

# Les domaines linguistiques de la subjectivité du locuteur: étude empirique avec données de corpus<sup>1</sup>

**Cristina GRISOT**

Université de Neuchâtel, Université de Genève

In this article we report the results of an empirical study with corpus data targeting the identification of the linguistic domains of the speaker's *subjectivity* – broadly understood as referring to the expression of the speaker's perspective, emotions and attitudes. In order to attain this aim, we carried out an annotation study on corpus data (randomly selected from a literary text and from a journalistic text), formulated hypotheses on the basis of the state of the research and analysed the data with respect to three variables: the register, the source of subjectivity (speaker or third party) and the types of linguistic cues indicating the speaker's subjectivity. Our quantitative analyses showed that (i) the register does not influence the expression of subjectivity, (ii) the main linguistic domains of subjectivity are the *affective-evaluative lexicon* and *syntactic structures*, (iii) that *deictic subjectivity* is in fact non-subjective and that speaker's subjectivity refers only to *affective-evaluative* subjectivity. More generally, this article makes a methodological contribution about how to study an almost ineffable phenomenon, speaker's subjectivity, in a neutral and empirical manner.

## 1. Introduction

Depuis la "vague humaniste" (Finegan 1995:2) ayant marqué la recherche en linguistique dans les années 1980, l'intérêt des chercheurs pour la notion de *subjectivité* a fortement augmenté, le langage n'étant plus considéré comme la stricte expression objective de la pensée propositionnelle mais comme l'expression de sujets qui parlent, qui ressentent des émotions, qui perçoivent le monde (cf. Finegan 1995)<sup>2</sup>. Cette idée n'était pas nouvelle (cf. par exemple, le *Précis de stylistique* de Bally en 1905, le volume *Problèmes de linguistique générale*, 2 de Benveniste en 1966), mais à cette période elle a connu un renouvellement particulièrement important. Les chercheurs se sont donc intéressés à la subjectivité et ont analysé l'expression du *soi* et de la représentation du *point de vue* et de la *perspective* d'un sujet parlant ou d'un narrateur, à travers d'autres formes que la forme propositionnelle des énoncés. Dans ce contexte, la subjectivité a été abondamment étudiée dans divers cadres théoriques tant en linguistique théorique qu'en linguistique appliquée (comme la grammaire cognitive, la critique littéraire, la linguistique générale, la

---

<sup>1</sup> Article rédigé dans le cadre du projet FNS VTS (Verbal Tenses and Subjectivity, n° 100015\_1700086093, 2017-2019). Je tiens à remercier Jacques Moeschler pour la direction scientifique de ce projet de recherche et pour son soutien. J'aimerais également remercier les évaluateurs pour leur lecture attentive, leurs commentaires et les suggestions qui m'ont permis de préciser certains propos. Pour finir, je remercie Léa Farine pour sa relecture.

<sup>2</sup> Pour d'autres chercheurs (cf. Sarfati 2008), il s'agirait d'autres tournants ayant marqué l'intérêt croissant vers le sujet parlant, des tournants autant interdisciplinaires (par exemple, de la philosophie vers la linguistique) qu'à l'intérieur de la discipline (par exemple, de la pragmatique vers la linguistique énonciative).

linguistique énonciative, la pragmatique et la linguistique computationnelle, parmi d'autres).

Pour Lyons (1982: 102), la subjectivité est "la manière dont les langues naturelles, dans leur structure et leur fonctionnement naturel, donnent accès à l'expression du soi du locuteur, de ses attitudes et croyances<sup>3</sup>" (notre traduction). Ce changement vers le locuteur a été particulièrement marqué en syntaxe (Giorgi & Pianesi 2004, Giorgi 2010), en sémantique (Potts 2005, 2007), en pragmatique cognitive (Sperber & Wilson 1986/1995, 2015; Moeschler 2013) avec l'intégration des effets non-propositionnels au sens du locuteur (Wilson & Carston 2019) et a même été appliqué à d'autres capacités cognitives que le langage, comme le *raisonnement* (Mercier & Sperber 2017). Dans le domaine de la linguistique énonciative, Kerbrat-Orecchioni (1999/2014) documente les domaines linguistiques de la subjectivité en français. Cependant, il nous semble manquer dans la littérature une approche empirique et quantitative de la subjectivité, qui viserait à l'identification des domaines linguistiques de la subjectivité par des locuteurs autres que les chercheurs eux-mêmes dans des données de corpus.

Afin de combler ce manque, nous avons mené la présente étude, qui n'est, pour le moment, que le point de départ d'une plus vaste étude systématique de l'expression linguistique de la subjectivité. Dans cette étude, la subjectivité du locuteur, c'est-à-dire l'individu dans le monde qui prononce une phrase, est comprise d'une manière large comme se référant à l'expression de l'état intérieur du locuteur, c'est-à-dire l'expression, explicite ou implicite, de la perspective, des émotions et des attitudes du locuteur. Afin d'explorer les domaines linguistiques de la subjectivité, nous avons construit et analysé un corpus littéraire et un corpus journalistique. Des annotateurs francophones ont annoté les phrases du corpus comme exprimant ou non la subjectivité du locuteur ou d'une tierce personne. Dans une deuxième phase, les annotateurs ont identifié, dans les phrases subjectives, les domaines linguistiques de la subjectivité du locuteur.

Notre article est structuré de la manière suivante. La section 2 est dédiée à une brève revue de la littérature, dans la section 3 nous décrivons l'étude de corpus avec la tâche d'annotation – à savoir la méthodologie, les objectifs, les participants, le matériel, les résultats et leur discussion. Nous concluons notre propos dans la section 4.

## 2. Revue de la littérature

La notion de *subjectivité* a attiré l'attention de chercheurs venant de cadres théoriques et de sous-domaines variés de la linguistique. Cette variété d'approches a également conduit à une grande variabilité dans la définition de

---

<sup>3</sup> "the way in which natural languages, in their structure and their normal manner of operation, provide for the locutionary agent's expression of himself and of his attitudes and beliefs."

la subjectivité, entraînant un manque de comparabilité entre les différentes études. Par exemple, dans l'approche cognitive des relations du discours (*Cognitive approach to discourse relations, CCR*) développée par Sanders et al. (1992, 1993; Sanders 2005), la subjectivité est considérée comme un *principe cognitif et discursif fondamental*. Ce principe rendrait compte non seulement de l'usage des relations causales et de leurs marqueurs linguistiques (les connecteurs causaux) mais aussi de la complexité cognitive des relations du discours, telle qu'elle a été montrée par les études d'acquisition et du traitement cognitif du langage. Dans la Grammaire Cognitive de Langacker (1999), la subjectivité est liée au type d'interprétation (une construction mentale, *construal*) qu'un destinataire est amené à construire d'une expression ou d'une phrase. Si le point de vue du locuteur est intégré dans la construction mentale de manière implicite, alors l'interprétation est subjective. Si, au contraire, le point de vue du locuteur est exprimé de manière explicite, alors l'interprétation est objective. Dans ce cadre théorique, le pronom de première personne *je* représente une référence explicite au locuteur et déclenche une interprétation objective alors que dans d'autres cadres théoriques linguistiques et pragmatiques comme la linguistique de l'énonciation (Benveniste 1966; Kerbrat-Orecchioni 1999/2014) et la pragmatique intégrée (Ducrot 1972, 1980, 1984; Anscombe & Ducrot 1983), la référence directe au sujet parlant, que ce soit le locuteur ou l'énonciateur, est une composante cruciale de la subjectivité.

En linguistique de l'énonciation, la relation privilégiée entre le langage et la subjectivité a été mise en avant dans la littérature francophone par Benveniste (1966), pour qui la subjectivité est *la capacité du locuteur à se poser comme sujet de l'énonciation*, et qui définit l'énonciation comme la "mise en fonctionnement de la langue par un acte individuel d'utilisation." (Benveniste 1966: 259). Le sujet de l'énonciation est le sujet de conscience, qui intègre dans la description d'une situation ses évaluations, ses émotions, ses jugements, ses incertitudes, ses croyances, ses attitudes, etc. Pour Benveniste, le point d'appui principal de la subjectivité dans le langage est le pronom personnel *je*, qui pointe vers le locuteur d'un énoncé. Le deuxième point, toujours chez Benveniste, est l'expression de la temporalité, avec deux plans d'énonciation structurés autour du *je* du locuteur. Il s'agit notamment du plan du *discours*, où le "maintenant" du locuteur au centre et où le passé simple est exclu, et du plan de *l'histoire*, qui exclut le locuteur.

Dans le cadre de la pragmatique intégrée et dans la théorie de la polyphonie (Ducrot 1980, 1984; Anscombe & Ducrot 1983), le postulat de l'unicité du sujet parlant – se référant au fait que le responsable des activités psychologiques liées à l'énonciation est le sujet parlant de cet énoncé – est mis en doute<sup>4</sup>. Ainsi,

<sup>4</sup> Nous retrouvons une distinction similaire dans la théorie narratologique de Genette (1972). Genette distingue entre l'*auteur*, qui est la personne dans le monde qui a écrit l'œuvre littéraire, le *narrateur*, qui est un être théorique responsable de la narration et désigné par les marques de la première personne, et le *point de vue*, qui désigne le personnage dont le récit exprime les opinions, sans que ce personnage soit nécessairement confondu avec le narrateur. Comme dans la théorie de

il est distingué entre le *locuteur* et l'*énonciateur* d'une phrase, qui sont des êtres théoriques distincts du *sujet parlant* d'une phrase. Si le locuteur<sup>5</sup> est le responsable de l'énonciation, l'énonciateur est responsable d'un acte illocutionnaire. En ce qui concerne le sujet parlant, il est écarté de la théorie de Ducrot. Avec cette large palette d'êtres théoriques, il nous semble que selon le modèle de Ducrot, la subjectivité serait liée soit à ces différents êtres impliqués dans la construction du sens (pragmatique) d'un énoncé comparé à la signification (linguistique) d'une phrase, soit à aucun de ces différents êtres. Ainsi, comme Reboul (1994b) le propose, la théorie de Ducrot ne permettrait pas de traiter du problème de la subjectivité.

Pour Kerbrat-Orecchioni (2014) également, la subjectivité est étroitement liée à l'*énonciation*, définie comme "le mécanisme d'engendrement d'un texte, le surgissement dans l'énoncé du sujet d'énonciation, l'insertion du locuteur au sein de sa parole." (Kerbrat-Orecchioni 2014: 34) et aux *faits énonciatifs*, c'est-à-dire aux unités linguistiques qui fonctionnent comme indices de l'inscription d'un des paramètres suivants: les protagonistes du discours (locuteur et interlocuteur(s)), la situation de communication, les circonstances spatio-temporelles et les conditions générales de la production/réception du message. Ainsi, pour Kerbrat-Orecchioni, la subjectivité et sa manifestation dans le langage se réfèrent aux "traces linguistiques de la présence du locuteur au sein de son énoncé." (2014: 36). Les indices linguistiques les plus manifestes sont les déictiques (personnels, temporels et spatiaux) et le lexique (conceptuel) affectif-évaluatif (substantifs, adjectifs, verbes et adverbes). En lien avec ces deux catégories d'indices linguistiques, Kerbrat-Orecchioni parle de la *subjectivité déictique* (se référant à la situation d'énonciation à strictement parler), et la *subjectivité affective-évaluative* (se référant à l'attitude du locuteur et à son évaluation d'une certaine personne ou d'une certaine situation). Comme Kerbrat-Orecchioni le suggère suite à son analyse qualitative de données de corpus, il semble que la subjectivité déictique ne soit pas réellement subjective mais plutôt non-subjective (cf. discussion dans la section 3.5). Sur la base des résultats de notre étude quantitative, qui soutiennent la proposition de Kerbrat-Orecchioni, nous défendrons l'idée que la subjectivité déictique ne fait pas partie de la subjectivité du locuteur, qui correspond uniquement à la subjectivité affective-évaluative (cf. sections 3.4 et 3.5).

La *subjectivité déictique*, exprimée par une large classe de déictiques (appelés "shifters" par Jakobson (1957) car leur sens varie en fonction de la situation de communication), est étroitement liée au discours représenté (le *style indirect libre*). Le discours représenté montre certaines propriétés du discours direct (la

---

l'énonciation (Ducrot 1984), l'auteur (en narratologie) et le sujet parlant (en théorie de l'énonciation) sont exclus des développements théoriques (cf. Reboul 1994a). Or, dans cette recherche, nous nous intéressons justement au sujet parlant (ou l'auteur d'un texte, littéraire ou journalistique) pour étudier la manière dont il exprime sa subjectivité en utilisant le langage.

<sup>5</sup> Dans Ducrot (1984), il est précisé que le locuteur peut se référer à deux êtres théoriques distincts: le locuteur en tant que tel, et le locuteur en tant d'être du monde.

syntaxe) et certaines propriétés du discours indirect (les temps verbaux et les pronoms sont convertis). L'intérêt des chercheurs pour le style indirect libre (Banfield 1982; Reboul 1992; Schlenker 2004; Eckardt 2014; Reboul et al. 2016) est lié au comportement particulier des temps verbaux et des pronoms par rapport aux autres indexicaux, particularités expliquées en termes d'accès au sujet de conscience et donc à la subjectivité du locuteur ou d'une tierce personne.

Quant à la *subjectivité affective-évaluative*, elle est concernée par "les usages individuels du code commun" (Kerbrat-Orecchioni 2014:79) car:

Lorsqu'un sujet d'énonciation se trouve confronté au problème de la verbalisation d'un objet référentiel, réel ou imaginaire, et que pour ce faire il doit sélectionner certaines unités dans le stock lexical et syntaxique que lui propose le code, il a en gros le choix entre deux formulations: (i) le discours 'objectif', qui s'efforce de gommer toute trace de l'existence d'un énonciateur individuel; (ii) le discours 'subjectif', dans lequel l'énonciateur s'avoue explicitement ('je trouve ça moche') ou se pose implicitement ('c'est moche') comme la source évaluative de l'assertion. (Kerbrat-Orecchioni 2014:80)

La subjectivité affective-évaluative, selon Kerbrat-Orecchioni, est donc exprimée par des éléments lexicaux, comme: (i) les substantifs dérivés des verbes ou d'adjectifs (*amour, prétexte, beauté*, etc.), (ii) les adjectifs affectifs (*poignant, drôle, pathétique*, etc.), (iii) les adjectifs évaluatifs (*grand, chaud, bon, beau*, etc.), les verbes subjectifs (*prétendre, crier, détester, accuser*, etc.) et (iv) les adverbes subjectifs (*évidemment, vraiment, en effet, à peine, guère, seulement*, etc.). Toutefois, dans la littérature en linguistique générale, ce deuxième type de subjectivité est construit, de manière plus large, autour de l'*affect* du locuteur et du statut *épistémique* des propositions exprimées par des phrases. L'*affect* se réfère à l'expression linguistique des émotions, de l'empathie, de l'état d'esprit, etc. du locuteur. En effet, le langage a une fonction *émotive*, en opposition à sa fonction référentielle (comme cela a été mis en avant par Bühler 1934; Jakobson 1960; Lyons 1977; Ochs & Schieffen 1989, et bien avant par Bally 1905). Cette fonction émotive peut s'exprimer de manière linguistique, par exemple, par le lexique affectif (Kerbrat-Orecchioni 2014) ou la syntaxe (Kuno 1987; Kuno & Kaburaki 1977; Giorgi & Pianesi 2004; Giorgi 2010). Concernant les domaines de l'*evidentialité* et de la *modalité*, le premier réfère à la source et à la fiabilité de l'information détenue, et le deuxième à l'attitude et au degré d'engagement du locuteur envers la proposition exprimée. Les participants à une situation de communication peuvent exprimer différents degrés d'engagement et attitudes en utilisant des adverbes évaluatifs et évidentiels et leurs adjectifs associés (*sincèrement, heureusement, surprenant, clairement, probablement, sûrement*), des verbes modaux, des verbes lexicaux (du type *sembler, suggérer, avoir peur, s'ennuyer*), parmi d'autres moyens, sachant que les langues varient beaucoup quant aux moyens linguistiques utilisés pour exprimer cette catégorie (Talmy 1985; Biber & Finegan 1989; Smith 2003; Aikhenvald 2004, 2014).

Par ailleurs, certains énoncés peuvent être doublement marqués subjectivement. Dans l'exemple suivant de Kerbrat-Orecchioni, la formulation "Notre douce France" est marquée, premièrement par l'usage du déictique se référant au locuteur et à l'interlocuteur (le pronom *nous* inclusif), et deuxièmement par l'usage de l'adjectif affectivo-axiologique<sup>6</sup> *doux*, qui exprime un "jugement de valeur, et un engagement émotionnel du locuteur vis-à-vis de l'objet donné." (Kerbrat-Orecchioni 2014: 80). Le but de la présente recherche est d'utiliser l'*annotation linguistique* de données de corpus afin de faire émerger les domaines linguistiques de la subjectivité du locuteur et de confronter les résultats empiriques avec la catégorisation théorique (c'est-à-dire la subjectivité déictique et la subjectivité affectiv-évaluative) décrite ci-dessus.

### 3. Étude de corpus

La notion de subjectivité pouvant être définie et utilisée de manière différente selon les approches (grammaire cognitive, psychologie, linguistique énonciative, linguistique générale, etc.), il convient d'apporter des précisions sur la définition qui sera adoptée dans l'analyse empirique du présent article. À la suite de Hancil (2011) et de Fraser (2011), la *subjectivité* sera ici définie avec un sens large en tant qu'expression, explicite ou implicite, de la perspective, des émotions et des attitudes du locuteur. Cette définition correspond à une approche intuitive de la subjectivité, dans le sens où elle permet de rendre compte de l'implication du locuteur (de l'ensemble de son être) dans la description d'une personne ou d'une situation lors de la production d'un énoncé. La subjectivité déictique s'intègre à cette définition de manière basique: par le fait qu'un locuteur se pose comme sujet d'un énoncé qu'il produit dans un certain contexte à un moment donné. Selon Benveniste, il s'agit de la "mise en fonctionnement de la langue par un acte individuel d'utilisation" (Benveniste 1966: 259).

Finegan (1995) parle de l'expression du *soi* (self) et de la représentation de l'empreinte du locuteur (*speaker's imprint*) dans le discours. Il est bien sûr possible de définir la subjectivité de manière plus fine et de distinguer parmi ses différentes composantes, soit la perspective, le point de vue, les états mentaux, l'attitude ou les émotions. Ces définitions plus fines sont utiles afin de cibler des catégories linguistiques bien précises, comme par exemple les démonstratifs pour la notion de perspective (comme le font Keysar et al. 2000), l'aspect grammatical pour la notion de point de vue (comme nous l'avons fait dans Grisot 2017a) ou encore les adverbes d'attitude dans leur usage illocutionnaire pour la notion d'attitude. Toutefois, pour atteindre notre but (à savoir étudier de manière

<sup>6</sup> Les adjectifs *non-axiologiques*, comme *grand*, *loin*, *chaud*, sont relatifs à l'idée que le locuteur se fait de la norme d'évaluation pour une catégorie d'objets donnée. Les adjectifs *axiologiques*, comme *bon*, *beau*, *bien*, sont relatifs à l'évaluation faite par le locuteur mais les normes sont plus floues, la quantification est implicite, et ils communiquent un jugement de valeur, positif ou négatif. (Kerbrat-Orecchioni 2014).

empirique les divers domaines linguistiques de la subjectivité du locuteur), il nous semble approprié d'utiliser une définition large, intuitive, de la subjectivité, afin de permettre aux annotateurs de faire émerger, à partir des données de corpus, les indices linguistiques de la subjectivité. Notre description quantitative de la subjectivité est liée au fait que, malgré la compréhension intuitive de la subjectivité, plusieurs annotateurs, travaillant de manière indépendante, proposeront les mêmes types d'indices linguistiques de la subjectivité.

### 3.1 *Méthodologie et procédure*

La méthodologie mise en œuvre dans l'étude présentée ici est l'*annotation linguistique*. Cette méthodologie, largement utilisée dans le domaine de la linguistique de corpus, consiste à demander à deux ou plusieurs locuteurs natifs de catégoriser des phrases ou des passages de texte dans deux ou plusieurs catégories théoriques sur la base d'un ensemble préétabli d'instructions d'annotation. Les annotateurs passent par une phase d'entraînement, afin d'assurer une bonne compréhension de la tâche, des instructions d'annotation et du déroulement de l'étude d'annotation. Par la suite, les annotateurs annotent les phrases de manière indépendante.

Dans la présente recherche, les données annotées permettent d'identifier la subjectivité comprise par les annotateurs et les marqueurs qui portent cette subjectivité. En comparaison avec les études faisant usage uniquement de l'intuition linguistique des chercheurs, cette méthodologie offre un avantage conséquent, à savoir une collecte des données systématique, non-biaisée et fiable. La systématisme et la fiabilité sont dues au fait que les annotations sont réalisées par trois annotateurs avec un niveau d'éducation comparable, ayant reçu le même entraînement et ayant travaillé de manière indépendante. Les annotateurs ont des intuitions linguistiques objectives et non-biaisées, contrairement aux chercheurs, qui peuvent être biaisés par leur expérience ou par le cadre théorique dans lequel ils travaillent. Cette méthodologie est toutefois limitée à une exploration consciente et métalinguistique du phénomène étudié.

Afin d'assurer la fiabilité des annotations, il est nécessaire que l'accord entre deux annotateurs soit suffisamment élevé (par exemple, plus de 0.7 sur une échelle de 0 à 1 pour des coefficients mesurant l'accord inter-annotateur comme *Kappa*, où 0 signale l'absence d'accord autre que celui qui peut arriver par pure hasard et 1 signale un accord parfait; Carletta 1996). Afin d'augmenter les taux d'accord inter-annotateur, certaines études (Bayerl & Paul 2011; Spooren & Degand 2010; Scholman et al. 2016) ont proposé des solutions, comme par exemple (i) d'augmenter l'intensité de l'entraînement, quand les annotateurs sont des locuteurs non-expérimentés pour des tâches de réflexion consciente sur la langue, (ii) de réaliser l'annotation uniquement par des annotateurs expérimentés, (iii) de réaliser l'annotation dans plusieurs phases pour utiliser

les connaissances acquises par les annotateurs, (iv) d'utiliser des techniques pour éditer postérieurement les données annotées. Dans la présente étude, nous avons opté pour une autre solution: nous avons utilisé trois annotateurs au lieu de deux afin de pouvoir calculer, pour chaque phrase annotée, une majorité dans les réponses. Plus précisément, pour évaluer les résultats de la phase 1 visant à identifier les extraits subjectifs et non-subjectifs, nous avons mesuré pour chaque extrait la réponse la plus fréquente (la majorité dans les réponses) et nous avons calculé les taux d'accord entre les trois annotateurs pour chaque corpus en calculant le taux moyen du coefficient Kappa (entre annotateur 1 – annotateur 2, annotateur 2 – annotateur 3, et annotateur 1 – annotateur 3). En ce qui concerne l'évaluation des résultats de la phase 2, deux types d'analyses ont été réalisés sur les données subjectives de chaque corpus. La première est une analyse dite *souple*, dans laquelle nous avons compté la fréquence des types d'indices linguistiques (par exemple, lexicale, structure syntaxique, pronom, etc.) identifiés pour chaque extrait par un annotateur au minimum. La deuxième est une analyse dite *conservatrice*, dans laquelle nous avons compté la fréquence des types d'indices linguistiques identifiés par au moins deux annotateurs (accord partiel) ou les trois annotateurs (accord total). Cette analyse a montré quels indices linguistiques sont les plus stables parmi les trois annotateurs.

Cependant, dans Grisot (2017b), nous proposons que la méthodologie de l'annotation puisse également être utilisée pour tester les intuitions des locuteurs natifs au sujet d'un phénomène linguistique ou pragmatique, sans devoir atteindre un taux élevé de l'accord inter-annotateur comme cela est nécessaire pour produire des ressources textuelles. Par conséquent, les taux d'accord désignent les stratégies d'annotation développées par les annotateurs. Des taux bas d'accord peuvent indiquer que les annotateurs ont développé des stratégies différentes et donc qu'ils comprennent différemment les mêmes phénomènes. Ce fait, connu sous le nom de *biais de l'annotateur*, n'est pas problématique lorsque l'annotation vise précisément le processus de compréhension des locuteurs natifs. À ce sujet, Spooren & Degand (2010: 254) écrivent<sup>7</sup> (notre traduction):

---

<sup>7</sup> "Of course the coding will be subject to individual strategies developed by the coder, but these strategies will presumably be systematic and there is no reason to assume that such strategies will be conflated with the phenomenon of interest. [...] So if our research question is whether judgements occur more often with *want* than with *omdat*, an overcoding of judgments will not impede answer to the research question."



Bien entendu, l'annotation est sujette à des stratégies individuelles développées par l'annotateur, mais ces stratégies sont vraisemblablement systématiques et il n'y a pas de raison de penser que ces stratégies seront assimilées aux phénomènes d'intérêt. [...] Donc si notre question de recherche est de savoir si l'expression du jugement<sup>8</sup> est exprimée plus souvent avec *want* qu'avec *omdat*, avoir un surcodage de la catégorie jugement n'entravera pas une réponse à la question de recherche.

Les expériences d'annotation discutées dans cet article ont été menées dans le but d'étudier la compréhension par des locuteurs natifs du français, de la subjectivité exprimée dans des énoncés tirés des corpus. L'étude présente les résultats d'une tâche d'annotation qui a visé l'émergence naturelle et intuitive des indices linguistiques de subjectivité: les annotateurs ont eu la possibilité de proposer tout type de marque linguistique qui les a conduits vers une interprétation subjective des extraits annotés.

### 3.2 Objectifs, matériel, participants et hypothèses

Les trois objectifs de l'étude d'annotation sont: (i) l'identification des énoncés subjectifs et non-subjectifs dans deux corpus appartenant à deux registres stylistiques différents (journalistique et littéraire), (ii) l'identification de la source responsable de la subjectivité (à savoir le locuteur ou une tierce personne), et (iii) l'identification à partir d'énoncés subjectifs des indices linguistiques. Les deux premiers objectifs ont été atteints dans une première phase de l'expérience d'annotation (cf. Annexe 1 pour les instructions d'annotation), alors que le troisième objectif a été atteint par la suite lors d'une deuxième phase (cf. Annexe 2 pour les instructions d'annotation).

Le matériel est composé d'un total de 400 extraits: 200 extraits sélectionnés de manière aléatoire dans le roman *Rhum* de Blaise Cendrars pour le corpus littéraire, et 200 extraits sélectionnés de manière aléatoire dans le corpus *Le Monde* (année 2012) pour le corpus journalistique<sup>9</sup>. Nous avons utilisé ces deux types de textes<sup>10</sup> en raison des attentes d'objectivité différentes qu'un lecteur pourrait avoir pour le genre littéraire et pour le genre journalistique. En effet, dans le cas d'un texte journalistique comme celui du journal *Le Monde*, nous

<sup>8</sup> Ici, les auteurs se réfèrent à l'étude de Sanders & Spooren (2009), dans laquelle ils ont annoté les sens des deux connecteurs du néerlandais: *omdat*, qui est le plus fréquemment utilisé pour exprimer une relation causale objective (c'est-à-dire, la causalité entre des événements du monde), et *want*, qui est considéré comme le marqueur linguistique prototypique des relations causales subjectives (c'est-à-dire, la causalité liée aux conclusions et jugements faits par un locuteur sur la base d'un événement, comme dans *Le gâteau n'est pas bon parce que personne ne veut le manger*; cf. Degand & Pander Maat 2003; Pit 2003; Canestrelli et al. 2013).

<sup>9</sup> La sélection aléatoire se réfère au fait que les extraits n'ont pas été sélectionnés après avoir été lus ou analysés du point de vue qualitatif, pour voir s'ils peuvent ou non soutenir nos hypothèses. La sélection a également été faite avant la formulation des hypothèses de travail et de leurs prédictions.

<sup>10</sup> Ces deux textes ont été choisis de manière aléatoire parmi les œuvres littéraires et textes journalistiques existants, et nous ne voudrions pas prétendre qu'ils sont représentatifs pour l'ensemble du genre littéraire ou du genre journalistique. Une plus large palette de textes littéraires et journalistiques pourrait être étudiée dans de futures études à ce sujet.

pourrions attendre un degré d'objectivité dans la description des événements ou des personnes plus élevé que dans un texte littéraire. Toutefois, une limite possible de l'annotation d'un corpus écrit est celle des biais relatifs aux capacités d'analyse textuelle des annotateurs. Afin de minimiser ces biais, les annotateurs, six locuteurs natifs du français, sont des personnes ayant l'habitude de réaliser ce type de tâche, néanmoins sans être linguistes. Ils ont été rémunérés pour leur participation. Les annotateurs ont reçu avant chacune des deux phases un court entraînement afin d'assurer la bonne compréhension des instructions d'annotation, lors duquel nous leurs avons présenté les instructions d'annotations, ils ont annoté de manière indépendante 10 extraits de nos corpus (différents des extraits annotés dans l'étude elle-même) et ils ont pu expliquer de manière orale leurs interprétations des 10 extraits. Les six annotateurs ont travaillé de manière indépendante. Trois annotateurs ont travaillé sur le corpus journalistique et trois annotateurs sur le corpus littéraire.

Selon la discussion présentée dans la revue de la littérature (section 2) ainsi que selon les différences possibles entre les deux genres de textes explorés, nous pouvons formuler une série d'hypothèses et de prédictions concernant notre étude empirique (Table 1).

Variable	Hypothèses	Prédictions pour le corpus littéraire	Prédictions pour le corpus journalistique
Le genre du texte <sup>11</sup>	1. Acceptions différentes de la subjectivité selon le genre du texte	Extraits subjectifs plus fréquents que les extraits non-subjectifs	Extraits subjectifs moins fréquents que les extraits non-subjectifs
La source de subjectivité	2. Acceptions différentes de la source de la subjectivité selon le genre du texte	Subjectivité du locuteur plus fréquente que la subjectivité d'une tierce personne	Subjectivité du locuteur moins fréquente que la subjectivité d'une tierce personne
Types d'indices linguistiques	3. Les déictiques et le lexique sont les indices linguistiques les plus manifestes de la subjectivité du locuteur	Déictiques (personnels, spatiaux, temporels) aussi fréquents que le lexique affectif-évaluatif	Déictiques (personnels, spatiaux, temporels) aussi fréquents que le lexique affectif-évaluatif

Table 1 Hypothèses et prédictions

### 3.3 Résultats de la Phase 1

Pour le corpus littéraire, le taux moyen d'accord entre les 3 annotateurs a été de 69.3%, ce qui correspond à une valeur moyenne du coefficient Kappa de 0.36. Parmi les 200 extraits annotés, 108 extraits ont reçu la même interprétation (c'est-à-dire, subjective ou non-subjective) par les 3 annotateurs

<sup>11</sup> Nous remercions un des évaluateurs qui a pointé le fait qu'afin de préciser les hypothèses et les prédictions au sujet du genre du texte, il serait utile de prendre en considération dans une recherche future des travaux sur les genres du discours et du texte (cf. Rastier 1989; Adam 1997, 1999) pour anticiper sur les normes sémiotiques propres aux contraintes génériques.

(54% des données avec un accord total) et 92 extraits ont reçu la même interprétation par une majorité des 3 annotateurs (c'est-à-dire 2 des 3 annotateurs) (46% des données avec un accord partiel). En moyenne (calculée entre les cas d'accord total et les cas d'accord partiel) les annotateurs ont interprété 64.3% des extraits comme subjectifs et 35.7% comme non-subjectifs.

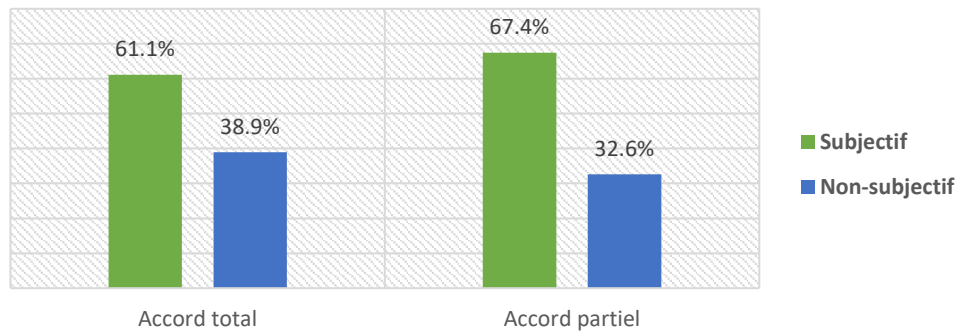


Figure 1: Fréquence des interprétations subjectives et non-subjectives dans le corpus littéraire

Le deuxième objectif de cette phase a été d'identifier, pour les extraits subjectifs, la source responsable de cette subjectivité. Le taux moyen d'accord entre les trois annotateurs a été de 87.9%, ce qui correspond à une valeur moyenne du coefficient Kappa de 0.36<sup>12</sup>. Parmi les cas d'accord inter-annotateurs, la source responsable de la subjectivité a été identifiée au locuteur dans 95.2% des cas (soit 88 extraits) et à une tierce personne dans 4.8% des cas (soit 8 extraits).

Pour le corpus journalistique, le taux moyen d'accord entre les trois annotateurs a été de 69.7%, ce qui correspond à une valeur moyenne du coefficient Kappa de 0.35. L'analyse des annotations a montré que 109 extraits ont reçu la même interprétation par les 3 annotateurs (54.5% des données avec un accord total) et 91 extraits ont reçu la même interprétation par une majorité des annotateurs (45.5% des données avec un accord partiel). En moyenne, les trois annotateurs ont interprété 65.3% des extraits comme subjectifs et 34.7% comme non-subjectifs.

<sup>12</sup>

Le taux très bas du Kappa comparé au pourcentage élevé (presque 90%) est expliqué par la distribution inégale des deux catégories (locuteur vs tierce personne). En effet, la valeur du coefficient Kappa est influencée par la distribution inégale des deux catégories car il nécessite que les données soient normalement distribuées (Artstein & Poesio 2008; Grisot 2017a)

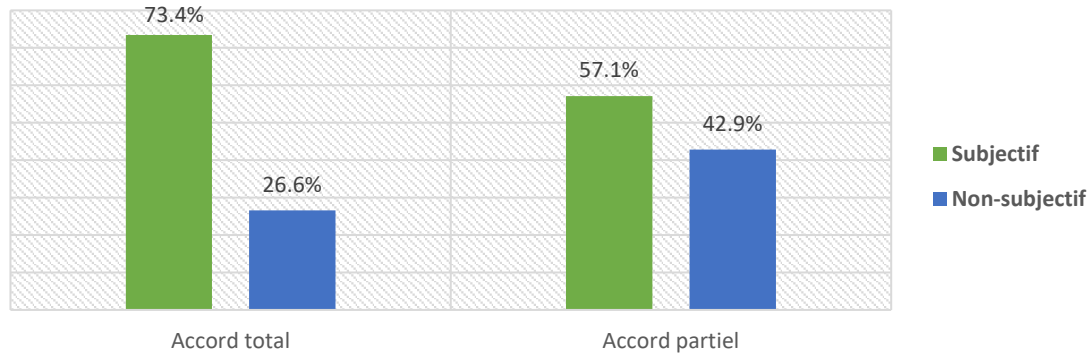


Figure 2: Fréquence des interprétations subjectives et non-subjectives dans le corpus journalistique

En ce qui concerne l'identification de la source responsable de la subjectivité, le taux moyen d'accord entre les 3 annotateurs a été de 84.4%, ce qui correspond à une valeur moyenne du coefficient Kappa de 0.7. Parmi les cas d'accord inter-annotateur, la source responsable de la subjectivité a été identifiée comme étant le locuteur dans 78.2% des cas (88 extraits) et une tierce personne dans 21.8% des cas (29 extraits).

### 3.4 Résultats de la Phase 2

L'objectif de la deuxième phase était d'identifier, à partir des extraits subjectifs, les indices linguistiques des interprétations subjectives. Dans cette étude, nous avons ciblé uniquement les extraits subjectifs pour lesquels la source responsable de la subjectivité est le locuteur (un total de 88 extraits pour chacun des sous-corpus).

Dans les instructions d'annotation (cf. Annexe 2), les annotateurs ont reçu des exemples de types d'indices linguistiques: les éléments lexicaux (mots et expressions du lexique), des éléments syntaxiques (c'est-à-dire des tournures de phrase comme le clivage, le parallélisme, etc.), des éléments morphologiques (temps verbaux, la construction *être en train de*, etc.), parmi d'autres. Toutefois, en raison du caractère exploratoire de cette deuxième phase et dans le but de faire émerger tous les types d'indices linguistiques, les annotateurs n'ont pas eu de contrainte particulière dans cette tâche.

#### 3.4.1 Corpus littéraire

Lors de l'analyse souple, pour les 88 extraits jugés par les annotateurs comme exprimant la subjectivité du locuteur, nous avons identifié un total de 318 occurrences de types d'indices linguistiques. Parmi ces éléments subjectifs identifiés par au minimum un annotateur, se trouvent d'abord le lexique affectif-évaluatif (41%) et les structures syntaxiques (19%). Ils sont suivis, dans un ordre décroissant de fréquence, par les figures de style (oxymore, répétition,

énumération, comparaison, etc.) (8%), les pronoms (8%), les temps verbaux (7%), les phrases interrogatives (6%), les points de suspension (5%), les phrases exclamatives (3%), et les guillemets marquant le discours direct (2%) (Figure 3).

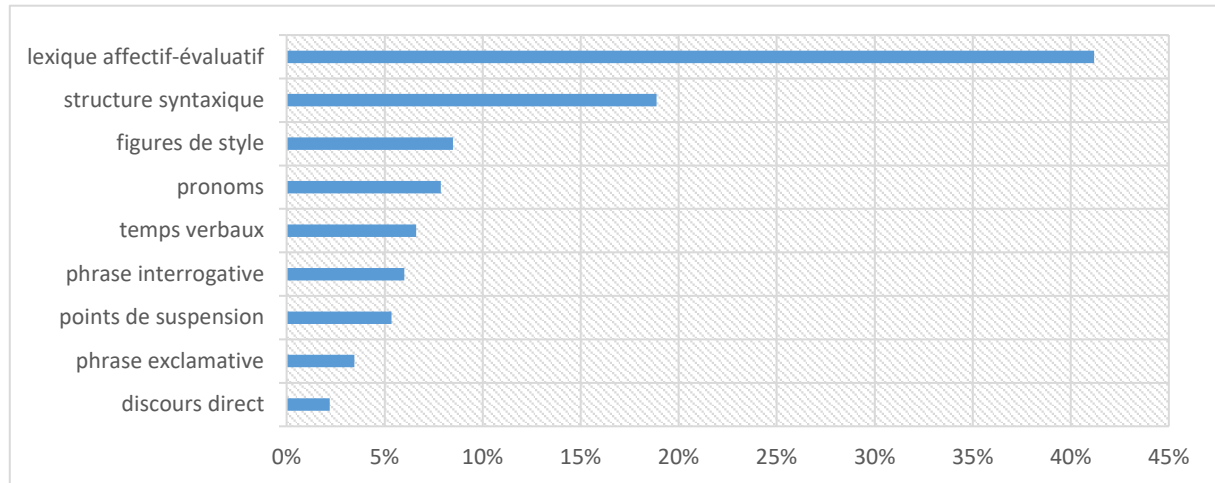


Figure 3: Types d'indices linguistiques de subjectivité du locuteur: analyse souple

Ci-dessous, nous donnons deux exemples annotés par les annotateurs.

- (1) Cela faisait partie de ce faisceau de calomnies dont on l'accabla alors. On l'accusait d'être un joueur, lui qui n'avait jamais plus fréquenté ni casino, ni cercle; un nouveau riche, lui qui avait mis plus de vingt ans pour édifier "sa maison"; un vulgaire fêtard, lui qui, d'après les constatations de M. Paul Benoît, syndic de faillite des Établissements Galmot, travaillait en moyenne seize heures par jour et menait une vie des plus simples. (Cendrars, Rhum) [BC188]
- (2) Comme tous les profiteurs, les spéculateurs, les nouveaux riches de France achetaient des châteaux, on en attribuait des douzaines à Galmot. (Cendrars, Rhum) [BC025]
- (3) Au prix de quelles souffrances est obtenu l'or des vitrines de nos bijoutiers et cette essence de bois de rose qui sert de base à la plupart des parfums ! (Cendrars, Rhum) [BC162]

Dans (1), les annotateurs ont relevé les éléments subjectifs suivants: (i) le lexique affectif-évaluatif *ce faisceau de calomnie*, *accabler*, *accuser*, *un nouveau riche*, *mettre plus de vingt ans*, *édifier*, *un vulgaire fêtard*; (ii) le pronom *on*; (iii) le passé simple dans *accabla* et l'imparfait dans *accusait*; (iv) les structures syntaxiques *ce faisceau de calomnies dont on l'accabla alors* et *lui qui*; (v) les guillemets autour de *sa maison* indiquant le discours direct. Dans (2), les annotateurs ont relevé les éléments subjectifs suivants: (i) le lexique évaluatif-affectif *profiteurs*, *spéculateurs*, *nouveaux riches*, *des douzaines* et (ii) la structure syntaxique antéposée *comme tous les profiteurs [...]*, *on en attribuait [...]*. Pour finir, dans (3), les annotateurs ont relevé (i) le lexique évaluatif-affectif *souffrances*, (ii) la structure syntaxique antéposée *au prix de quelles souffrances est obtenu [...]* et (iii) le point d'exclamation indiquant une phrase exclamative.

Contrairement à ce que ces exemples pourraient nous laisser croire, ces éléments subjectifs n'ont pas des fréquences similaires dans la totalité du corpus. L'analyse conservatrice menée sur les 88 extraits exprimant la subjectivité du locuteur a montré que seulement 173 occurrences des types d'indices ont été identifiées par au moins 2 annotateurs. Parmi ces types, les catégories du lexique évaluatif-affectif (54%) et des structures syntaxiques (16%) restent les plus fréquemment identifiées (en comparaison avec l'analyse souple). En effet, comme montré dans la Figure 4, l'analyse conservatrice fait ressortir que le lexique affectif-évaluatif et les structures syntaxiques sont les types d'indices les plus fréquents et les plus stables à travers l'analyse souple et l'analyse conservatrice.

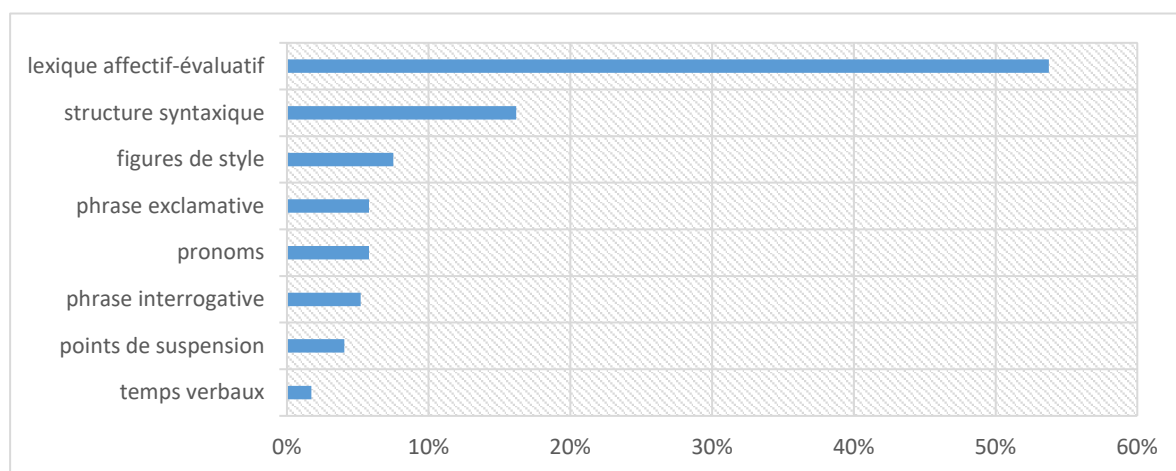


Figure 4: Types d'indices linguistiques de la subjectivité du locuteur: analyse conservatrice

Toutefois, les autres types d'indices, malgré leur fréquence plus réduite dans le corpus littéraire, restent également stables dans les deux types d'analyse. La catégorie des temps verbaux semble être la seule catégorie qui diminue drastiquement de fréquence entre l'analyse souple (7%) et l'analyse conservatrice (2%), ce qui pourrait montrer que cette catégorie fait partie d'une stratégie d'annotation individuelle. En d'autres termes, un des annotateurs a identifié cette catégorie comme indice de la subjectivité du locuteur plus fréquemment que les deux autres annotateurs.

### 3.4.2 Corpus journalistique

Lors de l'analyse souple, pour les 88 extraits jugés par les annotateurs comme exprimant la subjectivité du locuteur, nous avons identifié 457 occurrences de types d'indices linguistiques. Parmi ces éléments subjectifs identifiés par au minimum un annotateur, se trouvent en premier lieu le lexique affectif-évaluatif (70.2%) et les structures syntaxiques (18.8%), comme le montre la Figure 5.

Ces deux catégories sont suivies, dans un ordre décroissant, par les catégories de pronoms (5%), des temps verbaux (4.4%) et par d'autres catégories très peu fréquentes (moins de 1%).

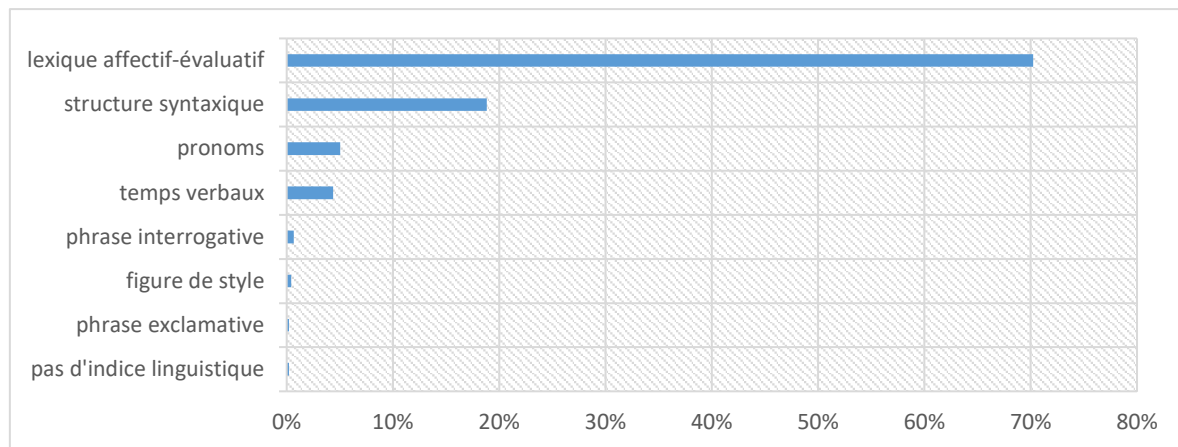


Figure 5: Types d'indices linguistiques de la subjectivité du locuteur: analyse souple

Ci-dessous, nous donnons trois exemples annotés par les annotateurs.

- (4) Il nous faut collectivement tout faire pour l'éviter, et, dans cette perspective, notre augmentation de prix contribuera directement à améliorer la rémunération de nos diffuseurs. [LM029]
- (5) C'est grâce à leurs talents, leur expertise et l'acuité de leur regard que Le Monde continuera à répondre à votre légitime exigence et restera le journal de référence que vous lisez sur tous nos supports, papier et digitaux. [LM031]
- (6) Une chose est certaine, là où la Corée du Sud, et plus spécifiquement sa vitrine, Séoul, donne l'impression de percer à l'échelle globale au détriment de Tokyo, c'est bien dans le domaine du "cool». [LM141]

Dans (4), les annotateurs ont relevé les éléments subjectifs suivants: (i) le lexique affectif-évaluatif *faut, tout, directement*; (ii) les pronoms *nous, notre*; (iii) la structure syntaxique *il nous faut* et (iv) le temps verbal du futur dans *contribuera*. Dans (5), les annotateurs ont relevé (i) la structure syntaxique *c'est [...] que [...]*; (ii) les pronoms *nos, leurs*; (iii) l'unité lexicale *journal de référence* et (iv) le temps verbal du futur dans *continuera*. Dans (6), les annotateurs ont relevé (i) le lexique affectif-évaluatif *donne l'impression* et (ii) la structure syntaxique *une chose est certaine, là où [...], c'est bien dans [...]*.

Dans l'analyse conservatrice, le lexique affectif-évaluatif, et les structures syntaxiques restent les types d'indices les plus fréquents, comme montré dans la Figure 6. Une analyse qualitative des données journalistique a mis en évidence deux stratégies d'annotation différentes développées par les annotateurs dans leur traitement des mots et de leur usage dans une phrase. Un des annotateurs mettait systématiquement l'accent sur la manière dont les mots ou syntagmes étaient placés dans une phrase, ce qui produisait une annotation de type *structure syntaxique*, alors que les deux autres annotateurs se focalisaient sur les mêmes mots mais considérés de manière individuelle, ce

qui engendrait une annotation du type *lexique affectif-évaluatif*. Malgré de telles différences, ces deux stratégies ont donné le même résultat en ce qui concerne l'interprétation subjective/non-subjective et la source responsable de la subjectivité (résultat qui confirme la suggestion méthodologique faite par Spooren & Degand (2010) à cet égard, cf. section 3.1).

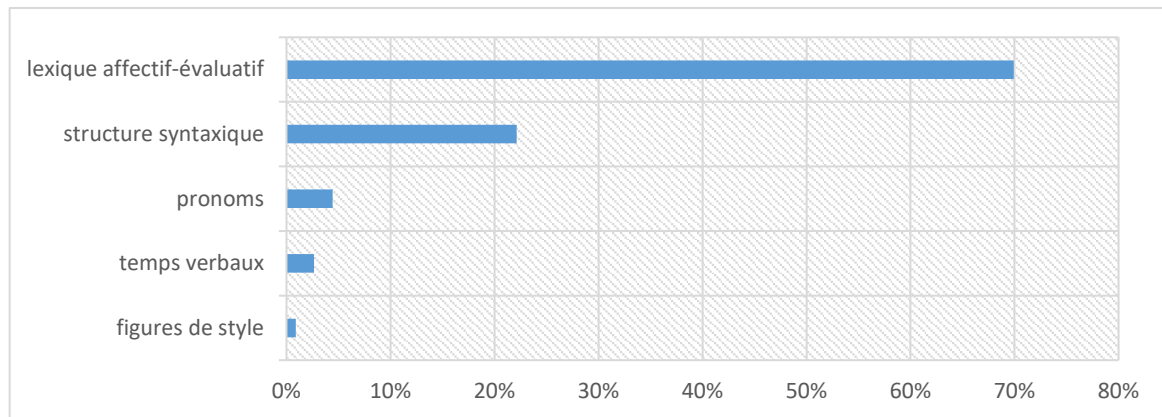


Figure 6: Types d'indices linguistiques de la subjectivité du locuteur: analyse conservatrice

Pour résumer, la deuxième phase de cette étude nous a permis de faire émerger, de manière non-contraignante pour les annotateurs, une série d'indices linguistiques de la subjectivité du locuteur. Dans les deux corpus analysés, les éléments lexicaux (du type affectif-évaluatif) et les éléments syntaxiques sont les indices les plus fréquents et les plus robustes (à travers les différents annotateurs), et ceci a été montré tant par l'analyse souple que par l'analyse conservatrice.

### 3.5 Discussion des résultats

Lors de la phase 1, les annotations ont montré que, dans les deux corpus, les énoncés interprétés comme subjectifs sont plus fréquents que les énoncés non-subjectifs, dans lesquels le locuteur décrit une situation ou une série de situations de manière objective. Les annotateurs ont identifié, en moyenne, 64.3% des extraits subjectifs dans le corpus littéraire et 65.3% dans le corpus journalistique. Ce résultat ne soutient donc pas l'hypothèse des acceptions différentes de la subjectivité selon le genre du texte (l'hypothèse 1 dans la Table 1) et montre que les textes littéraires et journalistiques peuvent manifester la subjectivité (du locuteur ou d'une tierce personne) de manière similaire. Les résultats de cette phase ont également mis en évidence le fait que, dans les deux corpus, le locuteur est le plus souvent la source responsable de la subjectivité. Ce résultat était prévu par l'hypothèse 2 (de la Table 1) pour le corpus littéraire uniquement. Si, dans le corpus journalistique, la fréquence des énoncés qui expriment la subjectivité du locuteur est de 78.2%, dans le corpus littéraire, leur fréquence est bien plus élevée (95.2%). Cette différence étant



statistiquement significative ( $X^2(1), 11.8, p < 0.05$ )<sup>13</sup>, nous concluons que, dans le genre journalistique, la subjectivité du locuteur est moins présente que dans le genre littéraire.

Cette étude a été menée dans le but de faire émerger à partir des données des corpus d'éventuels indices linguistiques de la subjectivité. Lors de la phase 2 de la présente étude, dans l'analyse souple, les résultats ont révélé une grande hétérogénéité des types d'indices linguistiques de la subjectivité du locuteur, visible surtout dans le corpus littéraire – les quatre premières catégories les plus fréquentes (c'est-à-dire le lexique affectif-évaluatif, les structures syntaxiques, les figures de style et les phrases exclamatives) occupent 76% du domaine linguistique selon l'analyse conservatrice. Dans le corpus journalistique, l'hétérogénéité des types d'indices est moins importante – les quatre catégories les plus fréquentes (c'est-à-dire, le lexique affectif-évaluatif, les structures syntaxiques, les pronoms et les temps verbaux) occupent 98% du domaine linguistique selon l'analyse conservatrice.

Les analyses menées dans les deux corpus ont mis en évidence des différences importantes en termes de fréquence entre *subjectivité affective-évaluative* (exprimée notamment par le lexique affectif-évaluatif selon la classification de Kerbrat-Orecchioni) et la *subjectivité déictique* (exprimée notamment par temps verbaux et pronoms), comme montré dans la Figure 7. Tant l'analyse souple que l'analyse conservatrice montrent que, dans les deux corpus, la subjectivité affective-évaluative est plus fréquente que la subjectivité déictique. En particulier dans le corpus littéraire, la fréquence des marqueurs linguistiques de la subjectivité affective-évaluative atteint 41.2% selon l'analyse souple et 53.8% selon l'analyse conservatrice. Dans le corpus journalistique, les taux de fréquence montent à 70% selon les deux types d'analyse. Ce résultat ne confirme pas l'hypothèse 3 (de la Table 1) selon laquelle les déictiques et le lexique sont les indices linguistiques les plus manifestes de la subjectivité du locuteur. En outre, un test du khi 2 montre que la fréquence de la subjectivité affective-évaluative est significativement plus haute dans le corpus journalistique que dans le corpus littéraire, et ceci tant selon l'analyse souple ( $X^2(1), 67.04, p < 0.05$ ) que selon l'analyse conservatrice ( $X^2(1), 8.18, p < 0.05$ ). Ces résultats montrent que dans le registre journalistique, la subjectivité affective-évaluative est particulièrement présente, ce qui est inattendu étant donné que ce registre est fortement associé à un degré d'objectivité élevé.

<sup>13</sup> Le test d'indépendance du khi-carré permet de calculer si la différence entre deux distributions de fréquence est attribuable à l'échantillonnage (le hasard) ( $p > .05$ ) ou est suffisamment grande pour être statistiquement significative ( $p < .05$ ). Dans ce dernier cas, l'écart entre les deux distributions est trop important pour être expliqué par le hasard seulement: une différence réelle existe entre les deux groupes testés.

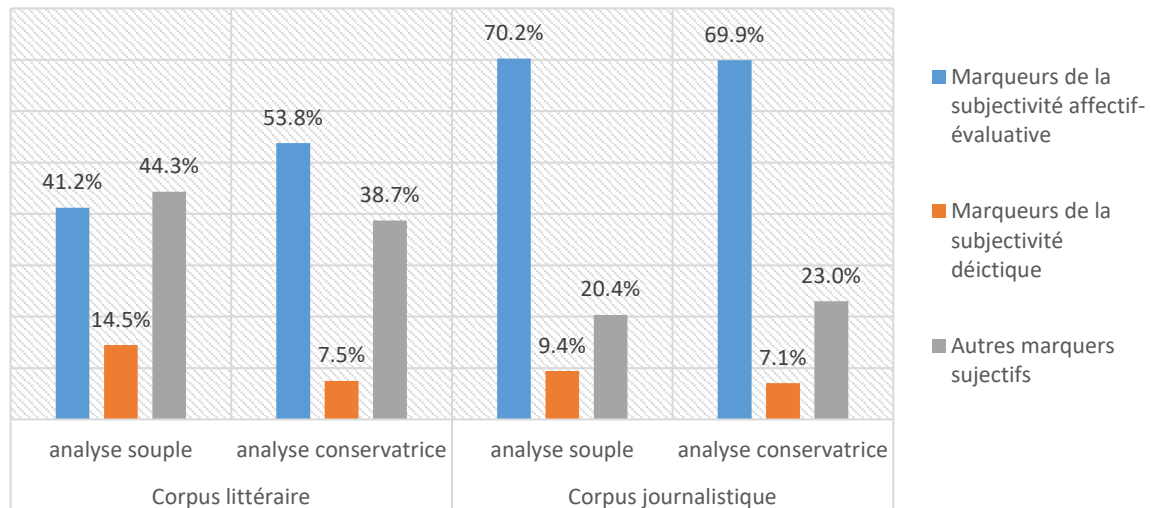


Figure 7 Subjectivité déictique et affective-évaluative (la classification de Kerbrat-Orecchioni)

Contrairement aux prédictions, les deux types d'analyse de nos données (souple et conservatrice) ont montré que dans les deux corpus (littéraire et journalistique) la subjectivité affective-évaluative est significativement plus fréquente que la subjectivité déictique. Pour nous, deux explications peuvent éclairer ce résultat. La première explication est que les annotateurs, en tant qu'usagers de la langue, ont intuitivement été particulièrement attentifs à la subjectivité affective-évaluative du locuteur. Ceci pourrait être dû au fait que ces deux types de subjectivité sont de nature très différente. En particulier, comme suggéré par Kerbrat-Orecchioni, la subjectivité déictique n'est pas subjective mais *non-subjective*, où le caractère subjectif ou non-subjectif d'un phénomène est lié à la présence ou l'absence d'un désaccord entre les personnes qui évaluent le phénomène en question. Concernant la *subjectivité déictique* et la *subjectivité affective-évaluative*, c'est uniquement dans le cas de la deuxième qu'il pourrait y avoir des évaluations différentes d'une personne à l'autre. Comme noté par Kerbrat-Orecchioni (2014: 165), l'"emploi des déictiques, tout en étant solidaire de la situation d'énonciation, repose sur un consensus incontestable: dans une situation donnée, tout le monde s'accordera à reconnaître que l'emploi d'un *ici* ou d'un *maintenant* est approprié ou inadéquat." Au contraire, l'emploi des évaluatifs peut toujours "être contesté, car il dépend de la nature individuelle du sujet d'énonciation." (Kerbrat-Orecchioni 2014: 165).

La deuxième explication possible est qu'en tant qu'interlocuteurs nous avons un accès facile et direct à la subjectivité affective-évaluative alors que la subjectivité déictique semble beaucoup moins accessible lors d'une tâche d'annotation (consciente et métalinguistique) comme celle que nous avons menée. Cette hypothèse est soutenue par deux faits. Le premier est que les pronoms ont été identifiés comme indices de subjectivité dans l'analyse conservatrice dans 6% des cas dans le corpus littéraire et dans 4% des cas

dans le corpus journalistique. Le deuxième est que la catégorie des temps verbaux ne semble pas être utilisée de manière systématique par les annotateurs comme indices de la subjectivité du locuteur. Au contraire, il est possible qu'il s'agisse plus d'une stratégie d'annotation individuelle que d'un phénomène systématique. Cette conclusion va dans la même direction que les résultats que nous avons obtenus dans Grisot (2017a), étude dans laquelle des locuteurs natifs du français ont annoté des occurrences du passé simple, du passé composé et de l'imparfait comme étant subjectifs ou non-subjectifs. L'analyse des résultats a montré que, dans les données de corpus analysées, ces temps verbaux ont autant d'usages subjectifs que non-subjectifs (cf. la proposition théorique étayée dans Moeschler et al. 2012 et Moeschler 2019). Ces résultats ne corroborent pas les hypothèses formulées dans les études linguistiques et pragmatiques au sujet du rôle privilégié de certains temps verbaux pour donner accès au point de perspective subjectif du locuteur ou d'une tierce personne (par exemple, le présent selon Benveniste 1966<sup>14</sup>; le passé simple selon Vuillaume 1990 et Tahara 2000; le passé composé futur et l'imparfait de rupture selon Saussure 2013; le présent historique selon Moeschler 2014). Toutefois, afin d'écarter cette hypothèse, des recherches expérimentales sont nécessaires afin de tester de manière systématique et contrôlée la capacité des temps verbaux visés (comme le passé simple, le passé composé futur, l'imparfait ou le présent historique) à donner accès au sujet de conscience d'un énoncé.

Sur la base de nos résultats, nous suivons Kerbrat-Orecchioni dans sa proposition de considérer la subjectivité déictique comme non-subjective et suggérons de limiter la subjectivité du locuteur à la subjectivité affective-évaluative. Cette suggestion remet en question l'approche classique (de la linguistique énonciative) de la subjectivité, définie comme la capacité du locuteur à se poser comme sujet de l'énonciation (cf. Benveniste 1966) et pointe vers une définition plus large, organisée autour de l'*affect* du locuteur et du statut *épistémique* des propositions exprimées par des énoncés. Il reste cependant encore un point à éclaircir dans nos données. Les annotations ont montré que le lexique affectif-évaluatif est le domaine linguistique principal de la subjectivité (affective-évaluative) du locuteur, suivi de la catégorie des structures syntaxiques, les figures de style et les types de phrases (exclamative, interrogative). Afin que notre proposition soit pleinement validée, nous devons par ailleurs tester si les trois autres catégories de marqueurs subjectifs sont également associées à la subjectivité affective-évaluative.

14

Pour Benveniste (1966: 239-243), certains temps verbaux, comme le passé simple, l'imparfait, le conditionnel et le plus-que-parfait, sont à privilégier dans le plan de l'*histoire* (qui est non-subjectif car "il exclut toute forme linguistique *autobiographique*"), alors que le présent est exclu de ce plan. En ce qui concerne le plan du *discours* (qui est subjectif car il est organisé autour du sujet de l'énonciation), tous les temps sont possibles, sauf le passé simple.

## 4. Conclusion

Notre but dans cet article était d'explorer les domaines linguistiques de la subjectivité du locuteur de manière empirique, en faisant usage de l'annotation linguistique des données de corpus. Cette étude avec une tâche d'annotation visait l'émergence naturelle et intuitive des indices linguistiques de subjectivité: les annotateurs ont eu la possibilité de proposer tout type de marque linguistique qui les a conduits vers une interprétation subjective des extraits annotés. Les annotations des données de corpus ont été classifiées et analysées (lors d'une analyse souple et d'une analyse conservatrice) afin d'en tirer des généralisations. L'interprétation des résultats concernant les trois variables considérées (le genre du texte, la source de la subjectivité et les types d'indices linguistiques de la subjectivité) a été faite en fonction des hypothèses formulées sur la base de la littérature et de leurs prédictions respectives.

Ainsi, cette étude a montré que, sur la base des deux corpus analysés, le genre du texte (littéraire vs. journalistique) ne semble pas influencer l'expression de la subjectivité – un taux similaire d'énoncés subjectifs (autour de 65%) a été trouvé dans les deux corpus. Des différences significatives ont été trouvées entre les deux corpus en ce qui concerne la source de la subjectivité: le locuteur est la source de la subjectivité moins souvent dans le texte journalistique (autour de 78%) que dans le texte littéraire (autour de 95%). Troisièmement, dans les deux corpus, les domaines linguistiques principaux de la subjectivité du locuteur sont le lexique affectif-évaluatif et les structures syntaxiques, avec une plus grande hétérogénéité des marqueurs subjectifs dans le corpus littéraire que dans le corpus journalistique. Enfin, les déictiques (subjectivité déictique) sont significativement moins fréquents que les éléments lexicaux (subjectivité affective-évaluative) tant dans le corpus littéraire (7.5% selon l'analyse conservatrice) que dans le corpus journalistique (7% selon l'analyse conservatrice). Ce dernier résultat est inattendu car il n'est prévu ni par l'approche de Benveniste (pour qui la subjectivité est définie autour de la capacité du locuteur à se poser comme sujet de l'énonciation) ni par l'approche de Kerbrat-Orecchioni, pour qui les déictiques et le lexique sont les indices linguistiques les plus manifestes de la subjectivité du locuteur. Malgré cela, Kerbrat-Orecchioni propose que la subjectivité déictique est non-subjective et nous la suivons en suggérant de réduire le domaine linguistique de la subjectivité du locuteur à la subjectivité affective-évaluative uniquement. Cependant, notre étude présente une lacune dans l'analyse des données, qui devra être corrigée dans une future étude: il nous reste à établir si les autres marqueurs subjectifs identifiés (notamment, les structures syntaxiques, les figures de style et le type de phrase – exclamative ou interrogative) sont des indices de la subjectivité affective-évaluative ou de la subjectivité déictique. En outre, ces résultats empiriques doivent être confirmés par des résultats d'études expérimentales.

## BIBLIOGRAPHIE

- Aikhenvald, A. (2004): *Evidentiality*. Oxford (Oxford University Press).
- Aikhenvald, A.Y. (2014): The grammar of knowledge: a cross-linguistic view of evidentials and the expression of information source. In A. Aikhenvald & R. M. W. Dixon (eds.), *The grammar of knowledge: A cross-linguistic typology*. Oxford (Oxford University Press), 1-51.
- Anscombe, J.-C., & Ducrot, O. (1983). *L'argumentation dans la langue*. Bruxelles (Editions Mardaga).
- Artstein, R. & M. Poesio (2008): Inter-coder agreement for computational linguistics. *Computational Linguistics*, (34)4, 555-96.
- Bally, C. (1905): *Précis de stylistique: esquisse d'une méthode fondée sur l'étude du français moderne*. Genève (A. Eggimann).
- Banfield, A. (1982): *Unspeakable sentences: Narration and representation in the language of fiction*. Boston (Routledge & Kegan Paul).
- Bayerl, P.S. & K.I. Paul (2011): What determines inter-coder agreement in manual annotations? a meta-analytic investigation. *Computational Linguistics*, (37)4, 699-725.
- Benveniste, E. (1966): *Problèmes de linguistique générale*. Paris (Gallimard).
- Biber, D. & E. Finegan (1988): Adverbial stance types in English. *Discourse Processes*, (11)1, 1-34.
- Bühler, K. (1934): *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Stuttgart (UTB, Stuttgart).
- Canestrelli, A.R., W.M. Mak, et al. (2013): Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive processes*, (28)9, 1394-1413.
- Carletta, J. (1996): Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, (22)2, 249-254.
- Ducrot, O. (1980). *Les mots du discours*. Paris (Minuit).
- Ducrot, O. (1984). *Le dit et le dire*. Paris (Minuit).
- Degand, L. & H. Pander Maat (2003): A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. *LOT Occasional Series* (1), 175-199.
- Eckardt, R. (2014): *The Semantics of Free Indirect Discourse: How Texts Allow Us to Mind-read and Eavesdrop*. Leiden (Brill).
- Finegan, E. (1995): Subjectivity and subjectivisation: an introduction. In D. Stein and S. Wright (eds.), *Subjectivity and Subjectivisation*. Cambridge (Cambridge University Press), 1-15.
- Fraser, B. (2011): The role of commentary pragmatic markers in conveying subjectivity. In S. Hancil (éd), *Marqueurs discursifs et subjectivité*. Mont-Saint-Aignan (Publications des Universités du Rouen et du Havre), 71-83.
- Genette, G. (1972). *Discours du récit*. Paris (Seuil).
- Giorgi, A. (2010): *About the speaker: Towards a syntax of indexicality*. Oxford (Oxford University Press).
- Giorgi, A. & F. Pianesi (2004): The temporal perspective of the speaker and the subject: from semantics to morphosyntax. In J. Guéron and J. Lecarme (eds.), *The syntax of time*. Cambridge MA (MIT Press), 129-152.
- Grisot, C. (2017b). A quantitative approach to conceptual, procedural and pragmatic meaning: Evidence from inter-annotator agreement. *Journal of Pragmatics*, (117), 245-263.
- Grisot, C. (2017a). Tense, Grammatical Aspect and Subjectivity: An Experimental Study Using Inter-Annotator Agreement Rates and Corpus-Based Data. *Corpus Pragmatics*, 1-29.
- Jakobson, R. (1957): *Shifters, verbal categories and the Russian verb*. Cambridge MA (Harvard University Press).
- Jakobson, R. (1960): Linguistics and poetics. In T. A. Sebeok (ed.), *Style in language*. New York (Wiley), 350-377.
- Kerbrat-Orecchioni, C. (1999/2014): *L'énonciation. De la subjectivité dans le langage*. Paris (Armand Colin).

- Keysar, B., D.J. Barr, J. A. Balin, & J. S Brauner (2000): Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, (11)1, 32-38.
- Kuno, S. (1987): *Functional syntax: Anaphora, discourse and empathy*. Chicago (University of Chicago Press).
- Kuno, S. & E. Kaburaki (1977): Empathy and syntax. *Linguistic inquiry*, 627-672.
- Langacker, R.W. (1999): *14 Grammar and conceptualization*. Berlin (Mouton de Gruyter).
- Lyons, J. (1977): *Semantics*. Cambridge (Cambridge University Press).
- Lyons, J. (1982): Deixis and subjectivity: Loquor, ergo sum. In R. J. Jarvella and W. Klein (eds.), *Speech, place, and action: Studies in deixis and related topics*. New York (Wiley), 101-124.
- Mercier, H. & D. Sperber (2017): *The enigma of reason*. Cambridge MA (Harvard University Press).
- Moeschler, J. (2013): Is a speaker-based pragmatics possible? Or how can a hearer infer a speaker's commitment? *Journal of Pragmatics*, (48)1, 84-97.
- Moeschler, J. (2014): Subjectivité et langage: l'exemple du présent historique. In B. Pavelin Lesic (éd.), *Francontraste: L'affectivité et la subjectivité dans le langage*. Mons (Ed. du CIPA), 29-40.
- Moeschler, J. (2019). Mais où la subjectivité se cache-t-elle donc? Une esquisse de réponse pragmatique. *Nouveaux Cahiers de Linguistique Française*, (33), 101-118.
- Moeschler, J., Grisot, C., & Cartoni, B. (2012). Jusqu'où les temps verbaux sont-ils procéduraux? *Nouveaux Cahiers de Linguistique Française*, (30), 119-139.
- Ochs, E. & B. Schieffelin (1989): Language has a heart. *Text-Interdisciplinary. Journal for the Study of Discourse*, 1(9), 7-26.
- Pit, M. (2003): How to express yourself with a causal connective: subjectivity and causal connectives in Dutch, German and French. Amsterdam (Rodopi).
- Potts, C. (2005): *The logic of conventional implicatures*. Oxford (Oxford University Press).
- Potts, C. (2007): The expressive dimension. *Theoretical linguistics*, 2(33), 165-198.
- Reboul, A. (1992): *Rhétorique et stylistique de la fiction*. Nancy (Presses Universitaires de Nancy).
- Reboul, A. (1994b). Polyphonie et énonciation. In J. Moeschler & A. Reboul (éds.), *Dictionnaire encyclopédique de pragmatique*. Paris (Seuil), 323-347.
- Reboul, A. (1994a). Narration et fiction. In J. Moeschler & A. Reboul (éds.), *Dictionnaire encyclopédique de pragmatique*. Paris (Seuil), 423-446.
- Reboul, A., D. Delfitto & G. Fiorin (2016): The Semantic Properties of Free Indirect Discourse. *Annual Review of Linguistics*, 1(2), 255-271.
- Sanders, T. (2005): Coherence, causality and cognitive complexity in discourse. *Proceedings/Actes SEM-05, First International Symposium on the exploration and modelling of meaning*, 105-114.
- Sanders, T. & W. Spooren (2009): Causal categories in discourse—Converging evidence from language use. In T. Sanders and E. Sweetser (eds.), *Causal categories in discourse and cognition*. Berlin (Mouton de Gruyter), 205-246.
- Sanders, T.J., W.P. Spooren & L.G. Noordman (1992): Toward a taxonomy of coherence relations. *Discourse processes*, 1(15), 1-35.
- Sanders, T.J., W.P. Spooren & L.G. Noordman (1993): Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics*, 2(4), 93-133.
- Sarfati, G. E. (2008). Pragmatique linguistique et normativité: Remarques sur les modalités discursives du sens commun. *Langages*, 2(170), 92-108.
- Saussure de, L. (2013): Perspectival interpretations of tenses. In K. M. Jaszczolt and L. Saussure de (eds.), *Time, Language, Cognition and Reality*. Oxford (Oxford University Press), 46-69.
- Schlenker, P. (2004): Context of thought and context of utterance: A note on free indirect discourse and the historical present. *Mind & Language*, 3(19), 279-304.
- Scholman, M.C., J. Evers-Vermeul & T. J. Sanders (2016): Categories of coherence relations in discourse annotation: Towards a reliable categorization of coherence relations. *Dialogue & Discourse*, 2(7), 1-28.

- Smith, C.S. (2003): *Modes of discourse: The local structure of texts*. Cambridge (Cambridge University Press).
- Sperber, D. & D. Wilson (1986/1995): *Relevance: Communication and cognition*. Blackwell, Oxford.
- Sperber, D. & D. Wilson (2015): *Beyond speaker's meaning*. *Croatian Journal of Philosophy*, 2(15) (44), 117-149.
- Spooren, W. & L. Degand (2010): *Coding coherence relations: reliability and validity*. *Corpus Linguistics and Linguistic Theory*, 2(6), 241-266.
- Tahara, I. (2000): *Le passé simple et la subjectivité*. *Cahiers de Linguistique Française*, 22, 189-218.
- Talmy, L. (1985): *Lexicalisation patterns: semantic structure in lexical forms*. In T. Shopen (éd.), *Language Typology and syntactic description, III: Grammatical Categories and the Lexicon*. Cambridge (Cambridge University Press), 57-149.
- Vuillaume, M. (1990): *Grammaire temporelle des récits*. Paris (Minuit).
- Wilson, D., & Carston, R. (2019). *Pragmatics and the challenge of 'non-propositional' effects*. *Journal of Pragmatics*, 145, 31-38.

## 5. Annexe 1: Instructions d'annotation, phase 1

Les locuteurs peuvent communiquer par le biais du langage leurs points de vue personnels, leurs états émotionnels et psychologiques ou les points de vue, les états émotionnels et psychologiques d'une autre personne.

Par exemple,

1. – Monsieur vous attend, madame, la soupe est servie. Et il fallut descendre ! Il fallut se mettre à table ! Elle essaya de manger. Les morceaux l'étouffaient. (Flaubert, *Madame Bovary*)

(1)

Nous sommes intéressés par la façon dont la langue peut nous donner accès à la subjectivité du locuteur ou à la subjectivité d'une autre personne mentionnée dans la phrase.

- Le locuteur exprime son point de vue à travers la description d'une situation ou une série de situations. → Le locuteur est la source de la subjectivité.
- Le locuteur exprime le point de vue d'une autre personne mentionnée par le texte à travers la description d'une situation ou une série de situations. → La source de la subjectivité n'est pas le locuteur lui-même mais une autre personne.

Selon cette définition, les phrases d'un texte peuvent être subjectives ou objectives:

- Une phrase est subjective quand la description d'une situation ou une série de situations est **centrée** sur le point de vue du locuteur ou le point de vue d'une autre personne mentionnée dans le texte.
- Une phrase est objective quand le locuteur relate une situation ou série de situations de manière objective et donc sans exprimer un point de vue personnel.

2. Une indemnité d'encouragement fut votée, séance tenante, en faveur du docteur Fergusson, et s'éleva au chiffre de deux mille cinq cents livres. (Verne, *Vingt mille lieues sous les mers*)
3. Or mon petit bonhomme ne me semblait ni égaré, ni mort de fatigue, ni mort de faim, ni mort de soif, ni mort de peur. (Saint-Exupéry, *Le petit prince*)
4. Sans résultat. Les catholiques, flairant le piège, se rallièrent autour de leur Souverain pontife, et quand on les contraignit à choisir entre leur foi et leur fidélité à l'Etat, penchèrent souvent en faveur de la première. (*Le Monde diplomatique*)
5. A Cuba, cela se traduit dès 1969 par une lettre pastorale dans laquelle l'épiscopat prenait ses distances avec l'exil. (*Le Monde diplomatique*)
6. Les dinosaures vivaient il y a des centaines de millions d'années.
7. Marie sauta dans le train. Cinq minutes plus tard, le train déraillait.



## 6. Annexe 2: Instructions d'annotation, phase 2

Les locuteurs peuvent communiquer par le biais du langage leurs points de vue personnels, leurs états émotionnels et psychologiques ou les points de vue, les états émotionnels et psychologiques d'une autre personne.

Par exemple,

1. – Monsieur vous attend, madame, la soupe est servie. Et il fallut descendre ! Il fallut se mettre à table ! Elle essaya de manger. Les morceaux l'étouffaient. (Flaubert, *Madame Bovary*)
2. Or mon petit bonhomme ne me semblait ni égaré, ni mort de fatigue, ni mort de faim, ni mort de soif, ni mort de peur. (Saint-Exupéry, *Le petit prince*)
3. Sans résultat. Les catholiques, flairant le piège, se rallièrent autour de leur Souverain pontife, et quand on les contraignit à choisir entre leur foi et leur fidélité à l'Etat, penchèrent souvent en faveur de la première. (*Le Monde diplomatique*)

Nous sommes intéressés par la façon dont la **langue peut nous donner accès à la subjectivité du locuteur ou à la subjectivité d'une autre personne mentionnée dans la phrase.**

Une phrase est subjective quand la description d'une situation ou une série de situations est **centrée** sur le point de vue du locuteur ou le point de vue d'une autre personne mentionnée dans le texte.

- Le locuteur exprime son point de vue à travers la description d'une situation ou une série de situations. → Le locuteur est la source de la subjectivité.
- Le locuteur exprime le point de vue d'une autre personne mentionnée par le texte à travers la description d'une situation ou une série de situations. → La source de la subjectivité n'est pas le locuteur lui-même, mais une autre personne.

L'interprétation subjective d'une phrase est déclenchée par un ou plusieurs indices linguistiques, qui sont donc des *éléments subjectifs*. Ces éléments subjectifs représentent la manifestation linguistique des états émotionnels et psychologiques du locuteur/d'une autre personne mentionnées dans le texte. Ci-dessous vous trouverez trois exemples d'indices linguistiques:

- Eléments lexicaux: certains mots ou groupes de mots
- Eléments syntaxiques: certaines tournures de phrase (le clivage, le parallélisme, etc.)
  - *Marie attend sur le bord de la route vs. C'est Marie qui attend sur le bord de la route.*
  - Marie part travailler, Jean part à la piscine.
- Eléments morphologiques: changement du temps verbal, la construction *être en train de*, etc.

Un test qui permet de vérifier ou confirmer la subjectivité d'un indice linguistique est de le supprimer ou de le remplacer par un autre. Si ce changement produit la conversion de la phrase de subjectif à objectif, alors cet indice est bien un élément subjectif.

Si aucun indice linguistique ne peut être identifié mais l'interprétation de la phrase est subjective, alors la subjectivité est inférée sur la base des connaissances contextuelles ou des connaissances générales du monde.