

# Modal forms expressing probability and their combination with concessive sequences in French and Italian<sup>1</sup>

Corinne Rossari, Claudia Ricci and Margot Salsmann

Department of French linguistics  
University of Neuchâtel  
E-mail: corinne.rossari@unine.ch

## Abstract

This paper deals with the combination between modal forms and concessive constructions in French and Italian, which we investigated in the framework of a tool-based approach to the linguistic analysis of corpora. To this end, we analyzed the frequency of forms expressing epistemic value and compared it to the frequency of their co-occurrence with the conjunction *mais/ma* ('but'), seeking regularities in the link between this co-occurrence and the ability of an epistemic form to convey new meanings.

## 1 Introduction

This contribution examines the combination between some modal forms and concessive constructions in two Romance languages, French and Italian<sup>2</sup>. Our purpose is to identify regularities in the correlation between how frequently an epistemic form occurs in the environment of *mais/ma* ('but'), and the ability of such a form to convey new meanings (e.g. a concessive value). To this end, we have analyzed the frequency of French forms expressing epistemic value in correlation with their co-occurrence with the conjunction *mais*. A parallel analysis has been carried out on the corresponding forms in Italian and their co-occurrence with the conjunction *ma*. Our analysis has been conducted in the framework of a tool-based approach to the linguistic analysis of corpora.

The research has been carried out from a synchronic perspective for French and Italian, with the aim of assessing the degree of convergence between two Romance languages, and from a diachronic perspective for French<sup>3</sup>, with the aim of assessing a possible evolution concerning the co-occurrence between *mais* and the epistemic forms discussed here.

## 2 Analysis

---

<sup>1</sup> This research has been carried out in the framework of a research project entitled *La représentation du sens modal et de ses tendances évolutives dans deux langues romanes: le français et l'italien* (Representation of modal meaning and of the trends in its evolution in two Romance languages: French and Italian, project n. 100012\_159458), directed by Corinne Rossari and financed by the Swiss National Science Foundation.

<sup>2</sup> For an overview of modal adverbs in French see Borillo [3].

<sup>3</sup> Our synchronic analysis covers both French and Italian corpora, while the diachronic analysis has been carried out on a French corpus only due to the unavailability of an equivalent corpus in Italian in our analysis tool.

The starting point of our research is the qualitative observation made in Rossari [9] and [11] that a number of French epistemic adverbs may take on a concessive value, which has become more or less conventionalized, while other adverbs of the same category do not acquire this kind of value<sup>4</sup>. We have observed that French adverbs *peut-être* ('maybe') and *sans doute* ('surely', lit. 'without (a) doubt') can either convey an epistemic value or be used by the speaker to cast doubt on the validity and/or relevance of a state of affairs, regardless of all evaluation of its degree of certainty. Let us look at the following case:

- 1) -Où est la directrice ?  
-Elle est *peut-être/sans doute* dans son bureau, son secrétaire vient de lui apporter le café  
- 'Where is the director?'<sup>5</sup>  
- 'Maybe/Surely she is in her office, her secretary just brought her coffee'

In such a context, the above adverbs are interchangeable with *probablement* ('probably') in terms of acceptability of the utterance:

- 2) -Où est la directrice ?  
-Elle est *probablement* dans son bureau, son secrétaire vient de lui apporter le café  
- 'Where is the director?'  
- 'She is probably in her office, her secretary just brought her coffee'

By contrast, in a context where the speaker resumes a content which is already present in the discursive background – for instance, because it has already been uttered by another speaker – with the intention of casting doubt on the validity and/or relevance of such a content, the adverb *probablement* is less natural, while *peut-être* and *sans doute* trigger a different interpretation of the utterance, less strongly rooted on an epistemic evaluation:

- 3) -Je trouve que la nouvelle directrice est sympa  
-Elle est *peut-être/sans doute/probablement* sympa, *mais* elle est super exigeante  
- 'I think our new director is really cool'  
- 'Maybe/Surely/Probably she is cool, but she is extremely demanding'

Within this kind of context, the epistemic adverbs *peut-être* and *sans doute* seem to be taking on a concessive value which can go so far as to entirely rule out epistemic evaluation in the interpretation of the utterance. Such is the case for *peut-être*, which can be associated with contents whose degree of certainty cannot be questioned:

- 4) Je suis *peut-être* italienne, *mais* je n'aime pas la pizza

---

<sup>4</sup> The framework concerning the conventionalization of a value is related to the grammaticalization theory developed by Hopper and Traugott ([6]). In Traugott [13] the epistemic adverbs are analyzed in relation to their position in the sentence.

<sup>5</sup> Our translations are intended to render the basic meaning of the French and Italian examples and do not necessarily reflect the relevant values discussed here for French or Italian adverbs and tenses.

‘I may be Italian but I don’t like pizza’

The other epistemic adverbs introduced above are either less natural or even incompatible with this type of context:

5) Je suis <sup>?</sup>*sans doute*/*\*probablement* italienne, *mais* je n’aime pas la pizza

‘I am surely/probably Italian, but I don’t like pizza’

*Sans doute* is less natural than *peut-être* in such a context, whereas *probablement* sounds particularly odd. However, in contexts where the state of affairs expressed is less self-evident, such as in the following example:

6) Le débat est *sans doute* français, *mais* il intéresse tout le monde (Mei Duanmu, Hugues Tertrais (eds.), *Temps croisés I*, Paris, Maison des Sciences de l’Homme, 2010)

‘The debate is surely French, but it is of interest to everyone’

*sans doute* may occur (as could *peut-être*), whereas *probablement* in the same context would exclusively convey an epistemic reading (‘it is not sure that the debate is French’).

When used in the above way, *peut-être* expresses a concessive value which has become conventionalized. The adverb does not provide epistemic quantification concerning the states of affairs ‘being cool’ or ‘being Italian’, but rather takes in its scope the very act of enunciation, as if it were merely potential. In Rossari [10], such a use or value is labelled “enunciative” as opposed to a content/semantic use or value. While not sharing the same contexts of use as *peut-être*, *sans doute* appears to follow a similar trend.

The values these French adverbs convey are comparable to those of their Italian counterparts<sup>6</sup>, as shown in the Italian rendering of the above examples:

7) -Dov’è la direttrice?

-*Forse/Probabilmente* è in ufficio, il segretario le ha appena portato il caffè.

-‘Where is the director?’

-‘Maybe/probably she is in her office, her secretary just brought her coffee’

Similarly to French, when used in a concessive context, the Italian adverbs show different degrees of acceptability. *Probabilmente* is less natural as it only conveys an epistemic meaning:

8) -Penso che la nuova direttrice sia proprio simpatica

-*Forse*/*Probabilmente* è simpatica, *ma* è molto esigente

-‘I think our new director is really cool’

-‘Maybe/probably she is cool, but she is extremely demanding’

As *peut-être*, *forse* can also be used in a context in which epistemic evaluation is completely ruled out, while we observe that *probabilmente* shows the same behavior as its French counterpart:

---

<sup>6</sup> Except for *sans doute*, which has no unique equivalent in Italian. To remain as close as possible to the French forms, we have restricted our analysis to the forms that convey very similar values in Italian. Therefore only *forse* (equivalent to *peut-être*) and *probabilmente* (equivalent to *probablement*) are analyzed here.

- 9) Sono *forse* Italiano per l'anagrafe perché ahimè sono nato lì, *ma* appena ho avuto l'opportunità, me ne sono andato il più lontano possibile.  
(www.australianboardcommunity.com)

'I may be Italian for the registry office because, alas, I was born there, but as soon as I had the opportunity I went as far away as possible'

- 10) Sono *\*probabilmente* Italiano per l'anagrafe perché ahimè sono nato lì, *ma* appena ho avuto l'opportunità, me ne sono andato il più lontano possibile.

'I am probably Italian for the registry office because, alas, I was born there, but as soon as I had the opportunity I went as far away as possible'

The same type of value conveyed by *peut-être* and *forse* can be taken on by the Italian future<sup>7</sup>, which can also function at the level of enunciation, when it precedes *ma*:

- 11) *Sarò* italiana *ma* non mi piace la pizza

'I may be Italian but I don't like pizza'

Such a future could not be used in French:

- 12) *\*Je serai* italienne, *mais* je n'aime pas la pizza

On the basis of these data, evaluated on a qualitative level, we make the assumption that there is an evolution pattern in the conventionalization of the concessive value of epistemic adverbs, or of a tense as the future in Italian. We claim that there is a correlation between this conventionalization process and the frequency of these forms in a context in which they precede *mais/ma*. The more frequently an epistemic form or the future occurs in such a context, the more likely such a form will be to lose its literal meaning and acquire a concessive value, thus allowing the speaker to cast doubt on the validity and/or relevance of a content already present in the discursive background.

Our purpose is: (i) from a synchronic perspective, to compare French and Italian in order to verify the convergence between two Romance languages as for the ratio: frequency of the correlation between *mais/ma* and an epistemic form (or the future) / conventionalization of a concessive value; (ii) from a diachronic perspective, to verify if such a frequency is stable or has evolved along the centuries. The latter analysis is focused on French data but we intend to extend it to Italian data.

### 3 French and Italian data

Our quantitative study<sup>8</sup> has been conducted on an annotated French corpus of over twenty million words. We have used the BTLC platform conceived by Sascha Diwersy (see Diwersy [5]), which allows, among other, extracting concordances, calculating frequencies and frequency-related specificities within the corpus, as well as extracting lexical-syntactic co-occurrences. This

---

<sup>7</sup> For a description of such a use of the Italian future see Berretta [1].

<sup>8</sup> See Legallois [7] for the theoretical background taken into account for our quantitative research.

platform includes synchronic corpora representing different genres of speech as well as a diachronic database gathering corpora from the 16<sup>th</sup> to the middle of the 20<sup>th</sup> century, calibrated with respect to the genres of speech (theatre, romance novels, essays) and selected in the framework of a French and German project on the diachronic evolution of French prepositions (PRESTO) (see <http://presto.ens-lyon.fr/> for a description of the project). For our synchronic analysis we used a corpus of press speech from the 21<sup>st</sup> century (*Le Monde* 2008). This corpus includes different subgenres (such as editorials, books reviews), which have been treated separately to verify if the frequencies observed in the whole media press corpus are also relevant in the subgenres, avoiding, as far as possible, corpus biases. For our diachronic analysis, we have used the corpora selected in the frame of the Presto project, restraining them to the period 17<sup>th</sup>-19<sup>th</sup> century, as data for the 20<sup>th</sup> century are only available until 1941.

As regards Italian, we have used the same platform, which contains an annotated Italian press corpus of over thirty million words (*La Stampa* 2002), equivalent to the French corpus in terms of genres, subgenres and synchronic span. Since BTLC does not include a diachronic corpus for Italian, such a corpus will be further constituted in order to conduct a parallel analysis of the diachronic evolution of these forms in Italian.

## 4 Discussion of results for French

The results including the synchronic and diachronic data for the epistemic adverbs support our initial assumption concerning the correlation between their frequency in the environment of *mais* and their ability to convey or stress a concessive value.

Corpus *Le Monde* 2008 (20 410 766 words)

Adverbs	Total number of occurrences	Relative frequency (occurrences per million words)	Number of occurrences followed by <i>mais</i> <sup>9</sup>	Relative frequency (occurrences per million words)	Rate of co-occurrences with <i>mais</i> on all occurrences of the adverb
Peut-être	2904	142.27	432	21.16	14.87 %
Sans doute	2483	121.65	323	15.82	13.00 %
Probablement	724	35.47	60	2.93	8.28 %

Figure 1: results for the association between epistemic adverbs and *mais* – synchronic perspective

These synchronic data bring out a significant fact: in the case of *probablement*, the rate of the adverb co-occurring with *mais* on its overall number of occurrences is lower by almost half than the same rate for *peut-être* and by 1.5 times than the same rate for *sans doute*.

<sup>9</sup> We have taken into account a text portion in which the adverb co-occurs with *mais* within a span of 0 to 20 words.

**Corpus Presto – 17th-19th century**

(17th: 7 040 233 words; 18th : 7 756 598 words; 19th : 8 020 248 words)

<b>17<sup>th</sup> century</b>	<b>Total number of occurrences</b>	<b>Relative frequency (occurrences per million words)</b>	<b>Number of occurrences followed by <i>mais</i></b>	<b>Relative frequency (occurrences per million words)</b>	<b>Rate of co-occurrences with <i>mais</i> on all occurrences of the adverb</b>
Peut-être	957	135.93	118	16.76	12.33 %
Sans doute	601	85.34	76	10.79	12.6 %
Probablement	21	2.98	0	–	0 %
<b>18<sup>th</sup> century</b>	<b>Total number of occurrences</b>	<b>Relative frequency (occurrences per million words)</b>	<b>Number of occurrences followed by <i>mais</i></b>	<b>Relative frequency (occurrences per million words)</b>	<b>Rate of co-occurrences with <i>mais</i> on all occurrences of the adverb</b>
Peut-être	2493	321.40	306	39.45	12.27 %
Sans doute	1411	181.79	278	35.81	19.7 %
Probablement	129	16.61	16	2.06	12.4 %
<b>19<sup>th</sup> century</b>	<b>Total number of occurrences</b>	<b>Relative frequency (occurrences per million words)</b>	<b>Number of occurrences followed by <i>mais</i></b>	<b>Relative frequency (occurrences per million words)</b>	<b>Rate of co-occurrences with <i>mais</i> on all occurrences of the adverb</b>
Peut-être	2774	345.87	360	44.88	12.9 %
Sans doute	2053	255.82	523	65.17	25.47 %
Probablement	362	45.09	22	2.74	6 %

Figure 2: results for the association between epistemic adverbs and *mais* – diachronic perspective

The diachronic data<sup>10</sup> show a continuous progression of the overall frequency of *peut-être*, coupled with a parallel progression of its frequency in co-occurrence with *mais*. The rate of occurrences of *peut-être* with *mais* with respect to the overall occurrences of the adverb is remarkably stable throughout three centuries. The same does not hold true for *sans doute*, whose use with *mais* progresses strongly compared to the overall progression of the adverb's frequency during the three centuries. On the contrary, while the frequency of *probablement* increases steadily through the three centuries considered, its co-occurrence with *mais* decreases by half from the 18<sup>th</sup> to the 19<sup>th</sup> century.

These data are consistent with the assumption that *peut-être* had the same ability to take on an enunciative value for the three centuries considered, whereas *sans doute* appears to develop this ability throughout the centuries.

<sup>10</sup> The data in the diachronic table cannot be compared to those in the synchronic one because each table results from a different corpus sample.

## 5 Discussion of results for Italian

The same synchronic quantitative analysis as for French has been conducted for the Italian epistemic adverbs corresponding to *peut-être* and *probablement*, namely *forse* and *probabilmente*. Our aim is to verify if, in Italian also, the value of the adverbs can be linked to their presence in a context containing *ma*.

**Corpus La Stampa 2002** (31 369 484 words)

Adverbs	Total number of occurrences	Relative frequency (occurrences per million words)	Number of occurrences followed by <i>ma</i>	Relative frequency (occurrences per million words)	Rate of co-occurrences with <i>ma</i> on all occurrences of the adverb
Forse	11773	375.30	1494	47.62	12.69 %
Probabilmente	2771	88.33	273	8.70	9.85 %

Figure 3: results for the association between epistemic adverbs and *ma* – synchronic perspective

Although the gap between the ratios for the different adverbs is smaller, it is a significant fact that *probabilmente* shows a lower rate of occurrences with *ma*, which is comparable to the results obtained for its French counterpart *probablement*.

Since the Italian future can also convey a concessive value when it precedes *ma*, besides conducting a quantitative analysis on the epistemic adverbs, we have also analyzed the frequency of the future with *ma/mais* in Italian and in French. We aimed at verifying if such a use is correlated to a higher frequency of the co-occurrence ‘future [...] *ma*’ in Italian as compared to the frequency of the co-occurrence ‘future [...] *mais*’ in French.

**Corpus Le Monde 2008** (20 410 766 words) / **La Stampa 2002** (31 369 484 words)

Future	Total number of occurrences	Relative frequency (occurrences per million words)	Number of occurrences with <i>ma/mais</i>	Relative frequency (occurrences per million words)	Rate of occurrences with <i>ma/mais</i> on all occurrences of the future
Italian future	143 063	4560.58	9902	315.66	6.9 %
French future	53139	2603.48	2754	134.92	5.1 %

Figure 4: results for the association between the future tense and *ma/mais* – synchronic perspective

Although the co-occurrence with *ma/mais* is lower in the case of the future than it is in the case of the adverbs, it is significant that its frequency is still higher in Italian than it is in French<sup>11</sup>. Consistently with the qualitative

<sup>11</sup> The difference between Italian and French does not appear to be so spectacular due to the fact that, within a span of 20 words, there are many cases in which the use of the future is

analysis, which shows that the future in French does not convey a concessive value, we observe that this tense is less likely to appear with *mais*.

## 6 General discussion

Besides the correlation between the frequency of occurrences of the above forms in the environment of *ma/mais* and their ability to convey a concessive value, the quantitative results show another peculiarity.

We can observe that there is a remarkable disproportion between the overall frequency of the future in French and that of the future in Italian. Our data show half the rate of occurrences of the future in French as compared to Italian. In a previous study (Ricci, Rossari and Siminiciuc [8]) we had observed that the non-temporal uses of the future (which we consider as resulting from an enunciative use of the tense and of which the concessive future is an instance) are much more frequent in Italian than they are in French.

We have verified this on a quantitative level by analyzing the rate of futures expressing probability<sup>12</sup> in relation to the overall frequency of the future. Below are the results of this research on the totality of our corpus:

**Corpus *Le Monde* 2008 (20 410 766 words) / *La Stampa* 2002 (31 369 484 words)**

Future	Total number of occurrences	Relative frequency (occurrences per million words)	Number of occurrences with an epistemic value	Relative frequency (occurrences per million words)	Rate of occurrences with an epistemic value on all occurrences of the future
<b>Italian future</b>	143 063	4560.58	25 in a sample of 1500	76.00	1.66 %
<b>French future</b>	53139	2603.48	4 in a sample of 1500	6.94	0.26 %

Figure 5: results for the epistemic use of the future tense in the global French and Italian corpora

The data in our table confirm that the overall frequency of occurrences of the future is linked to the rate of its occurrences as a future of probability. The relative frequency of the future is 1.75 times higher in Italian than in French. Correspondingly, the rate of occurrences of the future of probability on all occurrences of the future is 6.4 times higher in Italian.

---

independent of the use of *ma/mais*. Within a span of 10 words, the gap between the rate of occurrences with *ma* in Italian as compared to French increases up to a ratio of 2:1.

<sup>12</sup> Such a future, which we also consider as being an enunciative use in Rossari, Ricci and Siminiciuc [12] has been studied, a.o., by Dendale [4] for French and is referred to in the *Grande grammatica italiana di consultazione* as “epistemic future” (see Bertinetto [2], 118-120). This type of future exists both in French and in Italian. Below is an example of its use:  
 Non telefonargli! A quest’ora *dormirà* / Ne l’appelle pas! À cette heure-ci, il *dormira*  
 ‘Don’t call him. He’ll be sleeping at this hour’

The same research has been conducted on a subsection of each corpus, respectively *Le Monde des Livres* and *Inserto tuttolibri* (comparable subsections featuring book reviews) in order to make sure that the difference in rates is not due to corpus bias. The results show a very similar disproportion:

**Corpus La Stampa 2002, subsection *Inserto tuttolibri* (1 421 663 words)**  
**Corpus Le Monde 2008, subsection *Le Monde des livres* (1 095 656 words)**

Future	Total number of occurrences	Relative frequency (occurrences per million words)	Number of occurrences as a future of probability	Relative frequency (occurrences per million words)	Rate of occurrences as a future of probability on all occurrences of the future
Italian future	3384	2380.31	80	56.27	2.36 %
French future	2162	1973.24	18	16.42	0.83 %

Figure 6: results for the epistemic use of the future tense in a subsection of each corpus

The overall data make visible not only the correlation between the frequency of the adverb occurring in the environment of *mais* and its ability to convey a concessive value – a frequency which has been observed in synchrony, and diachrony for French – but also the link between the frequency of the future and its ability to have a non-temporal probability use.

## 7 Conclusion

Quantitative data have allowed: (i) for French, in a synchronic and in a diachronic perspective, to verify the influence of *mais* on the meaning of the adverbs, showing that the more frequently a marker appears in the left environment of *mais*, the more it will be prone to develop enunciative meanings consisting in stressing or taking on concessive values; (ii) to highlight the fact that enunciative meanings such as probability and concession for the future in Italian are correlated to the absolute frequency of the form.

As consequence of the latter observation, we could make another assumption which should be tested in a further analysis: the more frequent an epistemic form is, be it grammatical or lexical, the more it is prone to develop enunciative values.

## References

- [1] Berretta, Monica (1997) Sul futuro concessivo: riflessioni su un caso (dubbio) di de/grammaticalizzazione. *Linguistica e Filologia* 5, 7–41.
- [2] Bertinetto, Pier Marco (1991) Il verbo. In Renzi, Lorenzo, Salvi, Gianpaolo (eds), *Grande grammatica italiana di consultazione*, vol II, pp. 13–161. Bologna: il Mulino.

- [3] Borillo, Andrée (1976) Les adverbes et la modélisation de l'assertion. *Langue française* 30, 74–89.
- [4] Dendale, Patrick (2001) Le futur conjectural versus *devoir* épistémique : différences de valeur et restrictions d'emploi. *Le Français Moderne* 69, 1, 1–20.
- [5] Diwersy, Sascha (2014) La plateforme Primestat.BTLC et l'exploitation lexico-statistique de corpus diachroniques. Lecture given at the Journée d'études à l'Institut des Sciences du Langage et de la Communication, University of Neuchâtel.
- [6] Hopper, Paul and Traugott, Elisabeth C. (1993) *Grammaticalization*. Cambridge, UK: Cambridge University Press.
- [7] Legallois, Dominique (2012) La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ? *Corpus* [Online] 11, <http://corpus.revues.org/2202>.
- [8] Ricci, Claudia, Rossari, Corinne and Siminiciuc, Elena (in press) La représentation des sens modaux dans trois langues romanes : le français, l'italien et le roumain. Du qualitatif au quantitatif et retour. *Syntaxe et sémantique* 17.
- [9] Rossari, Corinne (2014) How does a concessive value emerge? In Ghezzi, Chiara and Molinelli, Piera (eds.) *Pragmatic markers from Latin to Romance languages*. Studies in diachronic and historical linguistics. Oxford: Oxford University Press.
- [10] Rossari, Corinne (2015) Une concession implique-t-elle une opposition ? In Ferrari, Angela, Lala, Letizia and Stojmenova, Roska (eds.), *Testualità. Fondamenti, unità, relazioni*, pp. 189–203. Firenze: Franco Cesati.
- [11] Rossari, Corinne (in press) La présupposition dans les structures concessives. In Biglari, Amir and Bonhomme, Marc (eds.) *La Présupposition entre théorisation et mise en discours*. Paris: Classiques Garnier.
- [12] Rossari, Corinne, Ricci, Claudia and Siminiciuc, Elena (in press) Les valeurs du futur modal en français, italien et roumain. In Baranzini, Laura, Sanchez-Mendez, Juan Pedro and De Saussure, Louis (eds.) *Le futur dans les langues romanes*. Bern: Peter Lang.
- [13] Traugott, Elizabeth C. (in press) On the function of the epistemic adverbs surely and no doubt at the left and right periphery of the clause. In Beeching, Kate and Detges, Ulrich (eds.), *The role of left and right periphery in semantic change*. Amsterdam: Benjamins.