

Measuring the Degree of Subjectivity Conveyed by Connectives in Two Discourse Genres and Languages¹

Corinne Rossari, Cyrielle Montrichard, Aylin Pamuksaç,
Claudia Ricci, Linda Sanvido, Chloé Tahar

ABSTRACT

This paper makes use of two distinct statistical approaches with the aim of studying the behavior of different categories of connectives in relation to their co-occurring adverbs. We analyze the behavior of these connectives across several written corpora from different informative genres (newspaper and encyclopaedic discourse) in French and Italian. The first method mobilizes co-occurrence specificity scores to identify the collocation profile of each connective based on their subjectivity charge. The second method makes use of additive tree clustering to graphically represent groupings of connectives. The results obtained through the latter method can be aligned with the co-occurrence specificity scores, but they can also reveal characteristics that the first method does not show, thus highlighting the interest of distinct but complementary approaches.

Keywords: connectives, adverbs, subjectivity markers, co-occurrences, additive tree clustering

1. Introduction

As pointed out in our previous research, cf. Rossari et al. (2022), connectives are a particularly interesting class because of their contribution to the argumentative structure of the text. Our paper adopts a textometric approach combining different statistical methods in order to investigate the behavior of this functional word class based on corpora belonging to two distinct informative genres.

Our hypothesis is that the behavior of connectives is not only guided by the cognitive relation they convey, but also by other factors, among which the degree of subjectivity conventionally attached to their semantics. By degree of subjectivity, we mean the strength of the speaker's presence in the relation

¹ This article is a contribution to the research project *The argumentative patterns of modal forms. A corpus-based study comparing French and Italian* (n. 100001F_192247) funded by SNF.

that the connective conveys: for instance, the speaker's presence is stronger in the relation conveyed by *mais* ("but") than it is in the relation conveyed by *pourtant* ("although"), as will be shown in the next section.

This hypothesis will be tested on two Romance languages and two discourse genres – encyclopaedia and newspaper discourse – in order to assess what is attributable to the linguistic code itself and what is attributable to the genre in the behavior of connectives. It will be shown that the subjective load of connectives is less pronounced in encyclopaedic discourse than it is in newspaper discourse. The subjective load of connectives is intrinsically related to their argumentative dimension, since the argumentative purpose of a discourse, in the sense of its convincing force, is correlated to its subjective load (there is no need to convince with information that does not involve personal stances or judgments). Therefore, the use of connectives in political discourse (which can be reported in newspaper texts) tends to express a higher degree of subjectivity than their use in encyclopaedic discourse, which primarily transmits informative content. Our newspaper corpora consist of a selection of articles from regional daily newspapers – *L'Est Républicain* for French and *L'Adige* for Italian – as our encyclopaedic corpora consist of a selection of entries from French and Italian Wikipedia. We add to the genre contrast a diachronic contrast for French encyclopedic discourse, as we also investigate *L'Encyclopédie de Diderot et D'Alembert*, which allows us to shed light on the use of connectives in an encyclopaedia from a different diachronic span, the 18th century². The contrast between the Wikipedia and the Diderot et D'Alembert encyclopedia is particularly interesting in that Wikipedia is compiled by free contributors, many of which are not professionals in academic writing. Wikipedia's editing policy therefore involves relatively strict constraints, such as the ban on the use of the first-person

² *L'Est Républicain* (ER), 25'825'949 tokens, all articles published between 2010.01.02 and 2010.02.26 and between 2010.03.17 and 2010.04.30; *L'Adige* (AD), 20'343'148 tokens, 38'496 articles covering the period 1996-2006; French *Wikipedia* (FR-WIKI), 18'715'455 tokens, 10'571 articles, from 2019 (without bibliographic sections); Italian *Wikipedia* (IT-WIKI), 18'394'987 tokens, 15'905 articles from 2019 (without bibliographic sections). All articles have been randomly selected. For the 18th century, we study *L'Encyclopédie de Diderot et d'Alembert* (EDDA), 24'239'478 tokens. We thank the members of the DISCO project, and more particularly D. Vigier and A. Brenon, for giving us access to this corpus.

singular pronoun³; it contrasts with Diderot et D’Alembert’s work, where the authors’ subjectivity is claimed as a guarantee of their expertise on the topic of the article (Doutreix 2020:49–67).

The paradigms of connectives that we have selected represent different classes of primitive cognitive relations: cause, purpose, opposition, and temporality. We investigate the degree of subjectivity of each class of connectives by examining the way in which they combine with the class of adverbs ending in *-ment/-mente*. Such a class of adverbs also involves subjectivity insofar as of that the evaluation they convey involves the speaker’s voice (e.g., modal adverbs such as *nécessairement*, “necessarily”). It and therefore they greatly contribute to the argumentative/subjective dimension of the text (cf. Rossari et al. 2022).

Our methodology is based on the combination of two statistical methods⁴:

- Co-occurrence specificity⁵ scores (cf. Lafon 1980) measuring the association between connectives and adverbs, in order to identify the collocation profile of each connective;
- Additive tree clustering (ATC), whose results can show, in a single graph, distance and proximity between connectives based on their profile—that is, their specific cooccurrences with the most frequent adverbs in our corpora.

The first part of our article (section 2) presents the classes of connectives, as well as the classes of adverbs involved in our quantitative analysis. The second part (sections 3 and 4) presents the results of the two selected statistical methods; the third part (section 5) brings out the behavior of a similar set of connectives in another diachronic span (18th century), based on the *Encyclopédie de Diderot et D’Alembert* (EDDA).

³ <https://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines>

⁴ All queries and statistical calculations were made with TXM (Heiden et al. 2010), and Hyperbase was used to build the additive tree clustering (Brunet 2012).

⁵ The specificity score indicates a probability higher than 99% that the association of two items is not due to chance. It is not impacted by the overall frequency of the forms considered.

2. Our Paradigm of French and Italian Connectives and the Class of Selected Adverbs

Connectives are usually considered as playing a major role in the argumentative construction of discourse (cf. Ducrot et al. 1980; Anscombre 2002; Roulet et al. 1987). We therefore assume that their use is sensitive to the subjectivity of the speaker to varying degrees: an opposition with *mais* brings out, more strongly than with *pourtant*, the presence of the speaker as the person responsible for conceiving the opposition. This is the reason why a relation of opposition can be conveyed by *mais* with any kind of conjoined contents, whereas *pourtant* requires an opposition to be easily retrievable from the conjoined propositional contents. This difference can be observed with the following two invented examples:

Example 1. Tous les étudiants ont réussi, mais/? pourtant l'examen était facile. ('All students passed, but/although the exam was easy.')

⁶

We can observe similar differences with causal connectives: *donc* allows causal connections involving personal judgments, whereas *de ce fait* seems less prone to express such links:

Example 2. Peu d'étudiants ont réussi cet examen, donc/? de ce fait il était trop difficile. ('Few students passed this exam, so/as a result, it was hard.')

The following contexts, where many forms conveying the speaker's judgment are used, are prototypical contexts for *mais* or *donc* uses:

Example 3. On ne peut *donc* malheureusement pas être certains de la provenance de chaque pièce. ('Therefore, we cannot unfortunately be sure of the origin of each piece.')

Example 4. *Mais* franchement, il faudrait une bonne dose de mauvaise foi pour constater que [l'équipe de foot] méritait cent fois de prendre le quart (de finale). ('But honestly, it would take a fair amount of bad faith to contest the fact that [the football team] really deserved to win the quarterfinals.')

We hypothesize that if some connectives are more prone to be used with propositional content conveying personal judgment contrary to others, this is

⁶ The English translations are given for readers' convenience, but are not part of the analysis.

due to their property of conveying a certain degree of subjectivity. The higher their subjectivity load, the more likely they are to attract items conveying the speaker's point of view.

This relation between subjectivity and connectives has already been largely investigated within the frame of coherence discourse relations (Degand & Pander Maat 1999; Pander Maat & Degand 2001; Degand & Pander Maat 2003). The analysis proposed by these authors, which is essentially focused on causal connectives, relies on factors such as the type of discourse relation (fact matters, epistemic, illocutionary) conveyed by the connective and the type of causality (volitional or non-volitional) involved. These factors are tested on the basis of the annotation of a list of occurrences retrieved in corpora. Our approach shares the idea that connectives involve a subjectivity load with these studies, but differs from them in the selected diagnostics.

To test our hypothesis, we will look at the combinatorial properties of connectives with adverbs in *-ment*, which we divide into two groups:

- (i) the sentence adverbs of enunciation and modality—categorized in the inventory by Molinier and Levrier (2000) as *adverbes disjonctifs de style* (enunciation) and *adverbes modaux* (modality). These adverbs systematically convey the speaker's point of view, since they are specialized in commenting on her enunciative act (for instance *franchement*, “frankly”) or judgment (for instance, *réellement*, “really”);
- (ii) the other adverbs; even though they do not systematically reveal the presence of the speaker—for instance, if *agréablement* (“pleasantly”) is strongly marked by the speaker's voice, it is more difficult to detect this voice in the use of *techniquement* (“technically”)—they contribute, to a lesser extent, to the subjectivity of connectives.

We will label the first group enunciative group (EG) and the second non-enunciative group (Non-EG).

The connectives tested represent the main cognitive categories at play to conceptualize a relationship between utterances. We selected them by avoiding forms with a strong degree of polysemy or presenting homonymy cases. For instance, *alors*, which is particularly polysemous (it can be used as a temporal marker, consequence connective and discourse marker), or *perché*, which can be used both as a connective and as an interrogation adverb

(“why?”) while also displaying polysemy (as a causal or a purpose connective), have not been retained.

The French and Italian connectives are listed in the table below.

	French	Italian
consequence	donc, de ce fait (‘so’, ‘therefore’)	dunque, quindi
cause	car, puisque (‘because’, ‘since’)	poiché, visto che
purpose	afin que, pour que (‘in order to’, ‘so that’)	affinché
opposition	mais, cependant, pourtant (‘but’, ‘however’, ‘yet’)	ma, però, eppure
contrast	alors que, tandis que (‘whereas’)	mentre
temporality	puis, ensuite (‘then’, ‘after that’)	dopo, poi

Figure 1: Paradigm of selected connectives in French and Italian.

The set of Italian connectives has been selected based on the French list, by consulting Europarl’s parallel corpus as well as Reverso.com⁷ as a reference for translation in context, paying attention to their predominant use: *mentre* and not *quando* has been retained as the Italian counterpart of *alors que* and *tandis que*, because the high frequency of the contrastive value of *mentre* makes it closer to *alors que/tandis que* than *quando*.

3. Co-occurrence Specificity Scores between Connectives and Categories of Adverbs

Our research has been conducted in three steps: firstly, we searched for all of the adverbs ending in *ment/-mente* that are specific co-occurents⁸ of each of the selected connectives; secondly, we divided the extracted adverbs into the two groups mentioned above (EG and Non-EG); thirdly, for each

⁷ <<https://www.statmt.org/europarl/>>; <www.reverso.com>

⁸ Since the relation between the connective and the adverb is not annotated, we tested multiple query spans. We choose ten items before the connective and five after it, in order to maximize the chances of collecting adverbs in first position within the sentence and to minimize the chances that the adverb following the connective is not relevant.

connective, we merged⁹ all of its specific co-occurrences pertaining to each group, treating them as if they constituted a single macro-lexical item. This allows us to measure the specificity score of each connective not in relation to one single adverb, but in relation to all the adverbs constituting the macro category.

The results of the attraction of each connective with the EG and the Non-EG are graphically represented below¹⁰. The connectives are sorted as follows: the ones that are the most attracted by the enunciative group are displayed on the left side of the graph, while the ones that are the least attracted to the enunciative group of adverbs are displayed on the right side of the graph. Besides, the graph first takes into account the score of the EG and then the score of the Non-EG group, thus allowing us to rank them on a scale.

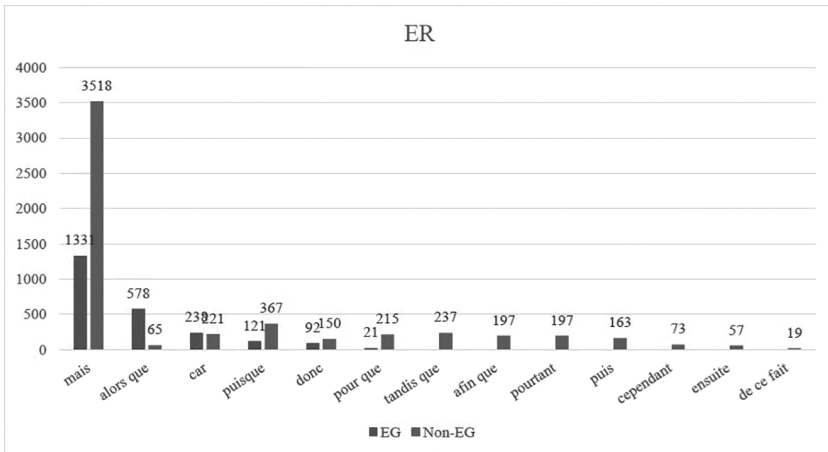


Figure 2: Specificity scores (LL) between connectives and the EG and non-EG in ER.

⁹ By merging we mean that we added up the frequencies of all the adverbs within each macro-category as well as their co-occurrences with connectives to calculate the strength of the association. We used a Log-Likelihood score for which the threshold of 10.83 indicates a higher than 99.9 % probability for the association to be due to chance.

¹⁰ The scores below the threshold are not shown in the graphs; if a connective shows no significant association with any group of adverbs it does not appear in the graphs either.

MEASURING THE DEGREE OF SUBJECTIVITY CONVEYED

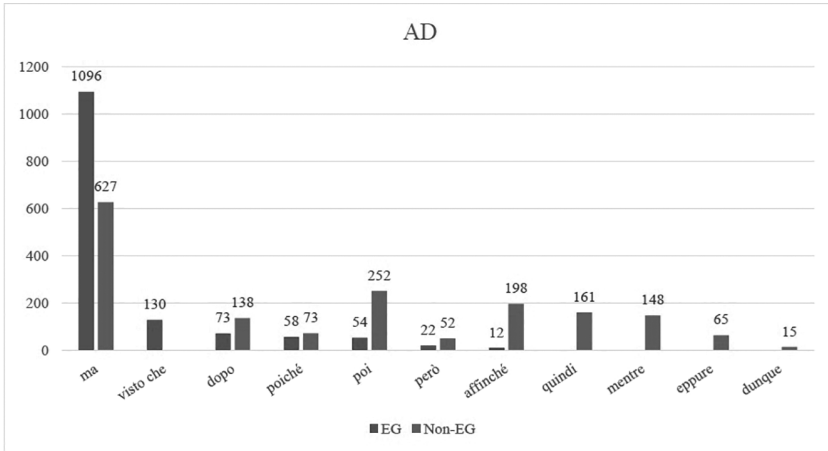


Figure 3: Specificity scores (LL) between connectives and the EG and non-EG in AD.

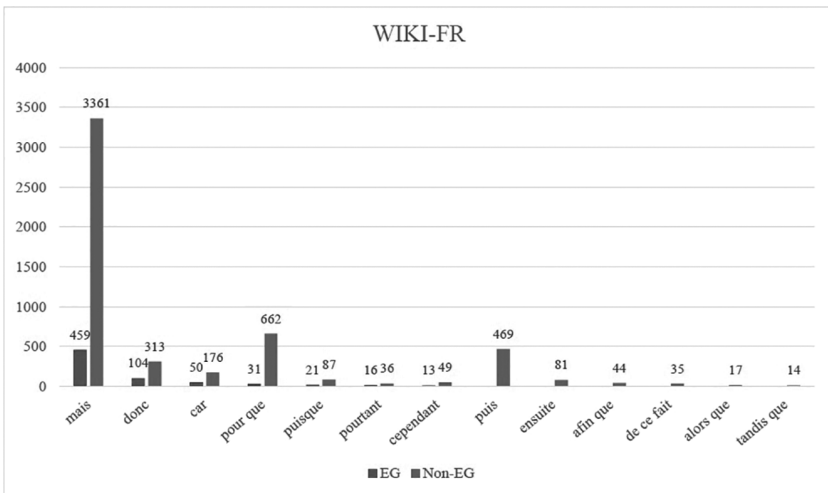


Figure 4: Specificity scores (LL) between connectives and the EG and non-EG in WIKI-FR.

At a first glance, these graphs show two striking results. First of all, *mais* and *ma* display very similar behaviors: they obtain the highest scores in both categories of adverbs in both corpora. *Ma* shows more attraction to the EG than to the Non-EG in both corpora, whereas it is the contrary for

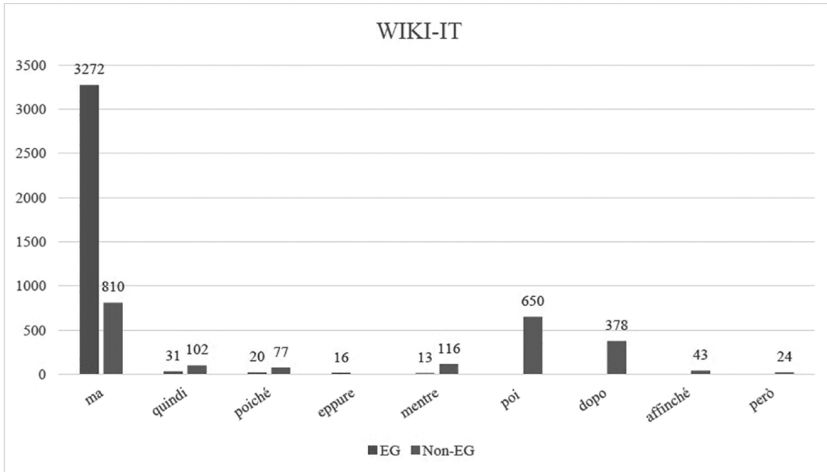


Figure 5: Specificity scores (LL) between connectives and the EG and non-EG in WIKI-IT.

mais. Secondly, the connectives attracted to the EG are also attracted to the Non-EG in both corpora of both languages, with an exception for *eppure* in WIKI-IT, only attracted to the EG.

Looking more in detail, we notice that, contrary to *ma* and *mais*, the morphological equivalents *donc* and *dunque* behave very differently: *donc* attracts forms pertaining to both groups of adverbs in both corpora, whereas *dunque* shows no attraction to either kind of adverbs in WIKI-IT and shows attraction only to non-enunciative adverbs in AD. In contrast, *quindi*, which shares the same cognitive relation as *dunque*, appears as more subjective than the latter: it is attracted to both kinds of adverbs in WIKI-IT and to a higher degree to non-enunciative adverbs in AD than *dunque*.

In French, we observe that *car* and *puisque* are very close in their behavior in both corpora: their observed similarity is in line with the fact that they share the property of connecting only speech acts (cf. Groupe Lambda-1 1975) and their tendency to appear in subjective causal contexts as shown in particular by the study of Degand (2005). It is likely that a connection between speech acts involves the subjectivity of the speaker more than a connection between the contents. This is corroborated by the behavior of *de ce fait* (a connective that only links contents), which only appears to be attracted to

non-enunciative adverbs in both corpora, and with the lowest score in ER. Such a result is convergent with our initial theoretical observation according to which *de ce fait* is less likely to convey the speaker's voice than *donc* (cf. examples 2 and 3). In this respect both methodologies (ours, and that based on the frame of coherence relations) show similar results, which means that the trait of subjectivity attached to a connective should definitely be part of their semantic profile.

While *pourtant* and *eppure* encode a cognitive relation of opposition, their behavior is in sharp contrast with that of, respectively, *mais* and *ma*. Both *pourtant* and *eppure* have much lower scores with the EG and the Non-EG than *mais* and *ma*. The latter result is convergent with our initial theoretical observation according to which *mais* is more akin to convey the speaker's voice than *pourtant* (cf. example 1).

Some observed differences are also dependent upon the genre parameter. *Alors que* occupies opposite positions in the ER graph and the WIKI-FR graph: in the former, it appears just after *mais* as a highly subjective connective, whereas in the latter it appears as clearly less subjective, close to *tandis que*. Such a difference seems to be due to the value of the connective itself: *alors que* functions more as a counter-argumentative connective in ER (similarly to *mais*), see example 5, than in WIKI-FR, where its predominant use seems to be dedicated to underline a contrast between contents, see example 6.

Example 5. Cette histoire présente les chrétiens comme les oppresseurs, *alors que* bien souvent ils sont les victimes (ER). ('This story presents Christians as the oppressors, while they are often the victims.')

Example 6. *Alors qu'*il est encore étudiant à Dauphine, Claude Harmel lui confie la gestion de la bibliothèque de l'Institut Supérieur du Travail (WIKI-FR). ('While he was still a student at Dauphine, Claude Harmel entrusted him with the management of the library of the Institut supérieur du travail.')

The analysis of the specificity scores gives an indication concerning the global behavior of the connectives in relation to the two classes of adverbs. However, it does not give us an idea of the proximity or distance of the connectives between each other. We can use the additive tree clustering method (ATC) to observe if certain connectives tend to form clusters with respect to their cooccurrence with the most frequent adverbs.

4. Additive Tree Clustering on the Matrix Connectives x Adverbs

We will now focus on the additive tree clustering method. This method allows us to look at connectives in relation to their proximity or distance between each other based on a different set of data than the one used with the specificity scores method. We extracted the thirty most frequent-*ment/-mente* adverbs of each corpus with no differentiation between enunciative or non-enunciative adverbs¹¹. For French, we worked on 28 adverbs, excluding *également* and *seulement* because of their strong attraction to *mais*, due to the adverbial expression *mais également* (but also) and *non seulement...mais* (“not only...but also”).

4.1 Results for French

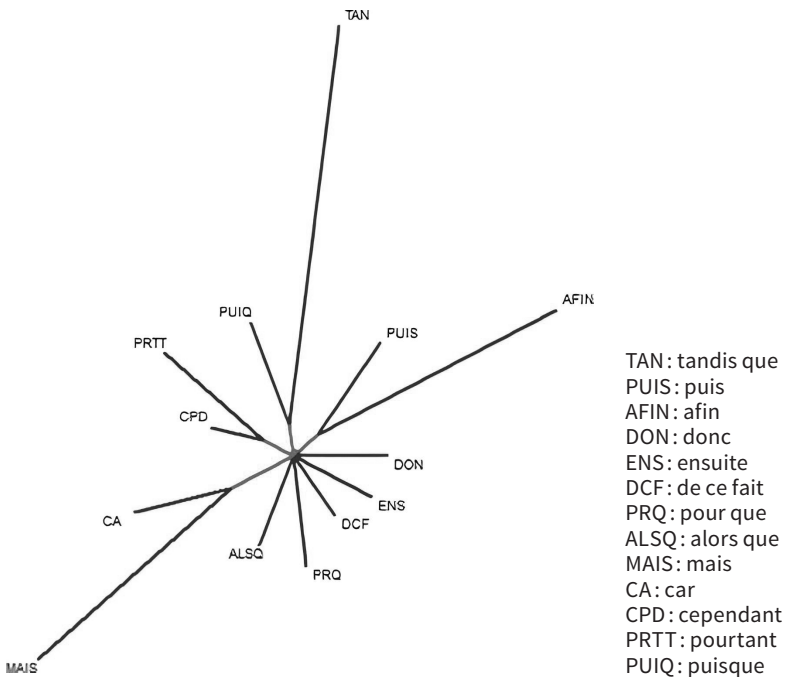


Figure 6: Additive Tree Clustering of connectives in ER.

¹¹ The matrix is based on the co-frequency between the adverbs and the tested connectives, using the same span as for the co-occurrence specificity scores.

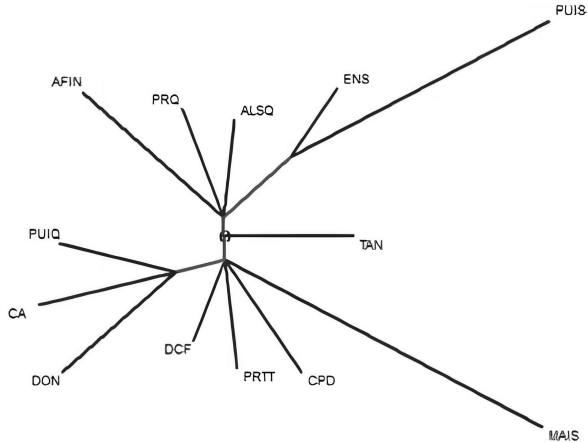


Figure 7: Additive Tree Clustering of connectives in WIKI-FR.

The additive tree clustering representation allows us to see the distance between connectives according to their profiles¹². The information in these graphs is conveyed mainly by nodes, and secondarily by the distances between leaves (Mayaffre and Luong 2003). This representation also allows observing similarities between the two corpora.

At a macro-level of analysis, we can notice that *puisque*, *car*, *pourtant*, *cependant*, *mais* stand in opposition to *pour que*, *alors que*, *ensuite* in both corpora. This opposition is very clearly noticeable in WIKI-FR, and is visible as well in ER, if we look at the proximity between *alors que*, *pour que*, *de ce fait*, *donc*, *ensuite*, which are opposed to four independent pairs of connectives: *car* and *mais*, *pourtant* and *cependant*, *puisque* and *tandis que*, *puis* and *afin*.

At a micro level, we observe proximity between the same pairs of connectives sharing a node in ER and in WIKI-FR: *cependant* and *pourtant*, *pour que* and *alors que*, and to a lesser extent *donc* and *de ce fait*, since they pertain to the same node although it is not at the same level in the tree – the node being at a higher level in WIKI-FR.

¹² To calculate the profiles, we used the z-score, which is measured considering the cooccurrences of the adverbs and the connectives.

The division between the macro-groups in WIKI-FR follows in the broad lines the partition of connectives in relation to their ability to attract or not enunciative adverbs, except for *de ce fait* – which is positioned in the group of connectives attracting enunciative adverbs, whereas its specificity scores show no attraction with them – and *pour que*, which presents the opposite behavior.

The subgroups noticeable in ER show consistency in relation to the linguistic properties of connectives: the pairs *cependant*, *pourtant* and *donc*, *de ce fait* both share the same kind of cognitive relation, respectively opposition and consequence; the pair *mais* and *car* share the property to link speech acts; the pairs *alors que*, *pour que* and *puis*, *afin* all share the property to link contents.

4.2. Results for Italian

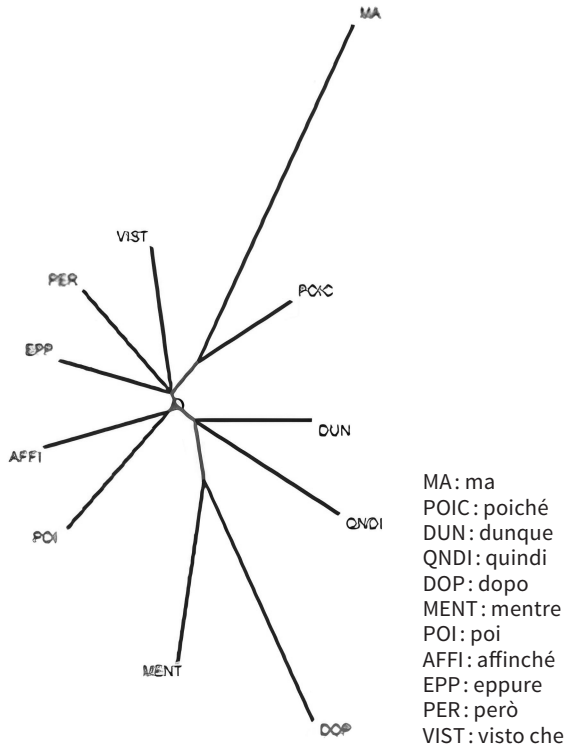


Figure 8: Additive tree clustering of connectives in AD.

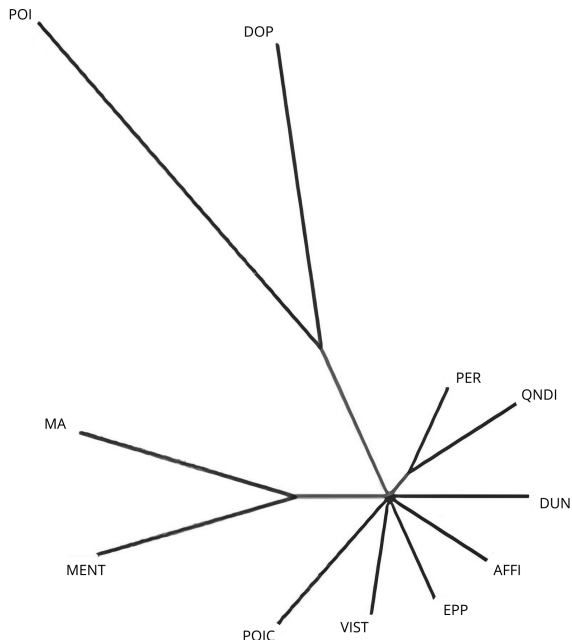


Figure 9: Additive tree clustering of connectives in WIKI-IT.

The additive tree clustering representation allows us to observe a division between connectives matching their property to attract or not enunciative adverbs in both corpora.

In AD, we find *eppure*, *però*, *ma*, *visto che*, *poiché* grouped together. Except for *eppure*, we notice that they all attract enunciative adverbs in this corpus. A second group emerges with *dunque* and *quindi*, which share the same node. This proximity, which matches the cognitive relation they express (consequence), can be linked to their non-attraction to the EG and not to the difference between their degree of attraction to the Non-EG (cf. section 3).

In WIKI-IT an opposition is noticeable between three pairs of connectives (*quindi* and *però*, *poi* and *dopo*, *ma* and *mentre*) and all the others (*poiché*, *visto che*, *eppure*, *affinché*, *dunque*). Two of these pairs share common properties: *ma* and *mentre* both show attraction with enunciative adverbs and they both convey a cognitive relation of opposition; *poi* and *dopo* strongly attract non-enunciative adverbs and convey a temporal relation.

5. The Behavior of French Connectives in the *Encyclopédie de Diderot et D'Alembert*

In this section we examine the results of the two statistical methods applied to an encyclopaedic corpus representing another diachronic span (18th century), the *Encyclopédie de Diderot et D'Alembert* (EDDA).

Investigating this additional corpus gives us a diachronic insight into the subjectivity profiles of connectives in the encyclopaedic genre. For the sake of comparability, we will primarily focus on a comparison between the subjectivity profile of the connectives in the corpus EDDA for 18th century French with their corresponding profile in WIKI-FR for present-day French. Before going any further, we would like to point out that two of the connectives investigated in present-day French were not yet commonly used in 18th century French. These connectives, *alors que* on one hand and *de ce fait* on the other, develop at later stages of the history of French according to the TIF¹³ (*alors que* developing during the 19th century, and *de ce fait* over the following century). For this reason, they are not included in our investigation on EDDA.

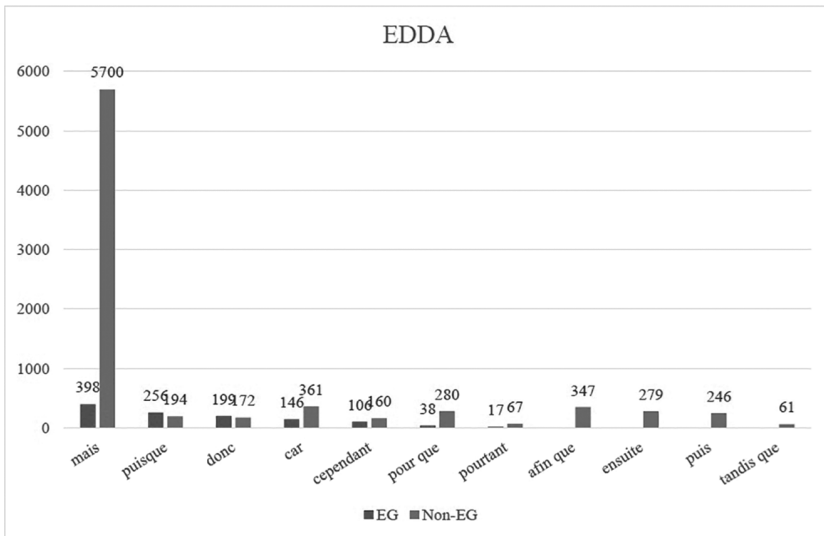


Figure 10: Specificity scores (LL) between connectives and the EG and non-EG in EDDA.

¹³ *Trésor de la Langue Française informatisé*, available online on the following page: <<http://atilf.atilf.fr/>>.

In Figure 10, the results of the specificity scores are presented. Overall, we can observe that the subjectivity profile of several connectives maintains a striking stability across the EDDA and the WIKI-FR corpora.

This observation applies to causal connectives, such as *car*, *donc* and *puis-que*. This group of connectives maintains a statistically significant attraction to both EG and Non-EG adverbs, a fact that we suggest interpreting as a manifestation of their subjectivity charge.

This observation also applies to purposive connectives such as *afin que* and *pour que*. *Pour que* has a significant LL score with both EG- and Non-EG adverbs in the EDDA corpus, unlike *afin que*, which only attracts Non-EG adverbs. The two connectives behave in a similar way in the WIKI-FR corpus. This suggests that the use of *pour que* is more likely to convey the subjective attitude of the speaker than its counterpart, *afin que*, see examples 7 and 8 from EDDA.

Example 7. Il se meut en effet de cette manière sans que rien ne l'en détourne: par conséquent pour qu'il se meuve en ligne courbe, il faut nécessairement qu'il soit poussé au moins par deux forces à chaque point de cette courbe. ("It does indeed move in this way without anything diverting it: consequently, for it to move in a curved line, it must necessarily be pushed by at least two forces at each point of this curve.")

Example 8. Si on veut se faire entendre à grande distance par le moyen d'un porte-voix, il faut prononcer chaque parole bien distinctement, afin que le bourdonnement ne cause aucune confusion. ("If one wants to be heard at a great distance through a megaphone, it is necessary to pronounce each word clearly, so that the buzzing does not cause any confusion.")

Among the group of oppositional connectives, the behavior of *mais* concerning the forms it attracts remains stable. In EDDA and WIKI-FR corpora, *mais* is the connective which most attracts EG and non-EG adverbs. However, the remaining two oppositional connectives here investigated, *cependant* and *pourtant*, do not seem to behave in the same way in the 18th century encyclopaedic genre and in the present-day encyclopaedic genre. What has changed? Firstly, *cependant* shows a strong attraction to both EG-adverbs (with a LL score of 106) and Non-EG adverbs (with a LL score of 160) in the EDDA corpus, which makes it close to *mais*, while its attraction to both groups of adverbs appears to be much weaker in the WIKI-FR corpus (with, respectively, a LL score of 49 and 13). The connective thus seems

to have become less subjective over time, undergoing a change contrasting with the well-known diachronic trend of *subjectification* (Traugott 1995). The connective *pourtant* underwent a similar trend as *cependant*, although to a lesser extent. The attraction of *pourtant* with respect to EGs remains stable across EDDA and WIKI-FR (respectively, a score of 17 and 16). However, its attraction with respect to Non-EGs is significantly more important in EDDA (with a score of 67 *contra* 36 in WIKI-FR). Based on the close inspection of corpus examples, we observe that both *cependant* and *pourtant* had a concessive value in the 18th century, next to their oppositional value. In this respect, they appear in a frequent argumentative pattern (with more than 625 occurrences, 125 with *pourtant* and 500 with *cependant*) involving the connective *quoique* (“although”) in EDDA. The content introduced by the *quoique*-clause is endorsed by a subjective stance different from the speaker, whereas the content introduced by the *cependant/pourtant*-clause is endorsed by the speaker themselves. Such a construction (illustrated by examples 9 to 12) is no longer productive in present-day French.

Example 9. Quoique leurs observations ne s'accordent pas parfaitement entr'elles, cependant la différence en est si petite, qu'elle peut être négligée. (EDDA) (“Although their observations do not agree perfectly with each other, nevertheless the difference is so small that it can be disregarded.”)

Example 10. Quoiqu'il n'ait point prétendu au caractere de législateur, il écrivit cependant des livres de lois. (“Although he did not claim to be a legislator, he nevertheless wrote books of laws”.)

Example 11. Quoique les écrits de Pomponace fussent élégants, eu égard aux écrits d'Achillinus, ils se ressentent pourtant de la barbarie où l'on étoit encore. (“Although Pomponace's writings were elegant, compared to those of Achillinus, they nevertheless reflect the barbarism still ongoing at this time.”)

Example 12. Quoique vous ne soyez pas du nombre des élus ou des favoris, vous n'avez pourtant pas sujet de vous plaindre. (“Even if you're not one of the chosen few or of the favorites, you have nevertheless nothing to complain about”.)

The observation that both *cependant* and *pourtant* had a concessive value in EDDA is corroborated by the fact that these two connectives have a significant attraction with the conjunction *quoique* (sometimes spelt as two words) in the EDDA corpus (a specificity score of 803 for *cependant* and of 231 for *pourtant*). The fact that the argumentative pattern involving *quoique* is no

longer productive in present-day French is reflected by the null score of this association in WIKI-FR (-4 for *cependant* and 0 for *pourtant*).

Globally, we notice that connectives are more attracted by EGs as well as non-EGs in EDDA than in WIKI-FR: in particular *tandis que*, which is located at the end of the scale in both corpora, has a significant score (61) with non-EGs in EDDA whereas its score is not significant in WIKI-FR (0). However, these score differences do not necessarily imply a change of value for the connective. If such is the case for *cependant* or *pourtant*, it is not for *tandis que* although the difference in score is noticeable. A closer inspection of corpus examples reveals that there is no difference between the use of *tandis que* in EDDA and in WIKI-FR: in both corpora, the connective has oppositional as well as temporal uses. See examples 13 and 14 for oppositional uses of *tandis que* in EDDA and WIKI-FR, and examples 15 and 16 for temporal uses.

Example 13. On a voulu faire une science de la nomenclature des plantes, *tandis que* ce ne peut être qu'un art, & seulement un art de mémoire. (EDDA) ("There has been an attempt to develop plant nomenclature into a science, whereas it can only be an art & only an art of memory.")

Example 14. On délaye la farine & l'amydon avec de l'eau tiède [...] *Tandis que* la colle se cuit, on la remue bien avec un balai, afin qu'elle ne se brûle pas au fond de la chaudière: on la laisse bouillir environ une bonne heure; on la retire ensuite, & elle est faite. (EDDA) ("We thin down the flour and the starch with tepid water [...] As the glue bakes, we stir it up with a broom, so that it does not burn at the bottom of the boiler; we let it boil about a good hour; we then remove it & it is done.")

Example 15. Les écarts socio-économiques étant devenus considérables: une part de la population s'est fortement enrichie, tandis que l'augmentation du coût de la vie a aggravé les conditions de vie des plus modestes. (WIKI-FR) ("Socio-economic disparities have become considerable, with a proportion of the population becoming much richer, while the rising cost of living has worsened the living conditions of the most modest.")

Example 16. Tandis que les Francs rhénans pillent une première fois Trèves, les Francs saliens protègent les provinces romaines de Belgique. (WIKI-FR) ("While the Rhenish Franks looted Trier for a first time, the Salian Franks protected the Roman provinces of Belgium.")

In short, we have observed that *cependant* and *pourtant* as well as *tandis que* are more attracted to EG-Adverbs in EDDA than they are in present days. This could seem surprising insofar as such a difference shows that they endorse a

greater subjective load, suggesting that the process of subjectification conceptualized by Traugott (1995) is bidirectional, since these connectives seem to have lost their property to attract subjective forms in nowadays encyclopaedia. However, the difference in the way encyclopaedic discourse is written cannot be ignored: EDDA absorbs much more subjectivity than WIKI-FR, as already mentioned, integrating personal stances, use of first-person pronouns (not admitted in Wikipedia). This can partly explain that these three connectives appear with more EG-Adverbs in EDDA.

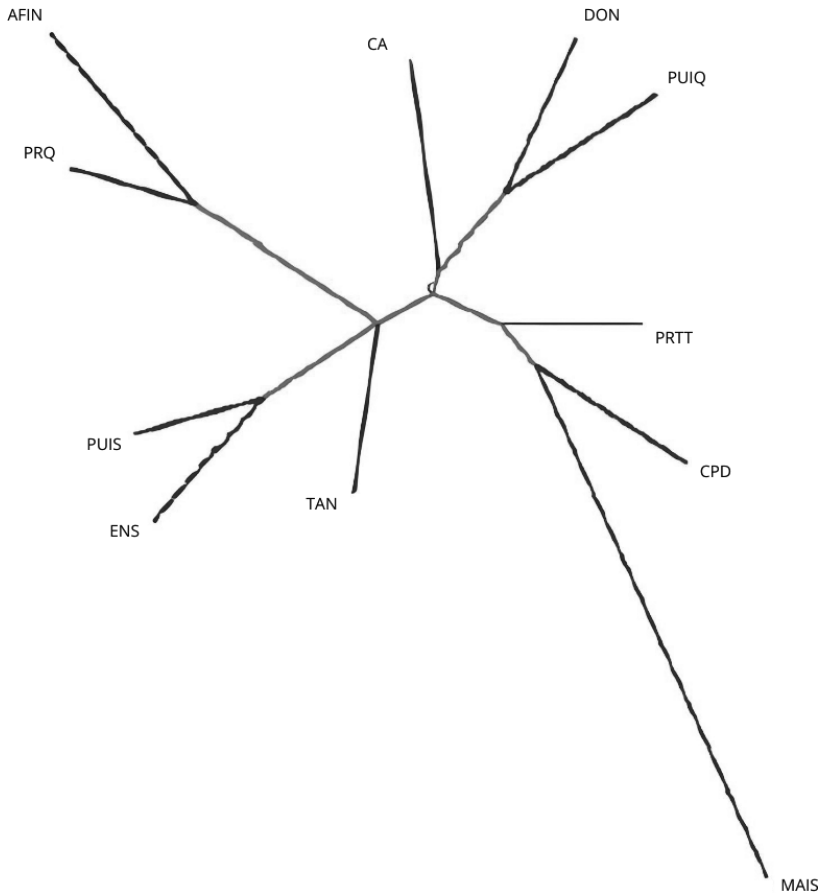


Figure 11: Additive Tree Clustering of connectives in EDDA.

By looking at the additive tree clustering, we can corroborate these results. It is interesting to note that, overall, the graph resembles the one we obtained with WIKI-FR in Figure 6. In both cases, the nodes that the connectives share are linked to their cognitive relations (e.g., causality, purpose, contrast). *Cependant* and *mais*, which both express concession relations, are located on the same branch. This confirms the observations made above concerning a certain closeness in their attraction to enunciative adverbs. In WIKI-FR, *mais* and *cependant* share the same node with *pourtant* and *de ce fait*, whereas in EDDA, only *mais* and *cependant* are brought together. We can interpret this configuration as a sign of close proximity between these two connectives in EDDA. This is less so in WIKI-FR (they appear with two other connectives on the same branch).

Temporal connectives such as *puis* and *ensuite* also share the same node. Halfway between them and the group of concessive connectives is *tandis que*. *Tandis que*, as noticed above, can function both as a temporal and oppositive connective, which is probably why it lies halfway between the junction of these two categories of connectives. The attraction with adverbs particularizing *tandis que* in EDDA does not impact its position, which is very similar to the one it has in WIKI-FR in figure 7.

Donc, *puisque* and *car*, causal connectives, are also very close to each other. This confirms both their proximity in terms of the relationship they express, and also the attractions we observed in the graph above. In fact, these three connectives have very similar specificity scores in EDDA, both with the Non-EGs and the EGs, but also in WIKI-FR. This information shows us a certain stability of these connectives over time.

To conclude our observations on the EDDA corpus, we can see that both the specificity scores and the tree analysis provide interesting information on the evolution of connectives. We were able to note that, for example, most of them had the same properties linked to subjectivity as in the contemporary encyclopaedic WIKI-FR corpus. The ones that seem to have evolved the most – by losing their subjective load – are *cependant* and *pourtant*; *tandis que* has also changed similarly, although less drastically; the change does not appear in its position in ATC.

6. Conclusion

The goal of this study was to evaluate the behavior of connectives in relation to their combinatory profile with *-ment/-mente* adverbs, so as to detect to what extent their subjectivity load, in addition to their relational properties, guide their behavior. The two statistical methods applied to two different corpus genres in terms of argumentative and subjective dimension sustain the hypothesis that the behavior of connectives is not only guided by their relational properties but also by their subjective load. Since the latter parameter is relevant in the two languages on which we have been working and can be observed in different corpora and different diachronic spans, it makes sense to address it as a conventional trait of connectives. Connectives should therefore be classified in relation to the amount of subjectivity that they convey in addition to the cognitive relation they display.

References

- Anscombre, J.-C. (2002). Mais/pourtant dans la contre-argumentation directe : raisonnement, généralité, et lexicque. *Linx*, 46, 115–131.
- Brunet, E. (2012). Nouveau traitement des cooccurrences dans Hyperbase. *Corpus*, 11, 219–248.
- Degand, L. & Pander Maat, H. (1999). Scaling Causal Relations in Terms of Speaker Involvement. In *Levels of Representation in Discourse, Working Notes of the International Workshop on Text Representation*, Edinburgh University, 45–54.
- Degand, L. (2005). De l'analyse contrastive à la traduction : le cas de la paire *puisque – aangezien*. In G. Williams, *La linguistique de corpus*. Rennes : Presses universitaires de Rennes, 155–168.
- Degand, L. & Pander Maat, H. (2003). A Contrastive Study of Dutch and French Causal Connectives on the Speaker Involvement Scale. *Lot Occasional Series*, 1, 175–199.
- Doutreix, M.-N. (2020). *Wikipédia et l'actualité : qualité de l'information et normes collaboratives d'un média en ligne*. Presses Sorbonne Nouvelle.
- Ducrot, O. & et al. (1980). *Les mots du discours*. Paris : Minuit.

- Groupe Lambda-1 (1975). Car, parce que, puisque. *Revue Romane*, 10, 248–280.
- Heiden, S., Magué, J.-P. & Pincemin, B. (2010). TXM: Une plateforme logicielle open-source pour la textométrie – conception et développement. *JADT 2010*. Rome, Italie.
- Lafon, P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots. Les langages du politique*, 1 (1), 127–165.
- Mayaffre, D. & Luong, X. (2003). Les discours de Jacques Chirac (1995-2002). *Histoire & mesure*, XVIII 3-4, 289–311.
- Molinier, C. & Levrier, F. (2000). *Grammaire des adverbes. Description des formes en -ment*. Genève: Librairie Droz.
- Pander Maat, H. & Degand, L. (2001). Scaling Causal Relations and Connectives in Terms of Speaker Involvement. *Cognitive Linguistics*, 12, 211–245.
- Rossari, C., Montrichard, C., Ricci, C. & Sanvido, L. (2022). Subjectivity and Connectors: Study of the Parameters of Discourse Genres and Languages. *16th JADT*, 772–779.
- Roulet, E. & et al. (1987). *L'articulation du discours en français contemporain*. Bern: Peter Lang.
- Traugott, E.C. (1995). Subjectification in Grammaticalisation. In Stein, D. & Wright S. (Eds.). *Subjectivity and Subjectivisation: Linguistic Perspectives*. Cambridge: Cambridge University Press, 31–54.