

Le pouvoir attracteur de *mais* sur le paradigme des adverbes épistémiques : du quantitatif au qualitatif*

Corinne Rossari,
Annalena Hütsch, Claudia Ricci, Margot Salsmann, Dennis Wandel

Université de Neuchâtel – Suisse

Abstract

We propose to evaluate the influence of *mais* on a paradigm of epistemic adverbs chosen based on the calculation of the *log-likelihood* for all adverbs co-occurring in the left context of *mais* in a span of 20 items. This calculation has enabled us to select the following six adverbs: *certes*, *peut-être*, *sans doute*, *certainement*, *assurément*, *sûrement*. A qualitative approach shows that some of these adverbs feature a concessive value, which is more or less conventionalised, whereas others do not. By calculating the frequencies of these adverbs in the left context of *mais*, we have tested if the frequency of this conjunction next to these adverbs could be considered to play a decisive role in the conventionalisation of the concessive value of some among the latter.

Key words: modality; epistemic; concession; *mais*; quantitative analysis; annotated corpus

Résumé

Notre propos est d'évaluer la force d'attraction de *mais* sur un paradigme d'adverbes épistémiques identifiés au moyen du calcul de la valeur *log-likelihood* appliquée aux adverbes cooccurrents dans le contexte gauche de *mais* dans un empan maximum de 20 items. Ce calcul a permis de sélectionner les six adverbes suivants : *certes*, *peut-être*, *sans doute*, *certainement*, *assurément*, *sûrement*. Une approche qualitative fait ressortir que certains de ces adverbes ont une valeur concessive plus ou moins conventionnalisée alors que d'autres non. En calculant la fréquence de ces adverbes dans le voisinage gauche de *mais*, nous avons testé si la présence de cette conjonction dans l'environnement de ces adverbes joue un rôle sur la conventionnalisation de la valeur concessive de certains d'entre eux.

Mots clés : modalité ; épistémique ; concession ; *mais* ; analyse quantitative ; corpus annoté

1. Introduction

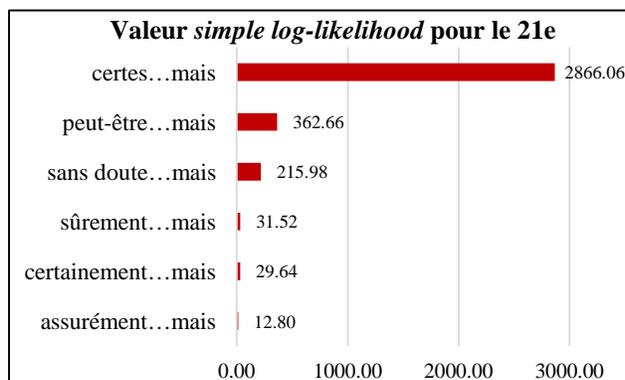
Notre propos est d'évaluer l'influence de *mais* sur les valeurs sémantiques d'un paradigme d'adverbes épistémiques¹. Les adverbes ont été identifiés en fonction de leur spécificité en tant que cooccurrent gauche de *mais* dans un empan maximum de 20 items (comprenant mots et signes de

* Cette recherche 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.

¹ Par adverbe épistémique, nous considérons uniquement ceux qui concernent le dit en exprimant un jugement quant au 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*.

punctuation). Cette spécificité est calculée au moyen de la valeur *simple log-likelihood*², qui fait ressortir une liste de six adverbess épistémiques pour lesquels la probabilité d'indépendance³ par rapport à *mais* est extrêmement faible.

Dans le cadre des études sur la grammaticalisation, qui considèrent le facteur de la fréquence comme préalable à une grammaticalisation (cf. entre autres Bybee 2003), notre propos est de vérifier si ce facteur entre en ligne de compte dans le fait que certains de ces épistémiques ont acquis une valeur concessive, considérant que cette valeur est le résultat d'un processus de subjectivation. Notre hypothèse est que la fréquence de l'adverbe en général, ainsi que sa fréquence dans l'environnement gauche de *mais*, jouent un rôle non négligeable sur son aptitude à véhiculer une valeur concessive.

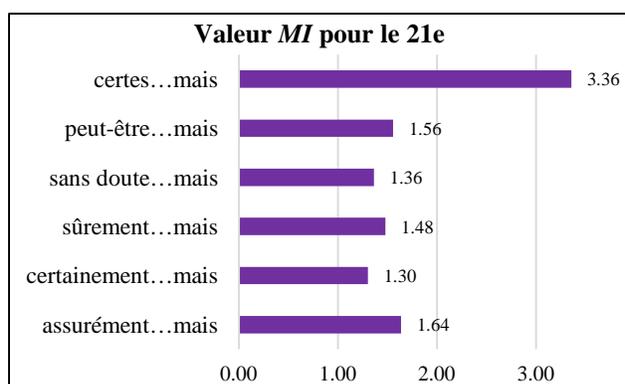


Pour vérifier cette hypothèse, nous avons évalué la fréquence de ces adverbess dans l'environnement gauche de *mais* (Fréquence 1), leur fréquence relative – indépendamment de l'emploi de *mais* – (Fréquence 2), ainsi que la valeur *MI*⁴ (valeur qui mesure la force d'attraction entre deux items, cf. Evert 2007) qui leur est attachée.

Ces calculs sont appliqués sur un corpus représentant une tranche synchronique du français contemporain (*Le Monde 2008*) et sur un corpus s'étalant sur une période de trois siècles (17^e au 19^e) permettant de donner une image de l'évolution diachronique possible quant à l'utilisation de ces adverbess en relation ou non avec *mais*. Ce dernier corpus a été établi dans le cadre du projet *Presto*⁵, qui réunit un échantillonnage homogène de textes en termes de genre et de taille.

2. Étude quantitative en synchronie

Les fréquences 1 et 2 font ressortir trois adverbess 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* est *sans doute* en ce qui concerne sa fréquence relative. La valeur *MI* confirme la force d'attraction entre les deux mots (*certes* et *mais*) : elle le met en tête de classement. Celle attachée aux autres adverbess ne fait pas apparaître de différence remarquable entre eux.



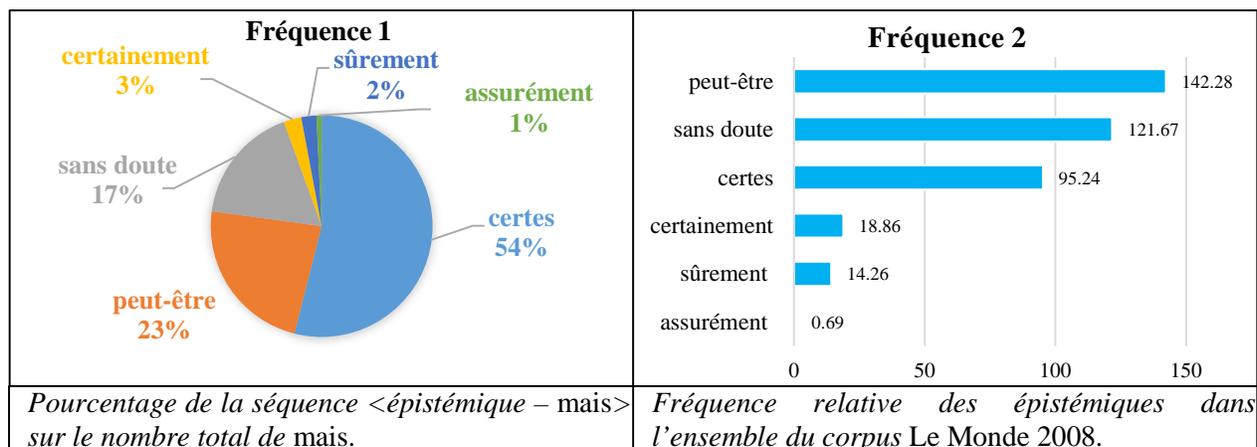
² Calculée au moyen de la formule suivante : $simple-ll = 2 \left(O \log \frac{O}{E} - (O - E) \right)$ (Evert 2007 : 22).

³ L'indépendance signifie que la distribution conjointe du pivot et du cooccurrent est proche du produit des distributions marginales (voir Evert 2007).

⁴ Calculée au moyen de la formule suivante : $MI = \log_2 \frac{O}{E}$.

⁵ Projet franco-allemand sur l'évolution diachronique des prépositions du français (*Presto* : <http://presto.ens-lyon.fr>).

Peut-être et *sans doute*, qui sont les cooccurrents les plus fréquents de *mais* après *certes*, ne sont pas associés à une valeur *MI* qui les particularise. Il reste à voir si la fréquence relative élevée de *peut-être* et de *sans doute* dans l'environnement gauche de *mais* (ils arrivent loin devant *certainement*, *sûrement* et *assurément*) est en phase avec une valeur concessive plus aboutie par rapport à celle des autres adverbess de notre paradigme.



3. Problématique issue des données quantitatives en relation avec une analyse qualitative

Ces données quantitatives, qui isolent *peut-être* et *sans doute* après *certes*, demandent à être mises en relation avec les propriétés distributives de ces adverbess (relevées dans Rossari 2015). Suivis par *mais*, *peut-être* et *sans doute* sont susceptibles d'intervenir dans des énoncés dans lesquels l'état de choses exprimé ne saurait être mis en doute. En revanche, *sûrement*, *certainement* et *assurément* ne partagent pas ce type d'emploi.

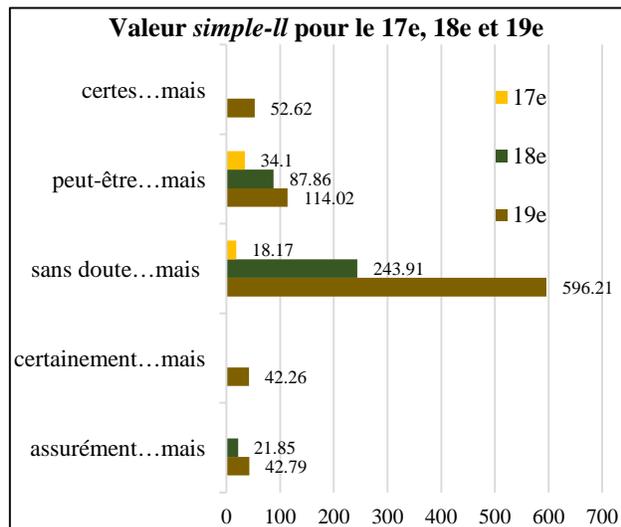
- (1) Je suis *peut-être* une femme, *mais* mes pouvoirs sont grands. Je n'hésiterai pas à vous y faire goûter si vous continuez à me poursuivre de vos assiduités.
(Hervé, J. (2013). *Le Monde d'Isaure*. Publibook, page 216)
- (2) Le débat est *sans doute* français, *mais* il intéresse tout le monde.
(Tertrais, H. (2010). Introduction. In Duanmu, M. et Tertrais, H. (éds), *Temps croisés I*. Maison des Sciences de l'Homme, page XXII)

Dans ce type d'emploi, la valeur épistémique que véhicule l'adverbe n'est plus au cœur de ce qui est communiqué par ce dernier. Il ne s'agit pas d'évaluer le degré de certitude de l'état de choses « être une femme » ou « être français ». Les trois autres adverbess de notre paradigme semblent continuer à évaluer le degré de certitude d'un état de choses et par conséquent peuvent produire un effet surprenant quand ce dernier ne se prête aucunement à une interprétation épistémique.

- (3) ??Je suis *sûrement*, *certainement*, *assurément* une femme, *mais* mes pouvoirs sont grands. Je n'hésiterai pas à vous y faire goûter si vous continuez à me poursuivre de vos assiduités.

Il s'en suit que *peut-être* et *sans doute* ont développé une forme de polysémie avec des emplois dans lesquels ils ne véhiculent pas un jugement épistémique. On considérera donc qu'ils ont acquis une valeur concessive en cours de conventionnalisation.

4. Étude quantitative en diachronie



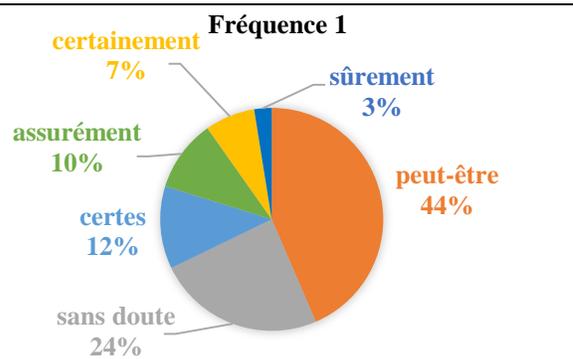
Nous allons maintenant examiner trois tranches diachroniques successives pour voir s'il y a évolution dans la fréquence de ces adverbes avec ou sans *mais* ainsi que dans leur degré d'attraction avec *mais*.

Nous commençons, comme en synchronie, par relever les adverbes pertinents au moyen de la valeur *simple log-likelihood*. 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 cooccurrents pouvant être considérés comme spécifiques de *mais* : *sans doute* et *peut-être*.

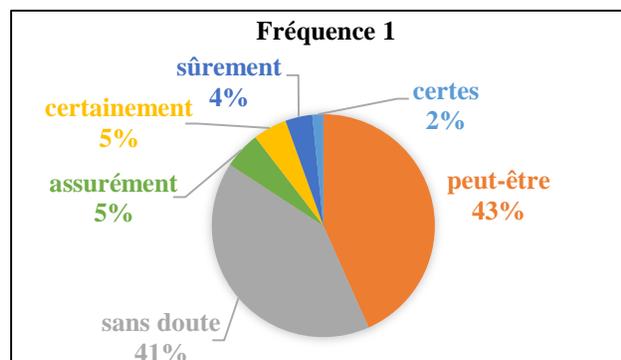
Certes n'apparaît qu'au 19^e. Aucun autre adverbe non pris en compte en synchronie n'apparaît comme spécifique de *mais*.

Ces fréquences font ressortir des résultats tout à fait inattendus par rapport à ceux établis en synchronie. La fréquence de *certes* chute drastiquement au 18^e avant de remonter au 19^e. Ce qui reste stable est le fait que *peut-être* et *sans doute* sont parmi les cooccurrents les plus fréquents de *mais* pendant les trois siècles.

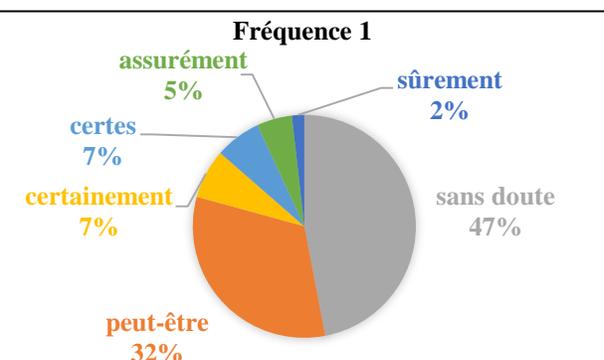
La valeur *MI* fait apparaître que l'attraction



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

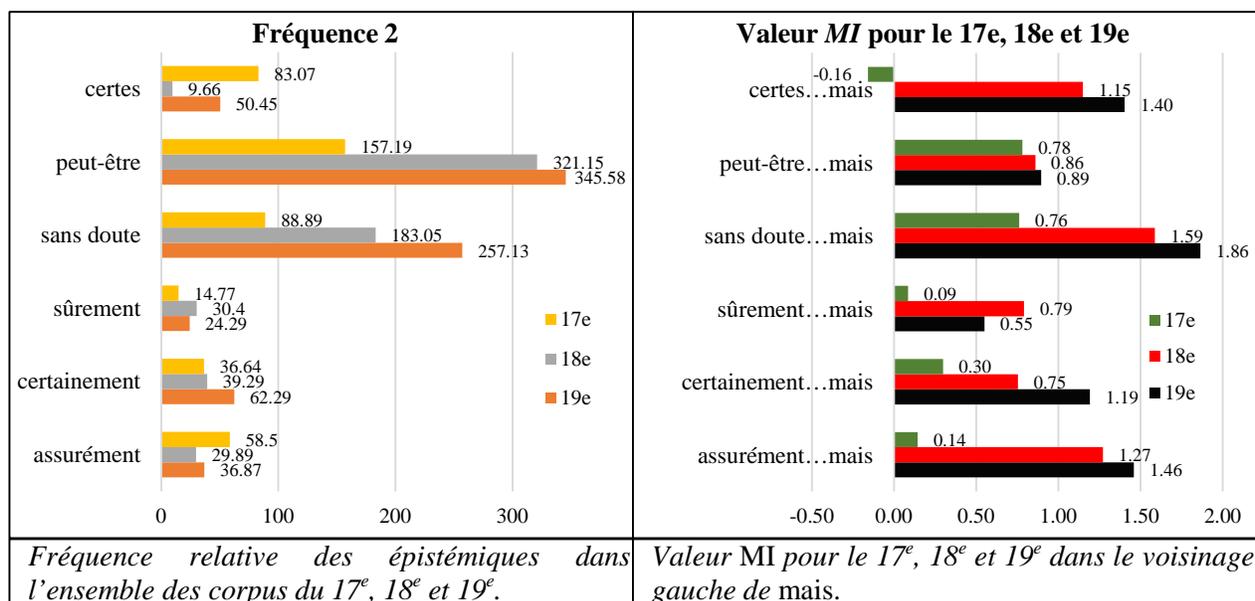


Pourcentage de la séquence < épistémique - mais > sur le nombre total de *mais* au 18^e siècle.



Pourcentage de la séquence < épistémique - mais > sur le nombre total de *mais* au 19^e siècle.

entre *mais* et *certes* est très récente, au 17^e on observe même une valeur négative qui signifie que l'association entre les deux formes est en dessous de la fréquence attendue. On rappelle d'ailleurs qu'à la même époque le *simple-ll* ne fait pas ressortir *certes* comme cooccurrent spécifique de *mais*.



5. Conclusions

Le calcul des fréquences 1 et 2 permet donc de voir apparaître tantôt une correspondance entre les deux tantôt une divergence. Il semble que la divergence prenne effet dans les cas où l'adverbe a une valeur concessive conventionnalisée ou en voie de conventionnalisation. En synchronie, c'est le cas de *certes*, qui a une fréquence relative inférieure à *peut-être* et *sans doute*, mais dont la proportion des emplois avec *mais* est supérieure à celle de ces deux adverbes. Cette attraction entre *certes* et *mais* est confirmée par la valeur *MI*. En diachronie, on relève cette divergence à propos de *sans doute* au 19^e : sa fréquence relative est en dessous de celle de *peut-être* alors que la proportion de ses emplois avec *mais* est au-dessus de celle de *peut-être*. La valeur *MI* confirme également cette attraction : au 19^e *sans doute* a la valeur *MI* la plus élevée avec 1.86.

Peut-être de son côté reste en tête de classement en diachronie et en synchronie en ce qui concerne sa fréquence relative. Cela le conduit à être fréquemment employé aux côtés de *mais*, sans pour autant qu'il puisse être considéré comme particulièrement attaché à *mais*. Il devrait donc sa valeur concessive à sa fréquence relative élevée qui induit une fréquence attendue plus haute dans l'environnement de *mais*. Parallèlement, la valeur *MI* qui indique sa force d'attraction avec *mais* reste très stable pendant les trois siècles en diachronie (aux alentours de 0.8), ce qui confirme qu'il ne peut pas être considéré comme particulièrement attaché à *mais*.

Références

- 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*, Oxford, Blackwell, 602-623.
- Diwersy, S. (2014). La plateforme Primestat.BTLC et l'exploitation lexicostatistique de corpus diachroniques. Journée d'études à l'Institut des Sciences du Langage et de la Communication, Université de Neuchâtel.
- Evert, S. (2007). Corpora and collocations. Manuscrit, Université d'Osnabrück.
- Rossari, C. (2015). Une concession implique-t-elle une opposition ? In Ferrari, A., Lala, L. et Stojmenova, R. (éds), *Testualità. Fondamenti, unità, relazioni*, 189-203.