

Discursive Functions of French Modal Forms: What Can Correspondence Analysis Tell Us About Genre and Diachronic Variation?



Corinne Rossari, Ljiljana Dolamic, Annalena Hütsch, Claudia Ricci, and Dennis Wandel

Abstract Our aim is to describe discursive functions of a set of French modal forms by establishing their combinatory profiles based on their co-occurrence with different connectors. We then compare these profiles using correspondence analysis in order to find evidence of genre and diachronic variation. The use of these forms is explored in contexts of informative discourse within two distinctly different genres—contemporary written press and encyclopedic discourse—as well as within two diachronic spans.

1 Introduction

Our aim is to analyze the genre and diachronic variation of discursive functions of French modal forms (MFs) in different discourse genres: written press and encyclopedia. Our methodology uses quantitative analysis based on statistical tools to detect similarities and differences in the use of terms in different corpora, as previously done for prepositions in Blumenthal [3], Charolles et al. [5] and Vigier [15]. By discursive function, we mean the rhetorical aim of the utterance in which a form occurs: counter-argument, argument, or conclusion (cf. [14]). We name the above-defined sequences *discourse movements*. We would like to see if the values that forms of a same paradigm can endorse are sensitive to diachronic or genre variation. On the other hand, we aim to verify whether the core meaning of a form entails a different behavior in relation to discursive sequences, and if so, if this behavior is in turn sensitive to diachronic and genre variation. For this purpose, we use two statistical tools: an association measure, log-likelihood [8], and correspondence analysis (CA) (see Désagulier [6],

This paper is an extended version of the paper *Discursive Functions of Epistemic Adverbs: What can Correspondence Analysis tell us about Genre and Diachronic Variation?* By the same authors, published in *Actes JADT 2018*, Roma: UniversItalia, pp. 668–675.

C. Rossari (✉) · L. Dolamic · A. Hütsch · C. Ricci · D. Wandel
University of Neuchâtel, Neuchâtel, Switzerland
e-mail: corinne.rossari@unine.ch

Greenacre [9]). Firstly, we take into account a paradigm of MFs conveying different nuances linked to the core meaning of possibility, which consists of the following epistemic adverbs¹: *certainement*, *certes* ('certainly'), *peut-être* ('maybe'), *probablement* ('probably'), *sans doute* ('without a doubt') and *sûrement* ('surely'). Secondly, we compare forms conveying two different core meanings: possibility (P) and necessity (N). Our paradigm consists of the following items: *devoir*, *falloir* ('must'/'have to,' 'be necessary') (N meaning) and *pouvoir* ('can'/'may'), *peut-être* ('maybe')² (P meaning). The functions of all MFs are explored in contexts of informative discourse within two distinctly different genres: contemporary written press and encyclopedic discourse. The former is represented by three daily newspapers: *Le Monde* (2008, 20,410,766 tokens), *Le Figaro* (2008, 10,795 373 tokens) and *Sud-Ouest* (2002, 29,763 988 tokens). For the latter, we consider two diachronic spans: the eighteenth century, represented by Diderot & D'Alembert's *Encyclopédie* (DDA, 29 940 181 tokens) and the twenty-first century, represented by the 2005 edition of *Encyclopédie Universalis* (UNI, 49 859 864 tokens) and by a random sample of the 2015 version of *Wikipédia* (WIKI, 50 396 345 tokens).³

We first carry out an analysis based on the combinatory profile of each MF constituting our first paradigm of forms (Sect. 2) within our corpus of contemporary written press, and, after having pinpointed what such an analysis can and cannot show, we use a more holistic approach based on correspondence analysis (Sect. 3). The same approach is chosen for our second paradigm. Our purpose is to verify if this type of approach allows to extend the analysis of modals, bringing out differences in behavior between the modals conveying necessity (N-MFs) and those conveying possibility (P-MFs).

2 Analysis of Combinatory Profiles of Epistemic Adverbs (E-ADVs)

In order to identify the discursive functions of the E-ADVs considered here, we searched connectors (C) specifically co-occurring with each of these E-ADVs within a 20-token span. We have chosen a 20-token span rather than a sentence span because a connector's combinatory profile can go beyond sentence boundaries. We define

¹Our selection of forms is based on Roulet's [13] paradigm of epistemic assertive adverbs. The English translations of all modal forms and connectors considered in this paper do not necessarily represent all the possible renderings of the French forms. They are literal translations, which we provide for the sake of clarity.

²*Peut-être* belongs to both paradigms because, for reasons of representativeness, we decided to include two forms for each core meaning: two N forms (*devoir*, *falloir*) and two P forms (*pouvoir*, *peut-être*).

³All the corpora used were supplied by the platform *BTLC* (*Base Textuelle Lexico-statistique de Cologne*), conceived by Diwersy [7], and were constituted within the French-German projects *Presto* (<http://presto.ens-lyon.fr>) and *Emolex* (<http://emolex.u-grenoble3.fr>).

connectors as linguistic forms linking segments of discourse. Such a functional category is not part of the tagset of the platform used. We therefore made our query by searching for three different categories which are tagged: adverbs, subordinating conjunctions, and coordinating conjunctions. We then manually filtered the resulting forms by keeping those which proved to function as connectors.

For all our subcorpora, each E-ADV is thus specifically assigned a series of connectors within constructions of the type ‘E-ADV...C1/C2/Cn’ and ‘C1/C2/Cn...E-ADV,’ which, taken as a whole, represent their *discursive combinatory profile*.⁴ We call each sequence within a combinatory profile a *discourse movement* as we consider it to have specific, rhetorically motivated discursive aims. These aims (mentioned in Sect. 1) are signaled by the connectors co-occurring specifically with an E-ADV: *néanmoins* (‘however’) and *mais* (‘but’) present the content of the utterance preceding them as a counter-argument to that of the utterance they introduce, *donc* (‘therefore’) and *finaleme*nt (‘finally’/‘after all’) signal that the utterance they introduce is a conclusion, and *car* and *parce que* (‘because’) signal that the utterance they introduce is an argument in favor of the utterance preceding them.

The tables below show the discursive combinatory profiles in three subcorpora of contemporary press (*Le Monde 2008*; *Le Figaro 2008*; *Sud-Ouest 2002*). The significance of each co-occurrence of a connector with an E-ADV is calculated using the significance measure log-likelihood (LL) (Table 1).

The data lead to the following observations: (i) Although the E-ADVs belong to the same semantic class, each of them has its own specific combinatory profile. (ii) Certain E-ADVs share comparable combinatory profiles: *sans doute* and *peut-être* share an almost identical set of specific connectors, whereas more frequently E-ADVs essentially only share one or more specific connectors (for instance, the connector *mais* for *certainement*, *sûrement*, *peut-être*, and *sans doute*). (iii) Certain E-ADVs stand out for their unique combinatory features: *certes* is almost exclusively associated with *mais*, but only with *mais_R*, and with a notably higher log-likelihood score than the other E-ADVs. *Probablement* is also associated with only a few connectors, but with a low log-likelihood score, close to the threshold of 10.83. (iv) There is homogeneity in the significant associations for each E-ADV in the three subcorpora of contemporary press. However, preceding studies—Rossari et al. [11] and Rossari and Salsmann [12]—show that an E-ADV’s combinatory profile varies depending on different genres and diachronic periods. This finding contrasts with what is observed here for the press genre: in the encyclopedias DDA and UNI, the association *peut-être...mais* is less significant than the association *mais...peut-être*; in addition, in DDA, no significant association *certes...mais* is observed, while the association *sans doute...mais* in the same corpus proves to be highly significant.

The analysis of combinatory profiles (based on log-likelihood scores; cf. [4]) allows for one-to-one comparison of different sequences of the type E-ADV...C and C...E-ADV. Thus, the associations of each E-ADV with each connector can easily be compared across subcorpora representing different newspapers, but also across different genres and diachronic periods. It is also possible to compare the associations

⁴We adapt the term *combinatory profile* used by Blumenthal et al. [4] and Blumenthal [1, 2].

Tables 1 Log-likelihood scores (threshold: 10.83; all scores equal or above are marked in bold)

Le Monde (2008)	Certainment (385)		Certes (1943)		Peut-être (2900)		Probablement (723)		Sans doute (2482)		Sûrement (307)	
	L	R	L	R	L	R	L	R	L	R	L	R
Car (6 706)	0.08 (3)	0.08 (3)	-0.27 (11)	0 (13)	49.30 (57)	-0.37 (12)	18.99 (17)	0.92 (7)	16.09 (35)	0.02 (17)	1.51 (4)	-0.64 (1)
Donc (8 276)	2.09 (6)	2.09 (6)	-2.47 (10)	2.9 (23)	-1.45 (18)	24.09 (51)	-0.68 (4)	1.43 (9)	0.03 (21)	4.18 (30)	0.10 (3)	0.10 (3)
Finalemnt (1 559)	-1.18 (0)	-0.01 (1)	-1.77 (1)	1.14 (5)	3.60 (9)	15.43 (15)	-0.01 (1)	0.58 (2)	0.01 (4)	6.96 (10)	-0.94 (0)	6.08 (3)
Mais (51 544)	29.83 (47)	27.94 (46)	3248 (979)	-22.55 (52)	371.45 (423)	107.28 (284)	10.30 (57)	10.30 (57)	205.88 (310)	32.85 (193)	30.90 (41)	65.68 (55)
Néanmoins (968)	-0.73 (0)	-0.73 (0)	5.84 (6)	14.22 (9)	0.02 (3)	-0.23 (2)	0.12 (1)	-1.38 (0)	-0.06 (2)	-0.06 (2)	-0.58 (0)	-0.58 (0)
Parce que (2 514)	0.88 (2)	2.81 (3)	1.78 (8)	-4.47 (1)	62.15 (37)	9.75 (17)	2.03 (4)	-3.58 (0)	86.58 (41)	0.12 (7)	3.78 (3)	0.07 (1)
Le Figaro (2008)	Certainment (268)		Certes (1084)		Peut-être (1851)		Probablement (441)		Sans doute (1240)		Sûrement (211)	
	L	R	L	R	L	R	L	R	L	R	L	R
Car (3 922)	-0.57 (1)	-0.57 (1)	3.89 (14)	-2.35 (4)	28.08 (37)	-0.48 (11)	14.27 (12)	1.95 (6)	8.45 (19)	4.43 (16)	7.53 (6)	-3.08 (0)
Donc (4 763)	-1.02 (1)	-1.02 (1)	-0.03 (9)	1.81 (14)	-20.39 (2)	3.16 (24)	-0.22 (3)	6.74 (10)	-0.37 (9)	0.37 (13)	-3.74 (0)	-0.48 (1)
Finalemnt (1 150)	-1.14 (0)	-1.14 (0)	-0.04 (2)	-0.95 (1)	-3.15 (1)	6.52 (10)	0 (1)	0 (1)	1.67 (5)	-1.34 (1)	-0.90 (0)	-0.90 (0)

(continued)

Tables 1 (continued)

	Certainement (268)		Certes (1084)		Peut-être (1851)		Probablement (441)		Sans doute (1240)		Sûrement (211)	
	L	R	L	R	L	R	L	R	L	R	L	R
Le Figaro (2008)												
Mais (28 552)	36.23 (41)	14.25 (30)	1757.55 (545)	-1.39 (49)	93.30 (204)	245.20 (281)	0.56 (27)	2.38 (31)	86.88 (151)	34.59 (117)	17.23 (27)	87.11 (52)
Néanmoins (580)	1.07 (1)	-0.58 (0)	10.02 (6)	0.49 (2)	0.44 (3)	0.44 (3)	-0.95 (0)	-0.95 (0)	-2.67 (09)	-2.67 (0)	-0.45 (0)	-0.45 (0)
Parce que (1 435)	0.10 (1)	-1.43 (0)	-1.65 (1)	-0.30 (2)	31.90 (22)	-2.25 (2)	6.88 (5)	-0.03 (1)	4.79 (8)	0.14 (4)	0.28 (1)	2.22 (2)
Sud-Ouest (2002)												
	Certainement (1277)		Certes (2795)		Peut-être (4950)		Probablement (812)		Sans doute (3930)		Sûrement (684)	
	L	R	L	R	L	R	L	R	L	R	L	R
Car (12 434)	10.78 (23)	6.53 (20)	2.89 (32)	5.92 (36)	44.71 (91)	-0.29 (38)	1.34 (10)	-1.35 (4)	45.13 (78)	-0.26 (30)	6.87 (13)	-0.09 (5)
Donc (19 185)	-3.00 (10)	5.72 (27)	-6.45 (22)	0.10 (38)	-1.98 (53)	1.55 (74)	-1.32 (7)	9.77 (22)	-1.61 (42)	9.50 (74)	-0.41 (7)	0.15 (10)
Finalement (2 858)	-0.09 (2)	0.11 (3)	3.19 (10)	0.07 (6)	7.35 (19)	3.68 (16)	-0.23 (1)	-0.23 (1)	2.23 (12)	0.72 (10)	-0.08 (1)	0.31 (2)
Mais (77 108)	28.87 (113)	64.77 (139)	6962.42 (1778)	-5.72 (118)	520.54 (682)	211.15 (513)	10.35 (64)	9.48 (63)	209.38 (434)	123.59 (376)	7.09 (52)	162.80 (130)
Néanmoins (1 698)	-0.16 (1)	-0.16 (1)	9.25 (10)	-0.01 (3)	5.39 (12)	-0.54 (4)	0.93 (2)	-1.85 (0)	6.72 (11)	1.20 (7)	0.06 (1)	0.06 (1)
Parce que (5 981)	-4.77 (11)	-13.45 (6)	12.51 (15)	8.48 (13)	814.50 (135)	104.89 (33)	16.58 (13)	-0.57 (2)	233.04 (108)	0.00 (16)	-1.04 (8)	-0.48 (9)

'L' ('left') and 'R' ('right') indicate on which side of the pivot a collocate appears

of different E-ADV with one or a few connectors. However, this method has certain insufficiencies when it comes to simultaneously comparing all of these variables in a holistic view. This type of analysis of combinatory profiles never takes into account all variables at the same time (e.g., frequencies, log-likelihood scores, paradigm of E-ADV, and paradigm of connectors). Moreover, using a threshold (in our case 10.83) in order to decide whether an association is significant is useful for traditional collocation analysis. But our goal is to also represent the use of each E-ADV in its typical discourse movements in contrast to its non-typical discourse movements. It thus seems counterproductive that all sequences (E-ADV...C/C...E-ADV) which are not statistically significant for certain E-ADV as such are not taken into account when establishing their combinatory profiles, since these non-significant cases play an important role in characterizing the overall use of the E-ADV and connectors. In order to allow for a holistic approach, we propose to use correspondence analysis (CA) [9].

3 Correspondence Analysis (CA) of E-ADV

The correspondence analysis presented in this section was performed using the R software and the package ‘CA’ [10].

- (1) In DDA, representing the eighteenth century, *certes* has a use which stands out. *Certes* left and right of *mais* differs clearly from all other E-ADV as to its association with the connectors. *Certes* is not typically used with any other connector analyzed and, most importantly, its association is not stronger with *mais* on its right than it is with *mais* on its left. Conversely, in all other five corpora (encyclopedic and press corpora), which represent the twenty-first century, there is an important difference between the use of *certes* right and left of *mais*: while *certes_L* is strongly linked to *mais*, *certes_R* is not.
- (2) In all six corpora, *mais* appears to be in contrast to all other connectors when it comes to its associations with E-ADV. Its central position appears to be linked to its high frequency, indicating its important contribution to the horizontal axis, which is confirmed by the correspondence analysis indicators.
- (3) An association between *sans doute_L* and *parce que* can be observed in DDA and WIKI, whereas in UNI, the adverb and the connector appear to be in the inverted relation. This behavior indicates that variation has to be expected even among subcorpora within the encyclopedic genre, concerning at least two parameters.
- (4) The results of the correspondence analysis show that in all subcorpora of one particular genre, in most cases, the same E-ADV are strongly associated with the same connector or group of connectors (*donc* and *finaleme*; *car* and *parce que*; *mais*). This phenomenon is particularly pronounced in the subcorpora representing written press. The connector *mais* differs the most from the other connectors with regard to the strength of its associations. Although *mais* is associated with most E-ADV, its association appears to be strong with only a few of them

in all corpora (*certes_L* being the only constant), while most other connectors have a higher number of strong associations. This indicates that certain discourse movements (such as E-ADV...*car/parce que*) seem to be rather regular, whereas *certes...mais* proves to be a special association, although only in the twenty-first century corpora (Fig. 1).

4 Correspondence Analysis (CA) of Necessity and Possibility Modal Forms

Before presenting the plots for these cases, some preliminary remarks are necessary. Interpreting the association of connectors with verbs requires manual sorting for the connectors which are mobile, i.e., whose position within the sentence may vary. A left syntactic position can in many cases correspond to a right semantic position: for instance, in a clause such as *il faut donc agir* (*so we must act*), the right scope of the connector *donc* includes in fact the modal verb. In our analysis, such a syntactic position is treated like a right semantic position of the modal form in relation to the connector, and thus is counted among the occurrences of the modal form in the right position. For *donc*, all the occurrences have been sorted. The other mobile connectors, *finaleme*nt and *néanmoins*, which assume the same discursive function as *donc* and *mais* and would require the same sorting, have not been taken into account for this research.

We present symmetric plots for visual clarity, but our interpretations are based on the asymmetric plots, which maintain the distances between the forms (Fig. 2).

- (1) The results of the correspondence analysis show that in all corpora some modal forms are grouped in relation to their core meaning. *Falloir* and *devoir*, which have a core meaning of necessity, are systematically grouped with regard to their right and left position. This shows that they have a very similar behavior in relation to the connectors considered here. The fact that such a clustering is transversal to genres and diachronic spans indicates that the core meaning of the forms is preponderant for the selection of the discursive sequence by the speaker.
- (2) Connectors with the same function are grouped: *car* and *parce que* show a similar behavior in relation to the four modal forms taken into account. This confirms that the modal forms are sensitive to the function of the connectors, and consequently to the discursive sequence they indicate, independently of diachronic span or genre represented by the corpus.
- (3) We cannot see regularity in the behavior of modal forms conveying a core meaning of possibility. *Pouvoir* and *peut-être* in the left position are grouped in some corpora, such as DDA and UNI, but not in others. This is due to the fact that there is a strong attraction between *peut-être* and *parce que*, because the modal can modify the causal relation in itself, in examples such as: *Paul est parti tôt, peut-être parce qu'il ne connaissait personne* (Paul left early, maybe because he didn't know anyone). In such a case, *peut-être* modifies the causal relation

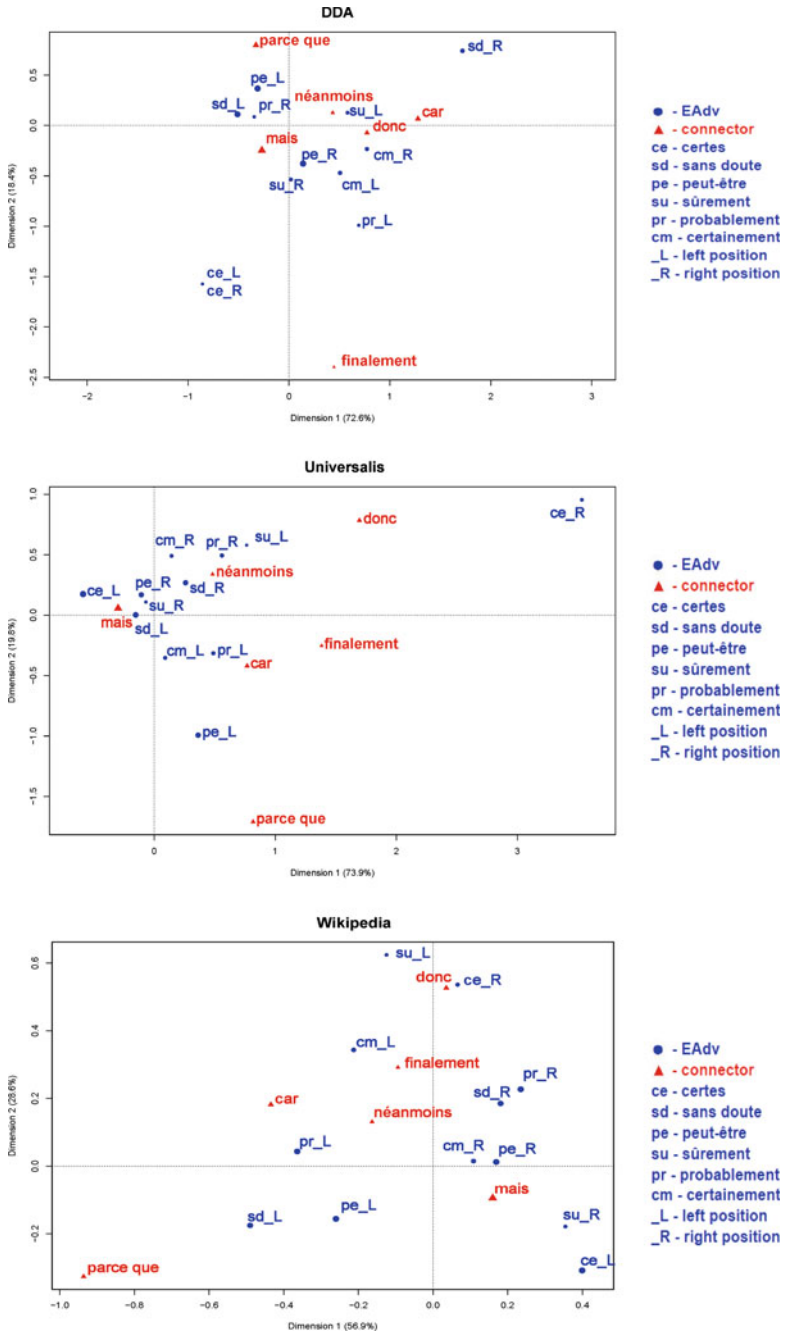


Fig. 1 Correspondence analysis scatter plots for the six corpora

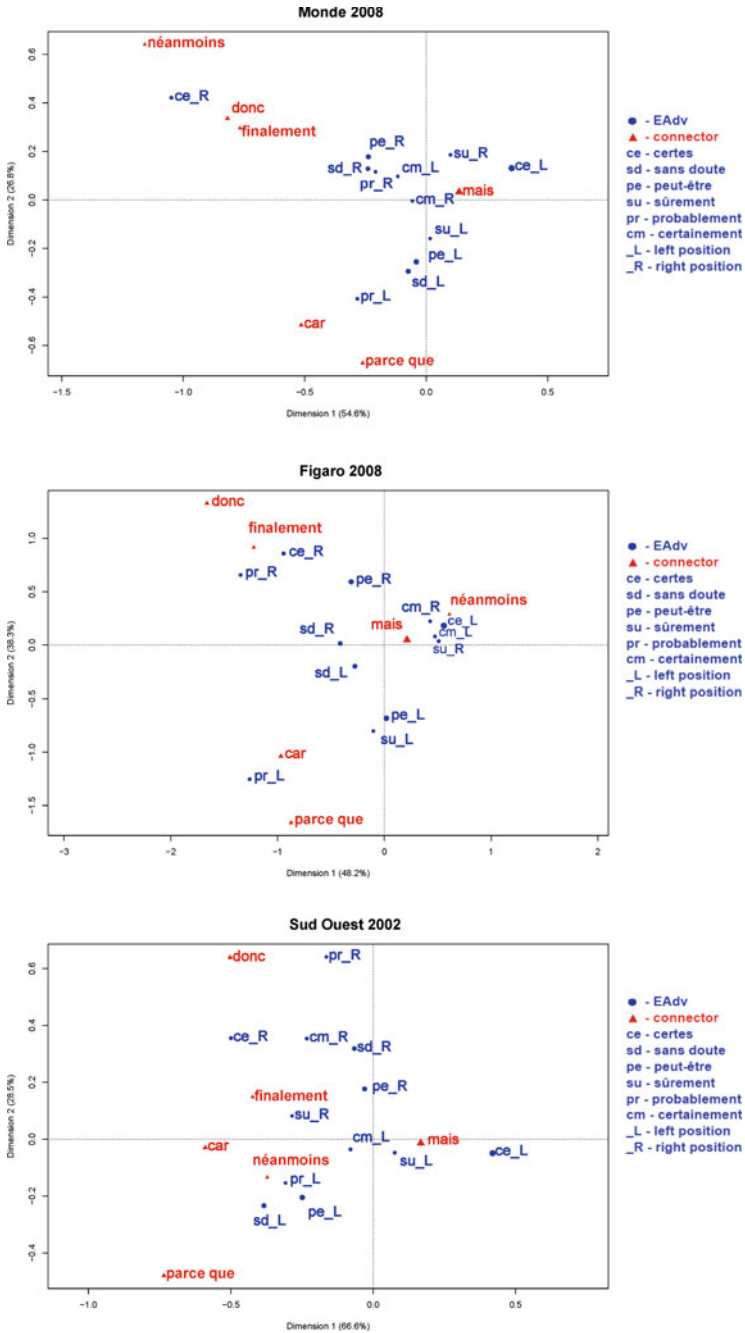


Fig. 1 (continued)

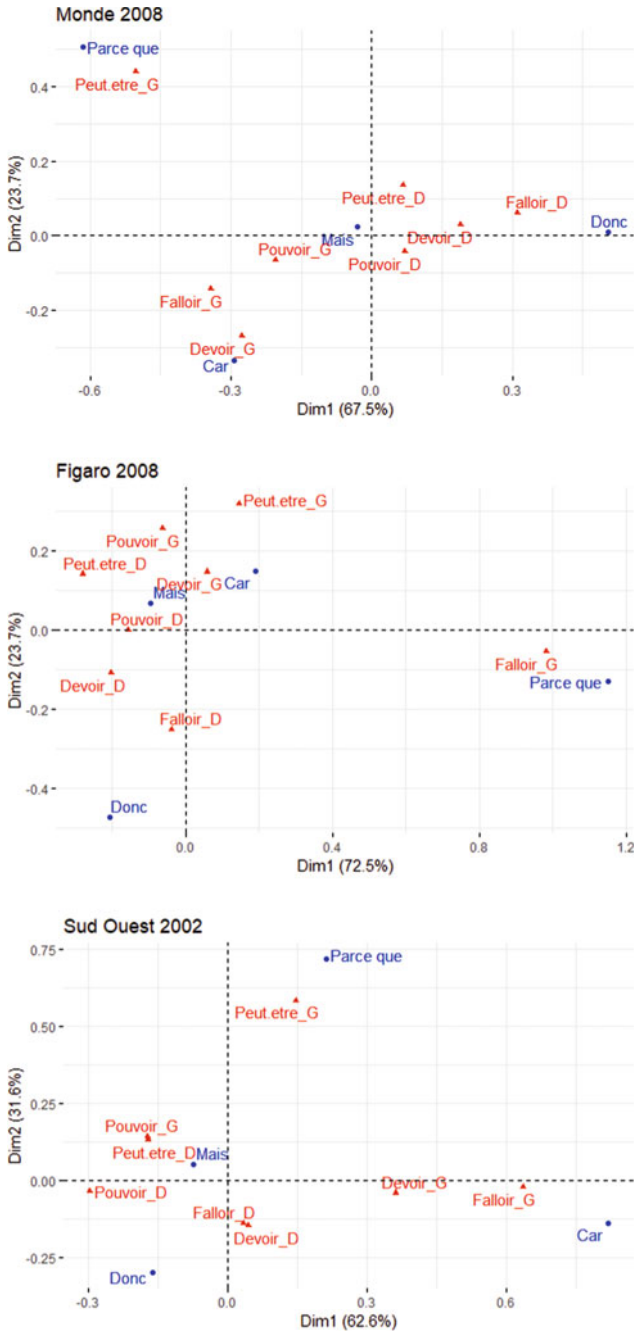


Fig. 2 Correspondence analysis scatter plots for modal forms of necessity and possibility within the six corpora (_G: left position; _D: right position)

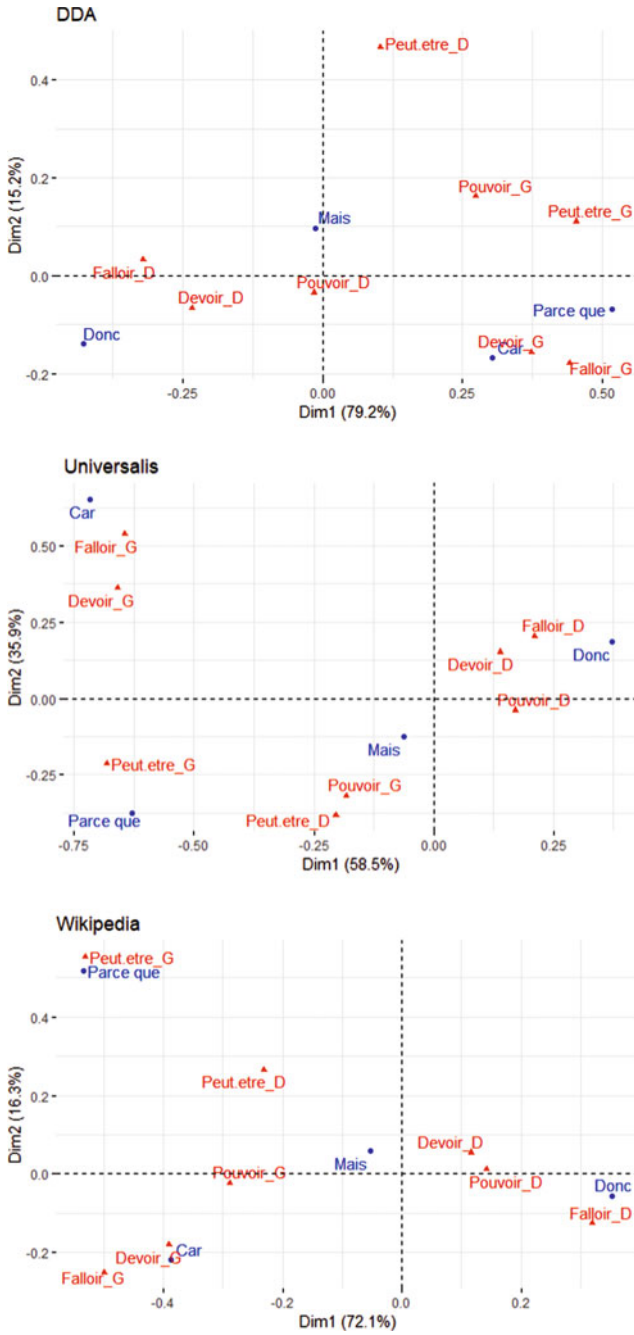


Fig. 2 (continued)

- marked by *parce que* and not the content *il ne connaissait personne*. These uses are very frequent and explain the proximity between *peut-être* and *parce que*, proximity that distances the connector from the other modals conveying a core meaning of possibility: *pouvoir* cannot have such a use.
- (4) One connector shows remarkable homogeneity in all corpora in relation to its association with modals: it is *mais*, which is systematically associated with *pouvoir* in the left position and with *peut-être* in the right position. The result concerning *peut-être* cannot be taken into account because of the modal's attraction to *parce que*, as indicated above. But the result concerning *pouvoir* is meaningful: it shows that, within a concessive sequence, a modal conveying possibility as its core meaning is preferred. This is consistent with the concessive interpretation that the modal verb can trigger (*il peut être riche, mais il n'est pas intelligent*, he may be rich, but he is not smart). Such an interpretation is not accessible with modal verbs having a core meaning of necessity: *il doit être riche, mais il n'est pas intelligent/il faut qu'il soit riche, mais il n'est pas intelligent*.
- (5) No corpus indicates a particular use of a N-MF or of a P-MF; conversely, we have seen that the plots concerning epistemic adverbs make emerge some change concerning *certes* and *sans doute*, (see § 1 and 3 in Sect. 3). The interpretation that we can advance is that the parameter relating to the core meaning, i.e., the difference in behavior between P modal forms and N modal forms, is less subject to change—we notice that the same clustering of N-MFs can be observed in all corpora—than the parameter relating to the different values that the forms of a same paradigm can convey—we notice that some E-ADV's show different degrees of attraction in relation to connectors.

5 Perspectives

Our attempt to use correspondence analysis to study different discursive movements has provided promising results regarding the genre and diachronic variation of discursive functions of French modal forms. It illustrates that the genre and time factors play a role in the use of forms conveying different values stemming from the same core meaning, as observed in Sect. 3. But the core meaning in itself shows strong stability and determines the nature of the discursive sequence in which the form appears, as observed in Sect. 4.

It seems therefore that the values that the modal forms of a paradigm can convey are context-dependent, but the core meaning attached to the form is context-free and can thus be considered as a cognitive primitive.

To test this hypothesis, we can go in two directions: (i) we can extend the analysis to other languages to assess stability and variation in relation to values or to the core meaning of the form; (ii) we can extend the analysis to other corpora representing different diachronic spans as well as to corpora resorting to different communication

channels such as those that favor a spontaneous use of language (spoken language, chat, blogs).

Acknowledgments We would like to thank Mitra Akasereh for her precious help in the constitution and the analysis of the modal verbs plots, as well as Jessica Chessex for her help in collecting and sorting data.

References

1. Blumenthal P (2008) Combinatoire des prépositions: approche quantitative. *Langue française* 157:37–51
2. Blumenthal P (2012) Particularités combinatoires du français en Afrique: essai méthodologique. *Le français en Afrique* 27:55–74
3. Blumenthal P (2017) D'une encyclopédie à l'autre. L'ascension des locutions prépositionnelles. *Langages* 206(2):123–138
4. Blumenthal P, Diwersy S, Mielebacher J (2005) Kombinatorische Wortprofile und Profilkontraste. Berechnungsverfahren und Anwendungen. *Zeitschrift für romanische Philologie* 121:49–83
5. Charolles M, Diwersy S, Vigier D (2017) Evolution des emplois des marqueurs de topiques de discours dans *Le Figaro* de la fin du XIXe et du début du XXIe siècles. *Langages* 206:85–104
6. Désagulier G (2017) *Corpus linguistics and statistics with R. Introduction to quantitative methods in linguistics*. Springer, New York
7. Diwersy S (2014) Corpus diachronique de la presse française: base textuelle créée dans le cadre du projet ANR-DFG PRESTO. Institut des Langues Romanes, Université de Cologne
8. Evert S (2008) Corpora and collocations. In: Lüdeling A, Kytö M (eds) *Corpus Linguistics. An international handbook*. Mouton de Gruyter, Berlin, New York, pp 1212–1248
9. Greenacre MJ (2017) *Correspondence analysis in practice*, 3rd edn. Chapman, Boca Raton
10. Nenadić O, Greenacre MJ (2007) Correspondence Analysis in R, with two- and three-dimensional graphics: the ca package. *J Stat Softw* 20(3):1–13
11. Rossari C, Hütsch A, Ricci C, Salsmann M, Wandel D (2016) Le pouvoir attracteur de *mais* sur le paradigme des adverbes épistémiques: du quantitatif au qualitatif. In: Mayaffre D et al (eds) *Proceedings of 13th international conference on statistical analysis of textual data*, II, pp 819–823
12. Rossari C, Salsmann M (2017) Étude quantitative des propriétés dialogiques des adverbes épistémiques. *Actes des 9èmes Journées Internationales de la Linguistique de corpus* 87–93
13. Roulet E (1979) Des modalités implicites intégrées en français contemporain. *Cahiers Ferdinand de Saussure* 33:41–76
14. Roulet E, Auchlin A, Moeschler J, Schelling M, Rubattel C (1991) *L'articulation du discours en français contemporain*, 3rd edn. Lang, Bern
15. Vigier D (2017) La préposition dans au xvie siècle. Apports d'une linguistique instrumentée. *Lang*, 206(2):105–122