

Similarities and disparities among French modal verbs through co-occurrence correspondence analysis

Aylin Pamuksaç

University of Neuchâtel – aylin.pamuksac@unine.ch

Abstract

The French modal verbs *devoir*, *pouvoir*, *falloir* and *vouloir* are rarely considered in one same study although they can all express the same main modalities which are the epistemic and the deontic values. In this paper, we want to show that despite the fact they are polysemous and seem to be interchangeable in some context, they are quite different in their use. To do so, we show how close or distant they are through correspondence analysis (CA). This study is innovative because this methodology will be used to determine proximities and distances between words and not between parts of the text, as is usually the case. The CAs are constructed on co-occurrences that modal verbs have with their verbal complementation and the subject pronouns that govern them. Not only do we compare two parameters to see if the properties of the modal verbs are stable but, we also conduct this research in four different corpora, two journalistic, and two encyclopedic to show regular patterns between genres.

Keywords: modality, French modal verbs, co-occurrences, specificity, correspondence analysis, corpora studies.

1. Introduction

Studies on modal verbs in French are numerous but often focus on only one or two verbs and mainly on *devoir* and *pouvoir*. For example, they compare the deontic and the epistemic values of these two verbs (Barbet 2012; de Saussure 2014; Sueur 1977, 1979; Vettters 2004, 2012). However, we know that *falloir* and *vouloir* are also considered modal verbs in French and can also express the two modalities mentioned:

Deontic modality:

- 1a. Tu **dois** ranger ta chambre.
- 1b. **Peux-tu** ranger ta chambre ?
- 1c. Il **faut** ranger ta chambre.
- 1d. Je **veux** que tu ranges ta chambre.

Epistemic modality:

- 2a. Il **doit** être bête pour croire que la terre est plate.
 2b. Il se **peut** qu'il soit bête pour croire que la terre est plate.
 2c. Il **faut** être bête pour croire que la terre est plate.

In another context and in some French-speaking regions, we can also hear¹ this epistemic value with *vouloir*:

3. Il **veut** pleuvoir ce soir.

As we can see from examples 1 to 3, not only *devoir* and *pouvoir*, but also *falloir* and *vouloir* are polysemous and fit into different modal categories. These classifications, although accepted and still widely studied in modality studies, do not allow us to dissociate the linguistic forms from each other. These qualitative studies therefore do not grant us to determine whether a verb form is preferred according to the modality expressed.

In this article, we want to disregard these modal categories in order to study how French modal verbs are similar or different in their actual functioning. To do so, we will study their associations with the 50 most frequent verbs of our corpora but also with their subject pronouns through correspondence analysis (now CA). These two variables were selected for our study because modal verbs are among the only verbs that must be associated with other verbs to have a meaning (Abeillé et al. 2021: 131) and because all verbs do have a subject in French (Lazard 2009: 152-153). Thus, these elements, because there are common for the verbs studied here, are particularly relevant to considerate.

We will conduct this research in four different corpora, all of which are informative. This variety of corpora will allow us to see if regularities of properties occur for French modal verbs across sub-genres of informative text but also across time as we are working on three contemporary corpora and one from the 18th century.

2. Methodology

Two journalistic corpora, one national (*Le Monde* 2010, 17895009 tokens, now LM) and one regional (*L'Est Républicain* 2010, 25825949 tokens, now ER) as well as two encyclopedic corpora, one contemporary (*Wikipedia* 2019, 18715455 tokens, now WIKI) and one from the 18th century (*L'Encyclopédie de Diderot et d'Alembert*, 24239478 tokens, now EDDA) were used for this study². All these

¹ We use here the verb to hear because this kind of construction appears mostly in the spoken language.

² All the corpora, with the exception of EDDA, were constructed by ourselves or by members of the team in which we are employed. For EDDA, we thank the members of the DISCO project, and more particularly D. Vigier and A. Brenon.

corpora were explored in the TXM software, developed at the ENS of Lyon (Heiden et al. 2010) to count cooccurrences between our four modal verbs and their verbal complements and their subject pronouns.

2.1. The CA methodology

Developed by Benzécri (1973), the goals of CAs are to summarize a large amount of data of a table in a graphical visualization form in order to bring out contrasts (Poudat & Landragin 2017: 104) but also proximities between the various represented elements (Leblanc 2015: 44). CA is not a new tool in the field of textual data analysis. However, it has rarely been used to study co-occurrences. Usually, this method is used to bring out features of a speech³. In our research, we will use this method, not to make sense of parts of a text, but to study the meaning of closely related linguistic forms, in this case French modal verbs, to see whether differences or commonalities emerge in their actual use.

To do so, we produced the CAs in the software Hyperbase (Brunet 2011). TXM can provide CAs but these do not concern co-occurrences but frequencies of lemmas, words or grammatical categories such as verbs, nouns, adjectives, etc. These CAs would show the distribution of these different elements in relation to the partitions of a corpus and thus give us information about the construction of the text rather than the functioning of the words⁴.

For our article, because we want to see the proximities or distance between verbs, we counted the number of co-occurrence for the four modal verbs with each verb from our list in every corpus. Our counting was made on TXM and rather manual. Indeed, we went through the KWIC tool of the software to see if the co-occurring verb was really modalised by one of our modals. For example, we omitted uses such as:

4. *Corneille dût même se résoudre à demander l'aide à trois auteurs* (LM)

where *demande* is linked to *se résoudre* and not the verb *devoir*.

³ See for example Mayaffre et al. (2019) for a study on the similarities and differences between speeches by French presidents.

⁴ For example, TXM could have provided us with the distribution of the various verbs we are studying here across the different EDDA authors or whether one uses more adjectives than another but not how close or distant two verbs are in their actual uses.

3. Discussion on the results

3.1 *The co-occurrences between the modal verbs and their verbal complements*⁵

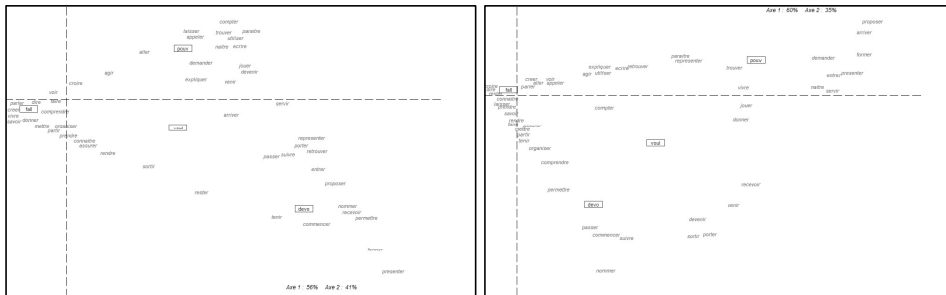


Figure 1. CA of co-occurrences between the four French modal verbs and the 50 most frequent verbs of our corpora. Left LM, right WIKI.

Falloir shows a very stable behavior as in all corpora it is found in the center of the two axes and surrounded by verbs such as *croire*, *parler* and *savoir*. *Vouloir* also adopts a regular behavior in all corpora. Indeed, there seems to be a “void” around it, except in ER, where it is surrounded by *expliquer*, *nommer*, *arriver*, *naître* and *commencer*.

Another regularity is that in all CAs, two scales seem to stand out from *falloir*. In fact, some co-occurents make the link between *falloir* and *pouvoir* towards the top of the graph and others move under, to *devoir*. *Falloir* thus seems to be at the intersection between *devoir* and *pouvoir*.

Indeed, while the vertical axis is similar in the four corpora, this is not the case for the horizontal axis. From the left, we have three possible patterns:

1) *Falloir* → à *pouvoir-vouloir* → *devoir* in the two press corpora: here, we can see that the verbs of necessity are very far from each other. This information is even more important as the horizontal axis contains more information than the vertical axis, so this distinction is all the more important to note. *Pouvoir* and *vouloir*, on the other hand, share more or less the same position on this axis, point that has not been found in studies on modality yet. The qualitative studies on modality did not allow us to identify such proximity between these two verbs because they were never studied together, as we said in our introduction, most

⁵ For reasons of space, we have included here only the CAs of the LM and WIKI corpora. These two CAs as well as the two missing CAs are visible in optimal quality under: <https://drive.switch.ch/index.php/s/vil2d3FnYX2hzdq>

of the time, *pouvoir* is compared to *devoir*. Our methodology demonstrates that this point deserves to be analyzed, or at least considered, in qualitative research.

II) *Falloir* → *devoir* → *vouloir* → *pouvoir* in WIKI: this scale seems much more predictable than the previous one if we consider the qualitative research. Indeed, *falloir* and *devoir* are both verbs of strong necessity, mostly used to express an obligation. This proximity between *devoir* and *falloir* in WIKI can be explained by the fact that *devoir* expresses more often the deontic modality in this corpus than an epistemic or quasi-certainty modality that we were able to identify in a previous study (Rossari et al. 2021). Thus, we should see in the other corpora whether the distance between *devoir* and *falloir* is not simply because *devoir* does not express deontic modality as often as one might imagine.

III) *Falloir* → *vouloir* → *devoir* → *pouvoir* in EDDA: in the 18th century corpus, another pattern seems to emerge. Volition seems to link two modal verbs expressing, *a priori*, necessity and *pouvoir* and *devoir* seem closer than in the other corpora. In addition, many verbs are found between *pouvoir* and *devoir* on the vertical axis which shows a certain sharing of co-occurrences between these two verbs.

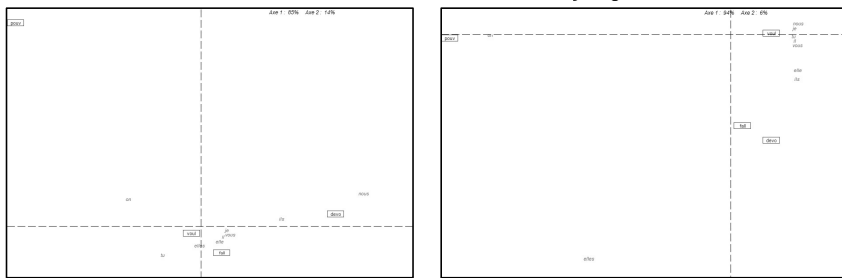


Figure 2. CAs of co-occurrences between the four French modal verbs and their subject pronouns in LM (left) and WIKI (right)

3.2 The co-occurrences between the modal verbs and their subject pronouns

We wanted to see if the subject pronouns of these modal verbs can also make appear some similarities or differences that the classic modal classification does not make come up. We did the CA here only for two corpora, one in each genre, to see if the differences and similarities already denoted before with the verbal complementation emerge again or not.

First of all, it could be surprising to see *falloir* in this kind of results because it is an impersonal verb, so it is always used with the third person singular male, i.e. *il* in French. But our primary goal was to consider all modal verbs for our study, regardless of their syntactic specificity or qualitative analyses about them. This is why we have taken *falloir* into account but counting the occurrences of

the complement pronouns that appear before the verb as in

5. Il **me** fallait gagner plus de temps. (LM)

In which *me* was counted like an occurrence of *je*. We thus considered all the cases where *falloir* was preceded by a pronoun in the dative, and even sorted the occurrences of *lui* and *leur* to see if they represented feminine or masculine nouns in order to count the number of occurrences of *elle* and *il*. Incidentally, we have therefore omitted *il* pronoun from our calculations⁶.

In the journalistic corpus, *vouloir* is very central and demonstrates some proximity with first-, second- or third-person pronouns. In contrary, in WIKI, this verb stands out from the others by attracting mainly first or second person pronouns, which seems, in this corpus, to be a specific feature of *vouloir*, which is on the extreme right of axis 1 with all these pronouns.

On the other extreme, in both corpora, *pouvoir* seems to have a behavior that stands out strongly from the others. The pronoun which put *pouvoir* in such position is without doubt *on*. This phenomenon could explain the proximity between this verb and *falloir* we already identified in the press corpora with the verb complements in §3.1. This closeness also appeared to us in a previous study, in which we only considered one semantic category, the *dicendi* and *putandi* verbs as complements to the modal verbs in our CAs (Pamuksaç 2022). *Pouvoir*, with *on*, could denote some “abstract effect”, already noted by Larsen (in Fløttum et al. 2007: 22-23) which put it closer to *falloir*, an impersonal verb.

Finally, we also note that in WIKI, again, *falloir* and *devoir* are closer in their associations with subjects than in LM. This proximity between these two verbs seems to be a stable aspect of this corpus, whatever the parameters considered.

4. Conclusion

In this study, we wanted to compare all four French modal verbs (*pouvoir*, *devoir*, *falloir* and *vouloir*) in a quantitative study to see if there were similarities or differences between them on various parameters. Indeed, as explained in the introduction, modal categories in qualitative studies allow us to see that modal verbs are all polysemous but not if the closeness of meaning between them. For this purpose, we used the CA representation to see whether in their actual use the modal verbs tended to resemble each other or to denote certain tendencies. It appeared that by studying their verbal complementation, several scales were

⁶ In our research, the study of pronouns is still in its infancy. For example, we do not know what to make of cases without a pronoun in the dative with *falloir*. Perhaps we can consider them as an impersonal *on*, and thus count these occurrences as the pronoun *on*. For the moment, we have considered that this pronoun does not appear with *falloir*.

emerging, ranging from a modality denoting a strong obligation to a possibility, i.e. a much weaker necessity. Even if certain regularities seemed to emerge, differences between corpora, and therefore perhaps between genres and periods, were apparent. But, we also denoted that some properties of these modal verbs are stable, despite the parameters we study: the proximity between *falloir* and *devoir* has been made salient in WIKI, both with verbal complementation and with subject pronouns but not in other corpora for example.

Our methodology can thus fill a gap in the study of the polysemy of French modal verbs. For now, we noticed some proximity or some distance between them with two criteria studied in four corpora, but we could deepen this study and determine clearer proximities or disparities between modal verbs, if, in the future, we label them with their values. Thus, some complement verbs might appear more often with *falloir* and *devoir* when both express necessity, and others would be closer to *pouvoir* and *devoir* when they denote an epistemic modality. In that case, we could therefore show that the modal verbs are similar or note in their uses according to their values.

References

- Abeillé A., Koenig J. P., Godard D. and Bonami O. (2021). Qu'est-ce qu'un verbe ? In A. Abeillé, D. Godard, A. Delaveau, & A. Gautier (Eds.), *La Grande Grammaire du Français* (Vol. 1, pp. 127-147). Actes Sud.
- Barbet C. (2012). Devoir et pouvoir, des marqueurs modaux ou évidentiels? *Langue française*, 1: 49-63.
- Benzécri J. P. (1973). *L'analyse des données*. Dunod.
- Brunet E. (2011). *Hyperbase, Logiciel hypertexte pour le traitement documentaire et statistique des corpus textuels, Manuel de Référence*. Université de Nice.
- de Saussure L. (2014). Verbes modaux et enrichissement pragmatique. *Langages*, 193 (1): 113-126.
- Fløttum K., Jonasson K. and Norén C. (2007). *On : pronom à facettes*. De Boeck/Duculot.
- Lazard G. (2009). Qu'est-ce qu'un sujet? *La linguistique*, 45 (1): 151-158.
- Leblanc J. M. (2015). Proposition de protocole pour l'analyse des données textuelles : pour une démarche expérimentale en lexicométrie. *Nouvelles perspectives en sciences sociales*, 11 (1): 25-63.
- Mayaffre D., Pincemin B. and Poudat C. (2019). Explorer, mesurer, contextualiser. Quelques apports de la textométrie à l'analyse de discours. *Langue française*, 3: 101-115.
- Pamuksaç A. (2022). Les verbes modaux du français : ressemblances et dissociations à l'aide de leurs associations avec les verbes *dicendi* et *putandi*. *8^{ème} Congrès Mondial de Linguistique Française 2022 (CMLF2022)*, <https://doi.org/10.1051/shsconf/202213811014>
- Poudat C. and Landragin, F. (2017). *Explorer un corpus textuel : Méthodes, pratiques, outils (Champs linguistiques. Recherches)*. De Boeck Supérieur.

Rossari C., Montrichard C., Pamuksaç A., Ricci C. and Sanvido L. (10 novembre 2021). Connecteurs et modalités : des traces pour identifier les choix rhétoriques dans un discours. *Colloque international ARGAGE 2021 (Argumentation & Langage)*, 10-12 novembre 2021.

Sueur J. P. (1977). À propos des restrictions de sélection : les infinitifs devoir et pouvoir. *Linguisticae investigationes*, 1 (2): 375-409.

Sueur J. P. (1979). Une analyse sémantique des verbes devoir et pouvoir. *Le Français Moderne Paris*, 47 (2): 97-120.

Vetters C. (2004). Les verbes modaux pouvoir et devoir en français. *Revue belge de philologie et d'histoire*, 82 (3): 657-671.

Vetters C. (2012). Modalité et évidentialité dans pouvoir et devoir : typologie et discussions. *Langue française*, 1: 31-47.