Studii de lingvistică 6, 2016, 127 - 142

Les fluctuations de sens dans quelques formes modales à la lumière d'une approche quantitative et qualitative

A quantitative and qualitative approach to the study of the fluctuations in meaning of certain modal expressions

Corinne Rossari[1]

Abstract: The expressions discussed in this article – *certes*, *peut-être* and *sans doute* – share the property of conveying a modal meaning (to be understood as non-denotational), which oscillates between an epistemic and a more or less conventionalized concessive value. To take this type of variation in meaning into account, the author on the one hand combines a qualitative and quantitative approach and on the other uses a theoretical framework, called the *triadic model*, which is based on the representation of the semantic contribution of the marker on three different levels: the *semantic*, the *enunciative* and the *rhetorica*l level.

Key words: modal expressions, quantitative method, collocation, statistics, corpus.

0. Introduction[2]

Le but de cet article est de présenter une méthode d'analyse alliant étroitement une démarche qualitative et quantitative et un cadre théorique pour l'étude de formes modales. Les formes que nous prendrons en considération – *certes*, *peut-être* et *sans doute* – ont la propriété de véhiculer un sens modal (à entendre comme non dénotationnel), qui oscille entre une valeur *épistémique* et une valeur *concessive* plus ou moins stabilisée.

L'analyse quantitative est menée au moyen de calculs statistiques de fréquences tirés de Evert (2008), qui servent à évaluer

[1] Université de Neuchâtel; corinne.rossari@unine.ch.

[2] Cet article s'inscrit dans le cadre d'un projet financé par le Fonds national suisse de la recherche scientifique intitulé « La représentation du sens modal et de ses tendances évolutives dans deux langues romanes : le français et l'italien », no. 100012_159458, auquel collaborent Annalena Hütsch, Claudia Ricci et Dennis Wandel. Une des étapes de ce projet est présentée dans Rossari, Hütsch, Ricci, Salsmann, Wandel (2016). Cet article est un développement de cette étape.

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

le degré de spécificité de collocations, à savoir des mots avec lesquels une forme donnée apparaît le plus souvent. Pour extraire nos données et pour calculer automatiquement certaines de ces fréquences, nous utilisons la plateforme BTLC[3] à partir de laquelle il est possible de faire des recherches automatiques sur les deux types de corpus que nous avons pris en compte, à savoir un corpus synchronique et un corpus diachronique. Le premier est constitué par une année du journal *Le Monde* (*Le Monde* 2008) qui correspond à 20 410 776 mots. Le corpus diachronique représente quatre tranches du français : le 16ème, le 17ème, le 18ème et le 19ème siècles et réunit en tout 25 881 338 mots. Il a été établi dans le cadre du projet PRESTO coordonné par Peter Blumenthal et Denis Vigier. Ce corpus rassemble sur les quatre siècles des textes semblables en genre et en nombre de mots.

Notre article est organisé ainsi : dans un premier temps (§1 et §2), nous présentons le cadre théorique que nous avons forgé, dit *modèle triadique*, qui nous permet de rendre compte dans une perspective qualitative des fluctuations des sens modaux inhérents à des adverbes comme ceux que nous avons sélectionnés, mais également à des formes temporelles comme le futur. Nous faisons allusion aux différents sens modaux de ce temps en français et en italien pour faire ressortir les convergences avec les sens modaux propres aux adverbes qui sont le focus de notre article. Dans une deuxième partie (§3 et §4), nous traitons spécifiquement de nos trois adverbes, que nous situons dans un paradigme, formé par *sûrement, certainement, assurément*, sélectionné au moyen d'un calcul (le *log-likelihood*). Nous avons utilisé ce calcul pour extraire, parmi les adverbes dont le sens est épistémique, ceux qui peuvent être considérés comme des co-occurrents spécifiques intervenant à la gauche de *mais*, marqueur qui est le plus fréquemment employé pour exprimer une concession[4]. Nous mesurons ainsi le lien entre la fréquence de l'adverbe à intervenir dans un contexte favorisant une interprétation concessive et sa propension à développer une telle valeur. Nous menons sur ce paradigme plusieurs calculs statistiques fondés sur notre corpus diachronique afin de voir si la valeur concessive dont ces adverbes peuvent se charger est sujette à évolution. Enfin, dans une troisième partie (§5 et §6), nous tirons un bilan de ces calculs statistiques en les mettant en perspective avec ce qui est ressorti de l'étude qualitative préalable (§ 1 et §2) portant sur les fluctuations des sens modaux.

---

[3] Cette plateforme et ses fonctionnalités sont présentées dans Diwersy (2014).

[4] Ci-après les chiffres des fréquences relatives (par mio de mots) des formes *mais, pourtant, toutefois, cependant, néanmoins* sur quatre corpus : *Le Monde* 2007 (LM7), *Le Monde* 2008 (LM8), *Le Figaro* 2007 (LF7), *Le Figaro* 2008 (LF8):

Mais	LM7 : 2380 ;	LM8 : 2525 ;	LF7 : 2521 ;	LF8 2644
Pourtant	LM7 : 240 ;	LM8 : 251 ;	LF7 : 252 ;	LF8 253
Toutefois	LM7 : 121 ;	LM8 : 118 ;	LF7 : 116 ;	LF8 123
Cependant	LM7 : 112 ;	LM8 : 109 ;	LF7 : 94 ;	LF8 124
Néanmoins	LM7 : 50 ;	LM8 : 47 ;	LF7 : 59 ;	LF8 53

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

1. La notion de sens modal

Pour définir la notion de sens modal, nous nous sommes fondée sur un modèle prenant en considération trois niveaux d'analyse, dit *modèle triadique*. Ce modèle est présenté dans Rossari, Ricci et Siminiciuc (à paraître) et Ricci, Rossari, Siminiciuc (2016) pour la description des emplois du futur en français, italien et roumain. Il répond aux deux idées clés suivantes : (i) les formes véhiculant des sens modaux ne sont pas intrinsèquement polysémiques ; (ii) les sens modaux ne sont pas pragmatiques, mais sont véhiculés de manière semi-conventionnelle.

(i) Par exemple, le futur, qui peut véhiculer un sens modal de nature épistémique ou concessive (ce dernier sens n'existe qu'en italien et en roumain) en plus de son sens temporel – voir Ricci (à paraître) pour une étude détaillée de ces différents sens –, n'est pas considéré comme une forme polysémique. Il est porteur d'une instruction de futurité dans tous ses emplois, mais cette instruction ne s'opère pas au même niveau d'analyse. Elle porte sur le prédicat quand le futur est interprété comme temporel stricto sensu, elle porte sur l'énonciation quand le futur est interprété comme épistémique ou concessif. La distinction entre les différentes valeurs modales se fait au niveau rhétorique. Ce dernier niveau entre en jeu pour la spécification des circonstances dans lesquelles une énonciation est présentée comme « futurale » ou différée. Par exemple, si ces circonstances sont que le locuteur n'a pas de connaissance concernant la réalisation d'un événement, le futur est interprété comme véhiculant une valeur épistémique, car l'indication de reporter une énonciation à plus tard est une façon pour le locuteur d'éviter d'endosser complètement le contenu de cette énonciation. Si les circonstances montrent que le locuteur a une pleine connaissance de l'état de choses qu'il énonce et que ce dernier ne peut pas être interprété comme ayant lieu au futur, comme c'est le cas pour les propriétés inaliénables, le futur peut être interprété comme concessif, quand l'indication de différer une énonciation est comprise comme une façon pour le locuteur de montrer le caractère « non pertinent » de cette énonciation.

(ii) Le fonctionnement énonciatif d'une forme est conventionnel, à savoir qu'il s'agit d'un trait systématique de cette forme. Certaines formes sont dotées de ce fonctionnement alors que d'autres non. En revanche, les circonstances dans lesquelles ce fonctionnement énonciatif est interprété intègrent les données contextuelles, comme par exemple le fait que l'état

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

> de choses décrit est interprété comme ayant déjà eu lieu ou non, ou comme évident ou non. Dans le premier cas le fonctionnement énonciatif du futur est interprété comme épistémique, et dans le second comme concessif. C'est la raison pour laquelle nous parlons de sens *semi-conventionnel*.

Il y a donc trois niveaux d'analyse pour une forme dotée d'une indication modale :

Le niveau sémantique : les indications données par la forme portent sur le contenu. Par exemple, le futur temporel (ex. 1) ou *peut-être* employé avec son sens de probabilité (ex. 2) portent sur la réalisation de l'état de choses évoqué dans ce contenu, en l'occurrence « être absent » :

(1) Demain il **sera** absent.

(2) Elle est **peut-être** absente.

Le niveau énonciatif : les indications données par la forme portent sur l'énonciation même qui résulte de l'apparition de l'énoncé. Par exemple, le futur italien peut signaler que l'énonciation d'un contenu est reportée à un moment ultérieur ; *peut-être* peut signaler que cette énonciation est potentielle (ou en d'autres termes, qu'il n'est pas sûr qu'elle ait lieu).

(3) **Sarà** una donna, ma non le piace truccarsi.

(4) C'est **peut-être** une femme, mais elle n'aime pas se maquiller.

Le niveau rhétorique : quand la forme porte sur l'énonciation, les indications qu'elle véhicule sont actualisées dans certaines circonstances, dont la détermination donne à cette dernière une valeur particulière. Dans les deux exemples précédents, l'énonciation est reportée à un moment ultérieur au moyen du futur ou présentée comme potentielle au moyen de *peut-être,* car elle s'inscrit dans des circonstances qui rendent son actualisation *hic et nunc* non pertinente (vu le caractère inaliénable de la propriété « être une femme »). Dans de telles circonstances, tant le futur italien que *peut-être* en français prennent une valeur concessive.

Au moyen de ce modèle, on rend compte du fait que des items différents véhiculent une même valeur modale (*peut-être* en français, *forse* en italien ainsi que le futur en italien et en roumain, toutes ces formes véhiculent une valeur concessive) et du fait que des valeurs modales qui semblent se rejoindre (comme la valeur épistémique du futur et la valeur de probabilité de *peut-être*) sont en fait le résultat de

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

modes d'activation bien distincts : *peut-être* agit au niveau sémantique en qualifiant la prédication alors que le futur à valeur modale épistémique agit au niveau énonciatif en qualifiant l'énonciation. Cette qualification intervient dans des circonstances où le locuteur n'a pas d'information sur la réalisation d'un événement qui s'est déjà produit ou qui est en train de se produire.

Le terme « sens modal » peut donc désigner des valeurs sémantiques semblables mais qui sont le fruit de modes de fonctionnement bien différents, comme la valeur épistémique du futur (qui résulte d'un mode de fonctionnement au niveau énonciatif) et la valeur de probabilité (dite également valeur épistémique) de *peut-être*, qui résulte d'un mode de fonctionnement au niveau sémantique.

2. Fluctuations des sens modaux

Les sens modaux fluctuent donc tant au niveau des formes, puisque le même sens peut être véhiculé par deux formes différentes, qu'au niveau des valeurs[5] qu'une même forme peut prendre, puisque celle-ci peut être vectrice de plusieurs valeurs modales.

Au niveau des formes : les mêmes sens peuvent se greffer sur des formes différentes d'une langue à l'autre, dans une même langue ou d'une époque à l'autre. Par exemple, pour ce dernier cas, la valeur concessive est rendue par *peut-être* en français contemporain. *Certes* a actuellement une valeur concessive, mais cet adverbe a eu un sens épistémique semblable à celui de *certainement* à un état antérieur, comme l'atteste l'exemple suivant :

(5) Si le Tiers État sait se connaître et se respecter, **certes**, les autres le respecteront aussi. (Sieyès, *Qu'est-ce que le Tiers état ?*, 1789, p. 48)

Ce sens est maintenu jusqu'au début du 20ème siècle mais devient actuellement de plus en plus rare.

(6) Il passe la main sur mon front. Il aperçoit ces deux enfants, raidis dans leur certitude... Ah, non, **certes**, ce n'est pas là qu'elle est, la vérité ! (Martin du Gard, *Jean Barois*, 1913, p. 515)

Au niveau des valeurs : une même forme peut prendre, au gré des circonstances dans lesquelles elle est interprétée, différentes

[5] Nous utilisons l'étiquette « sens » pour parler de façon générique des sens modaux (épistémique, concessif...) sans qu'ils soient attachés à une forme particulière et l'étiquette « valeur » pour parler de ces sens quand ils sont associés à une forme particulière (on parlera de la valeur épistémique ou concessive de *peut-être*).

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

valeurs que l'on identifie au niveau rhétorique du modèle triadique. Par exemple, une même forme (le futur) peut prendre une valeur concessive, épistémique, d'atténuation selon les langues : le futur français n'a pas de valeur concessive, mais a une valeur d'atténuation, comme relevé dans Rossari, Ricci, Siminiciuc (à paraître) :

(7) Vous **prendrez** bien encore un peu de vin ?

3. Les mécanismes responsables de ces deux types de fluctuation

Quels sont les mécanismes qui s'avèrent responsables de ces fluctuations ? Nous allons examiner le cas du sens concessif transmis par des adverbes qui véhiculent également un sens épistémique. Un adverbe comme *peut-être* a une valeur épistémique quand il agit au niveau sémantique et il prend une valeur concessive quand il agit au niveau de l'énonciation. Comme on l'a vu, ces deux valeurs ne sont pas le résultat d'une polysémie, mais celui d'une même instruction, celle d'indiquer la potentialité, instruction qui opère au niveau du prédicat, cas où l'interprétation de l'adverbe est épistémique, ou au niveau de l'acte d'énonciation en signalant que ce dernier n'est que potentiel, cas où l'interprétation de l'adverbe est concessive. A la lumière de ce cas, la question initiale de ce paragraphe peut être reformulée par la série d'interrogations suivante :

- Pourquoi certains adverbes modaux à valeur épistémique se mettent-ils à fonctionner au niveau énonciatif en prenant une valeur concessive ?
- En fonction de quels facteurs cette valeur se conventionnalise-t-elle, à savoir devient un composant conventionnel de la sémantique de la forme ?
- Pourquoi le sens concessif s'est-il greffé sur le futur en italien et sur *certes* ou *peut-être* en français ?

Plus concrètement, on relève que :

- *Probablement* n'a pas de sens concessif, *peut-être* en a un;
- *Certes* n'a plus qu'un sens concessif, *certainement* n'en a pas;
- *Sans doute* a un sens concessif, *assurément* n'en a pas.

3.1. Hypothèse de départ: le facteur de la fréquence dans l'environnement gauche de *mais*

Pour identifier un des facteurs qui joue un rôle sur l'acquisition par une forme d'une valeur ou de plusieurs valeurs modales, nous avons pris en compte le facteur de la fréquence, en nous fondant

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

sur les études de Bybee (2003) qui attribuent à la fréquence un rôle déterminant dans le processus de changement linguistique conduisant une forme à se grammaticaliser : "[…] It appears that one avenue by which *cunnan* begins to grammaticize as an auxiliary is determined by the fact that it was already frequent, and had already undergone some weakening of its semantic content" (Bybee 2003: 609).

La valeur concessive dont certains adverbes épistémiques sont vecteurs peut être considérée comme issue d'un processus de grammaticalisation fondé sur la tendance dite de « subjectivisation ». Pour rappel, la subjectivation/ subjectivisation est définie ainsi : « […] des signifiés fondés sur une situation externe ou interne évoluent vers des signifiés textuels : par exemple, les déictiques peuvent être employés comme "déictiques renvoyant au discours" » (Himmelmann 1996, *apud* Marchello-Nizia 2006 : 28).

La valeur concessive d'un adverbe peut être considérée comme résultant de ce processus dans la mesure où l'adverbe prend une signification textuelle en mettant en relation deux séquences : l'une qui correspond à l'état de choses concédé (celle qui précède *mais*) et l'autre à celle que le locuteur revendique (qui suit *mais*).

L'hypothèse que nous avons testée est de voir si la fréquence d'un adverbe dans l'environnement gauche de *mais* est un facteur qui favorise le développement d'un fonctionnement énonciatif conduisant à l'acquisition d'une valeur concessive. Pour tester cette hypothèse, nous sommes partie de la définition de la séquence concessive proposée dans Rossari (2016). Ce type de séquence est caractérisé par un contraste dans la prise en charge des deux énoncés qui forment la construction concessive. Le locuteur présente la prise en charge de l'énoncé qui correspond à ce qui est concédé comme étant en retrait par rapport à celle de l'énoncé qui suit. La prise en charge de ce dernier est présentée comme étant à l'avant plan discursif. Dans les séquences de type « X, mais Y », le *mais* opère ce contraste de prise en charge indiquant que l'énonciation de X est en retrait alors que celle de Y est à l'avant plan. Cette différence dans la prise en charge des deux énoncés induit un fonctionnement de l'adverbe épistémique au niveau énonciatif : ce dernier, intervenant dans le segment X, ne prend plus pour objet le prédicat de X, mais son énonciation, en favorisant une interprétation selon laquelle cette énonciation est présentée comme en retrait. L'instruction de « potentialité » propre à l'adverbe se reporte sur l'acte d'énonciation qui devient lui-même potentiel. Le locuteur se présente ainsi non pas comme disant X, mais comme susceptible de dire X, ce qui a pour effet direct de mettre en retrait l'énonciation de X.

En synthèse, plus un adverbe apparaît fréquemment (hypothèse de la fréquence propre à la grammaticalisation) et/ou plus il apparaît fréquemment dans l'environnement gauche de *mais*, plus il a de chances de développer une valeur concessive.

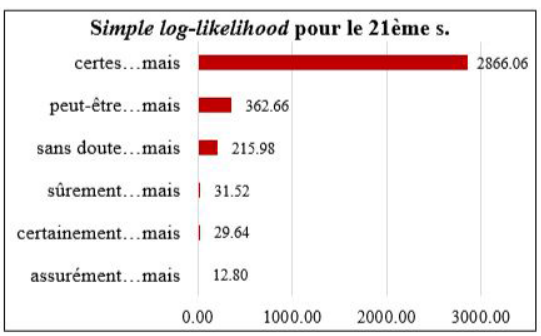
BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

3.2. Outils de mesures statistiques pour calculer la spécificité de la collocation adverbe épistémique + *mais*

Pour identifier les adverbes épistémiques (dorénavant adv.e) qui sont des co-occurrents spécifiques de *mais* dits *collocatifs*, nous avons utilisé deux calculs statistiques : l'un qui saisit le degré de non-indépendance du collocatif et du pivot (le mot auquel s'applique le collocatif, en l'occurrence *mais*). Ce calcul établit un seuil au-delà duquel le fait que ces deux mots interviennent ensemble ne peut pas être juste le fruit du hasard. L'autre calcul saisit le degré d'attraction entre le collocatif et *mais*. Il permet de comparer la force d'attraction entre plusieurs collocatifs avec le même pivot. Les gradations qui ressortent des diagrammes peuvent ainsi être interprétées comme différents degrés de force d'attraction. Le premier est désigné par la valeur *log-likelihood* ou *simple log-likelihood* et le second par la valeur MI (cf. Evert 2008).

Le premier calcul permet de sélectionner les adverbes à potentiel concessif. En effet, parmi les adverbes épistémiques, certains peuvent se retrouver devancer d'autres dans l'environnement gauche de *mais*, mais cela ne les rend pas des collocatifs spécifiques de la conjonction. Etant plus fréquents, il est attendu qu'ils soient également plus fréquemment associés à *mais*. Le calcul de la valeur *log-likelihood* (ou *simple ll*) permet d'éviter ce biais. Comme indiqué dans Rossari *et al.* (2016) cette valeur permet de faire ressortir une liste d'adverbes pour lesquels « la probabilité d'indépendance par rapport à *mais* est extrêmement faible ».

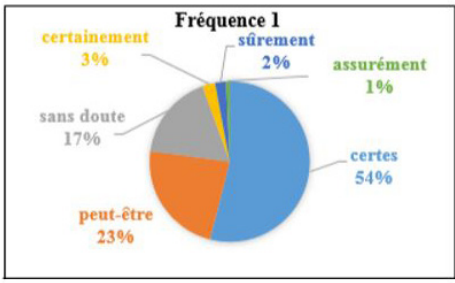
Parmi les adverbes que le calcul retient, nous avons sélectionné tous ceux qui répondent à une définition étroite d'adverbe épistémique, selon laquelle ils portent sur le dit en exprimant un jugement signifiant le degré d'adhésion du locuteur par rapport à la réalité de l'état de choses énoncé. Ce jugement doit pouvoir être paraphrasé par la tournure *Il est* adj. *que*. Ainsi nous avons retenu *certainement*, paraphrasable par *Il est certain que* et non *vraiment* qui n'est pas paraphrasable par *Il est vrai que*. Nous avons ainsi établi une liste de six adverbes (voir tableau issu de Rossari *et al.* 2016) : *certes, peut-être, sans doute, sûrement, certainement, assurément.*



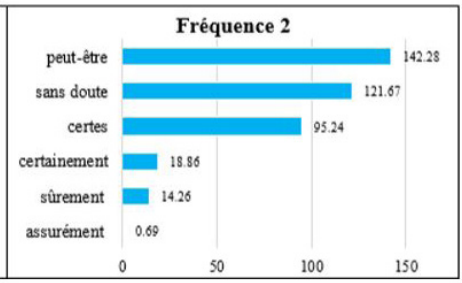
BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

De ces adv.e, nous avons testé la fréquence avec *mais* (fréquence 1) ainsi que celle de leur emploi indépendamment de *mais* (fréquence 2) pour mettre à l'épreuve notre hypothèse de départ (l'influence de *mais* sur la valeur concessive de ces adverbes).

Nous reprenons ici les résultats mentionnés dans Rossari *et al.* (2016) :



Pourcentage de la séquence <épistémique - *mais*> sur le nombre total de *mais*



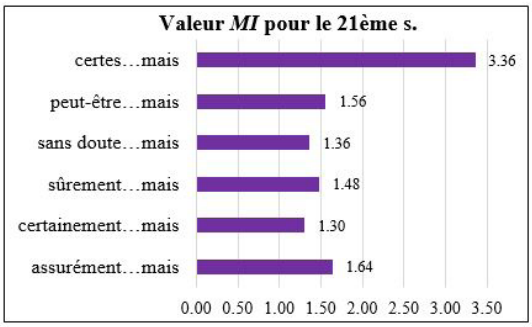
Fréquence relative des épistémiques dans l'ensemble du corpus *Le Monde* 2008

« Les fréquences 1 et 2 font ressortir trois adverbes en tête de classement : *certes*, *peut-être* et *sans doute*, avec une différence notable : *certes* est l'adverbe qui est le plus souvent utilisé avec *mais*, alors qu'il arrive après *peut-être* et *sans doute* en ce qui concerne sa fréquence relative » (Rossari *et al.* 2016 : 2).

Pour vérifier l'attraction effective de *certes* et de *mais*, nous avons employé la valeur MI qui calcule la force d'attraction du pivot et de son collocatif. En effet, la valeur *simple ll* nous garantit que l'association entre le collocatif et le pivot n'est pas aléatoire ; la fréquence 1 permet de classer les différents collocatifs en fonction du nombre d'occurrences de *mais* accompagné du collocatif ; la fréquence 2 permet de voir que le classement issu de la fréquence 1 n'est pas juste le reflet de la fréquence relative de chaque épistémique (par exemple, elle fait ressortir *certes* comme étant un collocatif plus spécifique de *mais* que *peut-être* : il est moins fréquent que *peut-être*, mais plus fréquemment associé à *mais* que ce dernier) ; enfin la valeur MI (qui teste la force d'attraction entre deux items) permet de vérifier que *certes* est bien plus spécifique de *mais* que les autres adverbes.

Cette valeur confirme le lien entre *certes* et *mais*. Nous reprenons ci-dessous le tableau de Rossari *et al.* (2016) :

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)



3.3. Convergence entre qualitatif et quantitatif

Les fréquences 1 et 2 font donc ressortir *certes* comme particulièrement spécifique de *mais* : il est moins fréquent dans l'absolu alors qu'il est plus fréquemment utilisé avec *mais*. La valeur MI confirme cette tendance. Cette divergence de fréquences pourrait être un indice quantitatif de la conventionnalisation de la valeur concessive. *Certes* a perdu son sens épistémique en français contemporain, ce qui se remarque par le fait qu'il est moins fréquemment employé que *peut-être* et *sans doute* dans l'absolu. En revanche, la plus grande fréquence de son emploi avec *mais* serait liée au fait qu'il s'est spécialisé dans le marquage d'une séquence concessive.

Une autre particularité que fait ressortir l'étude quantitative est la très forte fréquence de *peut-être* et *sans doute*. Cette fréquence élevée va de pair pour *peut-être* avec la présence d'une valeur concessive en plus de sa valeur épistémique, comme l'a fait ressortir l'analyse qualitative présentée sous le paragraphe 2. Il en va de même pour *sans doute*.

> (8) Le débat est sans doute [certes/peut-être] français, mais il intéresse tout le monde. (Tertrais, H., « Introduction », in Duanmu, M. et Tertrais, H. (dir.), *Temps croisés* I, Maison des Sciences de l'Homme, 2010, p. XXII)

Dans cet emploi, *sans doute*, comme *peut-être* ou *certes*, perd sa valeur épistémique. Il ne véhicule plus qu'une valeur concessive. Son usage est paraphrasable par :

il n'est pas pertinent de dire p (*le débat est français*)

et non par :

il ne fait pas de doute que p (*que le débat est français*).

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

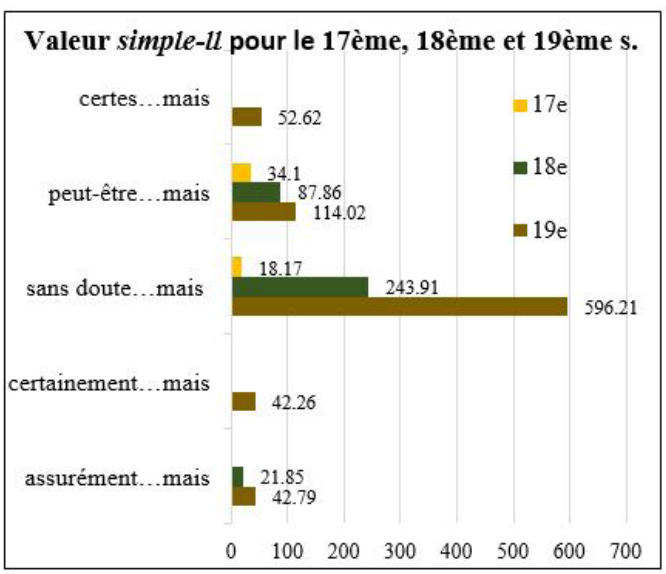
Dans cette construction, ce n'est pas l'état de choses exprimé dans l'énoncé qui est évalué par l'adverbe, mais l'énonciation même. L'acte d'énonciation dans lequel intervient l'adverbe en ressort ainsi affaibli au profit de celui qui est introduit par *mais* et qui porte le poids de l'argumentation.

Il ressort du calcul de ces différentes fréquences (fréquences 1, 2 et valeur MI) que (i) la fréquence 2 a une incidence sur la valeur concessive : ce sont les trois adverbes les plus fréquents qui ont la valeur concessive la plus développée (*certes*, *peut-être* et *sans doute*) ; (ii) la valeur concessive qui se conventionnalise fait apparaître une divergence entre la fréquence 1 et la fréquence 2, divergence qui peut être le signe d'une force d'attraction plus grande entre le collocatif et le pivot. Cela est confirmé par la valeur MI pour *certes*, qui est l'adverbe dont la valeur concessive est la plus aboutie.

L'étude quantitative en diachronie permettra de voir si l'existence constatée de la valeur concessive en synchronie pour *certes*, *peut-être* et *sans doute* correspond à une courbe évolutive ascendante concernant la fréquence de ces adverbes dans l'absolu ainsi que celle dans l'environnement de *mais*.

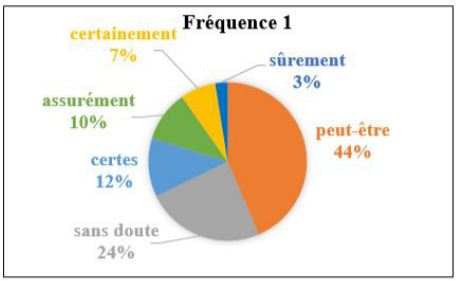
4. Étude quantitative en diachronie

Nous avons procédé comme en synchronie, en établissant la liste des adverbes épistémiques pouvant être considérés comme collocatifs de *mais*, au moyen de la valeur *simple ll*. Ci-dessous le tableau repris dans Rossari *et al.* (2016) :

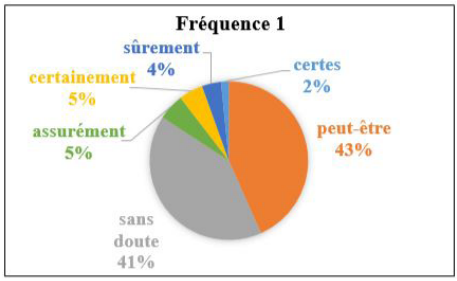


BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

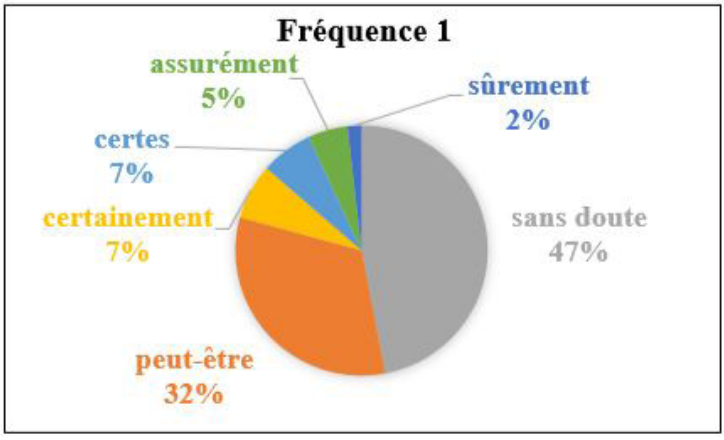
« Parmi les six adverbes que la valeur *simple-ll* nous a permis de retenir en synchronie, seuls deux se retrouvent pendant les trois siècles comme des co-occurrents pouvant être considérés comme spécifiques de *mais* : *sans doute* et *peut-être*. *Sûrement* ne ressort pour aucun de ces trois siècles comme spécifique de *mais*, raison pour laquelle il n'apparaît pas dans ce tableau. *Certes* n'apparaît qu'au 19ème siècle. Aucun autre adverbe non pris en compte en synchronie n'apparaît comme spécifique de *mais* » (Rossari *et al.* 2016). *Assurément* apparaît comme spécifique de *mais* aux 18ème et 19ème siècles et *certainement*, comme *certes*, au 19ème siècle seulement. Nous allons appliquer les calculs concernant les fréquences 1 et 2 sur les six adverbes (indépendamment de ce qui ressort du calcul *simple-ll*) afin de pouvoir faire état d'une possible évolution dans la fréquence des six adverbes retenus en synchronie.



Pourcentage de la séquence < épistémique - *mais*> sur le nombre total de *mais* au 17ème siècle

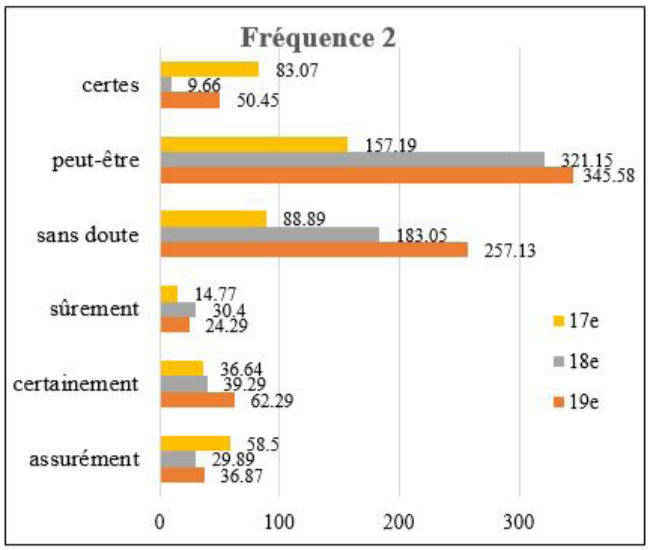


Pourcentage de la séquence < épistémique - *mais*> sur le nombre total de *mais* au 18ème siècle

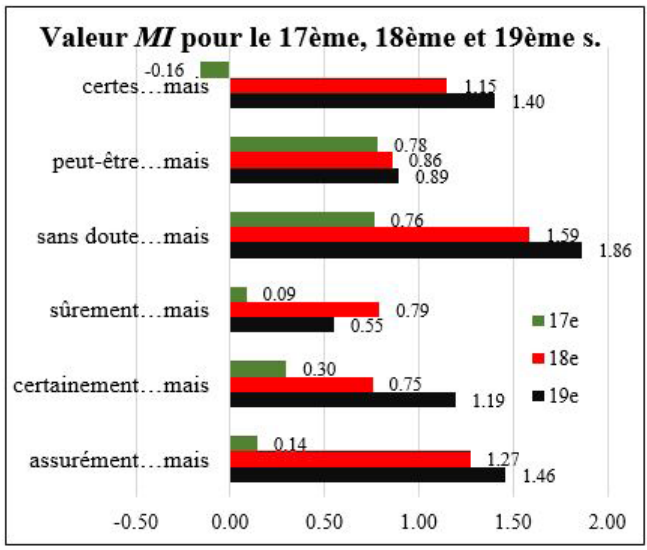


Pourcentage de la séquence <épistémique – *mais*> sur le nombre total de *mais* au 19ème siècle

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)



Fréquence relative des épistémiques
dans l'ensemble des corpus du 17ème, 18ème et 19ème siècles



Les trois types de fréquences font apparaître des résultats tout à fait inattendus par rapport à ceux qui ressortent de la synchronie. Le pourcentage de *mais* associé à *certes* chute drastiquement au 18ème siècle avant de remonter au 19ème siècle, mais ce pourcentage reste sensiblement inférieur à celui atteint au 17ème siècle. Cette chute concernant la fréquence de *certes* avec *mais* est aussi constatée pour sa

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

fréquence relative (entre le 17ème et le 18ème siècle). Les chiffres que l'on trouve sur la base Frantext concernant la fréquence relative de *certes* sont en harmonie avec ces données : on trouve 1668 occurrences de *certes* au 17ème, 343 au 18ème et 4131 au 19ème siècle. Ce qui reste stable est le fait que *peut-être* et *sans doute* sont parmi les co-occurrents les plus fréquents de *mais* pendant les trois siècles et ce aussi en synchronie.

Le statut de *certes* en tant que collocatif spécifique de *mais* est donc récent : aucun des deux calculs (valeur MI ou fréquence 1) ne le fait ressortir comme spécifique sur ces trois siècles. En revanche, les fréquences 1, 2 et la valeur MI mettent en évidence plusieurs particularités pour *peut-être* et *sans doute*. D'une part, *peut-être* est l'adverbe le plus fréquent sur les trois siècles, suivi dans des proportions semblables par *sans doute* (cf. fréquence 2). D'autre part, la fréquence 1 fait apparaître une variation sensible entre le 18ème et le 19ème siècle : le pourcentage de *mais* avec *peut-être* est semblable au 17ème et au 18ème siècle (env. 40%), mais il baisse sensiblement au profit de *sans doute* au 19ème siècle (32%). La valeur MI ne rend pas compte de cette baisse (elle est stable sur les trois siècles). L'augmentation du pourcentage de *mais* accompagné de *sans doute* n'est pas répercutée sur la fréquence relative de *sans doute* : l'adverbe reste derrière *peut-être*. La valeur MI en revanche fait apparaître une hausse à partir du 18ème siècle dans la force d'attraction entre *mais* et *sans doute*.

Selon ces trois calculs de fréquence, *peut-être* apparaît comme l'adverbe le plus stable sur ces trois siècles : il est l'adverbe le plus fréquent sur les trois siècles, la force d'attraction avec *mais* (valeur MI) ne varie presque pas d'un siècle à l'autre. La seule variation concerne son association avec *mais* qui baisse au profit de *sans doute* au 19ème siècle. Les fréquences associées à *sans doute* font état d'une variation au 19ème siècle : la fréquence 1 et la valeur MI montrent qu'il semble devenir plus spécifique de *mais* qu'il ne l'était auparavant.

5. Interprétation des données quantitatives en diachronie à la lumière de l'analyse qualitative

Ces données quantitatives semblent a priori ne pas cadrer avec l'analyse qualitative faite en synchronie, qui fait ressortir que les trois adverbes ont une valeur concessive et que *certes* est celui pour lequel cette valeur est la plus aboutie. En effet, aucune fréquence ne fait apparaître *certes* comme spécifique de *mais* sur les trois siècles pris en compte, aucune fréquence ne fait apparaître une évolution ascendante en ce qui concerne la spécificité de *peut-être* en tant que collocatif de *mais*. Les fréquences associées à *sans doute*, dont la valeur concessive n'est pas plus conventionnalisée que celle de *peut-être*, font en revanche ressortir une évolution ascendante à partir du 19ème siècle. Il ressort donc de ces calculs de fréquence que la valeur

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

concessive de ces adverbes ne s'est pas conventionnalisée de façon progressive et que ce processus n'a pas eu lieu aux mêmes périodes. Bien que la valeur concessive de *sans doute* soit moins aboutie que celle de *certes*, il semble que le processus de conventionnalisation ait démarré pour *sans doute* avant *certes* au 19ème siècle, puisque c'est à ce moment que la valeur MI, la fréquence 1 (et la fréquence 2 à cause du fait qu'il est toujours derrière *peut-être*) le font apparaître comme spécifique de *mais*. Pour *certes*, les trois calculs de fréquences font apparaître qu'il ne peut être considéré comme spécifique de *mais* avant la fin du 19ème siècle. Enfin, les calculs de fréquence concernant *peut-être* ne montrent pas de variation sensible. Celle relevée au 19ème siècle, qui montre une légère baisse pour la fréquence 1, peut être interprétée comme un effet de l'augmentation de la spécificité de *sans doute* comme collocatif de *mais* et non comme une tendance de *peut-être* à être moins utilisé avec *mais*, dans la mesure où la valeur MI ne fait pas état d'une régression concernant la force d'attraction entre *peut-être* et *mais*. Cela peut être le signe que le processus de conventionnalisation est déjà stabilisé pour *peut-être* au 17ème siècle et qu'il n'a pas sensiblement évolué.

Il y a donc trois cas bien distincts :

- *Peut-être* dont le processus de conventionnalisation de la valeur concessive est stabilisé dès le 17ème siècle.
- *Sans doute* dont le processus de conventionnalisation se met en place au 19ème siècle et se stabilise ensuite.
- *Certes* dont le processus de conventionnalisation se met en place après le 19ème siècle et évolue vite en se renforçant. Jusqu'au 19ème siècle toutes les fréquences le font apparaître comme étant nettement derrière *peut-être* et *sans doute* quant à sa spécificité en tant que collocatif, alors qu'en synchronie, toutes les fréquences le font apparaître comme plus spécifique de *mais* que *peut-être* et *sans doute*.

6. Conclusions

Nos analyses quantitatives ont fait ressortir trois aspects par rapport à notre analyse qualitative, qui a mis en relief la valeur concessive plus ou moins conventionnalisée des adverbes *certes, peut-être* et *sans doute*, en synchronie :

(i) Le processus de conventionnalisation se fait sur des empans diachroniques différents : stabilisé dès le 17ème siècle pour *peut-être*, à partir du 19ème siècle pour *sans doute* et après le 19ème siècle pour *certes*.

(ii) Le degré de conventionnalisation n'est pas lié à l'empan sur lequel le processus de conventionnalisation s'étale : l'empan est bref pour *certes* alors que la valeur concessive de cet adverbe est plus aboutie que celle de *peut-être* et *sans doute*.

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)

(iii) Le processus de conventionnalisation n'est pas évolutif : il ne suit pas une courbe ascendante à partir du moment où il se met en place. Il s'amorce au 19ème siècle pour *sans doute* et après le 19ème siècle pour *certes*, mais cela ne conduit pas *sans doute* à avoir une valeur concessive plus aboutie que celle de *certes*.

Nos analyses quantitatives font apparaître deux convergences entre les données quantitatives et le degré de conventionnalisation de la valeur concessive identifiée suite à l'analyse qualitative :

(i) Les adverbes les plus fréquents dans le voisinage de *mais* sont également ceux qui ont la valeur concessive la plus marquée.

(ii) L'aboutissement de la conventionnalisation de la valeur concessive ferait apparaître un décalage entre la fréquence de l'adv.e en tant que co-occurrent de *mais* et sa fréquence relative indépendamment de *mais*.

Références bibliographiques

Bybee, J. (2003), « Mechanisms of Change in Grammaticization: The Role of Frequency », in Joseph, B. D., Janda, R. D. (éds), *The Handbook of Historical Linguistics*, Blackwell, Oxford, p. 602-623.

Diwersy, S. (2014), « La plateforme Primestat.BTLC et l'exploitation lexico-statistique de corpus diachroniques », communication présentée à la Journée d'études organisée à l'Institut des Sciences du Langage et de la Communication, Université de Neuchâtel.

Evert, S. (2008), « Corpora and collocations », in Lüdeling, A., Kytö, M. (éds), *Corpus Linguistics. An International Handbook*, Mouton de Gruyter, Berlin, p. 1212-1248.

Marchello-Nizia, C. (2006), *Grammaticalisation et changement linguistique*, Duculot de Boeck, Bruxelles.

Ricci, C. (à paraître), *Analyse contrastive français-italien de formes temporelles à valeur modale dans une perspective de linguistique outillée de corpus*, thèse de doctorat, Université de Neuchâtel.

Ricci C., Rossari, C. et Siminiciuc, E. (2016), « 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 (Entre linguistique et pragmatique)*, 17, p. 93-113.

Rossari, C. (2016) « La concession sans opposition à la lumière de la théorie argumentative de la polyphonie », *Verbum*, XXXVIII/1-2, p. 151-168.

Rossari, C., Hütsch, A., Ricci, C., Salsmann, M. et Wandel, D. (2016) « Le pouvoir attracteur de *mais* sur le paradigme des adverbes épistémiques : du quantitatif au qualitatif », *Actes des 13èmes Journées internationales d'Analyse statistique des Données Textuelles* (en ligne : https://jadt2016.sciencesconf.org/82999).

Rossari, C., Ricci, C. et Siminiciuc, E. (à paraître), « Les valeurs rhétoriques du futur en français, italien et roumain », in Baranzini, L., de Saussure, L. et Sánchez Méndez, J. P. (éds), *Le futur dans les langues romanes*, Peter Lang, Berne.

BDD-A25147 © 2016 Facultatea de Litere din Oradea
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 08:53:10 UTC)