

# L'attribution de points de vue dans le genre encyclopédique : l'exemple de Wikipédia

Corinne Rossari<sup>1</sup>, Chloé Tahar<sup>2</sup>

<sup>1</sup>Université de Neuchâtel – corinne.rossari@unine.ch

<sup>2</sup>Université de Neuchâtel – chloe.tahar@unine.ch

*Soumis à « Sémantique, pragmatique, énonciation : nouvelles perspectives en sémantique argumentative », 05/2023*

## Résumé

Notre article établit une typologie sémantique des attributions de points de vue à un tiers (e.g., “X estime que p”) dans le corpus Wikipédia, en fonction du sujet grammatical qu’elles font intervenir. Ces sujets vont de l’indétermination complète (“on” ou “il” impersonnel) à la désignation précise d’un individu ou d’une collectivité avec un nom propre ou commun (“Lénine”, “les libertariens”), en passant par des cas intermédiaires avec des syntagmes nominaux renvoyant à des collectivités aux contours flous (“les scientifiques”, “certains chercheurs”). Nous analysons la répartition statistique de ces différents types de sujet selon les différents domaines de la connaissance (e.g., Sciences, Religions et croyances, Société) du corpus Wikipédia. L’originalité de notre contribution consiste, d’une part, à adopter un cadre théorique qui met l’accent sur la polyphonie intrinsèque de ces constructions où coexistent la voix du locuteur et celle d’un tiers et, d’autre part, à utiliser une méthodologie propre à l’Analyse des Données Textuelles pour mettre en relief l’incidence du choix du type de sujet grammatical sur la répartition statistique de ces constructions à travers les différents domaines de la connaissance représentés par l’encyclopédie Wikipédia.

**Mots clés :** attributions de points de vue, genre encyclopédique, transmission du savoir, textométrie

## 1. Introduction

Un des enjeux du genre encyclopédique moderne, un genre qui fait intervenir de manière non-triviale le discours d’autrui (voir aussi Doutreix (2020)), est de préserver autant que possible la neutralité des informations qui y sont transmises. L’exigence d’anonymat des contributeurs de l’encyclopédie collaborative Wikipédia manifeste cet enjeu de manière exemplaire. Selon les principes fondateurs de cette encyclopédie, la neutralité des informations repose sur l’attribution des savoirs transmis à des garants externes à l’instance auctoriale (“la neutralité de point de vue requiert l’attribution des points de vue”<sup>1</sup>). Nous avançons l’hypothèse que la transmission d’un savoir n’est jamais exempte de subjectivité et que, même dans un discours dénué d’une instance auctoriale, la neutralité n’est qu’une posture rhétorique.

---

<sup>1</sup> [https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Neutralit%C3%A9\\_de\\_point\\_de\\_vue](https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Neutralit%C3%A9_de_point_de_vue)

Notre contribution entend apporter un éclairage nouveau sur la question de la transmission du savoir encyclopédique, en se focalisant sur des cas de figure où le savoir transmis est médié par le discours d'un tiers, par le biais du repérage systématique et exhaustif des attributions de points de vue à un tiers dans le corpus Wikipédia. Dans notre étude, nous désignons comme attributions de points de vue les constructions à verbe introducteur de complétive ("X dit que p"), qui attribuent de manière explicite un point de vue (exprimé par le contenu de la complétive) à un sujet grammatical distinct de la première ou seconde personne. Notez toutefois que d'autres constructions peuvent être utilisées pour attribuer un point de vue à un tiers, allant des syntagmes cadratifs ("selon X", "d'après X", "pour X"), aux formes modales telles que le conditionnel évidentiel.

Une attribution de point de vue à un tiers est susceptible de recevoir deux types d'interprétations, selon que l'énoncé sert (i) à transmettre l'information exprimée par la complétive ou (ii) à caractériser le sujet du verbe introducteur, en lui attribuant un certain point de vue.

- (1) **Les spécialistes de l'organisation affirment que** la structure non hiérarchisée du réseau d'Al-Qaïda est à la fois sa force et sa faiblesse. En effet, la structure décentralisée permet à Al-Qaïda d'avoir une base mondiale ; cependant, les actions impliquant un haut degré d'organisation, comme les attaques du 11 septembre, prennent beaucoup de temps et d'efforts à mettre en œuvre. (WIKI, Al-Qaïda)
- (2) **Emmanuel Macron affirme qu'il n'y a eu aucun mort** "victime des forces de l'ordre", alors qu'une femme de 80 ans est morte le mois précédent à Marseille après avoir été touchée par une grenade lacrymogène. (WIKI, Mouvement des Gilets jaunes)

L'exemple (1) correspond à la première interprétation, l'énoncé ayant pour but de transmettre l'information, présentée comme faisant l'objet d'un consensus scientifique, que la structure décentralisée du réseau d'Al-Qaïda est à la fois sa force et sa faiblesse. La suite du discours enchaîne d'ailleurs sur ce contenu ("en effet, la structure décentralisée ..."). L'exemple (2) correspond à la deuxième interprétation, l'énoncé ayant pour but de caractériser Emmanuel Macron comme ayant tenu un discours mensonger.

Notre article poursuit deux objectifs. D'une part, il vise à établir une typologie sémantique des attributions de points de vue en fonction du sujet grammatical qu'elles font intervenir. Ce sujet va de l'indétermination complète ("on" ou "il" impersonnel) à la désignation précise d'un individu ou d'une collectivité avec un nom propre ("Lénine", "les libertariens"), en passant par des cas intermédiaires avec des syntagmes nominaux renvoyant à des collectivités aux contours flous ("les scientifiques", "certains chercheurs", "les statistiques"). S'inscrivant dans le cadre de l'argumentation dans la langue (Ducrot, 1984), notre analyse des attributions de points de vue s'appuie sur la distinction entre la voix du locuteur L et celle des énonciateurs E et s'intéresse au type de coexistence qui se joue entre ces instances. Ce cadre d'analyse nous permet de rendre compte du fait que la sémantique du sujet grammatical - associé à l'instance énonciateur - tend à favoriser l'une ou l'autre des deux interprétations identifiées. D'autre part,

notre article vise à analyser la répartition statistique des attributions de point de vue en fonction de l'instance d'énonciation qu'elles mettent en scène, selon les domaines de la connaissance auxquels appartiennent les articles du corpus Wikipédia. Nous montrerons que le type de connaissance destiné à être transmis - qu'il s'agisse de savoirs associés aux sciences dites "dures" (e.g., Sciences, Biologie), ou aux sciences humaines (Société, Religions et croyances) - est un facteur de variation important.

## 2. Cadre théorique

Dans la littérature précédente, l'interprétation ambiguë de l'attribution de points de vue à un tiers a fait l'objet d'études dans deux principaux cadres théoriques à l'interface sémantique-pragmatique : dans le cadre de la théorie des actes de langage (Faller, 2019) et dans celui de la sémantique du discours (Hunter, 2016). Le cadre de la théorie des actes de langage (issue de la philosophie du langage) a principalement pour but d'expliquer la force illocutoire des énoncés (Austin, 1962 ; Searle, 1968) au moyen d'une représentation de l'effet des actes de langage sur l'état informationnel des interlocuteurs (Stalnaker, 1978 ; Roberts, 2012 ; Farkas & Bruce, 2010). Faller (2019), propose qu'aux deux interprétations possibles des attributions de points de vue correspondent deux actes de langage distincts : un acte de langage d'assertion lorsqu'il s'agit de transmettre l'information que  $p$ , un acte de langage de présentation lorsqu'il s'agit de caractériser le point de vue du sujet modal sur  $p$ . Le cadre théorique de la sémantique du discours, comme celui de la SDRT (*Structured Discourse Representation Theory*) d'Asher et Lascarides (2003), permet d'analyser les relations pragmatiques qui existent entre de multiples énoncés au moyen de relations rhétoriques, conçues comme des relations entre des états informationnels (e.g., *Narration*, *Explication*, *Conséquence*). Dans ce cadre, Hunter (2016) (voir aussi Maier (2023)), propose de rendre compte des deux interprétations possibles des attributions de point de vue à partir des relations de discours que l'attribution entretient avec de précédents énoncés. Elle propose qu'il s'agit une relation de réponse à une question lorsque l'attribution sert à transmettre l'information que  $p$  et d'une relation de narration, lorsqu'il s'agit de caractériser le point de vue du sujet modal sur  $p$ .

Dans notre étude, nous nous situons dans le cadre de la théorie de la polyphonie telle qu'elle a été mise en place par Ducrot (1984) et Ducrot (1989). Dans ce cadre, un énoncé fait apparaître une pluralité de voix, dont la description relève de la sémantique. Nous commençons par préciser les notions fondamentales utilisées par Ducrot pour décrire la polyphonie des énoncés. Ducrot identifie l'instance Locuteur (L), qui n'est pas conçue en tant que sujet psychologique ou sujet parlant, mais en tant qu'instance de discours à laquelle la responsabilité du contenu de l'énoncé est attribuée. Le locuteur L met en scène d'autres instances de discours dits Énonciateurs (E), qui sont associés à des points de vue (PDV). Les énonciateurs E sont des êtres non-référentiels qui sont des parties intégrantes du sens de l'énoncé.<sup>2</sup> Dans ce cadre, un énoncé

---

<sup>2</sup> "J'appelle "énonciateurs" ces êtres qui sont censés s'exprimer à travers l'énonciation, sans que pour autant on leur attribue des mots précis ; s'ils "parlent", c'est seulement en ce sens que l'énonciation est vue comme exprimant leur point de vue, leur position, leur attitude, mais non pas, au sens matériel du terme, leurs paroles." (Ducrot 1984: 204).

d'attribution de point de vue à un tiers est analysé comme une mise en scène par L de la voix de E à qui est attribué un point de vue issu du contenu de la complétive p.

Selon Carel et Ducrot (2009), l'attribution d'un PDV à un E (tiers) peut être vue comme nécessitant une interprétation dite 'modale' ou une interprétation dite 'attributive'<sup>3</sup>. En effet, à propos de l'énoncé "Les philosophes nous assurent que les choses pesantes tombent d'elles-mêmes en bas" (Carel et Ducrot 2009 : 37) relèvent deux lectures. Dans le cas d'une lecture *modale*, "le locuteur soutient une certaine opinion "à travers" un énonciateur distinct de lui et assimilé aux philosophes. On ne doit pas dire qu'il les fait parler mais qu'il parle "à travers" eux, en les utilisant pour ainsi dire comme des masques". L'instance E sert donc à dissimuler la voix de L et le propos de l'énoncé est d'introduire le contenu de la complétive (contenu sur lequel porte l'enchaînement argumentatif ultérieur). Dans le cas d'une lecture *attributive*, "il s'agit au contraire de faire connaître le discours des philosophes", en mentionnant "le fait que quelqu'un, en l'occurrence les philosophes, a présenté l'opinion indiquée dans le discours que l'on rapporte". L'énoncé sert à attribuer un contenu à une instance E, qu'il s'agisse de caractériser un point de vue au moyen de E (par exemple, pour crédibiliser ce discours, lorsque E correspond à une source d'information fiable) ou de caractériser E par le point de vue qui lui est attribué, si le discours est focalisé sur cette instance.

L'approche polyphonique permet de penser l'incidence du choix du sujet dans l'interprétation de l'énoncé d'attribution de point de vue. En effet, le sujet grammatical est conçu comme un énonciateur, qui peut soit être assimilé au locuteur, soit en être distinct. En ce qui concerne le corpus Wikipédia, nous faisons l'hypothèse que les sujets indéterminés de la troisième personne ("on" ou "il" impersonnel) ont tendance à être assimilés au locuteur, tandis que les sujets déterminés de la troisième personne ("Elon Musk") ont tendance à être distincts du locuteur. Notre approche permet donc de concevoir le fait que le choix du sujet – selon qu'il est déterminé ou indéterminé - peut favoriser ou défavoriser l'une ou l'autre des deux interprétations des attributions de points de vue identifiées précédemment. En cela, nous fournissons un critère empirique pour étudier à l'échelle d'un grand corpus un phénomène d'ambiguïté contextuelle.

### 3. Définition de l'objet d'étude

L'attribution de point de vue ne consiste pas en la seule attribution d'un discours à un tiers au moyen d'un verbe de dire. L'étude du discours indirect et de son expression au moyen de la catégorie des verbes de dire ne saurait d'ailleurs fournir une vision adéquate de la diversité des constructions qui introduisent le discours d'un tiers en faisant apparaître une forme de polyphonie. Pour illustration, l'étude de Monville-Burston (1993) témoigne de cette difficulté. Bien que celle-ci se concentre sur la transmission des connaissances dans le genre médiatique au moyen des verbes d'activité de parole (*verba dicendi*), elle choisit d'inclure dans son objet d'étude des verbes tels que "estimer", dont le sens premier n'est pas celui de désigner "un acte

---

<sup>3</sup> Notez que la distinction entre ces deux interprétations recouvre la différence établie par Ducrot (1984) entre deux façons d'argumenter par l'autorité : le "raisonnement par l'autorité" et "l'autorité polyphonique".

de parole” (pp. 49), mais qui servent pourtant, selon elle, à rapporter le discours d’un tiers. Comment, compte tenu de la complexité de cette notion, circonscrire la définition des attributions de point de vue ? Nous délimitons les frontières de la classe des attributions de points de vue en utilisant un cadre pragmatique qui nous fournit un ensemble de critères pour circonscrire l’ensemble des verbes introducteurs de complétive pertinents.

Les constructions que nous prenons en compte sont celles qui permettent une mise à jour informative de l’état de connaissance commun (*Common Ground*), au sens de Krifka (2024). Dans le cadre de sémantique dynamique de Krifka, l’interprétation d’un énoncé affecte le contexte, c’est-à-dire l’état des connaissances communes des interlocuteurs. Une mise à jour informative par une proposition *p* consiste à ajouter la proposition *p* à l’état de connaissance commun. Pour nous, les énoncés d’attribution de points de vue dont la visée pragmatique est informative visent à faire connaître le contenu d’un discours ou d’une représentation mentale du sujet grammatical. Ces énoncés ne se limitent pas à ceux qui impliquent des verbes d’activité de parole assertifs tels que “affirmer”, “admettre” ou “prétendre”, bien que de tels énoncés remplissent de façon prototypique cette fonction, voir (3).

- (3) En janvier 2019, **Emmanuel Macron affirme qu’il n’y a eu aucun mort** « victime des forces de l’ordre » alors qu’une femme de 80 ans est morte le mois précédent à Marseille après avoir été touchée par une grenade lacrymogène (WIKI, Mouvement des gilets jaunes)

Pour autant, d’autres constructions mobilisant des verbes dits représentationnels chez Anand & Hacquard (2013) induisent une mise à jour du contexte informative, tels que les verbes de jugement (“penser”, “estimer”, “considérer”), de perception (“voir”, “s’apercevoir”, “observer”), d’activité épistémique (“savoir”, “découvrir”, “dédire”, “conjecturer”), d’argumentation (“expliquer”, “démontrer”, “établir”, “conclure”) ou de clarification (“ajouter”, “indiquer”, “préciser”), voir par exemple (4).

- (4) Le juriste **Jean-Éric Schoettl considère que** « l’expansion des droits fondamentaux, et plus précisément des droits subjectifs, opposables par un particulier à une personne publique, caractérise l’évolution du droit, en France comme partout en Occident, depuis un demi-siècle » (WIKI, Droits de l’homme)

Nous incluons également dans la catégorie des énoncés d’attribution de points de vue à visée informative ceux qui impliquent une construction à sujet impersonnel telle que (5). Ces constructions contribuent à ajouter à l’état de connaissances commun une information même si le responsable à qui cette information est imputée reste indéterminé.

- (5) **Il semble que** le déclin prononcé voire la disparition de la vente d’esclaves en France date du VII<sup>e</sup> siècle. La tradition établit un lien entre ce phénomène et les décisions de Bathilde, reine des Francs et régente du royaume (WIKI, Abolition de l’esclavage)

Les énoncés d’attribution de points de vue à visée informative diffèrent des énoncés d’attribution de préférence, qui mobilisent essentiellement des verbes désidératifs (“souhaiter”, “espérer”) ou émotifs-factifs (“regretter”, “se réjouir”). Ces verbes contribuent à l’expression

d'une préférence du sujet envers le contenu de la complétive p et à une mise à jour préférentielle du contexte, consistant à ordonner des alternatives, voir Starr (2020). Nous excluons donc les énoncés d'attribution de préférence, tels que (6), qui attribue à Brice Hortefeux une préférence pour une alternative où p par rapport à des alternatives où non-p, de notre étude.

- (6) Au cours de l'année 2008, **Brice Hortefeux souhaite que** des associations autres que la Cimade puissent intervenir dans les centres de rétention. (...) Les plaignants, dont le Gisti, voient dans cette réforme une tentative d'éloigner la Cimade des centres de rétention (WIKI, Brice Hortefeux)

Enfin, l'adoption d'un cadre pragmatique permet d'exclure de notre objet d'étude d'autres constructions à verbe introducteur de complétive dont la visée de mise à jour du contexte diffère essentiellement d'une visée informative, comme celles qui permettent une mise à jour performative. Les énoncés performatifs, à la différence des énoncés informatifs, ne décrivent pas un état de chose, mais exécutent une action. C'est le cas notamment des énoncés construits avec un verbe d'assertion ("annoncer", "déclarer") servant à rapporter l'accomplissement d'un acte langage de déclaration par le sujet grammatical, comme (7).

- (7) Le 2 juin 2021, **Emmanuel Macron annonce que** la vaccination est élargie aux adolescents ayant entre 12 et 18 ans, à compter du 15 juin. (WIKI, Pandémie de Covid-19)

## 4. Typologie des attributions de point de vue

Cette section établit une typologie des constructions d'attribution de point de vue en fonction de leur syntagme nominal sujet (SN). L'analyse polyphonique de ces constructions fait ressortir comme essentiel le type sémantique du SN sujet qui est la trace de l'instance énonciateur (E). Nous établissons une distinction entre les SN sujet qui favorisent une lecture attributive du point de vue introduit dans la complétive de ceux qui en favorisent une lecture modale. Dans cette perspective, nous avons distingué trois catégories de SN sujets : (i) les SN qui déclenchent quasi systématiquement une lecture attributive, (ii) ceux qui déclenchent quasi systématiquement une lecture modale et (iii) et une catégorie de SN susceptibles de déclencher une lecture modale ou attributive, dont seules les occurrences correspondant à une lecture modale ont été retenues.

### 4.1. Classe 1

La classe (i) inclut des constructions mobilisant un SN sujet ayant la forme d'un nom propre, d'un pronom personnel de troisième personne ou d'une description définie sélectionnée en fonction de sa sémantique (e.g., l'Eglise catholique).

Lorsque le syntagme nominal sujet renvoie à une personne ou à un groupe qui fait l'objet de l'article, le syntagme nominal reçoit de manière privilégiée une lecture attributive. Cette lecture attributive s'obtient en raison du fait que le point de vue exprimé par la complétive sert à caractériser la personne ou le groupe dont il est question, voir (8-9) et (10-11).

- (8) Les libertariens rejettent cette critique en s'appuyant sur les importants fonds privés des associations caritatives qui financent des œuvres de bienfaisance comme

l'éducation et la santé des démunis partout dans le monde, avec comme exemples courants le Fonds mondial pour la nature, la Fondation Rockefeller ou la fondation Bill-et-Melinda-Gates. **Les libertariens estiment que** le bénévolat privé est réduit d'autant plus qu'augmente la redistribution publique, et réciproquement. (WIKI, Libertarianisme)

- (9) **Les libéraux estiment que** les êtres humains, êtres rationnels, perfectibles et libres, possèdent des droits fondamentaux qu'aucun pouvoir n'a le droit de violer. En conséquence, les libéraux veulent limiter les obligations sociales imposées par le pouvoir et plus généralement le système social au profit du libre choix de chaque individu. (WIKI, Libéralisme)

Dans les exemples (8) et (9), le syntagme nominal renvoie à un groupe qui fait l'objet de l'article (e.g., les libertariens, dans l'article sur le Libertarianisme en (8) ; les libéraux dans l'article sur le Libéralisme en (9)). Le contenu de la complétive est utilisé pour caractériser l'idéologie du SN sujet.

- (10) Dans le contexte russe, **Lénine considère que** le parti doit ainsi se substituer à la bourgeoisie, qui n'existe pas au sens évolué des sociétés d'Europe occidentale (la Russie étant, à ses yeux, au stade de l'"arriération asiatique") (WIKI, Vladimir Illitch Lénine)

- (11) En 2017, [Candace Owen] a commencé à se classer parmi les partisans conservateurs de Donald Trump (...) **Elle a soutenu que** Trump ne s'était pas engagé dans une rhétorique nuisible aux Afro-Américains, n'avait pas proposé de mesures qui nuiraient aux Afro-Américains. (WIKI, Candace Owen)

Dans les exemples (10) et (11), le syntagme nominal renvoie à la personne qui fait l'objet de l'article (Lénine, dans l'article sur Vladimir Illitch Lénine, en (10); Candace Owen, désignée par un pronom, dans l'article éponyme, en (11)). La construction déclenche sans ambiguïté une lecture attributive.

## 4.2. Classe 2

La classe (ii) inclut des constructions mobilisant un SN sujet ayant la forme du pronom indéfini "on" ou du pronom "il" impersonnel. La théorie polyphonique que nous adoptons permet de rendre compte du fait que le locuteur L se dissimule derrière un être discursif indéterminé, "un certain ON" (Ducrot, 1982, p. 88) ou encore la voix de l'Absent chez (Carel, 2011, p. 343), pour avancer p, favorisant ainsi une lecture modale de l'ensemble de la construction, voir les exemples (12) et (13).

- (12) **On considère que** la culture des céréales a permis l'essor des grandes civilisations, car elle a constitué l'une des premières activités agricoles. En effet, en fournissant une alimentation régulière et abondante aux populations, les céréales ont permis l'organisation de sociétés plus denses et plus complexes. (WIKI, Céréales)

- (13) **On considère que** la création, cette même année, du *Budi Utomo* par de jeunes nobles javanais marque le début du mouvement national indonésien. Un « Serment de la Jeunesse » est prononcé en 1928, émettant le vœu de créer une patrie indonésienne. (WIKI, Indonésie)

Dans ces exemples, les constructions impersonnelles en “on” ont pour fonction d’introduire dans le discours le contenu de la complétive. Notez que le discours qui suit la construction enchaîne sur le contenu p de la complétive. En (12), la suite du discours étaye le contenu de la complétive par une justification (“en effet ... les céréales ont permis...”), en (13) la suite du discours porte également sur le contenu de la complétive (les débuts du mouvement national indonésien).

Les constructions en “il” impersonnel déclenchent de même une lecture modale, voir les exemples (14) et (15) :

- (14) **Il est nettement établi qu'**une dépendance à l'alcool est fortement accompagnée d'un haut niveau d'anxiété et de dépression qui amplifient encore davantage la consommation. Elle s'accompagne aussi fréquemment de perturbations dans l'identification des expressions émotionnelles d'autrui, notamment d'une hypersensibilité à la colère. (WIKI, Alcoolisme)
- (15) Par ailleurs, il s'agit d'un territoire riche en matières premières. Citons notamment la présence de pétrole dont l'exploitation a permis aux États-Unis de prendre part très largement à la deuxième révolution industrielle. En effet, **il est souvent considéré que** le premier puits de pétrole a été creusé sous la direction d'Edwin Drake à Titusville, Pennsylvanie, en 1859. Cela préfigure la domination américaine dans le domaine de la production pétrolière. (WIKI, Révolution industrielle)

Le discours qui suit les constructions (14) et (15) est focalisé sur le contenu de la complétive : il est question des effets de la dépendance à l’alcool en (14) et de l’impact qu’a eu le fait de creuser le premier puits de pétrole sur la domination américaine en (15).

### 4.3. Classe 3

La classe (iii) inclut des constructions mobilisant un SN sujet ayant la forme de descriptions indéfinies ou de descriptions définies désignant des collectivités aux contours flous (e.g., “*les spécialistes*”). Ces constructions permettent, dans certains cas, une lecture attributive et dans d’autres cas une lecture modale. Dans le cadre de notre étude, nous n’avons retenu que les cas où ces constructions autorisent une lecture modale, car d’un point de vue théorique, il est surprenant a priori que des noms renvoyant à des entités référençables (cf. notamment exemples (18) et (19) où il est question de spécialistes d’un domaine) soient utilisés pour véhiculer une lecture modale.

Les constructions appartenant à cette classe peuvent se présenter sous la forme de descriptions indéfinies, voir (16) et (17).



- (16) Une importante question de cosmologie porte sur la forme de l'Univers. Il peut être « plat », c'est-à-dire que le théorème de Pythagore pour les triangles droits y est valide à de plus grandes échelles. Actuellement, **la plupart des cosmologues pensent que** l'Univers observable est (presque) plat. (WIKI, Univers)
- (17) Depuis les années 1970, les migrations illégales se sont développées mondialement en raison des différences de vitesse de développement des économies des différents pays. Ceci a conduit à un nombre croissant d'études relatives aux migrations irrégulières. **Certains chercheurs considèrent que** les migrants irréguliers sont attirés par des employeurs qui préfèrent les migrants. (WIKI, Immigration illégale)

Les constructions appartenant à cette classe peuvent également se présenter sous la forme de descriptions définies comportant des noms renvoyant à des collectivités représentant un domaine de spécialistes (les géologues en (18), les historiens en (19)).

- (18) La densité de Mercure peut être utilisée pour déduire des détails sur sa structure interne. Bien que la haute densité de la Terre résulte sensiblement de la compression gravitationnelle, en particulier au niveau du noyau terrestre, Mercure est beaucoup plus petite et ses régions internes ne sont pas aussi comprimées. Par conséquent, pour qu'elle ait une densité aussi élevée, son noyau doit être volumineux et riche en fer. **Les géologues estiment que** le noyau de Mercure occupe environ 85 % de son rayon, ce qui représenterait ainsi environ 61, 4 % de son volume contre 17 % pour la Terre par exemple (Mercure (planète))
- (19) Édouard eut des maîtresses tout au long de sa vie. Il fréquenta l'actrice Lillie Langtry, Lady Randolph Churchill (la mère de Winston Churchill), la mondaine Daisy Greville, l'actrice Sarah Bernhardt, Lady Susan Vane-Tempest, la chanteuse Hortense Schneider, la riche humanitaire Agnes Keyser (en). **Les historiens présument que** le futur Édouard VII eut au moins 55 liaisons mais son niveau d'implication dans ces relations est inconnu. (WIKI, Edouard VII)

Dans ces deux exemples, le discours est organisé autour du contenu de la complétive. En (18), la responsabilité énonciative des géologues n'est pas l'enjeu de l'attribution. Il en va de même en (19), où c'est le contenu de la complétive qui constitue le cœur du discours. Le fait que ce contenu soit transmis par la voix des géologues ou des historiens, ne modifie pas l'enjeu du discours. Ces deux syntagmes nominaux, bien qu'associés à des spécialistes d'un domaine, déclenchent une lecture modale.

Pour illustrer l'ambiguïté de ces constructions, on peut observer qu'une description indéfinie, telle que "quelques historiens" peut aussi déclencher une lecture attributive si la fonction de la construction consiste à légitimer le contenu de la complétive au moyen de la voix de cette collectivité aux contours flous, voir l'exemple (20).

- (20) Ernst Gombrich estime en 2005 que « Louis XV et Louis XVI, les successeurs du Roi-Soleil [Louis XIV] étaient incompétents (...) » Toutefois, le roi a aussi des

défenseurs. **Quelques historiens soutiennent que** la mauvaise réputation de Louis XV est liée à une propagande visant à justifier la Révolution française. (WIKI, Louis VX)

En (20), dans le contexte d'un débat historiographique, le SN sujet sert à caractériser le point de vue de certains historiens sur Louis XV. L'enjeu de la construction porte sur le fait que le point de vue issu de la complétive est attribué à un groupe d'historiens, ce dernier venant défendre un jugement favorable de Louis XV.

## 5. Méthode et corpus

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 : Arts (950 759 mots), Biologie (720 980 mots), Géographie (824 021 mots), Histoire (1 388 046 mots), Politique (1 219 535 mots), Religions et croyances (955 924 mots), Sciences (771 712 mots), Société (1 079 422 mots), Sport (1 049 911 mots), Technologie (813 337 mots). Pour chacun de ces dix projets, nous avons extrait automatiquement et aléatoirement 200 articles. Nous avons ensuite 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.) et supprimé les doublons aléatoirement. L'exploration et l'analyse du corpus a été menée sur le logiciel TXM (Heiden et al., 2010).

La démarche d'analyse des données textuelles que nous avons adoptée a consisté à poursuivre deux objectifs distincts. Le premier de ces objectifs consiste à construire la requête CQL la plus englobante possible pour chaque classe d'attributions de points de vue, afin d'extraire de manière exhaustive et systématique toutes les occurrences de chacune de ces classes dans le corpus Wikipédia. Les requêtes associées à chacune des classes d'attributions de points de vue sont disponibles dans l'Annexe en ligne à cette adresse : <https://osf.io/4vwkt/files/osfstorage> ainsi que les tableaux des concordances (Concordances.csv) et des listes de fréquences (Lemmes.csv) associées à chaque requête. Pour construire chaque requête, nous avons dans un premier temps élaboré une requête initiale qui soit la plus englobante possible en vue d'exfiltrer par la suite des résultats les constructions à verbe introducteur de complétive n'ayant pas une fonction informative (mais performative ou préférentielle, voir la Section 3). Afin de procéder au filtrage des verbes introducteurs de complétives, nous avons exporté les résultats de la requête sous la forme de liste de fréquences (triée par lemmes) dans un classeur Excel. Sass (2022) nomme filtrage des résultats d'une requête le processus qui consiste à réduire l'ensemble des résultats obtenus à partir de la requête initiale, au moyen de plusieurs requêtes successives, afin d'exclure progressivement les occurrences non-pertinentes. Sass met en lumière la dimension incrémentale de ce travail de filtrage, qui suppose d'élaborer plusieurs requêtes, en se servant des résultats d'une requête donnée pour affiner la syntaxe de la requête suivante. Au terme de ce processus, nous avons pu élaborer, pour chaque classe d'attributions de points de vue, une liste des verbes (ou autres éléments, e.g., adjectifs) pertinents et déterminer en fonction

de cette liste une requête finale qui spécifie les éléments à inclure, plutôt qu'à exclure. Le second objectif consiste en l'analyse statistique de la distribution des trois classes d'attributions de points de vue, afin de connaître la sur- ou sous-représentation de chacune de ces classes à travers les sous-corpus (ou projets) de Wikipédia. Pour poursuivre cet objectif, nous avons d'abord consulté l'Index de TXM pour obtenir la fréquence brute de chaque classe par sous-partie du corpus. Les scores de spécificité de Lafon ont ensuite été calculés au moyen du logiciel R. Le script R sont également joints à l'Annexe.

## 6. Résultats et discussion

Les résultats sont présentés sous formes de graphiques qui donnent les scores de spécificité des trois classes d'attributions de points de vue par domaine. Les scores de spécificité, calculés à partir d'un script R, indiquent le sur- ou sous-emploi de chacune de ces classes, en comparant le nombre d'occurrences effectif d'une classe avec le nombre d'occurrences qui serait attendu si sa distribution était homogène entre les domaines. Nous présentons les résultats avec différents niveaux de granularité.

### 6.1. Vision d'ensemble

Le premier graphique donne une vision d'ensemble de la répartition des trois classes d'expressions en fonction de chacun des domaines.

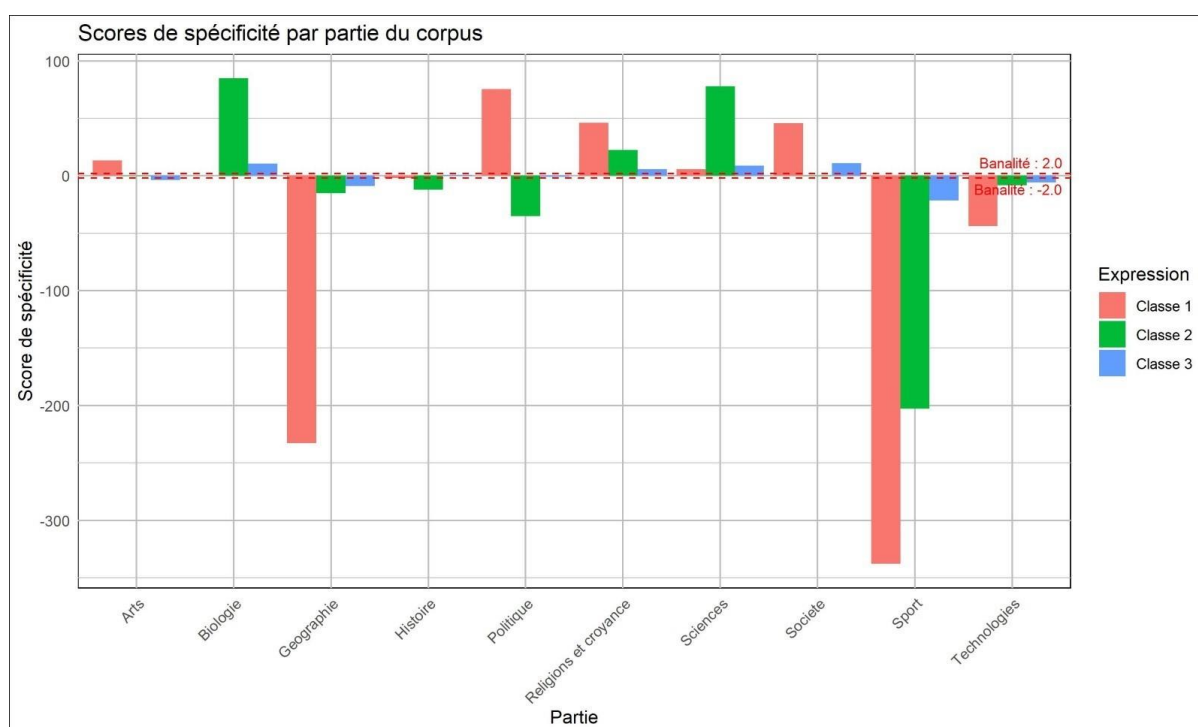


Figure 1 - Score de spécificités des trois classes d'attributions de points de vue dans Wikipédia

On observe d'abord l'irrégularité des tendances selon les domaines ce qui montre qu'ils ont bien un impact sur l'emploi de ces constructions. Les trois classes d'attributions de points de vue sont sur-représentées dans deux domaines : *Religions et croyances* et *Sciences*, mais leur distribution suit des tendances différentes. Le domaine *Religions et croyances* sur-représente la

Classe 1. En effet, les attributions de points de vue à une instance identifiable permettent d'organiser le discours autour des figures qui servent d'énonciateurs, voir l'exemple (21).

- (21) Le baptême effectué sur des nouveau-nés comme dans l'Église catholique est pour eux sans effet, car le baptême est imposé par la famille, le bébé n'étant alors pas encore une personne consciente et volontaire. De plus, **les témoins de Jéhovah estiment que** l'on doit connaître l'enseignement et se faire baptiser par immersion complète, ce qui n'est pas le cas lors d'un baptême de nouveau-nés. (Pratiques de Témoins de Jéhovah, Religions)

Au contraire, dans le domaine Sciences, la Classe 2 est sur-représentée.

On constate des contrastes entre la Classe 1 et la Classe 2 et 3 dans le domaine *Politique*. Dans ce domaine, la Classe 1 est sur-représentée et la Classe 2 sous-représentée. L'organisation du discours semble donc se faire autour des figures qui servent d'énonciateurs - E. Cela ressort d'autant plus que dans la Classe 2 (celle sous-représentée) certaines occurrences du pronom *on* censé déclencher une lecture modale, sont interprétables comme déclenchant une lecture attributive fondée sur une interprétation de *on* qui exclut le locuteur.

- (22) Sa biographie officielle diffère sur plusieurs points des biographies des observateurs occidentaux. Ainsi, **on prétend qu'il** serait né dans un « milyong » (camp secret) sur le mont Paektu, montagne sacrée et point culminant de la péninsule coréenne, le 16 février 1942 (tandis que le monde occidental fixe sa naissance un an plus tôt en Sibérie). (Kim Jong-II, Politique)

Dans l'exemple (22) le discours s'organise autour des débats qui animent les biographes de Kim Jong-II. Les voix s'affrontent et la construction "on prétend que" représente une de ces voix. Elle est ainsi caractérisée par le point de vue qui lui est attribué, point de vue qui s'oppose à celui qui est attribué à l'énonciateur 'le monde occidental'.

## 6.2. Attributions de point de vue à interprétation modale

Le second graphique se focalise sur une comparaison entre les deux types de classes d'attributions de points de vue qui déclenchent une lecture modale, à savoir celles qui impliquent un pronom indéfini ou impersonnel (Classe 2) et celles qui impliquent un sujet nominal (Classe 3) compatible avec une lecture modale.

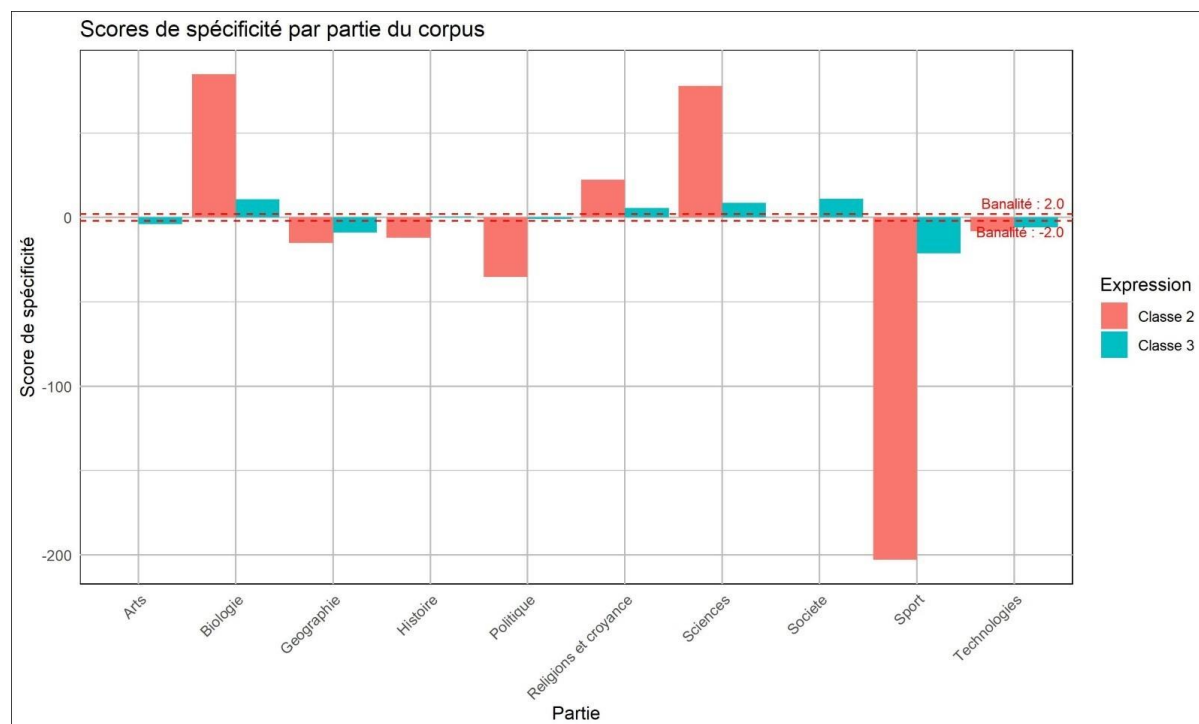


Figure 2 - Scores de spécificités des classes d'attributions de points de vue 2 et 3 dans Wikipédia

Le graphique permet de constater que les deux classes ne manifestent pas de tendances inverses selon les domaines: soit les deux sont sur-ou sous-représentées, soit les deux ou l'un d'elle est dans la fenêtre de banalité. Ces tendances non divergentes selon les domaines, étayant de manière empirique, le fonctionnement similaire des SN de cette classe et des pronoms indéfinis ou impersonnels montrant ainsi que les premiers sont susceptibles de déclencher autant que les seconds une lecture modale.

## 6.3. Zoom sur “il” impersonnel et “on”

Dans le troisième graphique, nous proposons de rentrer dans les détails de la Classe 2, en comparant les scores de spécificité du pronom indéfini “on” avec ceux du pronom “il” impersonnel au sein des attributions de point de vue. Cette comparaison permet de vérifier que ces deux pronoms, que nous avons considérées comme relevant du même type de catégorie, suivent les mêmes tendances dans les différents domaines.

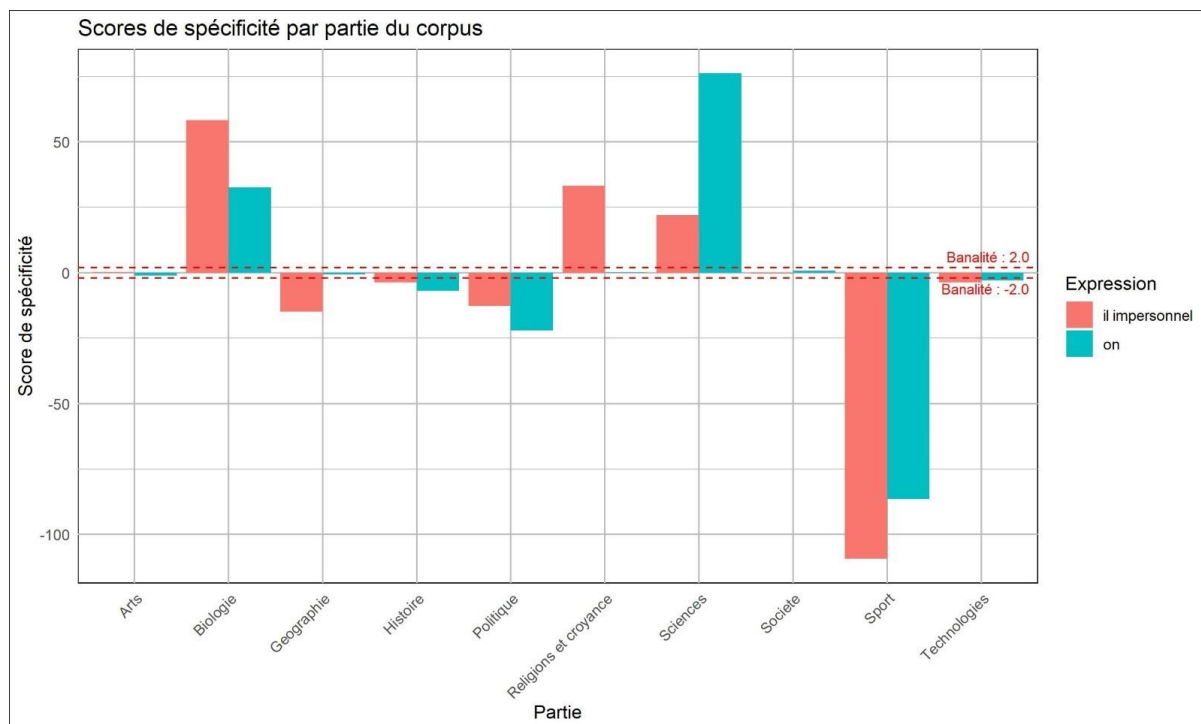


Figure 3 - Scores de spécificité des attributions de points de vue en "il" impersonnel et "on" dans Wikipédia

Le graphe confirme cette hypothèse, aucune tendance inverse ne peut être relevée selon les domaines. Toutefois, l'analyse qualitative que nous avons menée sur l'emploi du pronom *on* en contexte montre que selon les domaines ce pronom peut avoir des emplois qui donnent lieu à une lecture attributive. Une étape à venir sera de dissocier les emplois de *on* pour voir comment la distribution de ces différents emplois se profile en fonction des domaines.

## 7. Conclusion

Notre étude a montré que la façon de moduler les contenus qui sont rapportés dans Wikipédia est différente selon les domaines de la connaissance dans lesquels ils sont transmis. Le domaine suscite donc des manières plus ou moins stéréotypées de diffuser des connaissances. Notre étude s'est focalisée sur les liens qui unissent le contenu de la complétive au sujet du verbe introducteur en mettant en relief le type sémantique de ce sujet et ses retombées sur l'interprétation attributive ou modale de la construction à complétive.

Une prolongation souhaitable serait de comprendre comment la sémantique du verbe introducteur entre en jeu pour favoriser l'une ou l'autre des interprétations modale ou attributive. A priori nous avons considéré que tout verbe est susceptible de déclencher les deux interprétations, mais nous n'avons pas pris en compte le fait que le verbe peut être associé de façon privilégiée avec l'une des trois classes de SN sujet. Une telle étude permettra de prendre en considération le poids de la sémantique du verbe pour accéder à un type ou l'autre de lecture. Par exemple, si certains verbes sont associés de manière significative à des noms de la classe 1, alors que d'autres préfèrent les sujets à pronoms indéfinis (classe 2) on pourrait en déduire qu'ils sont prédisposés à déclencher respectivement une lecture attributive ou modale.

Pour conclure, nous pouvons retenir de cette étude que la question de la lecture modale ou attributive est davantage le fait de la sémantique des expressions qui composent la construction d'attribution de point de vue que celle d'une désambiguïsation contextuelle de l'énoncé dans lequel elles interviennent.

## Références

- Austin, J.L. (1962). *How to do things with words*. Oxford : Clarendon Press.
- Anand, P. et Hacquard, V. (2013). Epistemics and attitudes. *Semantics and Pragmatics* 6(8) : 1-59.
- Asher, N. et Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.
- Carel, M. (2011). *L'entrelacement argumentatif : Lexique, discours et blocs sémantiques*. Paris : Honoré Champion.
- Carel M. et Ducrot O. (2009). Mise au point sur la polyphonie. *Langue française* 164 (4): 33-43.
- Doutreix M.-N. (2020). *Wikipedia et l'actualité : qualité de l'information et normes collaboratives d'un média en ligne*. Presses Sorbonne nouvelle.
- Ducrot, O. (1984). *Le Dire et le Dit*. Les Editions de Minuit.
- Ducrot, O. (1989). *Logique, structure, énonciation. Lectures sur le langage*. Les Editions de minuit.
- Faller, M. (2019). The discourse commitment of illocutionary reportatives. *Semantics and Pragmatics* 12(8) : 1-53.
- Farkas, D. F. et Brice, K. B. (2010). On reacting to Assertions and Polar Questions. *Journal of Semantics* 27(1) : 81-118.
- 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.
- Hunter, J. (2016). Reports in Discourse. *Dialogue and Discourse* 7(4) : 1-35.
- Krifka, M. (2024). A framework for performative and assertive updates. *Proceedings of Sinn und Bedeutung* 28 : 490-508.
- Maier, E. (2023). Attribution and the discourse structure of reports. *Dialogue and Discourse* 14(1) : 34-55.
- Monville-Burston, M. (1993). Les verba dicendi dans la presse d'information. *Langue française* 98 : 48-66.
- Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6) : 1-69.
- Sass, B. (2022). Principles of corpus querying: A discussion note. *Acta Linguistica Academica* 69(4): 599-614.
- Searle, J. (1969). *Speech acts: an essay in the philosophy of language*. Cambridge: Cambridge University Press
- Stalnaker, R. (1978). Assertion. *Syntax and Semantics* 9. New York Academic Press: 315-332.
- Starr, W. (2020). A preference semantics for imperatives. *Semantics and Pragmatics* 13(6) : 1-60.