

PREPRINT

Corinne Rossari and Claudia Ricci

31 Exploring the behaviour of connectives within a textometric perspective

Abstract: This chapter proposes to deal with the issue of categorising connectives as a word class by adopting an innovative outlook, based on investigating the properties that emerge from the linguistic environment of these forms. After recalling the criteria used in traditional classifications of connectives in Romance languages and the various issues such criteria raise, we will take a selection of French and Italian connectives and observe their favourite linguistic environments – i.e. the forms associated with them in a statistically significant way – within a press corpus, using a textometric methodology. The data thus obtained will reveal properties contributing to the semantic profile of a given connective and suggest categorisations and sub-categorisations not always matching those based on traditional relational/functional criteria, but providing additional insight on the behaviour of these forms in our corpora.

Keywords: connectives, classification, statistical methods, specific associations, semantic profile

1 Introduction: connectives and their identification as a word class

Recognising connectives as a word class raises many issues.¹ If we consider the three criteria outlined by Salvi (2013, 15–24), (i) morphological, (ii) notional (or semantic), (iii) syntactic-functional, the class of connectives shows homogeneity as well as heterogeneity for each of its members.

Regarding the morphological criterion, connectives show overlapping with conjunctions and conjunctive adverbs, insofar as they link parts of text/discourse (see, *inter alia*, De Cesare 2019), but the proliferation of research studies on these forms since the Seventies has led to include more and more word classes in this category. To take French as an example, all the word classes listed below (together with some case studies among many others) have been investigated with regard to their function as connectives:

¹ This chapter 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 the Swiss National Science Foundation.

- Verbs: *disons* ‘let’s say’ (see, inter alia, Steuckardt 2018 within a more general analysis of markers formed on the verb *dire*), *mettons* ‘let’s suppose’ (Dostie 2004), *regarde* ‘look’ (Dostie 1998; Rossari 2011), *voyons* ‘come on’;
- Nouns: ~~*résultat*, *conclusion*~~ ‘result, conclusion’;
- Pronouns: *où* ‘where’ (Le Querler 2000);
- Adjectives: *pire* ‘worse’ (Rossari 2021, 1983);
- Interjections: *eh bien!* ‘well’ (see, inter alia, the interesting contrastive French-Romanian analysis in Pop 2006), *hélas!* ‘alas’, *ben* ‘well’ (see, inter alia, Bruxelles/Traverso 2001; Buridant 2006);
- ‘Polylexical fixed units’ (in the terms of Guimier 2009), such as adverbial phrases: *aussi bien* ‘as well’;
- Prepositional phrases: *d’où* ‘from which’ and noun phrases: *la preuve* ‘the proof’ (see Rossari 2005; Rossari/Cojocariu 2008);
- Main clauses: *il en résulte que* ‘the result is that’, *je veux dire* ‘I mean’ (for the latter, and for a whole family of propositional discourse markers such as *je crois*, *je pense* etc., see Andersen 1996; 2007) and subordinate clauses: *quoi qu’il en soit* ‘in any case’, among which relative clauses (*qui plus est* ‘what’s more’).

Although all these forms share a common morphological trait – invariability – they encompass not only traditional word classes, but also syntactic categories.

The notional criterion is not more homogeneous: even if we reduce it to the dichotomy lexical vs grammatical or, on a cognitive level, conceptual vs procedural (Blakemore 2002; Wilson 2011), the functioning of some connectives does not appear to be linkable to a single notional category, but has traits showing affinity with both: if *et* or *mais* are procedural, what about *il n’en demeure pas moins que* ‘the fact remains that’, or *pour mieux dire* ‘to put it better’? The decoding of the lexical items composing such connectives (*demeurer* ‘stay’, *dire* ‘say’) is not sufficient to access their contribution.

Only the functional criterion seems to be efficient to gather all the members of the class. Apart from very broad descriptions (such as, for instance, the one by Bazzanella 1995, which places Italian connectives within the macro-group of discourse markers, distinguishing their functions between *interactional*, *metatextual*, and *cognitive*), what most classifications have in common is the idea that all items being connectives have a linking function. However, on the one hand, such a function is not a prerogative of connectives: other classes ensure links between parts of a text or discourse, such as resumptive anaphora (*le problème* ‘the problem’), frame setters² (*en théorie* ‘in theory’, *en réalité* ‘in reality’, *en pratique* ‘practically’). On the other hand, many items pertaining to these ‘linking categories’ can also function as connectives: for instance, *le problème* can function both as a resumptive anaphora and as a connective conveying a relation compar-

² In the sense of Charolles/Lamiroy (2012, 93), namely adverbials in pre-verbal position extending their scope on the sentence following the one in which they appear.

able to that established by *mais* in a context such as: *L'économie repart, le problème est que l'inflation aussi / L'économie repart, mais l'inflation aussi* ('Economy is recovering, the problem is that inflation is too / Economy is recovering, but so is inflation'). Some forms fall in between two categories: if *en fait* ('in fact') is no longer considered a frame, but only a connective or a discourse marker, *en réalité*, usually conceived as a connective (Rossari 1997; Wandel 2020), is also considered a frame when it is used in correlation with *en théorie*.

If these traditional criteria are not helpful to define the class, that does not mean that the class is artificial. The items considered as being connectives still share a common trait of linking, even if the contours of such a trait are difficult to draw. A great number of categorisations of connectives allude to their linking function using notions such as *logical relations* (Ferrari 2014; Pons Bordería 1998), *coherence relations* (Asher/Lascarides 2003; Bras et al. 2021), *cohesion relations* (García Izquierdo 1998), *rhetorical relations* (Mann/Thompson 1987), *argumentative relations* (Ducrot 1980; Anscombre/Ducrot 1983, but also, more recently Borillo 2019; Rocci et al. 2021), *discourse relations* (Roulet et al. 2001). These notions allow to subdivide connectives into different classes according to the kind of linking function they show, based on their interchangeability properties: for instance, the substitution between *par conséquent* and *donc* or between *mais* and *toutefois* does not seem to affect the cognitive link that an interpretant establishes between two utterances – and this justifies their grouping within the same category –, whereas the substitution between *mais* and *donc* does, as shown by (1) and (2) below:

- (1) Marie est restée à la maison, **mais/toutefois** Lucie est sortie.
'Marie stayed home but/however Lucie went out.'
- (2) Marie est restée à la maison **donc/par conséquent** Lucie est sortie.
'Marie stayed home **so/consequently** Lucie went out.'

This does not mean that the similarity of semantic/pragmatic links motivating the grouping of *mais* and *toutefois*, to name but one case, is easy to define at a level of linguistic analysis. For instance, the link related to the idea of *opposition/concession*, called logical, semantic, or argumentative according to the theoretical framework under which it is scrutinised, is hardly definable, as it covers so many different discursive manifestations. The following utterances:

- (3) Paul est petit, Jean est grand.
'Paul is small, Jean is tall.'
- (4) Ton travail est bon; **toutefois** il aurait pu être meilleur.
'Your work is good; however, it could have been better.'

- (5) Ce que j'ai à dire a déjà été dit, **mais** je le répéterai brièvement.
 'What I have to say has already been said, but I will briefly repeat it.' (Ducrot 1999, 120, our transl.)

are all discursive manifestations that the opposition/concession label can cover. This is the reason why some classifications have resorted to the functional properties that connectives assign to the utterances they link. Using the argumentative framework by Anscombe/Ducrot (1983), Roulet (1997) and Roulet et al. (2001) distinguish connectives in relation to their discursive function: argumentative connectives introduce arguments in relation to a previous or following conclusion (*car* 'because', *parce que* 'because', *puisque* 'since'); conclusive connectives introduce the conclusion (*donc* 'therefore', *par conséquent* 'consequently', *de ce fait* 'therefore, as a result', *pour cette raison* 'for this reason'), counter-argumentative connectives introduce the conclusion (*mais* 'but', *cependant* 'nevertheless', *toutefois* 'however', *pourtant* 'yet') or the counter-argument of a sequence in which two utterances are anti-oriented (*bien que*, *quoi que*, *malgré* 'although, in spite of'), and reformulative connectives have the property of introducing an equivalent meaning (*c'est à dire* 'that is to say', *en d'autres termes* 'in other words') or a new perspective on a previous statement (*en tout cas*, *quoi qu'il en soit* 'in any case, anyway'). Nølke (1993) looks at the pragmatic function of connectives, differentiating between those whose function is to establish a connection to extralinguistic elements (*eh bien!*), to linguistic acts (*puisque*), or to propositional contents (*pourtant*), or to any of these elements depending on the context, like *donc* (Nølke 1993, 142).

Even though these functional classifications allow to grasp some common points in the relationship that the grouped connectives signal, they are still porous because of the polyfunctionality inherent to connectives. In other words, the same connective can integrate different classes; for instance, *donc* is reformulative as well as conclusive in classifications based on relation properties, but, as just mentioned (see Nølke 1993 above), it is also polyfunctional in terms of the nature of the arguments it introduces; *It. comunque* ('anyway') seems to be in between the class of reformulative connectives and that of counter-argumentative connectives. Also, some connectives can convey different relations according to their position in the utterance; for instance, *in effetti* ('indeed', 'actually') is analysed by Mandelli (2008) as tendentially conveying a reformulative relation when it is placed in a sentence-internal position, and a justification in an initial position. The fact that the function assigned by connectives to the utterance they introduce – and therefore their specific contribution – is not clear-cut is made apparent by the possibility, for connectives pertaining to different categories, to easily combine together. In a sequence as the following:

- (6) Forse credevo fosse "qualcosa di diverso", non so, **ma comunque** sembrava di parlare ad un amico e ricevere le solite rassicurazioni "di circostanza". (Internet)
 'Maybe I thought it was 'something different', I don't know, but anyway it felt like talking to a friend and getting the usual casual reassurances.'

in which *ma* co-occurs with *comunque*, what is the contribution of *ma* vis-à-vis that of *comunque* and vice versa? Is the utterance introduced by both connectives a counter-argument, a reformulation or both? The difficulty of grasping the contribution of connectives is even more apparent when one has to dissociate connectives pertaining to the same functional class. Many studies have dealt with this issue in Romance languages (see e.g. Manzotti 1998 on Italian exemplification markers; Portolés 1998 on pairs of Spanish markers of opposition and of conclusion; Lopes/Sousa 2014 for Portuguese opposition connectives; Rossari 1997 on French-Italian reformulation markers).

Our purpose is not to solve the issues that any categorisation of connectives will raise, but to open a new perspective based on the investigation of their properties emerging from their linguistic environments. Our hypothesis is that connectives are sensitive to the speaker's subjectivity due to the fact that, according to the connective used, the speaker's voice emerges more or less visibly in the relation it conveys. Using a textometric methodology including two different statistical calculations, we will observe the favourite linguistic environments of the connectives considered, namely their *specific co-occurrences*, i.e. forms whose co-occurrence with the connective is shown by statistical calculation to be significant. Such selections will be interpreted as revealing properties contributing to the semantic profile of a given connective. The way in which each connective combines with such statistically associated forms will reveal similarities but also differences between connectives, including those pertaining to the same relational/functional class.

In Section 2, we present the theoretical basis of the selection of items retained for our analysis (connectives as well as forms associated with them in a statistically significant way); in Section 3 we present the two statistical calculations used and the methodology followed for our queries to observe the behaviour of the selected connectives. In Section 4 we present our hypotheses concerning the groupings and de-groupings of the selected connectives as they emerge from the data resulting from the two different types of statistical calculations used. Before our concluding remarks, Section 5 gives the linguistic interpretation we draw from the groupings in terms of the type of semantic profile shared by connectives forming a cluster.

2 Selection of corpora and forms

Our investigation was conducted on two Romance languages, French and Italian. As for the selection of the corpora investigated, we chose the press, because it guarantees a large variability in terms of content, authors, genres. The focus of our study being on the use of connectives, we do not aim to conduct a comparative analysis between the press corpora of each language. Therefore, to maximise variability, we chose newspapers addressing respectively a more local audience – the Italian regional daily newspaper *L'Adige* – and a more national audience – the French daily

newspaper *Le Monde*.³ Independently of the diversity ensured by the press, we shall remind that any corpus research entails that the validity of the results is limited to the selected corpus. In the words of Blumenthal: “Nous ne cessons de souligner que la validité des résultats de cette méthode dépend du corpus choisi. Car le style de ‘la vie des mots’ [...] diverge souvent fortement d’un type de texte à l’autre [‘We keep emphasising that the validity of the results of this method depends on the corpus chosen. For the style of the ‘word’s life’ [...] often differs greatly from one type of text to another’]” (Blumenthal 2008, 35, our transl.).

Our approach is corpus-based, which means that we determine, upstream in the research process, what items we will be testing (selection of connectives) as well as what items are representative of the behaviour of the connectives we test (choice of the specific co-occurents of the selected connectives).

I. Selection of connectives. A paradigm of connectives has been investigated for each language, using parallel corpora – Europarl and Reverso – to substantiate the selection of Italian connectives as the most frequent counterparts of the French ones. The selection of French connectives is based on the diversity of the functional classes that they represent in the classification of Roulet et al. (2001), avoiding, as much as possible, homonymous forms.

In the four categories of connectives considered, presented below, we retain, for French and Italian respectively:

- 2 argumentative connectives: *car, puisque* ‘because’ / *poiché, visto che* ‘because’/‘since’;
- 2 conclusion connectives: *donc, par conséquent* ‘so, therefore’, ‘consequently’ / *quindi, di conseguenza* ‘therefore’, ‘consequently’;
- 2 counter-argumentative connectives: *mais, pourtant* ‘but, yet’ / *ma, però, eppure* ‘but, however, yet’;⁴
- 2 reformulation connectives: *c’est-à-dire, en d’autres termes* ‘that is (to say)’, ‘in other words’ / *cioè, in altre parole* ‘that is’, ‘in other words’.

To these four categories, we added the category of exemplification connectives (which was extensively investigated for Italian in Manzotti 1998), which have the property of introducing an example or an illustration of a previous content. For French, we selected

³ Our corpus *Le Monde* consists of 17,895,009 tokens, covering the year 2010; the corpus *L’Adige* consists of 20,343,148 tokens, corresponding to 38,496 articles covering the period 1996–2006.

⁴ We added *però* to our list as it is largely synonymous for *ma*. Also, we had to exclude *perché* from the group of connectives as it is too polysemous: it stands for *why, because*, and *in order to*, thus not always being a counterpart for *parce que*, as it does not always indicate a relation of cause. However, we did not exclude *perché* from the list of items co-occurring with our group of selected connectives; we listed this form among the forms signalling the discursive operation of ‘reasoning’ (see below), because we consider that it signals this kind of operation in all of its meanings.

par exemple ‘for instance’, *en particulier* ‘notably’, and for Italian *per/ad esempio*, *in particolare* (same transl.).

II. *Selection and categorisation of the specific co-occurrences of the connectives.* To initially access all the single possible items representing the favourite linguistic environment of the connectives in our corpora, we resorted to TXM, a platform for the analysis of structured and annotated textual corpora (Heiden/Magué/Pincemin 2010). TXM allowed us to automatically retrieve all the items being specific co-occurrences of each connective (in the sense explained in Section 1), within a right and left span of 10 items around it,⁵ by assigning each association a *co-occurrence specificity index*.⁶ Among all the specific co-occurrences thus retrieved, we wanted to bring out forms conveying the speaker’s subjectivity, on the basis of the hypothesis that, the more subjective a connective is, the more it will attract forms of this type. We selected the forms linked to the cognitive processes at stake when a relation is signalled by a connective: reasoning, evoking speeches or points of view, thinking, judging, showing personal views. These five cognitive processes take shape through *discursive operations*, identifiable via families of forms. Below is a description of each discursive operation, with examples to showcase the links between the articulations realised by the connectives and the forms revealing the operation, which we put in bold to better highlight them.

1. The *reasoning operation* corresponds to a process in which the speaker puts in relation contents through inferential links. Such a process must be visible through the linguistic item chosen, which can be modal forms, connectives, evidential adverbs, such as in the following examples: *Paul est parti, donc il peut s’être fâché/apparemment il s’est fâché* ‘Paul left, therefore he may have gotten upset/apparently, he got upset’; *Bien qu’elle ait dû voir qu’il était arrivé, elle n’a probablement pas voulu l’attendre* (‘Although she must have noticed that he arrived, she probably didn’t want to wait for him’); *Si Marie vient, Paul, alors, sera peut-être content* (‘If Marie comes, then perhaps Paul will be happy’). We thus distinguish cases like *Paul doit avoir agi de bonne foi* (‘Paul must have acted in good faith’), in which the presence of *devoir* qualifies the content introduced as being the result of inference, from cases like *Visiblement, Paul a agi de bonne foi* (‘Visibly, Paul acted in good faith’), in which – regardless of the mental process leading to the content – the adverb qualifies the state of affairs as being unmistakably evident. We classify this last case as a case of *judgement operation* (see below).

⁵ The span mentioned here was identified, after a series of tests, as corresponding to the best possible amplitude ensuring that the co-occurrence between a given form and the connective was relevant in our corpora.

⁶ See Lafon (1980). The specificity index indicates the probability that the association between two items is not due to chance. A score of 3 or above indicates attraction between the items considered.

2. The *enunciative operation* allows the speaker to put into perspective different points of view or speeches. Polyphonic markers (*certes* ‘certainly’/‘admittedly’, *il est vrai* ‘it is true’), *dicendi* verbs (*affirmer* ‘to state’, *prétendre* ‘to claim’) highlight the allusion to enunciative heterogeneity: ***Il est vrai*** *que Pierre a exagéré, mais tu lui as parlé sèchement* (‘It is true that Pierre overreacted, but you spoke to him curtly’) / ***Il m’a dit*** *que tu souhaites participer à la réunion, pour cette raison je t’ai inscrit* (‘He told me you wish to take part in the meeting, this is why I signed you up for it’); enunciative adverbs (*franchement* ‘frankly’, *sincèrement* ‘sincerely’, *honnêtement* ‘honestly’...) convey an indication of commitment: ***Franchement***, *il aurait pu te prévenir, étant donné que tu comptais sur lui* (‘Frankly, he could have informed you, since you were counting on him’).
3. The *thought operation* is explicitly signalled by the use of thought verbs (this is why we distinguish this operation from that of *reasoning*, in which different forms make appear the content as the result of an inference): ***J’ai toujours cru*** *que Paul était honnête, cependant maintenant je ne le pense plus* (‘I always believed that Paul was honest; however, now I no longer think so’); ***Je doute*** *que le cours ait lieu, car je ne vois personne dans la salle* (‘I doubt that the course will take place, because I cannot see anyone in the room’).
4. The *judgement operation* allows the speaker to make explicit that a content is the result of an activity of judgement, usually expressed by verbs, such as in: *Ce film me semble* excellent, *pourtant il a reçu de très mauvaises critiques* (‘I found this movie excellent; yet it got very bad reviews’); ***Visiblement***, *Marie a bien fait de changer de travail, puisqu’elle se sent mieux maintenant* (‘Visibly, Marie was right to change jobs, since she feels better now’).
5. The *personal view operation* is linked to appreciations expressed with forms as adjectives or adverbs conveying subjective meaning, i.e. based on personal standards (*bon* ‘good’ *mauvais* ‘bad’, *bêtement* ‘stupidly’). Such contents tend to be supported by the speaker with additional contents: *Julie a une mémoire impressionnante, elle peut, par exemple, réciter par cœur une page qu’elle vient de lire* (‘Julie has an impressive memory; she can, for instance, recite by heart a page she just read’); *Max s’est mal comporté, parce qu’il s’est emporté* (‘Max behaved poorly, because he overreacted’).

To select these forms, we proceeded as follows: (i) we screened all the specific co-occurrences of our connectives automatically retrieved by TXM. (ii) Among them, we selected and grouped the forms in accordance to the five discursive operations described above. (iii) We avoided, as far as possible, retaining forms presenting too much homonymy or polysemy issues. We thus proceeded to a verification of the uses of possibly polysemous items; if a large percentage of their uses did not correspond to the use manifesting one

of these five operations, we did not retain the form. For instance, we did not retain *comme*, as it can be a causal connective (*Comme il était fatigué, il n'est pas venu* 'As he was tired, he didn't come'), as well as a comparison or an exclamation marker (*Il a voulu s'habiller comme son frère* 'He wanted to dress like his brother'); neither did we retain a verb as *It. trovare*, which introduces a judgement only in some particular cases (*Trovo questo film noioso* 'I find this movie boring' (judgement) / *Ho trovato una soluzione al tuo problema* 'I found a solution to your problem' (≠ judgement)). (iv) When a form is susceptible to fall into more than one category, and the large majority of its uses on the total number of occurrences in our corpora fell into one of these categories, we assigned the form to the most represented category, since the remaining percentage has no impact on the statistical calculation. For example, *parere* ('to appear') can fall into the category corresponding to evoking a point of view, as in *Pare che Marco si sia sposato* ('Apparently Marco got married') or into the category corresponding to revealing a judgement, as in *Pare difficile dare un giudizio* ('It seems difficult to give a judgement'). We classified this form as revealing a judgement because the latter use is more represented in our corpora.⁷

3 Statistical calculations and methodology

The two statistical methods exploited are complementary: the first one, calculating a log-likelihood score (LL),⁸ allows to assess the attraction shown by connectives towards the groups of items representing the discursive operations listed above; the second one, Correspondence Analysis (CA), shows to what extent connectives are similar or different in relation to their co-frequency (the frequency of their co-presence) with some particular items within each discursive operation.

The first calculation (LL), made using the software R, concerns the behaviour of our connectives with each of the five discursive operations established as described in Section 2, each of which is taken as a whole and regarded as one global 'protoform'. The scores thus obtained represent therefore the degree of attraction of each protoform to our connectives, which are shown in the graphs in Section 4.1 below. Such a method allows us to have an overview reaching beyond the preferences of a given connective for a particular lexical unit.

The second calculation (CA), made using Hyperbase (Brunet 2012),⁹ was performed to graphically show the co-frequency of connectives with a number of single forms

⁷ 80 % on a sample of 100 occurrences.

⁸ Like the specificity index, log-likelihood provides a score expressing the probability that the association between two items is not due to chance; a threshold score of 10.83 indicates that such a probability is higher than 99.9 %.

⁹ Hyperbase is a downloadable academic software for documentary and statistical exploration of texts distributed by the CNRS and the University Nice Sophia Antipolis.

among those included in the protoforms representing our five operations. These forms were selected in relation to their frequency in our corpora. We retained the most frequent forms by discursive operation (between 3 and 5 items) using the list initially established via the TXM specificity index. This method allows us to test our data, this time in relation to specific lexical forms, independently of their grouping into discursive operations, thus having a fine-grained view on the behaviour of our connectives. In order to avoid a biased representation in the CA due to extremely low or high co-frequencies, we ruled out items which either co-occur only with less than five connectives, or co-occur particularly frequently with only one connective. The list of the selected forms is provided in Section 4.2.

4 Groupings and de-groupings of the connectives according to LL and CA calculations

4.1 LL calculations

The following graphs show how each connective combines with the protoforms representing each of our five discursive operations. Only columns representing LL scores above a threshold of 10.83 appear in the graph. Columns for LL scores above 500 were cut in order to avoid visual flattening of the remaining columns.

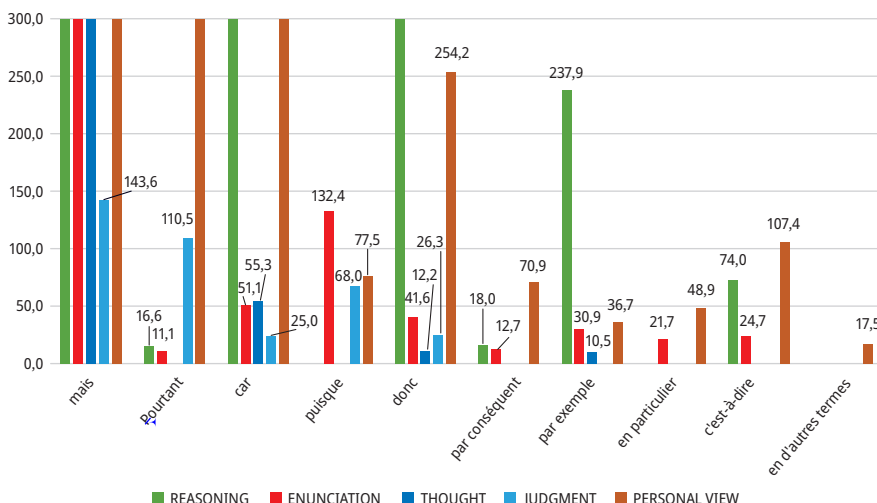


Figure 1: Associations of connectives with protoforms representing discursive operations (*Le Monde*)

At a first glance, we can divide our connectives into two groups: those – *mais*, *car*, *donc* – showing a strong sensitivity to forms representing discursive operations (high LL scores,

many protoforms) and those – *en d'autres termes, en particulier* – showing less sensitivity to these forms (low LL scores, few protoforms).

The same data can be presented through sets (Figure 2 below), which allows to better appraise the groupings/de-groupings of connectives in relation to discursive operations. The ensembles (coloured sets) represent groupings of connectives according to the protoform(s) they are attracted to. The mutual inclusion of sets corresponds to ensembles of connectives sharing attraction with the same discursive operations. Thus, the most central connectives, included in the greatest number of sets, are those specifically associated to the greatest number of operations, and vice-versa.

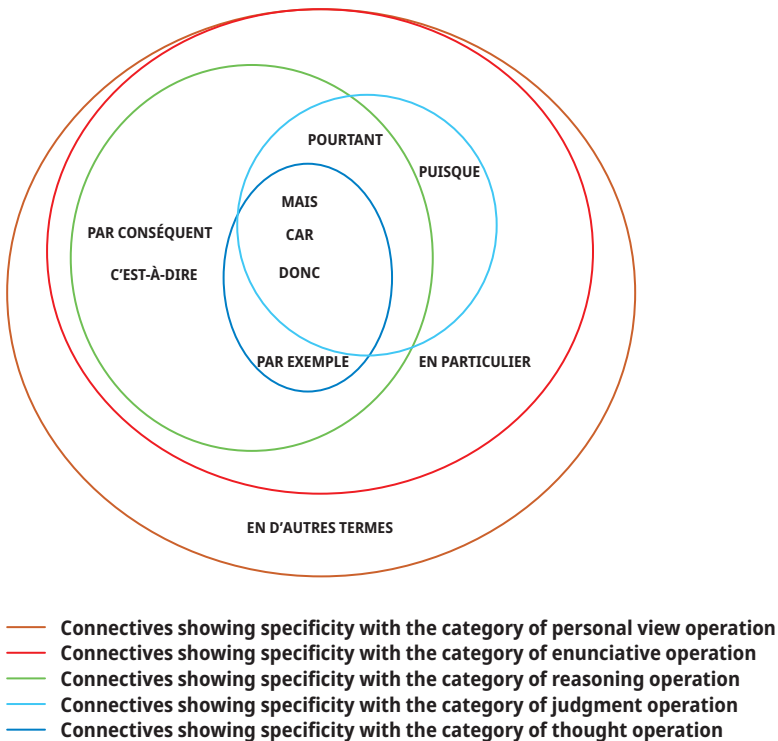


Figure 2: Representation in sets of different groupings of connectives (*Le Monde*)

The following groupings emerge:

- i. *mais*, *donc*, and *car* are attracted to all the categories considered, thus showing the strongest links with subjectivity.
- ii. By contrast, *en d'autres termes* and *en particulier* attract the fewest categories: respectively, personal view and personal view + enunciative operation.
- iii. The categories of personal view and enunciative operation are the ones to which our connectives are most sensitive: all connectives are attracted to items pertain-

- ing to the category of personal view, and all but *en d'autres termes* are attracted to items pertaining to the category of enunciative operation.
- iv. No grouping matches the traditional categories of connectives, based on logical/semantic/functional links: we can notice that the two connectives constituting each pair representing a functional class – *mais* and *pourtant*, *puisque* and *car*, *donc* and *par conséquent*, *par exemple* and *en particulier*, *c'est-à-dire* and *en d'autres termes* – are never in the same set.

The same research, using the same methodology, was carried out on our Italian press corpus, the newspaper *L'Adige* (see the results in Figure 3 and Figure 4).

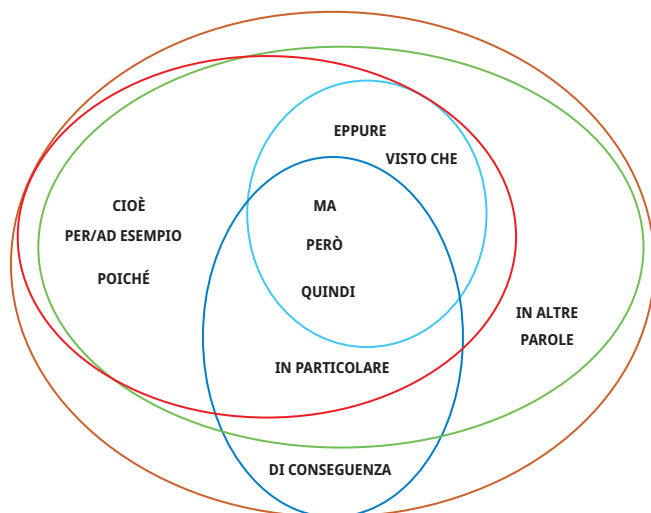


Figure 3: Associations of connectives with protoforms representing discursive operations (*L'Adige*¹⁰)

¹⁰ For visual clarity, we report here, and not directly in the graph, the LL values for *ma* and *però*:

	<i>ma</i>	<i>però</i>		<i>ma</i>	<i>però</i>	<i>ma</i>	<i>però</i>
Reasoning	5856,8	1894,9	Enunciation	4303,1	920,2	Thought	1042,1 350,1
Judgment	673,7	380,2	Personal View	8782,6	706,1		

All the LL values are stored in a repository, available at the following URL: <https://drive.switch.ch/index.php/s/JLYOfCwPrpePIg0>

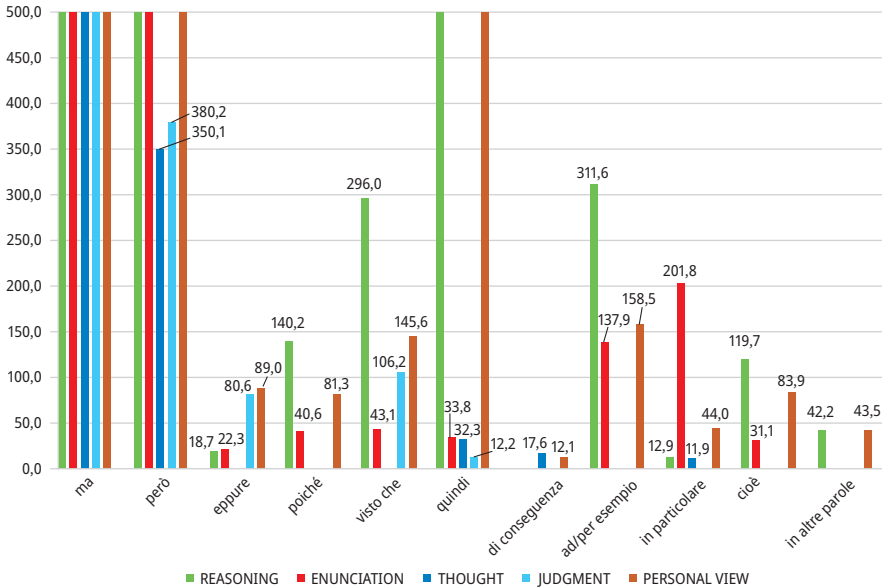


Figure 4: Representation in sets of different groupings of connectives (*L'Adige*)

The following facts are noteworthy:

- i. *ma*, *però*, and *quindi* are the connectives attracted to all the categories considered, thus showing the strongest links with subjectivity.
- ii. By contrast, *di conseguenza* and *in altre parole* attract the fewest categories: respectively, personal view and reasoning, and personal view and thought.
- iii. The categories to which the connectives are the most sensitive are personal view, to which all connectives are attracted (as it is the case for French) and reasoning, to which all connectives but *di conseguenza* are attracted (differently from French).
- iv. Similarly to French, no grouping matches the functional classifications of connectives: we can notice that the two connectives representing the same category are never in the same set. The only exception is constituted by *ma* and *però*, together in the same ensemble (which confirms their synonymy), both distinguishing themselves from *eppure*.

The above representations show the affinity of connectives or groups of connectives with respect to discursive operations, which can include, as mentioned in Section 2, different sub-categories of items. Our methodology allows us to appraise the sensitivity of connectives to sub-categories as well. For instance, we have seen the discursive operation of reasoning may be expressed by connectives, evidential forms, and also modal forms. This latter category of items can in turn be divided into clear-cut sub-categories, necessity and possibility, which can be expressed by modal verbs as well as adverbs. In Rossari et al. (2022), we noted that the lexical items representing verbs of possibility or

necessity can be differently attracted to *mais* and *donc*, *ma* and *quindi*. The following figure confirms this discrepancy considering a broader paradigm of necessity and possibility forms (verbs and adverbs) with the connectives which we took into account for this study.

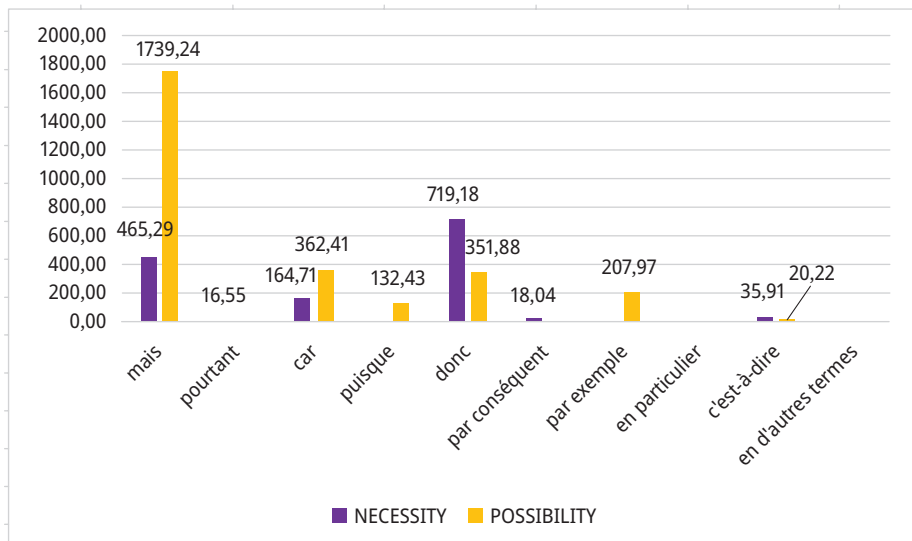


Figure 5: Association of connectives with the categories of necessity and possibility (*Le Monde*)

This representation allows to highlight the special relationship that *mais*, *donc*, and *car* have with both modal categories; the grouping of these three connectives seen above is, thus, not only due to their ability to attract all categories of discursive operations, but also to their ability to attract modal items particularly strongly. It also reveals affinities between connectives which are not visible when looking at the graph based on macro-categories. *Par exemple* shows very strong attraction to the modal category of possibility and, in this respect, is similar to *mais*, *donc*, and *car*; however, it also differs insofar as it is only attracted to the modality of possibility and not to that of necessity.

We have carried out the same search for Italian:

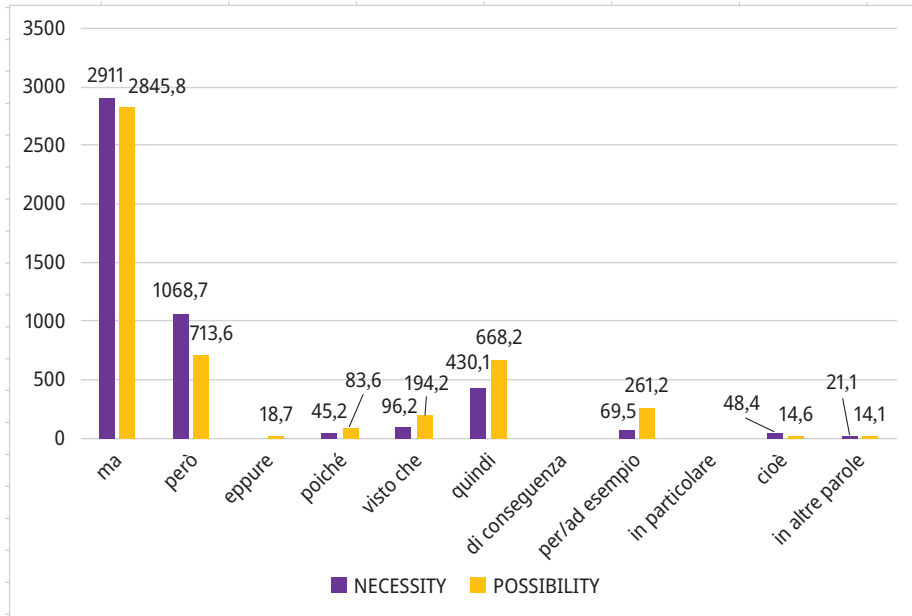


Figure 6: Association of connectives with the categories of necessity and possibility (*L'Adige*)

In this corpus, the forms showing a privileged relationship (specific co-occurrence) with both modal categories of necessity and possibility are *ma*, *però*, and *quindi*. They are followed by *per esempio*, which, differently from *par exemple*, shows an affinity with both categories, though with a higher score for possibility. Similarly to *en particulier*, *in particolare* does not show affinity with any of the two modalities considered. However, the Italian connective shows affinity with the discursive macro-category of reasoning, to which necessity and possibility pertain (see Figure 3). This means that *in particolare* is attracted to the other micro-category forming part of the discursive operation of reasoning: that of connectives. With this kind of intermediate analysis, we are able to notice that, although *in particolare* is involved in discursive operations manifesting inferential links, this involvement takes place through a proximity with connectives and not with modal forms.

4.2 Correspondence analysis

Resorting to correspondence analysis is essential to go deeper in the analysis of the behaviour of connectives, since, as already mentioned, this method allows us to grasp the lexical sensitivity of the selected connectives. We selected the most frequent forms of each discursive operation. Table 1 below lists the forms thus retained.

Table 1: Most frequent forms representing discursive operations in French and Italian

	<i>Le Monde</i>	<i>L'Adige</i>
Reasoning operation	<ul style="list-style-type: none"> • <i>devoir</i> 'must' • <i>pouvoir</i> 'can' • <i>si</i> 'if' • <i>parce que</i> 'because' 	<ul style="list-style-type: none"> • <i>dovere</i> 'must' • <i>potere</i> 'can' • <i>se</i> 'if' • <i>perché</i> 'why'/'because'/'in order to'
Enunciative operation	<ul style="list-style-type: none"> • <i>dire</i> 'to say' • <i>proposer</i> 'to propose' • <i>franchement</i> 'frankly' • <i>c'il est vrai</i> 'it is true' 	<ul style="list-style-type: none"> • <i>dire</i> 'to say' • <i>chiedere</i> 'to ask' • <i>parlare</i> 'to talk' • <i>è vero</i> 'it is true' • <i>certo</i> 'of course'
Thought operation	<ul style="list-style-type: none"> • <i>penser</i> 'to think' • <i>croire</i> 'to believe' • <i>imaginer</i> 'to imagine' • <i>doute</i> 'doubt' 	<ul style="list-style-type: none"> • <i>pensare</i> 'to think' • <i>credere</i> 'to believe' • <i>ritenere</i> 'to believe'/'to consider' • <i>dubbio</i> 'doubt'
Judgment operation	<ul style="list-style-type: none"> • <i>sembler</i> 'to seem' • <i>considérer</i> 'to consider' • <i>visiblement</i> 'obviously' 	<ul style="list-style-type: none"> • <i>sembrare</i> 'to seem' • <i>considerare</i> 'to consider' • <i>parere</i> 'to appear' • <i>giudicare</i> 'to judge'
Personal view operation	<ul style="list-style-type: none"> • <i>très</i> 'very' • <i>bien</i> 'good' • <i>peu</i> 'a little'/'few' • <i>bon</i> 'good' 	<ul style="list-style-type: none"> • <i>molto</i> 'much'/'very' • <i>grande</i> 'big'/'large' • <i>importante</i> 'important' • <i>diverso</i> 'different'

It is on these forms that we performed our CA calculation, whose results for French and Italian are shown in Figures 7 and 8 below. Before commenting on these results, it is necessary to present a brief explanation of how to read a CA graph.

A CA plot is obtained from a Chi-Square distance between rows and columns within a contingency table (in our contingency table, rows contain forms representing discursive operations and columns contain connectives).¹¹ The visual plot represents the dependence between the rows and the columns, which are displayed in the form of points.¹² The sum of the percentages resulting from the horizontal and vertical dimen-

¹¹ As for the LL values, the CA contingency tables and analyses are also stored in the repository available at the address mentioned in fn. 4.

¹² A more detailed explanation is given by Alberti (online), which we report here: "CA graphically represents the dependence between rows and columns of contingency tables. The visual display of data helps the interpretation and allows patterns to emerge. The technique reduces the number of dimensions needed to display the data points by decomposing the total inertia (i.e. the variability) of the table and defining the smallest number of dimensions capable of capturing the data variability. The graphical output of CA is a scatterplot where rows and/or columns are represented as points in a sequence of low-dimensional spaces. These spaces have the properties to retain a decreasing amount of the total inertia.

sions (77% for French, 74% for Italian) represent the amount of information rendered from contingency tables.

To read the plot, we appraise the distance between data points representing connectives, as well as the distance between data points representing single items expressing discursive operations. The proximity between such points indicates a similarity in their profile. The proximity between a row and a column point indicates that the combination row/column is more frequent than expected if the items were independent from each other. The more a form is off-centre, the less its profile is average with respect to the other forms of the same category.

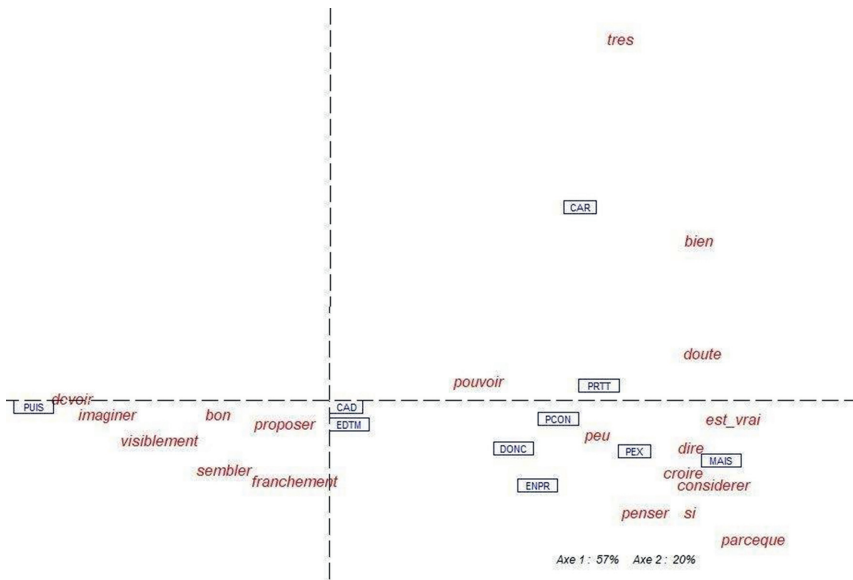


Figure 7: CA for French press corpus *Le Monde*. Abbreviations for the longer labels:

CAD = *c'est-à-dire*; EDTM = *en d'autres termes*; ENPR = *en particulier*; PCON = *par conséquent*; PEX = *par exemple*; PRTT = *pourtant*; PUIS = *puisque*

The following profiles emerge:

- *c'est-à-dire* and *en d'autres termes* both show no particular behaviour compared to what is expected in relation to all the forms considered here, which is represented by the fact that they appear in the centre of the graph. This grouping matches their functional categorisation as reformulatives.
- *puisque* stands out for being at the far-left end of the graph. Such a position, which particularises its profile with regard to that of the other connectives, may be due to

The first dimension will capture the highest amount, while the second will be associated with the second largest proportion, and so on".

its proximity with *devoir* and *imaginer*. This reveals that the semantics of the lexical form is also at play in determining the behavior of a given connective, since we can see that the profile of *imaginer* sets it apart from the profiles of other thought verbs, such as *penser* and *croire* (whose proximity in the plot reflects their semantic proximity). Such an observation also applies to connectives: the profile of *puisque* sets it apart from the profile of *car*, which reflects their differences in terms of assertion (*car*) vs presupposition (*puisque*) with respect to the content they introduce.

- *mais* is set apart from *donc* by its proximity with *il/c'est vrai que* and with *dire*, both items expressing dialogicity, which gives new information in relation to what we observed in the LL graphs, where both connectives show being associated to the enunciation category, with a higher score for *mais* than for *donc*.
- *pourtant* is in between *donc* and *mais*, which suggests that it is not the 'opposition' feature that leads to proximity between connectives.

The groupings brought out by the CA do not preclude the emergence of points of convergence between the two statistical methods: none of the connectives stands out from the others showing signs of proximity vis-à-vis forms pertaining to the category of personal view (which can be seen by the fact that the latter are scattered throughout the graph); this confirms what already observed with the LL scores (all connectives being specifically associated to this category). We also find a multiplicity of forms surrounding *mais*, which is in accordance with the highest LL scores that the connective has with all discursive operations.

Figure 8 shows the CA for our Italian corpus:

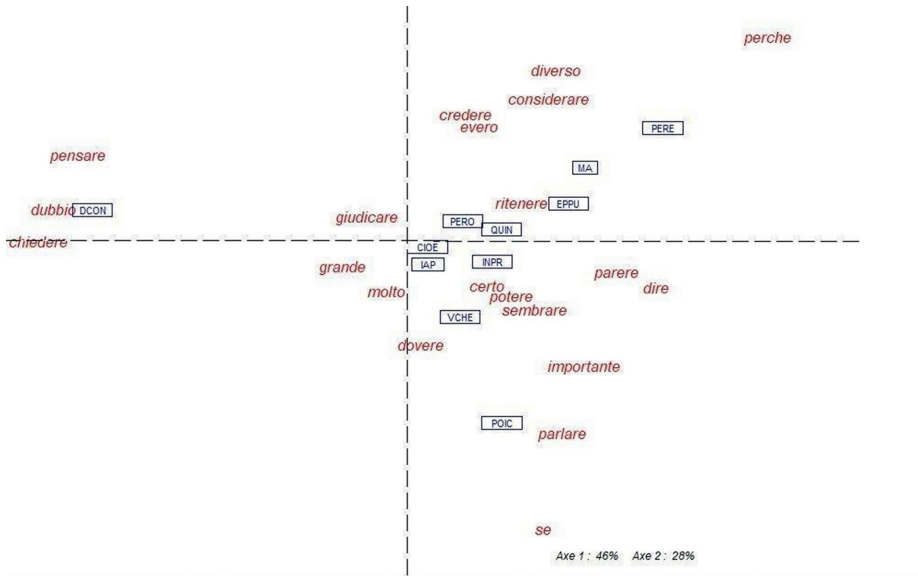


Figure 8: CA for Italian press corpus *L'Adige*. Abbreviations for the longer labels: DCON = *di conseguenza*; EPPU = *eppure*; IAP = *in altre parole*; INPR = *in particolare*; POICH = *poiché*; QUIN = *quindi*; VCHE = *visto che*

The following groupings emerging from the Italian CA match the functional categories of connectives:

- *ciò* and *in altre parole* display the same behaviour as their French counterparts *c'est-à-dire* and *en d'autres termes*: both are situated at the crossing of the two axes, which corresponds to average.
- Differently from *mais* and *pourtant*, *ma*, *però*, and *eppure* are situated in the same quadrant and all show proximity with *ritenere*. This may suggest that *eppure* is semantically less further away from *ma* (and *però*) than *pourtant* is from *mais*.
- Differently from *car* and *puisque*, *visto che* and *poiché* are in the same quarter, which shows a certain proximity in their behaviour; however, *visto che* distinguishes itself in that it attracts both modals and the dialogical expression *certo*.

On the other hand, as it is the case for French, we can also notice groupings not matching at all functional categories of connectives:

- *quindi* and *però* appear very close to each other, both accompanying an item revealing a thought operation (*ritenere* 'to believe'/to consider').
- *di conseguenza* differs from *quindi* as well as from all the other connectives in particular with respect to forms pertaining to the same category – thought operation – but showing nuances relative to uncertainty: *dubbio* and *pensare*, close to *di conseguenza*, vs *ritenere* – not conveying uncertainty –, close to *quindi*. We can thus high-

light differences concerning more specifically the sensitivity of the connectives to the lexical meaning of the forms.

Here too, we can find points of convergence between the two statistical methods adopted: *ma*, *però*, and *quindi*, the three connectives attracting all macro-categories, are located in the same quarter – with the particular proximity between *quindi* and *però* mentioned above – in the CA graph.

5 Theoretical interpretation of data

Two main theoretical observations can be drawn from these statistical analyses.

Firstly, both statistical methods add new information on connectives and their behaviour, sometimes different with respect to the traditional categories of connectives. The most evident clue of the fact that the privileged environment of a connective is independent from the relation it conveys is given by the pairs *puisque* / *car* and *di conseguenza* / *quindi*. Both are positioned in different quarters in the CA graph and show very different tendencies in the LL graphs: they pertain to different sets in the representation based on the macro-categories of discursive operations, and they do not share the same type of modality in the graph based on the micro-categories of necessity and possibility. The connective *puisque* does not attract the modality of necessity, whereas *car* attracts both modalities; the same contrast emerges for the pair constituted by *quindi* and *di conseguenza*: the former attracts both modalities, whereas the latter attracts neither of them. These results are in line with the fact that, although expressing the same functional relation (respectively, causality; and consequence), as it is well known these forms have their own peculiarities, like the different enunciative properties distinguishing *car* and *puisque*, which makes them not interchangeable:

- (7) Pourquoi veux-tu que je fasse des études supérieures **puisque** [*car] je vais hériter?

‘Why do you want me to go to university since I’m going to inherit?’ (*Le Monde*)

However, Italian connectives do show some matching between relational properties and their attraction properties: *ma* and *però* appear to be very similar in all graphs, and so do *poiché* and *visto che*; this similarity is in sharp contrast with the very different behaviour of *car* and *puisque*. This, too, shows that, in addition to the relational, linking properties of connectives, other properties have to be taken into account. For instance, in Italian, *poiché* shares with *visto che* the possibility to introduce given content. Such a proximity can be put in relation with analogous enunciative properties for these two pairs of markers: *poiché* and *visto che* can convey a low degree of *prise en charge* (‘endorsement’) of the content they introduce – in the same way as *puisque*; *ma* and

però, similarly to *mais*, both indicate that the content they introduce is more strongly endorsed ('pris en charge') by the speaker than the preceding content.

The second theoretical point that emerges is the various degree of sensitivity of connectives to items conveying the subjectivity of the speaker. Some connectives, and this in both languages, strongly attract items of all macro-categories of discursive operations. *Mais*, *donc*, and *car* on the one hand, and *ma*, *però*, and *quindi* on the other hand are associated to all groups of forms, lying in the intersection of all sets in Figures 2 and 4 respectively. Conversely, other connectives do not seem sensitive to such items: *di conseguenza*, *en d'autres termes* show attraction only with one of our macro-categories and no attraction to the modality of necessity or possibility. Such a different behaviour among connectives can be put in relation to an intrinsic property of subjectivity: the more the connective conveys a trait of subjectivity, the more it attracts forms conveying subjective meaning. This correlation is based on the observation that similar forms tend to attract each other, as Hütsch (2020) notices concerning *devoir* and *nécessairement*. By way of example, concerning connectives, we can notice that *mais* and *néanmoins* 'nevertheless' as well as *ma* and *bensi* 'rather', show attraction to each other. As for modality, in Italian as well *dovere* attracts *necessariamente*. Table 2 below shows the specificity index for each pair of items, which we searched in our corpora within a span of five items. Even if Italian shows a stronger attraction between *dovere* and *necessariamente*, all the specificity indexes are clearly above the threshold.

Table 2: Specificity index (TXM) for the associations of the pairs *ma/bensi*, *mais/néanmoins*, *dovere/necessariamente*, *devoir/nécessairement*

Pairs of connectives in French and Italian corpora	Specificity index
<i>mais</i> [...] <i>néanmoins</i>	10
<i>ma</i> [...] <i>bensi</i>	5
<i>devoir</i> [...] <i>nécessairement</i>	5
<i>dovere</i> [...] <i>necessariamente</i>	92

To sum up, the analysis of the linguistic environment of connectives shows that this environment often does not match with their relational properties, but it makes appear another property, linked to their capacity to convey subjectivity in the very relation they express.

6 Concluding remarks

Taking into account different relational categories of connectives, we have seen that observing their favourite linguistic environment in terms of forms representing discursive operations allowing to display the speaker's subjectivity can give an additional

insight to the analysis of their functioning, complementing the traditional analysis based on their relational properties. The capacity of connectives to attract items pertaining to different discursive operations shows that the trait of subjectivity is clearly relevant to appraise the way they are used in corpora, since such a trait can cross different relational categories. To go further in this respect, the number of connectives considered must be increased; furthermore, their behaviour needs deeper investigation notably in relation to the nature of the interrogated corpora, in order to assess more precisely what is stable and what is corpus-dependent in the way they behave.

7 References

7.1 Research literature

- Andersen, Hanne Leth (1996), *Verbes parenthétiques comme marqueurs discursifs*, in: Claude Muller (ed.), *Dépendance et intégration syntaxique: subordination, coordination, connexion*, Tübingen, Niemeyer, 307–315.
- Andersen, Hanne Leth (2007), *Marqueurs discursifs propositionnels*, *Langue française* 154, 13–28.
- Anscombre, Jean-Claude/Ducrot, Oswald (1983), *L'argumentation dans la langue*, Bruxelles, Mardaga.
- Asher, Nicholas/Lascarides, Alex (2003), *Logics of conversation*, Cambridge, Cambridge University Press.
- Bazzanella, Carla (1995), *I segnali discorsivi*, in: Lorenzo Renzi/Giampaolo Salvi/Anna Cardinaletti (edd.), *Grande grammatica italiana di consultazione*, vol. 3, Bologna, il Mulino, 225–257.
- Blakemore, Diane (2002), *Relevance and linguistic meaning. The semantics and pragmatics of discourse markers*, Cambridge, Cambridge University Press.
- Blumenthal, Peter (2008), *Histoires de mots. Affinités (s)électives*, in: Jacques Durand/Benoît Habert/Bernard Laks (edd.), *Actes du CMLF 2008, 1er Congrès Mondial de Linguistique Française*, Paris, EDP Sciences, 31–46.
- Borillo, Andrée (2019), *L'usage des expressions "à côté de ça", "en face de ça", "en dehors de ça", "(mis) à part ça"... dans une fonction discursive d'orientation argumentative*, in: Peter Blumenthal/Denis Vigier (edd.), *Études diachroniques du français et perspectives sociétales*, Bern, Lang, 47–64.
- Bras, Myriam, et al. (2021), *Vers un corpus de textes d'élèves annoté en relations de discours*, *Langue française* 211, 115–129.
- Brunet, Etienne (2012), *Nouveau traitement des cooccurrences dans Hyperbase*, *Corpus* 11, 219–246, <http://journals.openedition.org/corpus/2275> (access 1 August 2023).
- Bruxelles, Sylvie/Traverso, Véronique (2001), *Ben: apport de la description d'un "petit mot" du discours à l'étude des polylogues*, *Marges linguistiques* 2, 38–55.
- Buridant, Claude (2006), *L'interjection. Jeux et enjeux*, *Langages* 161, 3–9.
- Charolles, Michel/Lamiroy, Béatrice (2012), *Les adverbiaux. De la phrase au discours*, in: Jan Radimsky (ed.), *Actes du 31e Colloque International sur le Lexique et la Grammaire*, Université de Bohême du Sud, České Budějovice, 93–100.
- De Cesare, Anna-Maria (2019), *Le parti invariabili del discorso*, Roma, Carocci.
- Dostie, Gaétane (1998), *Deux marqueurs discursifs issus de verbes de perception. De "écouter" / "regarder" à "écoute" / "regarde"*, *Cahiers de lexicologie* 73, 125–146.
- Dostie, Gaétane (2004), *"Mettons". Mise en parallèle avec "admettons" et "disons"*, in: Gaétane Dostie, *Pragmaticalisation et marqueurs discursifs. Analyse sémantique et traitement lexicographique*, Bruxelles, De Boeck/Duculot, 157–179.
- Ducrot, Oswald (1980), *Les mots du discours*, Paris, Minuit.
- Ducrot, Oswald (1999), *Sémantique linguistique et analyse de textes*, *Littérature* 115, 104–125.

- Ferrari, Angela (2014), *Linguistica del testo. Principi, fenomeni, strutture*, Roma, Carocci.
- García Izquierdo, Isabel (1998), *Mecanismos de cohesión textual. Los conectores ilativos en español*, Castelló de la Plana, Publicacions de la Universitat Jaume I.
- Guimier, Claude (2009), *De la comparaison à l'énonciation. Le cas de "aussi bien"*, *Langue française* 161, 97–114.
- Heiden, Serge/Magué, Jean-Philippe/Pincemin, Bénédicte (2010), *TXM. Une plateforme logicielle open-source pour la textométrie – conception et développement*, in: Sergio Bolasco/Isabella Chiari/Luca Giuliano (edd.), *JADT 2010. Statistical analysis of textual data. Proceedings of the 10th International Conference, 9–11 June 2010 – Sapienza University of Rome*, Milano, LED, 1021–1032, http://halshs.archives-ouvertes.fr/docs/00/54/97/79/PDF/Heiden_al_jadt2010.pdf (access 1 August 2023).
- Hütsch, Annalena (2020), *L'usage des verbes modaux en français et en allemand. Étude contrastive de la combinatoire adverbiale sous l'éclairage quantitatif*, thèse de doctorat, Université de Neuchâtel.
- Lafon, Pierre (1980), *Sur la variabilité de la fréquence des formes dans un corpus*, *Mots* 1, 127–165.
- Le Querler, Nicole (2000), *Le connecteur "où"*, *Syntaxe et sémantique* 1, 113–132.
- Lopes, Ana Cristina Macário/Sousa, Sara (2014), *The discourse connectives "ao invés" and "pelo contrário" in contemporary European Portuguese*, *Journal of Portuguese linguistics* 13, 3–27.
- Mandelli, Magda (2008), *"In effetti" nel testo*, in: Emanuela Cresti (ed.), *Prospettive nello studio del lessico italiano. Atti del IX Congresso SILFI, Firenze, 14–17 giugno 2006*, Firenze, Firenze University Press, 439–444.
- Mann, William C./Thompson, Sandra A. (1987), *Rhetorical structure theory. Description and construction of text structures*, in: Gerard Kempen (ed.), *Natural language generation. New results in artificial intelligence, psychology and linguistics*, Dordrecht, Nijhoff, 85–95.
- Manzotti, Emilio (1998), *L'esempio. Natura, definizioni, problemi*, *Cuadernos de filología italiana* 5, 99–126.
- Nølke, Henning (1993), *Il y a connecteurs et connecteurs. La connexion analytique et synthétique en français moderne*, in: Henning Nølke (ed.), *Le regard du locuteur. Pour une linguistique des traces énonciatives*, Paris, Éditions Kimé, 133–142.
- Pons Bordería, Salvador (1998), *Conexión y conectores. Estudio de su relación en el registro informal de la lengua*, Valencia, Departamento de Filología Española, Facultat de Filologia, Universitat de València.
- Pop, Liana (2006), *Peut-on parler de style communicatif interjectif? Le cas du roumain*, *Langages* 161, 24–36.
- Portolés, José (1998), *Dos pares de marcadores del discurso: "en cambio" y "por el contrario", "en cualquier caso" y "en todo caso"*, in: María Antonia Martín Zorraquino/Estrella Montolío Durán (edd.), *Marcadores del discurso. Teoría y análisis*, Madrid, Arco Libros, 243–264.
- Rocci, Andrea, et al. (2021), *The significance of the adversative connectives "aber", "mais", "ma" ("but") as indicators in young children's argumentation*, *Journal of argumentation in context* 9, 69–94.
- Rossari, Corinne (1997), *Les opérations de reformulation. Analyse du processus et des marques dans une perspective contrastive français – italien*, Bern, Lang.
- Rossari, Corinne (2005), *Le leurre des significations lexicales. L'exemple de "la preuve"*, *Revue romane* 40, 65–78.
- Rossari, Corinne (2011), *Grammaticalization and persistence phenomena in two hybrid discourse markers: "la prévue" and "regarde"*, *Acta linguistica hafniensia: international journal of linguistics*, 38, 161–179.
- Rossari, Corinne (2021), *Les connecteurs*, in: Anne Abeillé/Danièle Godard (edd.), *La grande grammaire du français*, Arles/Paris, Actes Sud/Imprimerie nationale, 1983–1993.
- Rossari, Corinne/Cojocariu, Corina (2008), *Constructions of the type "la cause" / "la raison" / "la preuve" + utterance. Grammaticalization, pragmaticalization, or something else?*, *Journal of pragmatics* 40, 1435–1454.
- Rossari, Corinne, et al. (2022), *Structures argumentatives et analyses statistiques. Étude des combinaisons entre connecteurs et indications énonciatives*, *Humanidades & inovação* 9:4, 398–416.
- Roulet, Eddy (1997), *L'organisation polyphonique et l'organisation inférentielle d'un dialogue romanesque*, *Cahiers de linguistique française* 19, 149–179.
- Roulet, Eddy, et al. (2001), *Un modèle et un instrument d'analyse de l'organisation du discours*, Bern, Lang.
- Salvi, Giampaolo (2013), *Le parti del discorso*, Roma, Carocci.
- Steuckardt, Agnès (2018), *Les marqueurs de reformulation formés sur dire. Exploration outillée*, *Langages* 212, 17–34.

Wilson, Deirdre (2015), *The conceptual-procedural distinction. Past, present and future*, in: Victoria Escandell-Vidal/Manuel Leonetti/Aoife Ahern (edd.), *Procedural meaning. Problems and perspectives*, Leiden, Brill, 3–31.

Wandel, Dennis (2020), *Emplois propositionnels et énonciatifs des cadratifs à nom abstrait “en fait”, “en réalité”, “en vérité”, “en pratique”, et de leurs équivalents allemands*, thèse de doctorat, Université de Neuchâtel.

7.2 Web pages

Alberti, Gianmarco (online), *Correspondence analysis in archaeology*, <http://cainarchaeology.weebly.com/> (access 1 August 2023).

Europarl = <https://www.statmt.org/europarl> (access 1 August 2023).

Repository for the calculations and results presented in this article: <https://drive.switch.ch/index.php/s/JLYOfCwPrpePIg0> (access 1 August 2023).

Reverso = <https://context.reverso.net> (access 1 August 2023).