *je, nous, on*, impersonal *il* + verbal syntagm + *que* across different fields of knowledge specific to each encyclopedia. We establish a relation between variants of these constructions in terms of subject pronouns and semantic classes (i.e., the type of function the construction gives to the content of its complement) and these different domains. Our results show (i) that the type of knowledge mediation changes significantly depending on the thematic domain in both encyclopedias, and (ii) that there are points of convergence between the modes of mediation of the two encyclopedias.

Keywords: encyclopedic discourse, mediation markers, textometry.

Résumé

Le but de cette recherche est de mesurer la façon dont la connaissance est médiée dans les différents domaines thématiques de deux ouvrages encyclopédiques représentant les prémices du genre avec l'*Encyclopédie Diderot et d'Alembert* et son état le plus actuel avec Wikipédia. Nous utilisons une approche textométrique fondée sur différentes méthodes statistiques (indices de spécificité et AFC) pour mesurer la part que des structures du type : *je, nous, on, il* impersonnel + syntagme verbal + *que* prennent selon différents domaines de la connaissance propres à chaque encyclopédie. Nous mettons en relation des variantes de cette construction en termes de pronoms sujet et de classes sémantiques (i.e. le type de fonction que la structure donne au contenu de la complétive) avec ces différents domaines. Les résultats montrent (i) que le type de médiation de la connaissance change sensiblement selon le domaine dans les deux encyclopédies et (ii) que l'on retrouve des points de convergence dans les modes de médiation des deux encyclopédies.

Mots clés : discours encyclopédique, marqueurs médiatifs, textométrie.

1. Introduction

Toute connaissance est passible d'une transmission directe ou médiée. On peut transmettre un savoir sans médiation (*Ce mot vient du Latin*) ou avec (*Il a été démontré/On pense que ce mot vient du Latin*). Notre étude s'intéresse aux façons de transmettre un savoir dans deux discours encyclopédiques - représentant d'un côté la naissance du genre avec l'*Encyclopédie de Diderot et d'Alembert* (EDdA) et d'un autre la forme la plus récente de ce genre avec l'encyclopédie Wikipédia (Wiki) - en fonction de différents domaines thématiques (e.g., *sciences, société*). Cette étude, tout en s'inscrivant dans la lignée d'autres études mettant en perspective une

encyclopédie moderne avec EDdA¹, est la première à problématiser le rôle déterminant que joue le domaine thématique sur l’emploi des formes linguistiques de la médiation dans le genre encyclopédique.

2. Présentation des corpus et des domaines du savoir encyclopédique

Nous avons sélectionné un certain nombre de domaines dans chacune de nos encyclopédies, sans chercher à établir de correspondances entre eux, dans la mesure où chaque encyclopédie présente un mode d’organisation de la connaissance qui lui est propre.

2.1. Wikipédia

Le corpus Wikipédia que nous avons constitué compte 1 953 articles pour un total de 9 828 019 *tokens*. Il est composé d’articles listés sur les « Pages populaires » de dix grands projets thématiques de Wikipédia. Ces projets sont : *Arts, Biologie, Géographie, Histoire, Politique, Religions & Croyances, Sciences, Société, Sport et Technologies*. Pour chacun de ces dix projets, nous avons extrait automatiquement et aléatoirement 200 articles². Nous avons ensuite nettoyé les articles de façon à ne retenir que le texte à visée informative (en supprimant par exemple les notes bibliographiques, les indications de pages homonymes, les liens hypertextes, etc.) et supprimé les doublons aléatoirement. A partir de ce corpus, nous avons constitué un sous-corpus de travail de 4 626 260 tokens, composé de cinq projets permettant de représenter chacun un domaine du savoir différent (*Sport* : 1 049 911 tokens, *Biologie* : 720 980, *Sciences* : 771 713, *Politique* : 1 219 535, *Géographie* : 864 021).

2.2. Encyclopédie des Lumières

Le corpus EDdA au format XML-TEI dont nous disposons a été publié par ARTFL (Morrissey et Roe, 2020)³. Il contient 74 198 articles pour un total de 24 611 707 tokens. Pour opérer une classification thématique des articles de EDdA, nous nous sommes appuyées sur l’édition collaborative et critique ENCCRE, qui distingue 44 ensembles de domaines qui regroupent 327 domaines de connaissances et de pratiques abordées dans l’Encyclopédie (Guilbaud, 2017), regroupant eux-mêmes les articles (identifiés ou non par un titre)⁴. Les métadonnées associées aux articles encyclopédiques du corpus renseignent, pour la plupart d’entre eux, leur titre, mais seules les métadonnées de 61 435 articles en renseignent le domaine. Nous avons sélectionné 21 723 articles qui appartiennent à 20 ensembles de domaines, parmi les 44 identifiés par ENCCRE. Nous les avons regroupés en 4 grands domaines du savoir, en nous fondant sur la classification opérée par Diderot dans le *Système figuré des connaissances humaines* qui figure

¹ Par exemple, Blumenthal (2017) s’intéresse à la façon dont les savoirs sont présentés dans *Universalis*, qui privilégie, contrairement à EDdA, les locutions introductrices de point de vue (*dans la perspective de, du point de vue de*) ; Doutreix (2020) mentionne un rapport à la connaissance différent dans Wiki et EDdA : distinguer le vrai du faux dans EDdA et « présenter de manière impartiale les différents points de vue pertinents » dans Wiki, cf. l’article de Wikipédia -*Wikipédia : Neutralité de point de vue* (in Doutreix, 2020 : 64-65).

² L’extraction des articles a été faite le 16 novembre 2023. Les « pages populaires » des projets auxquels ils appartiennent sont mises à jour chaque début de mois par des bots.

³ Nous remercions l’équipe du projet DISCO (CollEx-Persée) pour le partage du corpus EDdA annoté (Vigier *et al.* 2020).

⁴ Pour exemple, l’ensemble de domaines « Religion » regroupe 7 domaines, dont le domaine « Critique Sacrée », qui regroupe lui-même 315 articles (e.g., Hérésie, Mystique, Pâque des Juifs).

au frontispice de l'Encyclopédie. Le sous-corpus de travail que nous avons constitué contient un total de 9 112 332 tokens. Nous avons ajusté les labels aux systèmes des connaissances du lecteur contemporain. Nous distinguons ainsi le domaine *Sciences* : 4 201 587 tokens (e.g. Mathématiques, Physique), *Société* : 3 715 324 (Philosophie, Politique, Grammaire, Economie), *Sciences Humaines* : 682 430 (Géographie, Histoire) et *Sciences Techniques* : 512 991 (Métiers du bois, Métier du papier).

3. Les constructions médiatives

Cette étude prend pour objet un sous-type de constructions médiatives, celles qui correspondent à un patron énonciatif où il est question de l'attitude du locuteur exprimée par un verbe sur un état de choses⁵. Dans le cadre de la théorie de la polyphonie linguistique présentée dans Carel et Ducrot (2009 : 37), les auteurs analysent ce type de construction en reprenant la distinction de Port Royal qui différencie, pour un énoncé du type *x dit que q* une lecture correspondant à une façon de dire *q* – dite lecture modale, et une lecture correspondant à une façon de faire connaître le discours de *x*, – dite lecture attributive. En nous inspirant de ce cadre, nous n'avons retenu que les constructions passibles d'une lecture modale telles que celles en *je/nous/on/il* impersonnel, en excluant les autres personnes grammaticales, davantage susceptibles d'être interprétées comme attributives. A noter que pour des raisons de contraintes concernant les consignes de rédaction de Wiki⁶, seules les constructions en *on* et *il* impersonnel ont été retenues dans cette encyclopédie.

3.1. Extraction des constructions médiatives

L'extraction des constructions médiatives a été réalisée en deux temps, à partir de TXM (Heiden et al., 2010). Une première requête a permis d'extraire les constructions dont la syntaxe correspond à un pronom (*je, nous, on, il* impersonnel) suivi d'un verbe, suivi d'un mot optionnel (adverbe, autre verbe, adjectif ou participe passé) suivi du complémenteur *que* (e.g. ; *on croit vraiment que ; je veux dire que ; il est possible que*). Afin de diminuer le nombre de verbes qui n'auraient pas de valeur médiative dans ces constructions, nous avons uniquement pris en compte les verbes au présent dans Wiki ; pour EDdA nous avons pris en compte toutes les flexions verbales sauf le passé simple – qui ne donne jamais lieu à une lecture modale. Les résultats ont été extraits et triés pour s'assurer de la valeur médiative des constructions, en ayant recours au concordancier pour les cas les plus ambigus. Après ce tri, une deuxième requête faite au moyen des verbes retenus a permis de sortir uniquement les constructions pertinentes (en excluant par exemple les pronoms objet, e.g. ; *il nous demande*).

3.2. Classification des constructions médiatives

Les constructions médiatives ainsi extraites ont fait l'objet d'une classification, fondée sur le type d'indication qu'elles peuvent donner sur le contenu qu'elles introduisent (i.e., la complétive *p*). Pour construire cette classification, nous nous sommes appuyées sur la typologie des *verba dicendi* dans la presse d'information établie par Monville-Burston (1993). Son étude

⁵ Les constructions que nous prenons en compte sont un sous-ensemble de formes de la médiation qui intègrent notamment des adverbiaux cadratifs d'énonciation dont la fonction est proche de celle de nos structures – *selon, d'après N* (voir Charolles et Vigier, 2005).

⁶ Le recours à la première ou deuxième personne est proscrit dans Wiki, cf. « Comment rédiger un bon article » https://fr.wikipedia.org/wiki/Aide:Comment_rédiger_un_bon_article

a dépouillé quarante articles du *Monde*, de *France-Soir* et des *Nouvelles Calédonniennes* entre décembre 1989 et juin 1991 en relevant toutes les « occurrences de *verba dicendi* (ou expressions verbales très lexicalisées) utilisées par les journalistes pour rapporter un message oral ou bien pour rendre compte d'un contenu d'un texte écrit » Monville-Burston (1993 : 49). La démarche et le genre du corpus étant proches de nos objectifs, nous avons utilisé sa typologie comme point de départ en intégrant des constructions qui dépassent les frontières des *verba dicendi*. Monville-Burston (1993 : 65) établit sa catégorisation sur la base de cinq fonctions liées à la valeur de ces verbes en contexte : primitif (*dire*), diffusion de l'information (*déclarer, affirmer, annoncer*), jugement impartial (*estimer*), clarification (*expliquer, ajouter, souligner, préciser, indiquer*). Notre catégorisation et les définitions associées étant fondées sur nos corpus, elles s'écartent en partie de sa classification. Nous prenons en compte la valeur qu'a l'ensemble de la construction pour lui attribuer une classe, ce qui nous permet d'intégrer le même verbe dans dans différentes catégories. Ainsi *il paraît que* est catégorisée comme une construction dont la valeur correspond à *diffusion de l'information*, voir (1), tandis qu'*il paraît certain que* est catégorisée dans *évaluation épistémique*, voir (2).

(1) Le peuple appelle cette fête Notre-Dame de Mars, à cause du mois où elle tombe. *Il paraît que* cette fête est de très-ancienne institution dans l'Eglise Latine (EDdA, volume01-2815, Annonciation)

(2) (...) Cependant *il paraît certain qu'*Hippocrate, à en juger par ses écrits, ne fit jamais usage que de peu de remèdes & des plus simples. (EDdA, volume08-1170, Hippocratism)

Pour chaque classe présentée ci-dessous, nous donnons sa définition et quelques exemples de constructions, qui permettent de faire ressortir de façon prototypique la valeur de la classe à laquelle elles appartiennent (liste complète en annexe). Les cinq premières classes sont communes aux deux encyclopédies ; les deux dernières sont uniquement présentes dans EDdA, étant formées essentiellement de verbes avec des sujets à la première personne.

Les constructions à valeur de (1) **clarification** ont pour fonction de mettre en évidence l'importance de dire *p* (*il est expliqué que, il est souligné que, il est spécifié que, etc*). Celles à valeur de (2) **jugement** ont pour fonction de décrire le processus cognitif ou perceptif qui conduit à établir *p* (*il résulte que, on peut penser que, je crois que, etc*). Celles à valeur de (3) **diffusion de l'information** ont pour fonction d'indiquer que *p* est mis à disposition (*on entend que, il est dit que, il paraît que, etc*). Celles à valeur de (4) **évaluation épistémique** ont pour fonction d'indiquer le crédit qui est accordé à *p* (*il paraît certain que, on est assuré que, il est douteux que, etc*). Celles à valeur de (5) **évaluation axiologique** ont pour fonction d'indiquer un jugement de valeur sur *p*, que ce jugement soit de nature appréciative ou indiquant que *p* est à réaliser (*il est malheureux que, il est normal que, il faut que, etc*). Celles à valeur de (6) **engagement** ont pour fonction d'indiquer qu'un agent assume *p* (*je dis que, j'avance que, je garantis que, etc*). Enfin, les constructions à valeur de (7) **réaction** ont pour fonction d'indiquer que *p* vient en réaction à un dire/faire antérieur ou à venir (*je dis simplement que, comment peut-on douter/nier, etc*).

Notez que la catégorie *engagement* intègre en partie les verbes dits « primitifs » de Monville-Burston, dans la mesure où les verbes *dicendi* les plus neutres (*je dis que*) servent à mettre en avant l'engagement de l'agent qui introduit *p*. Par ailleurs, la catégorie *réaction* comprend également des formules à questions rhétoriques avec le pronom *on* (*comment peut-on soutenir que*). Elles sont destinées à réagir à un savoir présupposé partagé, d'où leur intégration à cette classe.

4. Finalité des méthodes statistiques et résultats

4.1. Fréquence d'usage des constructions médiatives

Dans un discours encyclopédique, le savoir transmis peut être présenté comme relevant d'une source identifiable, quel qu'en soit le degré de généralité (*Les philosophes disent que ; Epicure dit que* ou *X dit que*) ou d'une source pour laquelle la question de l'identification ne se pose pas (y compris celles avec le pronom *je* qui ne sont pas vouées à faire ressortir l'instance responsable de la médiation). Nous comparons les fréquences relatives de ces deux types de constructions – avec et sans visée d'identification – en recherchant les constructions à sujets nominaux et sujets pronominaux à la troisième personne au moyen de la deuxième requête décrite sous § 3.1. Pour les constructions sans visée d'identification, nous différencions les fréquences du pronom *je* de celles des pronoms à valeur indéfinie (*on, il* impersonnel).

Constructions		Wikipédia 5 domaines (4 626 160 tokens)		EDdA 4 domaines (9 112 332 tokens)	
Visée d'identification	Pronom	# brutes	# relatives	# brutes	# relatives
Sans	<i>on/il</i> (impers.)	746	161,26	1 991	218,1
	<i>je</i>	-	-	1 052	115,4
	Total	746	161,26	3 043	333,5
Avec	3 ^{ème} personne/SN	1 431	309,33	2 651	290,9

Tableau 1 Fréquences relatives des constructions à vocation attributive par rapport aux autres constructions médiatives dans les sous-corpus

Les fréquences relatives montrent une différence de pratique entre les deux encyclopédies. Dans EDdA la proportion des constructions sans visée d'identification est importante, atteignant un total de 333 occurrences par million de mots lorsqu'on prend en compte les *je*⁷, surpassant celle des constructions à visée d'identification (290). On voit donc se dessiner une rhétorique qui reflète directement la subjectivité du scripteur avec *je*, ou qui dilue sa responsabilité avec les constructions en *on* et *il* impersonnel. Wiki en revanche fait une large part aux constructions à visée d'identification (presque le double de celles sans visée d'identification). Ces résultats peuvent être considérés comme attendus compte tenu des stéréotypes associés au savoir encyclopédique moderne, qui se doit de sourcer les informations. Toutefois, en allant plus loin dans l'étude de la façon dont ces constructions médiatives sans visée d'identification sont utilisées selon chaque domaine, on voit se dessiner des particularités par domaines manifestant à la fois des divergences et des convergences entre les deux encyclopédies.

4.2. L'instance responsable de la médiation à travers les domaines du savoir encyclopédique

⁷ Le pronom *nous* (non pris en compte pour des questions de tri – pour différencier le sujet de l'objet) ne ferait qu'accroître l'écart entre constructions. Pour une étude spécifiquement consacrée à ces pronoms nous renvoyons à Vigier (2016).

Les recherches présentées ici ont comme objectif de déterminer si le domaine impacte la part que prennent les constructions médiatives. Dans un premier temps, au moyen du calcul de spécificité sur TXM (Lafon, 1980 ; Lebart et Salem, 1994) nous relevons la sur- ou sous-représentation de ces constructions par domaine pour chaque encyclopédie. Dans un second temps, nous différencions ces constructions en fonction du pronom pour avoir une idée des sources de médiation privilégiées selon les domaines.

4.2.1. Wiki

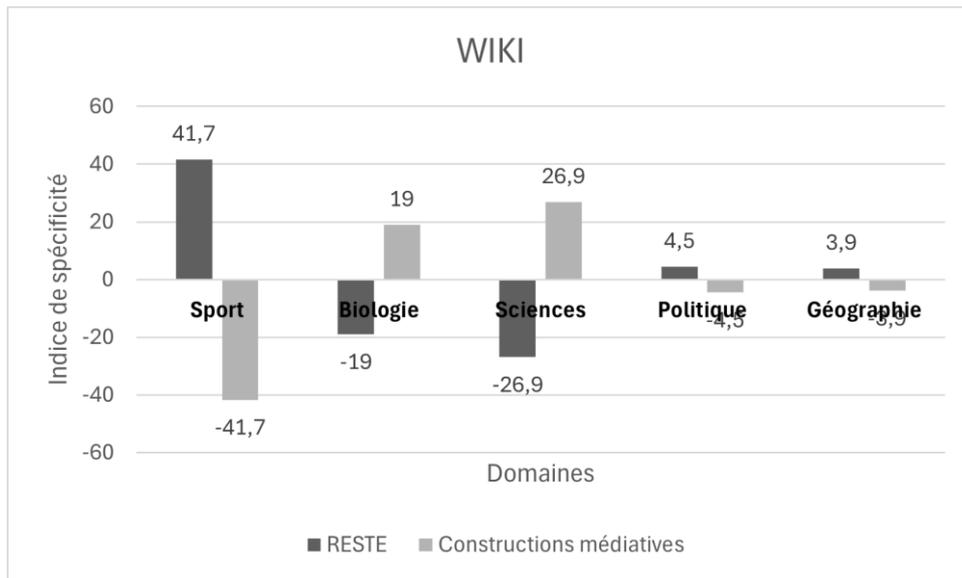


Figure 1 Sur- et sous-représentation des constructions médiatives par domaine de connaissance encyclopédique (Wiki)

La Figure 1 montre que l'usage des constructions médiatives est sensible au domaine. Le recours à ces constructions est privilégié dans les domaines *biologie* et *sciences*. L'usage de ces formes est très nettement évité dans le domaine *sport*, tout comme dans les domaines *politique* et *géographie*, dans une moindre mesure.

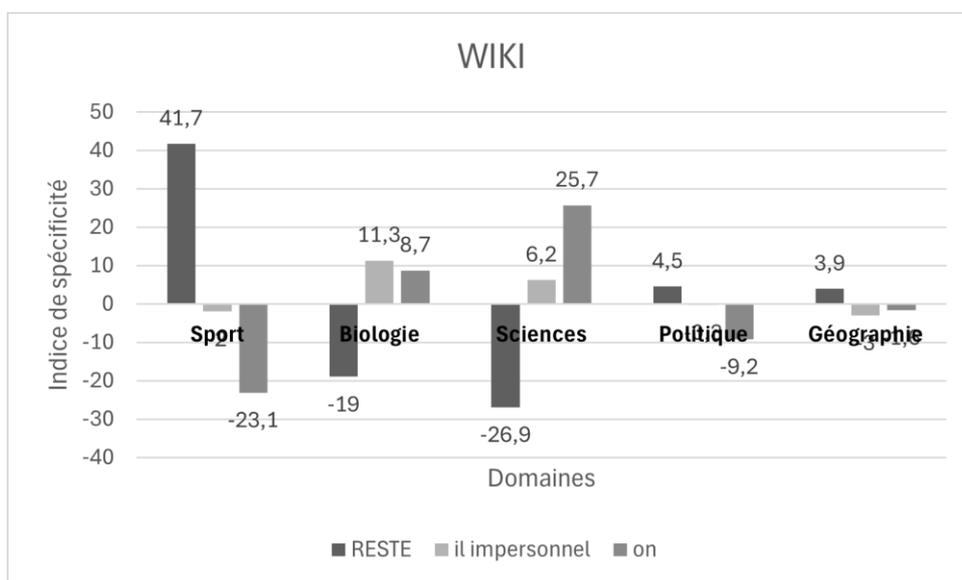


Figure 2 Sur- et sous-représentation des instances responsables de la médiation par domaine de connaissance encyclopédique (Wiki)

En ce qui concerne le choix du pronom, on observe dans la Figure 2 que le choix des pronoms *on* et *il* impersonnel suit une tendance d'usage similaire dans chaque domaine : les deux sont soit sur-représentés, soit sous-représentés, soit l'un des deux est banal, mais il n'y a aucun cas où l'un est sur-représenté et l'autre sous-représenté, ce qui laisse paraître une utilisation du *on* très semblable à celle du *il* impersonnel : les deux permettant d'effacer l'instance responsable de *p*.

4.2.2. EDdA

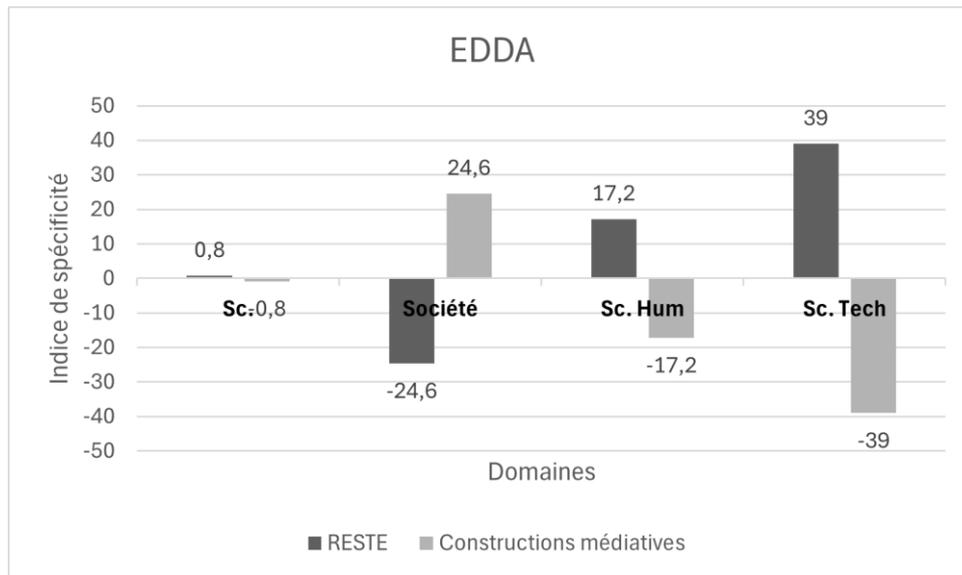


Figure 3 Sur- et sous-représentation des constructions médiatives par domaine de connaissance encyclopédique (EDdA)

Dans EDdA également, des contrastes apparaissent nettement selon les domaines. La distribution des constructions médiatives n'y est pas aléatoire : elles sont sur-employées dans le domaine *société*. Elles sont dans la fenêtre de banalité pour le domaine *sciences* et sous-employées en *sciences humaines* et *sciences techniques*.

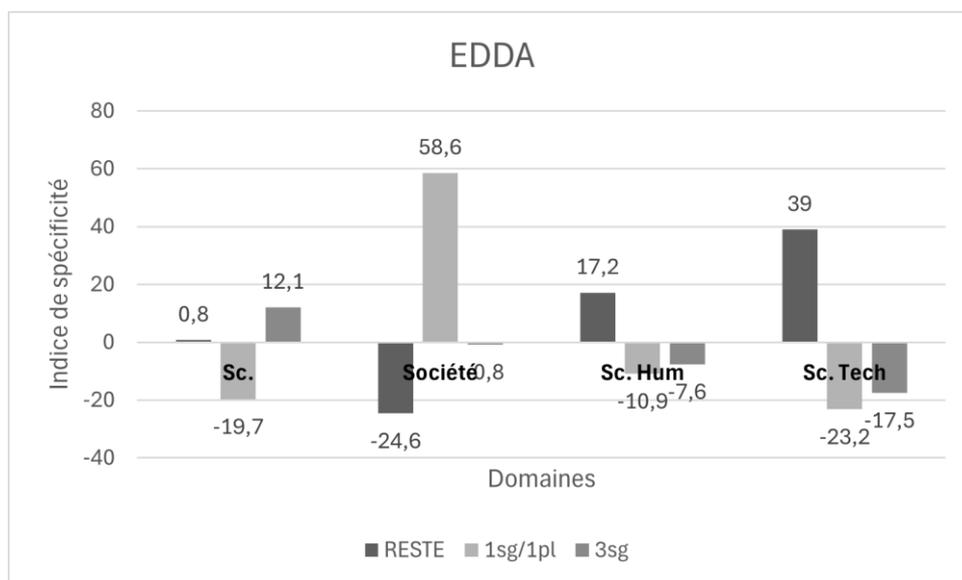


Figure 4 Sur- et sous-représentation des instances responsables de la médiation par domaine de connaissance encyclopédique (EDdA)

La Figure 4 permet de dissocier la première personne (*je* et *nous*) du *on* ou *il* impersonnel (étiquetés 3^e pers. médiative)⁸. Le contraste entre ces deux possibilités de médiation manifeste un choix rhétorique précis : soit la construction médiative est ‘personnalisée’ avec un *je* ou un *nous*, soit elle est ‘dépersonnalisée’ avec *on* ou *il* impersonnel. Dans le domaine *sciences*, la construction dépersonnalisée est sur-utilisée au détriment des constructions personnalisées. À l’inverse, dans le domaine *société*, la construction personnalisée est sur-utilisée et la construction dépersonnalisée est banale. Dans *les sciences humaines* et *les sciences techniques* ce sont les deux constructions qui sont sous-utilisées. Cette première étude montre donc que le choix des constructions médiatives est dépendant du domaine de connaissance qui est en jeu.

4.3. Classes sémantiques privilégiées à travers les domaines du savoir encyclopédique

Pour gagner en finesse d’analyse concernant la sensibilité des différentes constructions médiatives aux domaines du savoir encyclopédique, nous allons utiliser des analyses factorielles des correspondances – AFC (Brunet, 2012) pour chacun de nos corpus afin de visualiser la proximité ou l’éloignement entre les classes de constructions et les domaines. Cette méthode offre une visualisation des « ressemblances et [des] oppositions les plus significatives » (Mayaffre, Pincemin et Poudat, 2019 : 112) entre les sous-parties du corpus, permettant « d’en repérer les faits saillants en termes de proximités » (Leblanc, 2015 : 43-44). Pour l’interpréter, il faut surtout être attentif aux éléments excentrés, ceux qui se situent près du centre manifestant un comportement moins significatif (Poudat et Landragin, 2017 : 108).

4.2.1. Wiki

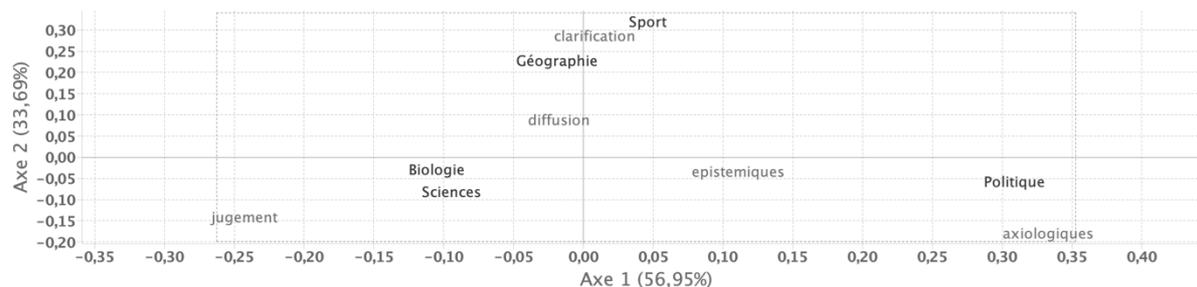


Figure 5 AFC des classes sémantiques des constructions médiatives selon les domaines de connaissance (Wiki)

L’AFC présentée dans la Figure 5 permet de révéler que certains domaines sont très similaires quant aux classes de constructions médiatives qui y sont mobilisées : le domaine *biologie* est très proche de *sciences* et *géographie* de *sport*. Le premier duo est associé à la classe *jugement* et le second aux classes *diffusion de l’information* et *clarification*. Le domaine *politique*, lui, se singularise des deux duos par sa proximité avec la classe des *axiologiques*.

⁸ Nous avons réuni *on* et *il* impersonnel pour les faire contraster avec *je/nous*. Une étude ultérieure prendra en compte les valeurs de *on* qui peuvent, dans cette encyclopédie, aller du côté du *nous* ou de l’impersonnel à la différence de Wikipédia où *on* dans la plupart des cas a une valeur proche à celle du *il* impersonnel (cf. les indices de spécificité qui laissent paraître une utilisation du *on* très semblable à celle du *il* impersonnel).

4.2.2. EDdA

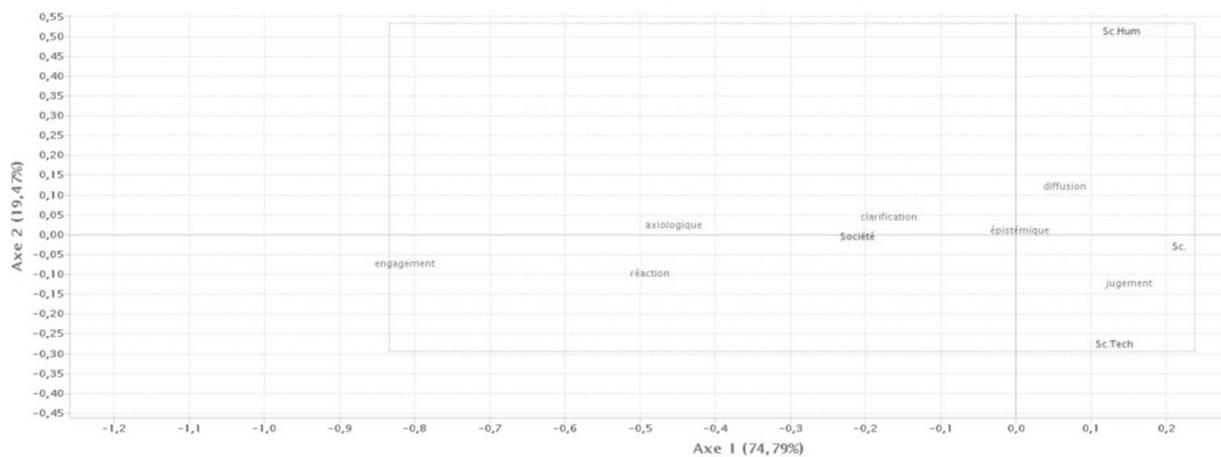


Figure 6 AFC des classes sémantiques des constructions médiatives selon les domaines de connaissance (EDdA)

L'AFC présentée dans la Figure 6 met en évidence trois tendances : (i) le domaine *société* se différencie des trois autres domaines qui sont du même côté de l'axe vertical : il est associé aux classes *clarification*, *axiologie*, *réaction* et *engagement*. (ii) La classe *épistémique* qui est au centre ne caractérise aucune des domaines en particulier. (iii) La classe *jugement* est plus du côté du domaine *sciences* et *sciences techniques* et la *diffusion* plus du côté *sciences humaines*.

4.2.3. Synthèse

Dans chaque encyclopédie, un domaine se singularise des autres : dans Wiki c'est le domaine *politique*, qui est proche de la classe *axiologie* et dans EDdA c'est le domaine *société*, également proche de la classe *axiologie*. Dans les deux encyclopédies, les classes *diffusion* et *épistémique* sont celles qui occupent la position la plus proche du centre, montrant une utilisation peu différenciée de ces constructions médiatives selon les domaines.

On remarque également des points de convergence entre les deux fenêtres synchroniques interrogées. Les stéréotypes concernant la diffusion des connaissances dans les domaines relevant des sciences dures semblent stables : dans les deux encyclopédies c'est la classe *jugement* qui se trouve être proche des domaines *sciences* et *sciences techniques* (EDdA) et des domaines *biologie* et *sciences* (Wiki), classe, qui, nous le rappelons, met en avant les opérations cognitives que le sujet fait pour introduire l'information.

5. Conclusion

Les deux méthodes statistiques montrent que le domaine a un impact sur les constructions médiatives et cela tant dans EDdA que dans Wiki. Le domaine a une incidence non seulement sur le sur- ou sous-emploi des constructions médiatives en général, mais également sur le choix des classes sémantiques. Ces choix montrent que la médiation de la connaissance est régulée selon des traditions épistémologiques dans la mesure où chaque scripteur a à sa disposition l'intégralité des formes quel que soit le domaine, presque aucun domaine n'étant dépourvu

d'une classe particulière de constructions médiatives⁹. Le savoir est donc toujours l'objet d'un choix de transmission lié à un domaine, choix qui induit une certaine façon d'appréhender la réalité qu'il est censé décrire.

Bibliographie

- Blumenthal P. (2017). D'une encyclopédie à l'autre. L'ascension des locutions prépositionnelles. *Langages*, 206 (2), 123-138.
- Brunet E. (2012). Nouveau traitement des cooccurrences dans Hyperbase. *Corpus*, 11, 219-246.
- Carel M. et Ducrot O. (2009). Mise au point sur la polyphonie. *Langue française*, 164 (4), 33-43.
- Charolles M. et Vigier D. (2005). Les adverbiaux en position préverbale : portée cadrative et organisation des discours. *Langue française*, 148, 9-30.
- Doutreix M.-N. (2020). *Wikipedia et l'actualité : qualité de l'information et normes collaboratives d'un média en ligne*. Presses Sorbonne nouvelle.
- Guilbaud, A. (2017). L'ENCCRE, édition numérique collaborative et critique de *L'Encyclopédie, Recherches sur Diderot et sur l'Encyclopédie*, 52, 5-22.
- Heiden S. Magué J.-P. et Pincemin B (2010). TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, Rome, Italie, 1021-1032.
- Lafon, P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. In: *Mots*, n°1: 127-165.
- Lebart L. et Salem A. (1994). *Statistique textuelle*. Paris : Dunod.
- Leblanc J.-M. (2015). Proposition de protocole pour l'analyse des données textuelles : pour une démarche expérimentale en lexicométrie. *Nouvelles perspectives en sciences sociales*, 11 (1), 25-63.
- Mayaffre D., Pincemin B. & Poudat C. (2019). Explorer, mesurer, contextualiser. Quelques apports de la textométrie à l'analyse du discours. *Langue française*, 203 (3), 101-115.
- Poudat C. et Landragin F. (2017). *Explorer un corpus textuel : Méthodes, pratiques, outils*. Louvain-la-Neuve : De Boeck Supérieur.
- Robert M. et Glenn R. (2020). "The ARTFL Encyclopédie: Towards a Hermeneutical Aesthetic", In Simon Burrows and Glenn Roe (eds.), *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies*, Oxford University Studies in the Enlightenment, Voltaire Foundation in association with Liverpool University Press.
- Vigier, D. (2016). Pronoms personnels JE, NOUS, ON et énonciation dans le discours encyclopédique. Une étude sur corpus instrumenté. Les phénomènes énonciatifs dans le discours encyclopédique : approche linguistique, littéraire et philosophique, Neuchâtel, Suisse.
- Vigier, D., Moncla, L., Brenon, A. McDonough, K & Joliveau, T. (2020). Classification des entités nommées dans l'Encyclopédie ou dictionnaire raisonné des sciences des arts et des métiers par une société des gens de lettres (1751-1772). *Congrès Mondial de Linguistique Française - CMLF 2020*.

⁹ Exception faite de la classe *engagement* dans les domaines *sciences humaines* et *sciences techniques* de EDdA.