

**Attribuer des connaissances à un tiers non-spécifié :
Une étude textométrique des assertions médiatives dans Wikipédia**

Ascribing knowledge to a non-specified source of information :
A textometric study of evidential assertions in Wikipedia

Corinne Rossari et Chloé Tahar

Université de Neuchâtel
corinne.rossari@unine.ch
chloe.tahar@unine.ch

Soumis le 30/08/2026

Creative Commons Attribution 4.0 International (CC BY 4.0)
© 2026 Prénom Nom de famille

Pour citer cet article

Nom de famille, Prénom 2026. Titre. Sous-titre éventuel. *Studia linguistica romanica* 2026.17, page-
page. <https://doi.org/10.25364/19.2026.17.4>.

Résumé

Dans la médiation du savoir encyclopédique, il n'est pas rare que des connaissances soient attribuées à une source d'information tierce (Doutreix, 2020). Cet article se concentre sur des cas où ces connaissances sont attribuées à un tiers non-spécifié, dans le Wikipédia français. Notre étude vise à comparer empiriquement les constructions qui attribuent des connaissances à une source indéfinie (le pronom *on*) et celles qui les attribuent à une source effacée (le pronom explétif *il*). Sur le plan méthodologique, nous montrons l'intérêt de prendre en compte les domaines du savoir encyclopédique comme dimension de variation pertinente permettant de faire ressortir, au moyen des méthodes statistiques de la textométrie (Lafon, 1980), des contrastes dans l'emploi de ces assertions médiatives (Faller, 2019).

Mots-clés

Assertions médiatives, discours encyclopédique, textométrie.

Abstract

In the mediation of encyclopedic knowledge, it is not uncommon for knowledge to be ascribed to a third-party source of information (Doutreix, 2020). This paper focuses on cases in which knowledge is ascribed to a non-specified third party, within the French Wikipedia. Our study aims to provide empirical evidence for a comparison between constructions that ascribe knowledge to an indefinite source (the pronoun *on*, "one") and those in which the source is left unexpressed (the expletive pronoun *il*, "it"). Methodologically, we highlight the value of using domains of encyclopedic knowledge as a relevant dimension of variation for bringing out, through the statistical methods of textometry (Lafon, 1980), contrasts in the use of evidential assertions (Faller, 2019).

Keywords

Evidential assertions, encyclopedic discourse, textometry.

Sommaire

1 Introduction	4
2 Corpus et méthodologie.....	6
2.1 Corpus	6
2.2 Extraction des constructions attributives.....	7
2.3 Analyses statistiques.....	8
3 Résultats	9
3.1 Résultats au niveau macro	9
3.2 Résultats au niveau micro.....	10
4 Analyse	12
5 Conclusion.....	13
Références bibliographiques	15

1 Introduction

[1] Le discours encyclopédique fait intervenir de manière non-triviale le discours d'autrui (Doutreix, 2020). De ce fait, dans la médiation du savoir encyclopédique, il n'est pas rare que des connaissances soient attribuées à une source d'information tierce. S'inscrivant dans le prolongement des travaux de Rossari et al. (2024) et de Rossari & Tahar (2025) sur les attributions de connaissances à un tiers dans le discours encyclopédique, cet article s'intéresse aux cas où des connaissances sont attribuées à un tiers non-spécifié, dans le Wikipédia français. Les attributions de connaissances sont ici définies comme des attributions d'attitude mentale ou de parole rapportée au moyen d'un verbe introducteur de complétive. Celles-ci attribuent au sujet d'ancrage du verbe la responsabilité épistémique du contenu p de la complétive.

[2] Notre étude compare empiriquement deux types de constructions qui attribuent des connaissances à une source qui n'est pas identifiable : les attributions dont le sujet est le pronom *on*, tel qu'en (1), et celles dont le sujet est un *il* explétif, tel qu'en (2).

(1) *Crise de la vache folle*, Biologie (Wikipédia)

En juillet 2012, **on estime que** la maladie a fait 224 victimes, dont 173 au Royaume-Uni (...)

(2) *Théorie du complot*, Société (Wikipédia)

Il a été démontré que le biais d'intentionnalité pouvait avoir un impact sur la croyance des individus dans la théorie du complot.

L'impossibilité d'identifier la source d'information repose sur des mécanismes distincts dans ces deux types de constructions. Dans la première, la non-identification de la source tient au fait que le pronom *on* désigne une source sous-déterminée ou indéterminée, comme relevé dans Fløttum et al. (2007) ; dans la seconde, elle tient au fait que le *il* étant, du point de vue de sa syntaxe, explétif, il ne renvoie à aucun référent. La source à laquelle peut être attribuée la responsabilité de la prise en charge du contenu p demeure ainsi non-identifiable dans les deux types de constructions, soit parce qu'elle est encodée par une forme indéfinie, soit parce qu'elle est effacée sur le plan grammatical.

[3] Les attributions de connaissances à un sujet non-spécifié sont des assertions médiatives (Faller, 2019), dans la mesure où, en masquant l'instance responsable de la prise en charge du contenu de la complétive, elles visent moins à présenter le point de vue d'un sujet identifiable sur p qu'à intégrer l'information véhiculée par p aux connaissances communes (*common ground* ; cf. Rossari & Tahar, 2025). Notre objectif, sur le plan méthodologique, est de montrer l'intérêt de prendre en compte les domaines du savoir encyclopédique comme dimension de variation pertinente permettant de faire ressortir, au moyen des analyses statistiques de

la textométrie (Lafon, 1980), des contrastes dans l'emploi de ces assertions médiatives.

[4] Afin de délimiter les verbes susceptibles de participer aux attributions de connaissances, nous prenons aussi bien en compte des verbes d'attitude mentale que des verbes de parole rapportée. En ce qui concerne les verbes d'attitude mentale, nous ne retenons que les verbes dits représentationnels, en nous appuyant sur la distinction sémantique proposée par Anand & Hacquard (2013) entre verbes représentationnels et verbes de préférence. Dans leur cadre, les premiers quantifient sur des états d'information, tandis que les seconds impliquent une comparaison entre différentes alternatives. Cette distinction nous permet d'isoler les verbes qui attribuent à une source la responsabilité épistémique d'un contenu *p*, par opposition à ceux qui attribuent une préférence au sujet grammatical (tels que les verbes volitifs, *souhaiter*, ou injonctifs, *ordonner*). En ce qui concerne les verbes de parole rapportée, nous ne retenons que les verbes rapportant un acte de langage assertif, en nous appuyant sur la distinction établie par Krifka (2024) entre actes de langage assertifs et déclaratifs. Dans le cadre de Krifka, les actes de langage assertifs visent à ajouter l'information véhiculée par *p* aux connaissances communes, tandis que les actes de langage déclaratifs visent à changer l'état du monde. Cette distinction nous permet ainsi de ne pas retenir des verbes qui, comme *déclarer* ou *annoncer*, rapportent un acte de langage déclaratif.

[5] À partir des seuls verbes attestés dans notre corpus, nous avons établi des catégories sémantiques correspondant aux différentes opérations cognitives ou discursives qu'ils expriment. Cette subdivision ne vise pas à proposer une modélisation de la classe des verbes représentationnels ou de celle des verbes assertifs, mais répond avant tout à un objectif méthodologique : limiter la dispersion lexicale des données en regroupant des verbes sémantiquement proches, afin de faire apparaître plus nettement des tendances dans leur répartition selon les domaines du savoir. Les dix catégories retenues, inspirées notamment d'Hacquard & Wellwood (2011) et de Monville-Burston (1993), sont les suivantes : les verbes de **savoir** (*savoir, apprendre*), les verbes de **jugement** (*estimer, considérer, penser*), les verbes de **conjecture** (*supposer, faire l'hypothèse*), les verbes d'**observation** (*voir, observer, s'apercevoir, remarquer*), les verbes de **démonstration** (*montrer, démontrer, établir, prouver*) les verbes de **fiction** (*imaginer*) et les verbes d'**anticipation** (*prédire*), ainsi que les verbes de **dire** (*dire, affirmer*), les verbes d'**acceptation** (*admettre, accepter*, ainsi que leurs contreparties négatives : *réfuter, nier*) et les verbes de **clarification** (*ajouter, préciser, indiquer, souligner*).

[6] La suite de cet article présente d'abord, dans la section 2, le corpus et la méthodologie. La section 3 examine ensuite la manière dont les deux types de constructions attributives à une source non-spécifiée se répartissent selon les domaines qui structurent l'architecture interne de l'encyclopédie Wikipédia, ainsi que les catégories de verbes mobilisées dans chacune de ces deux constructions. La section 4 donne une analyse synthétique des tendances observées. La section 5 conclut.

2 Corpus et méthodologie

[7] Nous présentons dans cette section le corpus ainsi que la méthodologie adoptée pour, dans un premier temps, extraire les constructions en *on* et de celles en *il* impersonnel, et d'analyser, dans un second temps, leur répartition dans les différents domaines de la connaissance de Wikipédia au moyen des méthodes statistiques de la textométrie. Le corpus, les scripts Python utilisés pour l'extraction des données, les scripts en R utilisés pour l'analyse statistique ainsi que les bases de données d'occurrences collectées sont disponibles dans une annexe en ligne déposée sur la plateforme OSF : <https://doi.org/10.17605/OSF.IO/A829P>.

2.1 Corpus

Domaine	Nombre de tokens	Nombre de textes
Arts	950'759	196
Biologie	720'980	195
Géographie	824'021	198
Histoire	1'388'046	194
Politique	1'219'535	193
Religions et croyance	955'924	196
Sciences	771'712	193
Société	1'079'422	191
Sport	1'049'911	200
Technologies	813'337	195

Tableau 1 : Architecture interne du corpus Wikipédia

[8] L'encyclopédie Wikipédia se particularise par son mode de rédaction collaboratif, sa gratuité et son actualisation continue, caractéristiques qui font d'elle un corpus où les usages sont les plus susceptibles d'être représentatifs de la langue actuelle. Le corpus Wikipédia a été constitué en 2024 par l'Institut des Sciences du langage de l'Université de Neuchâtel. Il compte 1952 articles pour un total de 9'813'647 tokens. Il est composé d'articles listés sur les « Pages populaires » de chacun des dix grands projets thématiques de Wikipédia¹ retenus dans le cadre de notre étude. Notez que les Pages Populaires sont établies automatiquement par l'encyclopédie en ligne, en fonction du nombre de vues qu'elles ont reçu. Celles-ci sont consultables sur une page dédiée, au sein de chaque projet thématique. Les dix projets retenus dans le cadre de notre projet sont : Arts (950'759 tokens), Biologie (720'980 tokens), Géographie (824'021 tokens), Histoire (1'388'046 tokens), Politique (1'219'535 tokens), Religions et croyances (955'924 tokens), Sciences (771'712 tokens), Société (1'079'422 tokens), Sport (1'049'911 tokens) et Technologie (813'337 tokens). Pour chacun de ces dix projets, nous avons extrait automa-

¹ <https://fr.wikipedia.org/wiki/Projet:Accueil>

tiquement et aléatoirement 200 articles associés aux « Pages populaires », en excluant ensuite les articles trop courts. Nous avons nettoyé les articles de façon à ne retenir que le texte à visée informative (en supprimant les notes bibliographiques, les indications de pages homonymes, les liens hypertextes, etc.). Les fichiers XML du corpus ainsi constitué sont annotés sur TXM avec TreeTagger. Les fichiers XML comportent, pour chaque token, sa forme graphique (“word”) ainsi que son lemme (“frlemma”) et sa forme morphosyntaxique (“frpos”).

2.1 Extraction des constructions attributives

[9] L'extraction des constructions attributives à sujet *on* et *il* impersonnel a été réalisée à l'aide de scripts Python appliqués au corpus Wikipédia. Deux ressources lexicales, comprenant au total 132 verbes, ont été utilisées pour repérer les verbes dans le corpus : (i) une liste de 83 verbes simples (e.g., *croire*, *penser*) et (ii) une liste de 49 verbes complexes (*être certain*, *avoir l'impression*). Ces listes ont été établies à partir d'une première requête englobante sur la plateforme TXM, visant à extraire les lemmes de tous les verbes du corpus suivis par le complémenteur *que* dans un empan maximal de deux tokens vers la droite.

[10] Pour le repérage des constructions en *on*, le patron général recherché est *on ... verbe ... que*. A partir de chaque occurrence d'un verbe appartenant à l'une des deux ressources lexicales, le pronom *on* est recherché dans un empan maximal de huit tokens vers la gauche et *que* ou *qu'* dans un empan maximal de huit tokens vers la droite. La recherche tient également compte de frontières définies à partir de l'étiquette morphosyntaxique “frpos” : elle est interrompue dès lors qu'un token étiqueté comme ponctuation forte ou comme guillemet est rencontré. Cette contrainte permet notamment d'exclure les segments dans lesquels un verbe de parole rapportée introduit une citation.

[11] Pour les constructions en *il* impersonnel, la procédure diffère principalement en ce qu'elle vise à distinguer les emplois explétifs du pronom *il* de ses nombreux emplois référentiels. Le script prend pour point de départ chaque occurrence de la forme *il* et examine les quinze tokens suivants, sans franchir de frontière étiquetée “SENT” ou “PUN:cit”. Dans cette fenêtre, seules les occurrences de *que* ou *qu'* portant l'étiquette “KON” sont considérées comme candidates, afin de cibler les emplois de *que* en tant que complémenteur. Pour chaque candidat, le segment compris entre *il* et *que/qu'* est soumis à une série de contraintes morphosyntaxiques destinées à identifier les configurations impersonnelles. Cinq configurations sont prises en compte : (i) *il ... être ... participe passé ... que*, comme dans *il est admis que* ou *il a pu être démontré que* ; (ii) *il ... falloir ... infinitif ... que*, comme dans *il faut préciser que* ; (iii) *il ... convenir/importer/s'agir ... de ... infinitif ... que* ; (iii) *il ... être à ... infinitif ... que* ; (iv) *il ... être ... adjectif ... de ... infinitif*, comme dans *il est important de remarquer que* ; (v) *il y a ... de ... infinitif ... que*, comme dans *il y a des raisons de douter que*. Dans chacune de ces configurations, le participe passé ou l'infinitif correspondant au verbe attributif doit appartenir à l'une des deux

ressources lexicales utilisées pour l'étude.

[12] Les bases de données résultant de l'extraction des occurrences retenues sont structurées selon les variables suivantes : nom du fichier XML, domaine du texte, verbe identifié, catégorie sémantique du verbe et segment retenu. Lors du tri manuel des occurrences extraites automatiquement, ont été écartés les segments où une construction en *on* ou en *il* impersonnel apparaît avec un adverbial cadratif énonciatif (cf. Schrepfer-André, 2005), tel que *selon X, conformément aux conclusions de X* ou *comme le dit X*. Ont également été exclues les constructions en *il* impersonnel avec un complément d'agent. Après ce tri manuel, la base de données de constructions en *on* comporte 974 lignes, correspondant à 996 occurrences au total. La base de données des constructions en *il* impersonnel comporte quant à elle 577 lignes, pour un total de 590 occurrences.

2.3 Analyses statistiques

[13] La distribution des constructions a été analysée au moyen d'analyses de spécificité (Lafon, 1980, voir aussi Poudat & Landragin, 2007) réalisées sous R avec le package "textometry". Deux niveaux d'analyse ont été distingués. L'analyse dite macro vise à déterminer si les constructions en *on* et en *il* impersonnel sont sur- ou sous-représentées dans les différents domaines du corpus. L'analyse micro porte quant à elle sur la distribution des catégories sémantiques des verbes pour chaque type de construction. Ces deux analyses reposent sur le même principe de calcul des spécificités, mais diffèrent par la définition du corpus de référence et, surtout, par la manière dont est calculée la taille des sous-corpus comparés.

[14] Pour l'analyse macro, l'unité dont on cherche à mesurer la spécificité est le type de construction : construction attributive en *on* ou construction attributive en *il* impersonnel. Les dix domaines thématiques de Wikipédia constituent les sous-corpus comparés. Ainsi, la taille du sous-corpus correspond ici à la taille du domaine dans le corpus Wikipédia. L'analyse micro change de niveau de référence. Il ne s'agit plus, en effet, de rapporter les constructions à l'ensemble des mots du corpus Wikipédia, mais d'étudier la répartition des catégories sémantiques des verbes au sein de chaque type d'attribution. Les deux collections d'occurrences, celle de 996 occurrences des constructions en *on* et celle des 590 occurrences des constructions en *il* impersonnel sont réunies pour constituer un référentiel commun de 1'586 attributions. La définition de la taille du corpus change donc fondamentalement par rapport à l'analyse macro. Le corpus de référence n'est plus constitué de 9'813'647 tokens de Wikipédia, mais des 1'586 attributions effectivement extraites. De même, pour chaque domaine, la taille du sous-corpus correspond au nombre total d'attributions présentes dans ce domaine, les constructions en *on* et en *il* impersonnel étant additionnées.

3 Résultats

[15] Nous présentons les résultats selon deux niveaux de granularité. Dans un premier temps, au niveau macro, nous examinons la répartition, dans les différents domaines pris en compte, des attributions de connaissances mettant en jeu le pronom *on* et de celles mettant en jeu le pronom *il* impersonnel. Dans un second temps, au niveau micro, nous distinguons, au sein de chacun de ces deux types d'attribution, les dix catégories sémantiques de verbes et nous examinons leur répartition selon les différents domaines.

3.1 Résultats au niveau macro

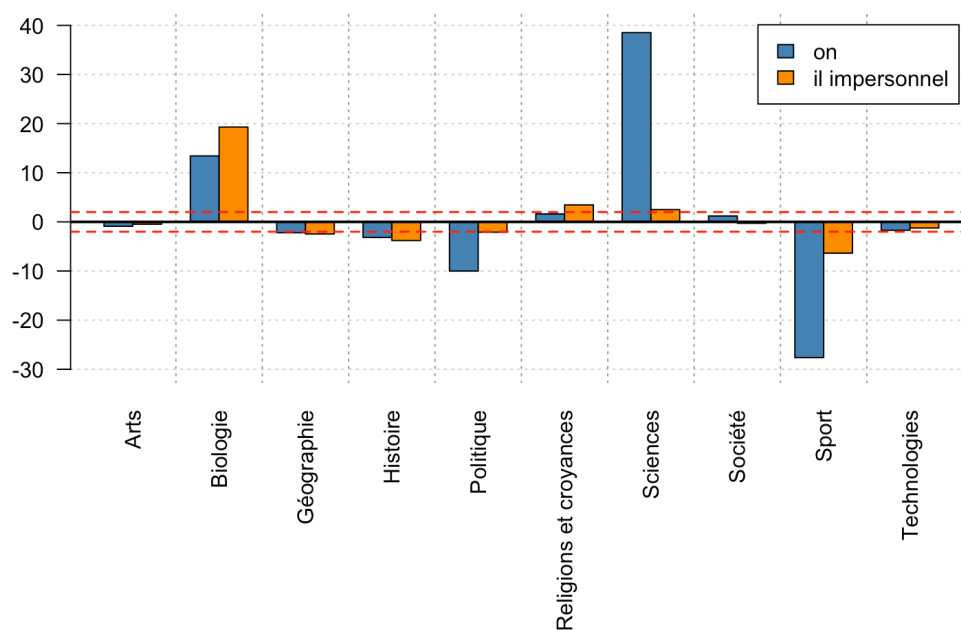


Figure 1 : Spécificité des attributions en *on* et *il* impersonnel

[16] On observe d'abord dans la Figure 1 que les deux types d'attribution en *on* et *il* impersonnel ont des tendances semblables dans tous les domaines, à l'exception de deux domaines : le domaine des Sciences, où les attributions au pronom *on* sont largement surreprésentées tandis que les attributions au pronom *il* impersonnel se situent dans le seuil de banalité ; le domaine de la Politique, où les attributions au pronom *on* sont largement sous-représentées tandis que les attributions au sujet *il* impersonnel se situent dans le seuil de banalité.

[17] On observe ensuite que les domaines se différencient en fonction de la sur- ou sous-utilisation de ces deux types d'attributions. Le contraste le plus saillant

oppose le domaine Biologie, où les attributions en *on* et *il* impersonnel sont toutes deux sur-utilisées, au domaine du Sport, où les deux types d'attribution sont au contraire sous-utilisées.

3.2 Résultats au niveau micro

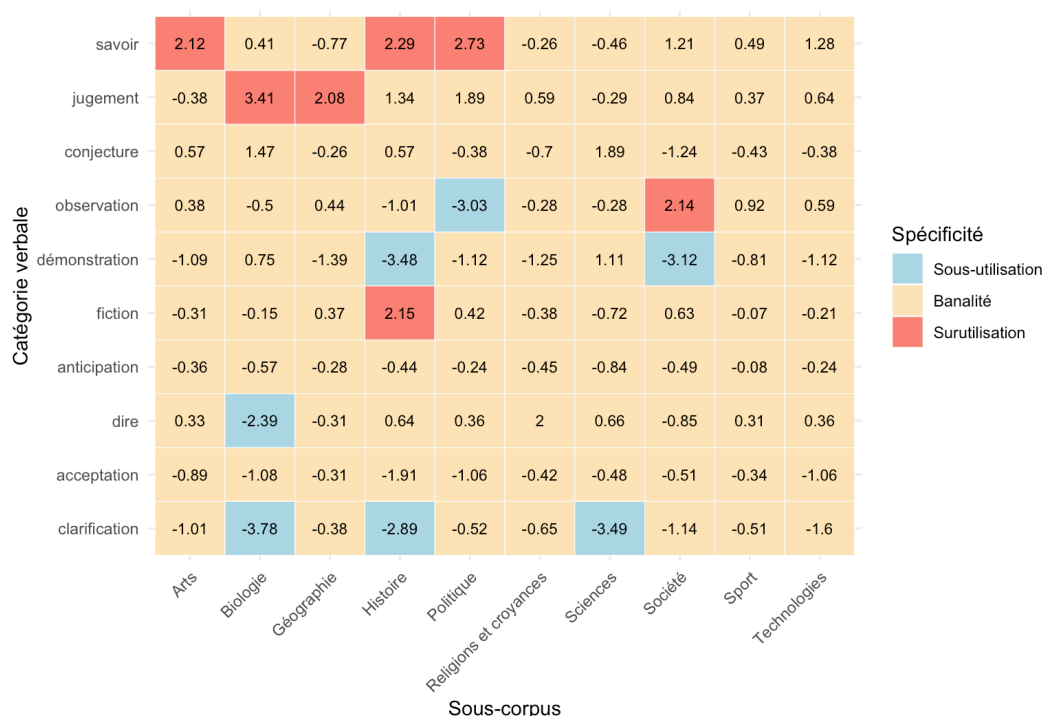


Figure 2 : Spécificité des catégories verbales dans les attributions en *on*

[18] Au sein des constructions en *on*, certaines tendances transversales aux domaines apparaissent nettement dans la distribution des catégories verbales. On observe notamment que les verbes de savoir et de jugement sont surreprésentés dans plusieurs domaines. Les verbes de savoir sont surreprésentés en Arts (2,12), en Histoire (2,29) et en Politique (2,73), tandis que les verbes de jugement sont surreprésentés en Biologie (3,41) et en Géographie (2,08). À l'inverse, les verbes de démonstration et de clarification tendent à être sous-représentés dans plusieurs domaines. Les verbes de démonstration sont sous-représentés en Histoire (-3,48) et en Société (-3,12), tandis que les verbes de clarification sont sous-représentés en Biologie (-3,78), en Histoire (-2,89) et en Sciences (-3,49).

[19] En ce qui concerne les contrastes entre domaines, on observe que la plupart des domaines privilégient des catégories sémantiques différentes de verbes. En Arts, les constructions en *on* se caractérisent par une surreprésentation des verbes de savoir (2,12). En Biologie, les verbes de jugement sont nettement surreprésentés

(3,41). En Géographie, seuls les verbes de jugement franchissent le seuil de surreprésentation (2,08). En Histoire, les verbes de savoir (2,29) et de fiction (2,15) sont surreprésentés, tandis qu'en Politique ce sont les verbes de savoir (2,73). Enfin, en Société, les verbes d'observation sont surreprésentés (2,14). Seuls les domaines Religions et croyances, Sciences, Sport et Technologies ne présentent pas de catégorie verbale nettement privilégiée. Ces résultats font ainsi apparaître des profils verbaux différenciés selon les domaines, certaines catégories de verbes étant spécifiquement associées à certains domaines dans les constructions en *on*.

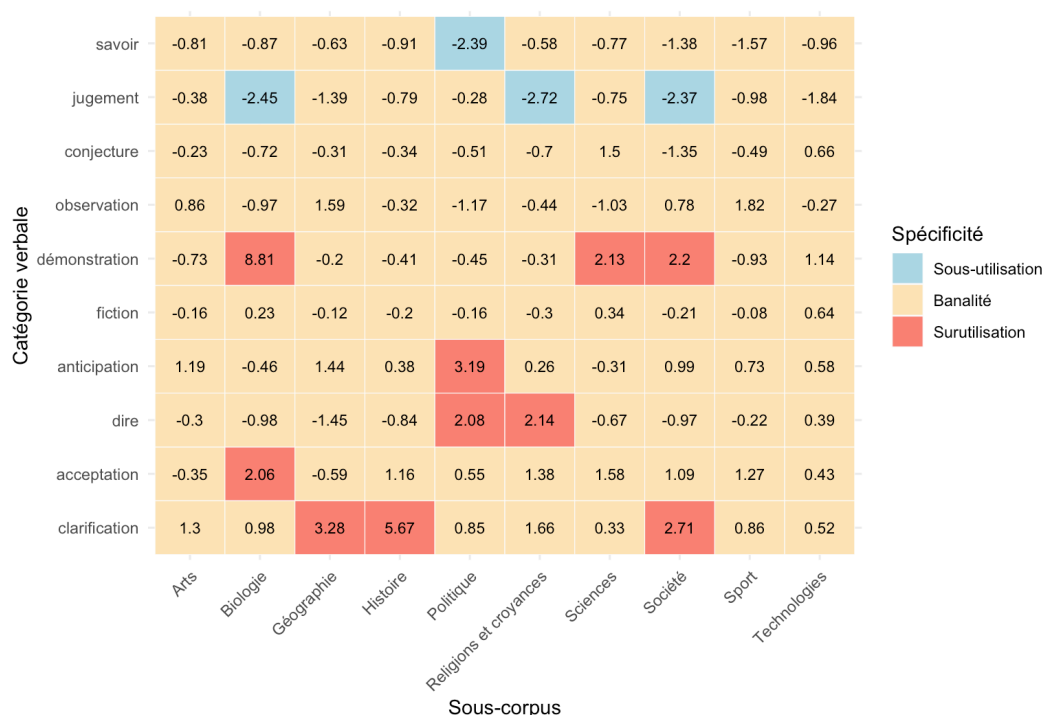


Figure 3 : Spécificité des catégories verbales dans les attributions en *il* impersonnel

[20] Au sein des constructions en *il* impersonnel, certaines tendances dans la distribution des catégories verbales apparaissent de manière transversale aux domaines. Les verbes de démonstration, de dire et de clarification tendent à être surreprésentés dans plusieurs domaines. Premièrement, les verbes de démonstration sont surreprésentés en Biologie (8,81), en Sciences (2,13) et en Société (2,2). Deuxièmement, les verbes de dire sont surreprésentés en Politique (2,08) et en Religions et croyances (2,14). Enfin, les verbes de clarification sont surreprésentés en Géographie (3,28), en Histoire (5,67) ainsi qu'en Société (2,71). Inversement, les verbes de jugement sont sous-représentés en Biologie (-2,45), en Religions et croyances (-2,37) et en Société (-2,37).

[21] En ce qui concerne les contrastes entre domaines, l'examen des domaines

met également en évidence des profils préférentiels. La Biologie se caractérise par une forte surreprésentation des verbes de démonstration (8,81). Les verbes d'acceptation sont également surreprésentés, mais de manière moins prononcée (2,06). En Géographie, ce sont les verbes de clarification qui sont surreprésentés (3,28). En Histoire, les verbes de clarification présentent une surreprésentation particulièrement importante (5,67). En Politique, les verbes d'anticipation (3,19) et de dire (2,08) sont spécifiques. Le domaine des Religions et croyances se caractérise par une surreprésentation des verbes de dire (2,14). En Sciences, les verbes de démonstration sont surreprésentés (2,13). Le domaine Société présente quant à lui deux catégories de verbes surreprésentés, les verbes de clarification (2,71) et de démonstration (2,2). Enfin, ni les Arts, ni le Sport et ni les Technologie ne présentent de catégorie de verbe spécifique.

4. Analyse

Domaine	Attributions en <i>on</i>		Attributions en <i>il</i> impersonnel	
	Sur-utilisées	Sous-utilisées	Sur-utilisées	Sous-utilisées
Arts	Savoir	–	–	–
Biologie	Jugement	Dire ; Clarification	Démonstration ; Acceptation	Jugement
Géographie	Jugement	–	Clarification	–
Histoire	Savoir ; Fiction	Démonstration ; Clarification	Clarification	–
Politique	Savoir	Constat	Anticipation ; Dire	Savoir
Religions et croyance	–	–	Dire	–
Sciences	–	Clarification	Démonstration	–
Société	Observation	Démonstration	Démonstration ; Clarification	Jugement
Sport	–	–	–	–
Technologies	–	–	–	–

Tableau 2 : Catégories de verbes sur-utilisées dans les attributions en *on* et *il* impersonnel

[22] Le Tableau 2 met en évidence qu'il existe peu de convergences entre les attributions en *on* et celles en *il* impersonnel. On constate même des tendances opposées entre ces deux types de constructions pour une même catégorie verbale, dans certains domaines. C'est notamment le cas en Biologie, en Histoire, en Politique et en Société.

[23] Si l'on regarde de plus près quels verbes sont utilisés dans les domaines qui montrent des tendances opposées, on relève qu'en Biologie, les verbes de jugement sont surreprésentés dans les constructions en *on*, principalement représentés

par *estimer* (n = 17 ; 36,1%), *considérer* (n = 13 ; 27,7%) et *penser* (n = 13 ; 27,7%) voir l'exemple (1), tandis qu'ils sont sous-représentés dans les constructions en *il* impersonnel. En Politique ce sont les verbes de savoir, principalement représentés par *savoir* (n = 9 ; 50%), qui sont surreprésentés dans les constructions en *on*, comme l'illustre l'exemple (2), et sous-représentés dans les constructions en *il* impersonnel.

(3) *Crise de la vache folle*, Biologie (Wikipédia)

En juillet 2012, **on estime que** la maladie a fait 224 victimes, dont 173 au Royaume-Uni (...)

(4) *Élisabeth Ire (reine d'Angleterre)*, Politique (Wikipédia)

En janvier et février 1554, Thomas Wyatt mena une révolte contre les politiques religieuses (...). Elisabeth fut convoquée à la cour pour y être interrogée sur son rôle (...). Même s'il est improbable qu'elle ait comploté avec les rebelles, **on sait que** certains d'entre eux l'avaient approchée (...)

[24] À l'inverse, dans les domaines de l'Histoire et de la Société, ce sont les constructions en *il* impersonnel qui sur-utilisent des catégories de verbes sous-utilisées dans les constructions en *on*. En Histoire, les verbes de clarification, principalement représentés par *rappeler* (n = 5 ; 35,7%) et *relever* (n = 4 ; 28,8%), sont surreprésentés dans les constructions en *il* impersonnel, voir l'exemple (3), tandis qu'ils sont sous-représentés dans les constructions en *on*. De même, en Société, les verbes de démonstration, principalement représentés par *démontrer* (n = 6 ; 50%) et *prouver* (n = 3 ; 25%), sont surreprésentés dans les constructions en *il* impersonnel, voir l'exemple (4), et sous-représentés dans les constructions en *on*.

(5) *Révolution américaine*, Histoire (Wikipédia)

Il faut rappeler que Washington D. C. n'a été construite qu'après la révolution ; la ville rend hommage aux Pères fondateurs par ses monuments.

(6) *Théorie du complot*, Société (Wikipédia)

Il a été démontré que le biais d'intentionnalité pouvait avoir un impact sur la croyance des individus dans la théorie du complot.

[25] Ces résultats montrent ainsi que le choix entre *on* et *il* impersonnel est associé à des catégories verbales distinctes, voire présentant des tendances opposées au sein d'un même domaine.

3 Conclusion

[26] Au terme de cette étude, trois principaux résultats peuvent être dégagés. Premièrement, les catégories de verbes ne se distribuent pas de la même manière

selon les domaines du savoir encyclopédique. Les différents domaines font ainsi apparaître des profils différenciés quant aux catégories verbales mobilisées dans les constructions étudiées. Deuxièmement, le choix des catégories verbales employées varie également selon le type de construction, en *on* ou en *il* impersonnel. Plusieurs tendances opposées ont ainsi été relevées entre les deux constructions selon les domaines, tandis qu’aucune convergence ne se dégage. Troisièmement, la comparaison des niveaux macro et micro montre que la sous-représentation globale d’une construction dans un domaine n’implique pas l’absence de spécificités au niveau des catégories verbales. Ainsi, dans le domaine Politique, bien que les constructions en *on* soient globalement sous-représentées, les verbes de connaissance et de jugement y sont sur-représentés à l’échelle micro. La faible mobilisation globale d’une construction dans un domaine peut donc coexister avec des préférences marquées pour certaines catégories verbales.

[27] Ces résultats ouvrent deux pistes qui mériteraient d’être approfondies. La première concerne l’articulation entre le type de construction et la sémantique du verbe d’attitude ou de parole rapportée mobilisé : les constructions en *on* et en *il* impersonnel sont associées à des profils de répartition différents des catégories sémantique de verbes, ce qui invite à examiner plus précisément les affinités entre le type de construction et le type de verbe mobilisé. La seconde concerne le rôle des domaines du savoir encyclopédique. Les régularités observées suggèrent que ceux-ci ne constituent pas simplement un principe de classement thématique des articles encyclopédiques du corpus, mais peuvent être envisagés comme des cadres discursifs au sein desquels se stabilisent certaines préférences dans la manière d’attribuer des connaissances à une source non-spécifiée. En ce sens, la prise en compte des domaines du savoir encyclopédique apparaît comme une dimension de variation pertinente pour mettre au jour, par les méthodes statistiques de la textométrie, des régularités dans les choix stylistiques des scripteurs.

Références bibliographiques

- Anand et Hacquard 2013 = Pranav Anand et Valentine Hacquard 2013. Epistemics and attitudes. *Semantics and Pragmatics* 6(8), 1-59.
- Doutreix 2020 = Marie-Noëlle Doutreix 2020. *Wikipedia et l'actualité : qualité de l'information et normes collaboratives d'un média en ligne*. Presses Sorbonne nouvelle.
- Faller 2019 = Martina Faller 2019. The discourse commitment of illocutionary reportatives. *Semantics and Pragmatics* 12(8), 1-53.
- Flottum et al. 2007 = Kjersti Flottum, Kerstin Jonasson et Coco Norén 2007. *On: pronom à facettes*. De Boeck Supérieur.
- Hacquard et Wellwood 2012 = Valentine Hacquard et Alexis Wellwood 2012. Embedding epistemic modals in English: A corpus-based study. *Semantics and Pragmatics* 5(4), 1-29.
- Krifka 2024 = Manfred Krifka 2024. A framework for performative and assertive updates. *Proceedings of Sinn und Bedeutung* 28, 490-508.
- Lafon 1980 = Pierre Lafon 1980. Sur la variabilité de la fréquence des formes dans un corpus. *Mots: Les langages du politique* 1, 127-165.
- Monville-Burston 1993 = Monique Monville-Burton 1993. Les verba dicendi dans la presse d'information. *Langue française* 98, 48-66.
- Poudat et Landragin 2017 = Céline Poudat et Frédéric Landragin 2017. *Explorer un corpus textuel : Méthodes, pratiques, outils*. De Boeck Supérieur.
- Rossari et al. 2024 = Corinne Rossari, Laura Aubry et Chloé Tahar 2024. Assumer ou diluer les responsabilités énonciatives dans la diffusion du savoir encyclopédique. In Anne Dister, Dominique Longrée (eds.) *JADT 2024, Mots Comptés, Textes Déchiffrés*, vol. 2, pages 803-812. Presses Universitaires de Louvain.
- Rossari et Tahar 2025 = Corinne Rossari et Chloé Tahar 2025. L'attribution de points de vue dans le genre encyclopédique : l'exemple de Wikipédia. Manuscrit Université de Neuchâtel.
- Schrepfer-André 2005 = Géraldine Schrepfer-André 2005. Incidence des formes de reprise du SN régime des SP en selon X énonciatifs sur leur portée phrastique et textuelle. *Langue française* 4(148), 80-94.