

Pour une approche sémantique des connecteurs au-delà de leurs propriétés relationnelles : étude sur des variations génériques et diachroniques dans des corpus écritsⁱ



Corinne Rossari^{1,*}, Cyrielle Montrichard¹, et Claudia Ricci¹

¹ Chaire de linguistique française, Université de Neuchâtel, 2000 Neuchâtel, Suisse.

Résumé. Cette contribution vise à documenter la propriété qu'ont les connecteurs non seulement d'agencer les contenus textuels, mais aussi de rendre ou non manifeste la voix à l'origine de cet agencement. Nous montrerons comment cette propriété est à l'œuvre dans des textes à visée informative représentant deux époques différentes. Ceci nous permettra de voir (i) dans quelle mesure les propriétés relatives aux connecteurs sont stables ou sujettes à des changements et, (ii) si ces propriétés font ressortir des similitudes/différences entre genres/époques. Après avoir présenté les propriétés des connecteurs d'un point de vue théorique, nous utiliserons une démarche de linguistique de corpus fondée sur des méthodes statistiques pour évaluer le comportement des connecteurs par rapport à la façon dont ils se combinent avec deux autres classes de mots. Notre étude fera ainsi ressortir le rôle essentiel des connecteurs et de leurs propriétés combinatoires pour cerner les contours d'un genre donné.

Abstract. For a semantic approach to connectors beyond their relational properties: a study on generic and diachronic variations in written corpora. This paper aims to document the property of connectors not only of organizing textual contents, but also of making visible or not the voice behind the textual organization. We will show how this property is at work in informative texts belonging to two diachronic periods. This will allow observing (i) to what extent the properties of the connectors are stable or subject to change and, (ii) whether these properties reveal similarities/differences between genres/epochs. After presenting the properties of connectors from a theoretical point of view, we will use a corpus linguistic approach based on statistical methods to evaluate the behavior of connectors in relation to the way they combine with two other classes of words. Our study will thus highlight the essential role of connectors and their combinatory properties in defining the contours of a given genre.

* Corresponding author: corinne.rossari@unine.ch

1 Introduction

Les connecteurs sont un des outils privilégiés pour agencer des constituants d'un texte. Ils servent autant de repères pour l'analyse de la cohésion – par leurs propriétés de mises en relation de constituants discursifs – que pour l'analyse de la cohérence – par leur capacité à véhiculer des liens logiques entre ces constituants.

Notre propos sera de documenter une toute autre propriété qui les caractérise. Celle-ci concerne non pas l'agencement discursif en tant que tel, mais la propriété qu'a le connecteur de faire ressortir le rôle que le locuteur joue dans cet agencement. En testant une sélection de connecteurs au moyen d'une démarche fondée sur des méthodes et calculs statistiques, nous allons montrer que certains ont la capacité de faire apparaître des contenus comme essentiels au détriment d'autres, et / ou de masquer ou de montrer la voix qui est à l'origine de l'agencement de ces contenus. Ces deux propriétés, que nous considérons comme sémantiques en cela qu'elles sont stables, permettent de différencier les connecteurs selon leur aptitude à rendre manifeste une instance de locution derrière leur emploi. Nous documentons cette caractéristique des connecteurs en prenant en compte la façon dont ils sont utilisés dans des corpus à visée informativeⁱⁱ. Trois corpus du début du 21^{ème} siècle et un corpus du 18^{ème} siècle sont pris en compte : *Le Monde* 2010 (LM), *L'Est Républicain* 2010 (ER) et *Wikipédia* 2019 (Wiki), et un échantillon de *l'Encyclopédie Diderot et D'Alembert* (EDDA)ⁱⁱⁱ. Le choix de ces corpus représentant des genres et des époques différentes permet une vue sur la stabilité des propriétés des connecteurs sélectionnés selon le genre et l'époque des textes interrogés pour cette étude. Notre corpus global est ainsi constitué de textes manifestant une variation d'époque concernant le genre encyclopédique et une variation de genre dans une même synchronie, entre discours de presse / discours encyclopédique, mais aussi au sein de la presse entre presse à vocation nationale et presse de proximité.

Nous présentons dans un premier temps chacune de ces propriétés d'un point de vue théorique en les adossant aux travaux qui, pour certains connecteurs, ont mis en relief des propriétés semblables. Dans un deuxième temps, nous énonçons notre hypothèse étayée par une démarche de linguistique de corpus outillée. Dans une troisième partie, nous approfondissons la démarche outillée en utilisant deux méthodes statistiques que nous appliquons à chacun des corpus pour cerner le comportement des connecteurs sélectionnés. Dans une dernière partie, nous tirons un bilan de ces analyses en évaluant la part liée au facteur générique et ou diachronique et celle liée à la stabilité des propriétés inhérentes à un connecteur donné, dans les différences et ressemblances que chacun des corpus fait apparaître. Nos analyses ouvrent ainsi des perspectives pour la caractérisation du genre, en pointant le rôle des connecteurs et de leurs propriétés combinatoires comme un des facteurs clés pour définir les contours d'un genre donné.

2 Cadrage théorique

Les connecteurs ont fait l'objet de nombreuses études en tant qu'items révélant les mécanismes cognitifs (Sperber et Wilson 1986), argumentatifs (Ducrot et al. 1980), inférentiels (Jayez & Rossari 2004) propres à l'interprétation de séquences discursives complexes.

Parmi ces études, nous retenons d'une part celles qui mettent en relief des propriétés concernant leur aptitude à hiérarchiser les contenus. L'étude d'Anscombe et Ducrot (1977) sur *mais* est la première à relever cette propriété, en montrant que la conclusion tirée de l'énoncé qui suit *mais* prime sur celle tirée de l'énoncé qui le précède. Des travaux s'intéressant à la structure du discours, comme ceux de Roulet et al. (1987), ont étendu cette idée de hiérarchie à d'autres connecteurs, qui sont utilisés comme indices pour attribuer le statut principal ou subordonné aux constituants d'un discours. Ainsi, un connecteur comme *mais* ou *donc* est considéré comme introduisant un constituant principal, alors que *certes* ou

bien que introduisent un constituant subordonné. Un connecteur comme *pourtant* en revanche n'a pas cette propriété ; il est décrit comme un connecteur symétrique ayant la propriété de permettre l'inversion des contenus (Roulet et al. 1987, Anscombe 2002). Toutefois, ces travaux s'appuient sur des tests intuitifs, comme la suppression de certains constituants sans que la cohérence d'un discours en soit affectée (cf. Roulet et al. 1987), ou sur des manipulations d'exemples, comme les possibilités d'enchaînement qu'une suite en *mais* permet (cf. Anscombe et Ducrot 1977).

Nous relevons d'autre part qu'il y a également des études, dont le postulat théorique s'est moins diffusé dans la littérature, qui prêtent aux connecteurs la capacité de faire intervenir la subjectivité de l'instance d'énonciation. Une d'entre elles, celle de Guimier (2000), concerne la différence entre *tandis que* et *alors que*. Concernant *alors que*, l'auteur y voit « le signe d'un certain regard porté par l'énonciateur sur la relation [entre les deux propositions] » (Guimier 2000 : 101), regard que *tandis que* ne ferait pas apparaître. Cette différence est mise en relief au moyen de l'analyse de l'utilisation effective des connecteurs dans différents corpus ainsi que de manipulations d'exemples. D'autres encore mettent en relief les caractéristiques polyphoniques des connecteurs : Sitri (2015 : 56-66) s'interroge sur la part dialogique de connecteurs concessifs tels que *certes... mais* ou *bien que* ; Mellet (2009), à propos de marqueurs concessifs, comme *quand bien même, quand même, si*, relève leur capacité à articuler une proposition *p* faisant entendre un point de vue *autre* « tout en assurant la prise en charge énonciative de [l'autre proposition] *p* » (Mellet, 2009 : 171). Dans une perspective plus large, les études d'Authier Revuz (1984 : 103-105) sur les marqueurs métadiscursifs mettent aussi en relief la capacité de ces formes à faire intervenir la voix du locuteur. Les constructions comme *pour le dire autrement, ou je veux/voulais dire, si on peut dire* permettent au locuteur de se dédoubler en revenant sur son dire pour le commenter.

La capacité qu'a *mais* à hiérarchiser des contenus peut être illustrée par des emplois où *mais* sauve une contradiction interne au discours du locuteur. Dans l'extrait qui suit, on comprend que la locutrice est consciente que ces histoires sont fausses, mais que l'essentiel pour elle est de croire en ces histoires, ce qui permet de dépasser la contradiction.

- | | |
|---|---|
| <p>1. J'me raconte des histoires
Des scénarios Chinois
C'est pas vrai ces histoires
Mais moi j'y crois</p> | <p>1.' J'me raconte des histoires
Des scénarios Chinois
C'est pas vrai ces histoires
#Pourtant moi j'y crois</p> |
|---|---|

(extrait de la chanson *Mon mec à moi*)

Si *pourtant* était utilisé dans cet extrait, l'effet serait tout autre : au lieu de résoudre la contradiction, le connecteur la souligne ; la locutrice se présenterait comme un être ayant la capacité de croire en des histoires qu'il sait être fausses.

Dans l'extrait suivant, c'est *pourtant* qui sauve la cohérence, en mettant l'accent sur l'incompatibilité des deux situations envisagées par le locuteur, celle qui est une projection de l'avenir, et ce qu'il ressent au présent ; le locuteur se montre ainsi conscient de cette incompatibilité.

2. Sans un remords, sans un regret je partirai
Droit devant moi sans espoir de retour
Loin des yeux loin du cœur j'oublierai pour toujours
Et ton cœur et tes bras
Et ta voix
Mon amour
Et **pourtant, pourtant**, je n'aime que toi

(extrait de la chanson *Et pourtant*)

- 2.' Sans un remords, sans un regret je partirai
(...)
#Mais, je n'aime que toi

L'emploi de *mais* annulerait l'interprétation fondée sur le fait que le locuteur est conscient de l'incompatibilité des situations pour indiquer qu'il ne retient que la situation décrite après *mais* ; du coup, l'interprétant ne comprendrait plus la raison d'être de la situation future présentée avant le connecteur où le locuteur se décrit comme n'étant plus amoureux.

La capacité qu'a *alors que* de faire entendre la voix du locuteur, à la différence de *tandis que*, peut être illustrée par l'exemple suivant :

3. Louise, 87 ans, reçoit un PV d'Italie **alors qu'**elle n'a pas quitté son village (Internet)
4. ?Louise, 87 ans, reçoit un PV d'Italie **tandis qu'**elle n'a pas quitté son village

La présence du locuteur pour dénoncer la situation absurde dans laquelle se trouve la protagoniste est nécessaire : *alors que* permet de faire apparaître une instance responsable de la mise en relation des deux états de choses ; *tandis que* ne permet pas de faire apparaître une telle instance, ce qui vide de sens la mise en relation des deux situations.

A l'inverse, dans l'extrait ci-dessous il s'agit de mettre en avant le fait que les deux situations se produiront, ce qui est compatible avec *tandis que* qui ne laisse pas apparaître une instance évaluant les deux situations. L'emploi de *alors que* ferait intervenir une instance d'énonciation, ce qui conduirait l'interprétant à devoir problématiser la mise en relation de ces deux situations.

5. Allons, je serai discrète, et charmée surtout de me rendormir **tandis que** tu seras le plus heureux des hommes.
(George Sand, *La dernière Aldini*)
6. ?Allons, je serai discrète, et charmée surtout de me rendormir **alors que** tu seras le plus heureux des hommes.

En prenant appui sur ces deux cadrages théoriques, nous attribuons à n'importe quel connecteur la capacité ou non de véhiculer de façon conventionnelle la propriété de donner des indications sur la hiérarchie des constituants qu'ils relie, propriété que nous nommons de *hiérarchie*, et la propriété de faire entendre la voix de l'instance responsable de l'agencement discursif inhérent à leur emploi, propriété que nous nommons *énonciative*.

3 Hypothèses et démarche heuristique

Nous partons d'une définition large de connecteur, inspirée de celle présentée dans la *Grande Grammaire du français* (Abeillé & Godard 2021 : 1983 et ss.), selon laquelle toute construction ou expression linguistique susceptible de donner des indications relationnelles fondées sur des primitives cognitives, comme la cause, l'opposition, l'addition et la temporalité, peut être considérée comme un connecteur. Les connecteurs que nous testons sont sélectionnés en fonction de deux critères : ils véhiculent des relations compatibles avec différentes macro catégories cognitives et ils sont au moins deux par catégorie. Notre sélection consiste donc en l'ensemble suivant : une paire pour la macro-catégorie temporalité : *puis/ensuite*, trois paires pour la macro-catégorie causalité : *car/puisque* ; *donc/de ce fait* ; *afin que, de/pour que*, en raison du fait que chacune de ces paires véhicule une relation qui se charge de différents sens : justification pour *car* et *puisque*, conclusion pour *donc* et *de ce fait* et finalité pour *afin que, de* et *pour que* ; et trois items pour la catégorie opposition : *mais, pourtant* et *cependant*^{iv}.

Notre hypothèse est que les connecteurs, indépendamment des indications relationnelles qu'ils véhiculent, peuvent être dotés ou non de la propriété de hiérarchie et/ou de la propriété énonciative.

Ces propriétés sont testables dans une approche de linguistique de corpus fondée sur des méthodes statistiques. A la suite de Firth qui considère que « You shall know a word by the company it keeps » (1957 :11), nous admettons qu'un mot peut être caractérisé sémantiquement par son entourage. Dans cette perspective, nous interrogeons les formes qui sont le plus significativement associées aux connecteurs et sélectionnons celles que nous

considérons être des indices de l'une ou l'autre des propriétés. Pour ce faire, nous mobilisons le calcul Log-likelihood (LL), qui rend compte du caractère non aléatoire de la rencontre entre le connecteur et l'une de ces formes dans un empan déterminé^v.

3.1 Les indices de la propriété de hiérarchie

Dans nos études précédentes (cf. Rossari *et al.*, 2018 et Rossari *et al.*, à paraître), nous avons relevé que les verbes modaux ne se distribuent pas aléatoirement avec certains connecteurs observant que les modaux de nécessité sont préférés dans les contenus indiqués par un connecteur donné comme relevant de l'avant-plan. Ainsi nous avons vu que, dans différents corpus, *mais* est systématiquement associé de manière spécifique aux verbes modaux qui indiquent une nécessité, *devoir* et *falloir*, quand ceux-ci interviennent dans le constituant qui suit *mais* (celui sur lequel l'accent est mis). D'autres connecteurs manifestent également cette capacité d'attirer dans une position donnée des verbes modaux de nécessité (cf. Rossari *et al.*, à paraître) : *afin que/de* attire les modaux de nécessité dans la principale (usuellement à gauche du connecteur), alors que dans la subordonnée, les mêmes verbes sont en dés-attraction ; c'est le verbe modal *pouvoir* qui est attiré dans la subordonnée et symétriquement à *devoir*, *pouvoir* est en dés-attraction quand il apparaît dans la principale.

Le caractère non aléatoire de la distribution des verbes modaux de nécessité par rapport aux connecteurs donnant une indication de hiérarchie peut être corroboré par les travaux concernant la façon dont les verbes modaux sont exploités dans les différentes phases d'une argumentation. Dans Toulmin (2003 [1958]), cité par Rocci (2017) il est mis en relief le rôle de la fonction argumentative des modaux, avec l'idée d'un parallélisme entre les phases d'une argumentation et la sémantique du verbe modal. Le modal de possibilité est exploité dans les phases préparatoires, comme les hypothèses, alors que les modaux de nécessité sont utilisés dans les phases conclusives, notamment dans la partie qui concerne la conclusion que le locuteur veut *in fine* retenir. Ces études permettent de faire un pont avec ce que l'on observe concernant l'attraction des verbes modaux de nécessité dans les contenus qui sont présentés par le connecteur comme étant essentiels.

3.2 Les indices de la propriété énonciative

Pour interroger la propriété énonciative, nous postulons que les connecteurs qui font apparaître la voix du locuteur auront tendance à se combiner plus naturellement avec des formes qui laissent également apparaître une telle instance. La catégorie des adverbes de phrase est l'une de celles qui permet le plus systématiquement de faire ressortir une appréciation du locuteur sur son dire ou sur son dit. Nous basant sur la classification de Molinier et Lévrier (2000), nous avons retenu les disjonctifs de style qui donnent une appréciation sur le dire et les adverbes d'habitude et adverbes modaux (issus de la macro-classe disjonctifs d'attitude) qui donnent une appréciation sur le dit.

4 Comparaison des quatre corpus en fonction des accompagnateurs spécifiques des connecteurs

Les données concernant les accompagnateurs spécifiques révélant chacune de ces deux propriétés sont rassemblées dans quatre tableaux (en annexe A) représentant les quatre corpus pris en compte. Le tableau suivant présente une vue d'ensemble résumant pour chaque connecteur son profil selon qu'il est doté ou non de la propriété de hiérarchie (H) et/ou de la propriété énonciative (E) dans chaque corpus.

Tableau 1 : Profil de chaque connecteur dans les quatre corpus

	EDDA	WIKI	ER	LM
afin que/de	H	H	H	H
pour que	H	H & E	H & E	H
mais	H & E	H & E	H & E	H & E
cependant	H & E	H	H	H
pourtant	H	H	H	∅
donc	H & E	H & E	H & E	H & E
de ce fait		∅	H	∅
car	E	E	E	E
puisque	H & E	H & E	E	H & E
alors que		∅	H & E	H
tandis que	∅	∅	∅	∅
ensuite	∅	H	H	H
puis	E	∅	∅	∅

Ce tableau permet de voir se dessiner des propriétés stables et des points d'instabilité qui font ressortir une sensibilité dans l'emploi des connecteurs au genre ou à l'époque. *Mais* semble imperméable à toute variation générique et diachronique. Il ressort comme doté de la propriété de hiérarchie et de la propriété énonciative dans tous les corpus : il se combine de manière spécifique avec les deux verbes modaux de nécessité dans l'énoncé qui suit le connecteur et il est associé à un grand nombre d'adverbes dont certains sont communs aux trois corpus du 21^{ème} (*certainement, forcément, évidemment* et *nécessairement*) et *certainement* se retrouve même dans le corpus EDDA. *Donc* représente un autre cas de stabilité : il est systématiquement associé à des accompagnateurs spécifiques révélant, comme pour *mais*, sa propriété de hiérarchie et sa propriété énonciative. La conjonction *afin que/de* est dotée de la seule propriété de hiérarchie qu'elle conserve du côté de la principale, la conjonction *car* est uniquement dotée de la propriété énonciative. Enfin *tandis que* représente aussi un exemple de stabilité dans la mesure où aucune de ces deux propriétés ne ressort dans les corpus interrogés. D'autres connecteurs manifestent des comportements moins homogènes selon les corpus : *cependant* est hiérarchiseur concernant l'énoncé qui suit le connecteur dans tous les corpus mais est doté de la propriété énonciative uniquement dans EDDA. *De ce fait* est hiérarchiseur dans un seul corpus (ER) concernant l'énoncé qui suit (à l'instar de *donc*) et jamais énonciatif, à noter qu'il n'est pas représenté dans le corpus du 18^{ème}. *Alors que* (non représenté non plus dans ce corpus) est un connecteur particulièrement instable : tantôt doté des deux propriétés (ER), tantôt doté uniquement de la propriété de hiérarchie (LM) ou encore d'aucune des deux (Wiki). Enfin *pourtant* est doté de la propriété de hiérarchie dans tous les corpus, sauf dans LM. C'est donc sur la propriété énonciative qu'il se singularise de *mais*, et non sur la propriété de hiérarchie contrairement à ce que manifeste l'analyse qualitative qui met en évidence sa propriété de symétrie. La consultation des usages de ce connecteur dans le corpus du 18^{ème} permet de voir encore mieux se dessiner une proximité avec *mais*, avec un emploi où il paraît souligner un mouvement concessif, au moyen de *quoique* significativement associé à *pourtant* (LL de 288 dans un empan de 15 items précédant *pourtant*).

- « car ce philosophe, quoiqu'assez laid, savoit **pourtant** si bien dépouiller la rudesse philosophique, & prendre les airs de la cour, qu'il étoit regardé comme un des hommes les plus aimables . » (EDDA, *Aristotelisme*)

On comprend ainsi mieux pourquoi il semble y avoir une variation en synchronie du 21^{ème} siècle concernant les indices qui révèlent la propriété de hiérarchie : ceux-ci seraient des retombées de son ancien usage dans un tour concessif. Un genre comme la presse de proximité paraît plus enclin à exploiter le potentiel hiérarchique de *pourtant* qu'un genre plus

neutre comme la presse nationale. Cela va de pair avec ce que l'on peut constater pour d'autres connecteurs : dans l'ER *alors que* et *pour que* ressortent comme dotés de la propriété énonciative, tandis que ce n'est pas le cas dans LM et Wiki. On peut y voir le signe selon lequel la presse de proximité exploite davantage les formes laissant apparaître la présence d'une instance énonciative.

Les statistiques fondées sur les indices de spécificité permettent donc de repérer des comportements plus ou moins stables des connecteurs selon les corpus, d'identifier les formes lexicales avec lesquelles ils se combinent significativement et de prendre en compte les régularités quant aux formes avec lesquelles un connecteur donné se combine dans plusieurs corpus. Toutefois, ce type d'indice ne permet pas d'avoir une vue d'ensemble faisant des re/dé-groupements entre connecteurs, car elle catégorise les connecteurs selon un principe dichotomique, à savoir s'ils sont dotés ou non de la propriété étudiée. Par exemple, on vient de voir que cette méthode permet de repérer une grande proximité entre *donc* et *mais*, mais elle ne donne pas d'information sur le fait que ces connecteurs ont ou non des comportements similaires du point de vue de leur combinatoire avec une classe particulière de formes. Nous allons donc utiliser une autre méthode statistique – l'Analyse Factorielle des Correspondances de cooccurrences (AFC) – en prenant en compte comme autre classe de formes, les adverbess issus des trois catégories d'adverbess de phrase retenues^{vi}. Nous verrons de la sorte si, indépendamment de l'aptitude des connecteurs à s'associer spécifiquement à certains de ces adverbess, ils manifestent des comportements qui les particularisent ou les rassemblent par rapport à la fréquence d'apparition de ces adverbess dans leur entourage. La possibilité d'avoir accès à ce type d'information est d'autant plus cruciale quand il s'agit d'évaluer des connecteurs qui montrent en synchronie du 21^{ème} de l'instabilité par rapport à l'une ou l'autre des propriétés analysées selon les corpus. Nous pourrons ainsi mieux appréhender l'impact du facteur genre / sous genre / époque sur l'emploi des connecteurs en utilisant une autre méthode statistique.

5 Les AFC de cooccurrences dans les quatre corpus

Les AFC^{vii} (en annexe B) représentent des écarts à l'indépendance sur un graphique à deux axes fondés sur un calcul statistique et probabiliste. Le graphique fait ressortir les informations principales contenues dans la matrice étudiée, qui est construite de la manière suivante : en titre de colonne on retrouve les connecteurs et en titre de ligne les adverbess. S'agissant d'une AFC de cooccurrences, à l'intersection de la ligne et de la colonne se trouve la co-fréquence brute de l'adverbe avec le connecteur considéré. Au moyen de cette méthode, nous pouvons ainsi relever des similitudes et oppositions entre connecteurs selon la présence dans leur entourage proche (15 items autour du connecteur) des adverbess sélectionnés. Cette sélection a été fondée sur deux objectifs principaux :

- (i) d'une part, être en mesure d'interroger une liste commune à tous les corpus en synchronie avec des adverbess dont la fréquence est d'au moins 80 dans chacun de ces corpus, soit les 9 adverbess suivants : *apparemment*, *certainement*, *effectivement*, *forcément*, *franchement*, *manifestement*, *peut-être*, *probablement*, *réellement*, *vraiment*. Cette liste a été adaptée pour l'Encyclopédie Diderot d'Alembert compte tenu de l'évolution de la langue et de la fréquence d'usage des adverbess interrogés. Nous avons donc exclu *forcément* (fréquence de 9), *franchement* (fréquence de 28), *manifestement* (fréquence de 54) et avons ajouté en raison de leur haute fréquence *vraisemblablement*, *assurément* et *sûrement*.
- (ii) d'autre part, tenir compte de la singularité des textes composant chacun des corpus en sélectionnant les adverbess énonciatifs qui ont montré une attraction pour au moins un connecteur lors de l'étape précédente, à savoir :
 - a. Pour Wiki : *évidemment*, *généralement*, *nécessairement*, *personnellement*, *potentiellement*, *simplement*, *véritablement*.

- b. Pour ER : *évidemment, nécessairement, personnellement, sérieusement, simplement, sûrement, visiblement, véritablement.*
- c. Pour LM : *assurément, éventuellement, évidemment, nécessairement, potentiellement, sûrement, visiblement.*
- d. Pour EDDA : *communément, évidemment franchement, généralement, manifestement, nécessairement, ordinairement, raisonnablement*^{viii} .

La comparaison des quatre AFC fait ressortir les points suivants :

- *Mais* est le connecteur qui se singularise le plus visiblement : il paraît éloigné des autres connecteurs dans tous les corpus, sauf dans l'ER, se dissociant non seulement de *donc*, - les indices de spécificité avaient fait paraître les deux connecteurs comme partageant les mêmes propriétés-, mais aussi des connecteurs relevant de la même classe cognitive d'opposition : *alors que* (dans une moindre mesure dans Wiki), *cependant* et *pourtant*. On voit donc au moyen de cette méthode également se dégager une différence entre LM et l'ER avec un emploi de *mais* particulier à ce dernier corpus.
- Dans l'ER, c'est *alors que* qui se particularise en étant éloigné de tous les autres connecteurs : il rejoint dans ce corpus la singularité qu'a *mais* dans les autres. Même si aucune proximité ne ressort avec *mais* selon l'AFC (encore que ces deux connecteurs paraissent relativement proches sur l'axe 2), nous notons que, selon les indices de spécificité, ils partagent les mêmes propriétés, ainsi que l'adverbe *véritablement* dans ce corpus uniquement (voir tableau 4 dans l'annexe).
- *Cependant* et *pourtant* partagent des profils similaires dans les trois corpus du 21^{ème}, à la différence du corpus du 18^{ème}. Dans EDDA, les trois connecteurs d'opposition occupent des places disjointes : *mais* et *pourtant* manifestent le contraste le plus important et *cependant* occupe une place intermédiaire entre les deux. On relève donc que, bien que *pourtant* ait un emploi concessif tout comme *mais* dans ce corpus, il se singularise par rapport à la façon dont il se combine avec les adverbes. Le graphique de EDDA semble d'ailleurs construit autour de ce contraste : *cependant* est positionné du côté des adverbes de certitude élevée (*certainement, assurément, vraiment, effectivement, manifestement*) *mais* du côté des adverbes d'habitude (*communément, ordinairement, généralement*) et *pourtant* du côté d'adverbes évidentiels ou épistémiques : *raisonnablement, réellement, nécessairement, probablement*.
- Les adverbes d'habitude qui entourent *mais* dans EDDA le particularisent par rapport aux trois corpus du 21^{ème}, dans lesquels on relève que *mais* est plutôt associé à des adverbes épistémiques ou évidentiels et jamais à des adverbes d'habitude : on retrouve dans l'ER et Wiki *véritablement, probablement* et *peut-être* (à noter que dans l'ER *peut-être* est proche de *mais* seulement sur l'axe 2) et dans LM *peut-être, sûrement* et *certainement*.
- *Alors que* et *tandis que* sont dissociés uniquement dans l'ER dans les corpus du 21^{ème}. Cela corrobore ce qui a été observé avec les indices de spécificité qui font ressortir la propriété énonciative pour *alors que* uniquement dans ce corpus.
- *Pour que* et *afin que/de* ne sont dissociés que dans Wiki, alors que les indices de spécificités montrent que *pour que* est doté d'une propriété énonciative qui le différencie de *afin que* non seulement dans Wiki, mais aussi dans l'ER. Le comportement qui ressort des AFC ne conduit donc pas à généraliser une dissociation des deux connecteurs.
- Enfin, la position centrale de *puis* et *ensuite* dans tous les corpus (y compris celui du 18^{ème}) laisse présager que les adverbes sélectionnés ont peu d'impact sur ces connecteurs. Cela peut être corrélé à leur propriété relationnelle : les deux sont des connecteurs qui indiquent des relations temporelles, qui sont peut-être moins mobilisées à des fins argumentatives.

6 Influence du facteur générique et/ou diachronique sur le comportement des connecteurs : ouverture sur la problématique des genres

Le facteur genre joue indéniablement un rôle dans le maniement des connecteurs, mais il ne découpe les corpus ni au niveau des macro genres – encyclopédie vs. presse – ni au niveau des tranches synchroniques représentées – textes du 21^{ème} vs. textes du 18^{ème}. On peut relever des similitudes entre des textes de presse et des textes encyclopédiques au 21^{ème} : dans LM et Wiki *mais* se singularise des autres connecteurs, alors que ce n'est pas le cas dans l'autre corpus de presse – l'ER. Les deux tranches synchroniques ne font pas non plus apparaître de différences quant à la singularité de *mais* : dans EDDA, comme dans les corpus du 21^{ème}, *mais* occupe une place excentrée. Toutefois, ces deux tranches synchroniques manifestent aussi des différences : EDDA est le seul corpus à disjoindre *pourtant* de *cependant*, ce qui rejoint l'observation sur l'emploi concessif de *pourtant* dans ce corpus. LM et Wiki sont proches quant à la position de plusieurs connecteurs : *alors que* et *tandis que* sont proches dans ces deux corpus et disjoints dans l'ER ; *pourtant* et *cependant* sont éloignés de *mais*, alors que dans l'ER ils s'en rapprochent.

Si les différences relevées dans les deux tranches synchroniques peuvent être interprétées comme manifestant des changements dans l'emploi même de certains connecteurs, (nous l'avons envisagé pour *pourtant*, par exemple), celles relevées dans les corpus de la même tranche peuvent être imputables à la représentation que le locuteur se fait des normes rédactionnelles du genre dans lequel il s'exprime : il est vraisemblable que le corpus de presse de l'ER qui représente une presse régionale favorise un style mettant en scène une forme de proximité avec son lectorat, ce qui n'est pas le cas dans un corpus de presse nationale comme LM. Enfin, en ce qui concerne le corpus encyclopédique Wiki, il se démarque par certains aspects tant du corpus LM que de ER : il est le seul à faire apparaître une grande proximité entre *donc* et *car* d'une part et *puisque* et *de ce fait* d'autre part, tout en montrant un regroupement entre les quatre connecteurs (positionnés dans le même cadran); bien qu'il s'agisse de connecteurs proches relationnellement (ils relèvent tous de la même catégorie cognitive causale), dans les deux corpus de presse on note des regroupements différents de ceux de Wiki, avec des paires qui sont non seulement positionnées dans des cadrans différents, mais aussi avec des appariements différents : *donc* est proche de *puisque* et *de ce fait* de *car*. Ces différences méritent indéniablement des approfondissements, mais elles sont déjà le signe de façons différentes de construire la structure argumentative d'un texte, manifestant une sensibilité plus ou moins accentuée des connecteurs vis-à-vis des adverbes. Cela peut être dû tant à la mobilisation autre que le locuteur fait de ces adverbes selon les contraintes rédactionnelles du genre qu'au fait que les connecteurs y réagissent différemment.

7 Conclusions

En croisant une approche qualitative et quantitative sur les connecteurs, notre étude a montré d'une part qu'au-delà des différences relationnelles qui les caractérisent, ils sont dotés ou non des propriétés de nature énonciative et hiérarchique qui déterminent des façons différentes de se combiner avec les verbes modaux et les adverbes dans les corpus ; d'autre part, en examinant des textes relevant de genres ou sous genres différents et d'époques différentes, elle a montré que ces modes combinatoires ne sont pas étrangers aux contraintes rédactionnelles d'un genre, ou aux modifications que le connecteur subit quant à ses emplois à travers le temps. Ces variations ouvrent des pistes de réflexion pour la caractérisation du genre dans une perspective de détection automatique de traits linguistiques pertinents comme celle conçue par Biber, Conrad & Reppen (1998), en pointant le rôle fondamental des connecteurs et de leurs propriétés combinatoires.

Annexes : tableaux et AFC de cooccurrences

A. Tableaux récapitulant les propriétés de hiérarchie et énonciative pour chaque connecteur testé dans les quatre corpus

Les tableaux ci-après montrent, pour chaque corpus, les connecteurs testés et la présence ou non d'une propriété de hiérarchie selon le(s) verbe(s) (en gras) qui en sont l'indice. Le tableau permet aussi de visualiser si cette mise à l'avant-plan se situe dans la proposition précédant le connecteur (1^{re} colonne) ou suivant ce dernier (2^{ème} colonne). La dernière colonne détaille les adverbes qui répondent des critères fixés et présentés ci-avant et qui présente une cooccurrence spécifique dans l'entourage des connecteurs.

Tableau 2 : Propriétés des connecteurs relevées dans EDDA

Connecteurs testés dans EDDA	Accompagnateurs spécifiques révélant la propriété de hiérarchisation				Accompagnateurs spécifiques révélant la propriété énonciative
	Verbe modal avant	LL	Verbe modal après	LL	Adverbes disjonctifs de style, d'habitude et modaux
puis					manifestement
ensuite					
donc			pouvoir	255	évidemment, nécessairement, raisonnablement, réellement
			devoir	396	
			falloir	1188	
puisque	devoir	87	pouvoir	79	apparemment, certainement, effectivement, évidemment, manifestement, nécessairement, réellement
	pouvoir	54			
car	falloir	14	falloir	180	assurément, certainement, franchement, nécessairement, réellement,
	devoir	27	pouvoir	222	
	pouvoir	43			
tandis que					
cependant			falloir	264	communément, ordinairement
			devoir	14	
			pouvoir	207	
mais	pouvoir	170	pouvoir	632	apparemment, assurément, certainement, peut-être, généralement
			devoir	222	
			falloir	1165	
pourtant			pouvoir	16	
			falloir	184	
afin que/de	devoir	101	pouvoir	592	
	falloir	62			
pour que	falloir	281	pouvoir	1094	
	devoir	97			

Tableau 3 : Propriétés des connecteurs relevées dans WIKI

Connecteurs testés dans WIKI	Accompagnateurs spécifiques révélant la propriété de hiérarchisation				Accompagnateurs spécifiques révélant la propriété énonciative <i>Adverbes disjonctifs de style, d'habitude et modaux</i>
	Verbe modal avant	LL	Verbe modal après	LL	
puis					
ensuite			falloir	29	
donc	devoir	21	devoir	267	nécessairement, potentiellement, forcément
	pouvoir	23	falloir	479	
			pouvoir	696	
de ce fait			pouvoir	42	
puisque	pouvoir	28	falloir	29	nécessairement
			pouvoir	73	
car	falloir	15	falloir	50	généralement, peut-être, réellement
	devoir	50	devoir	59	
	pouvoir	74	pouvoir	166	
alors que					
tandis que					
cependant			devoir	70	
			falloir	167	
			pouvoir	128	
mais	devoir	47	devoir	304	nécessairement, peut-être, forcément, probablement, vraiment, simplement, véritablement, certainement, réellement,
	pouvoir	44	falloir	270	
			pouvoir	745	
pourtant			falloir	16	
afin que/de	devoir	126	pouvoir	277	
pour que	falloir	1930	falloir	15	personnellement, vraiment
	devoir	72	pouvoir	686	

Tableau 4 : Propriétés des connecteurs relevées dans LM

Connecteurs testés dans LM	Accompagnateurs spécifiques révélant la propriété de hiérarchisation				Accompagnateurs spécifiques révélant la propriété énonciative <i>Adverbes disjonctifs de style, d'habitude et modaux</i>
	Verbe modal	LL	Verbe modal	LL	
puis					
ensuite			pouvoir	39	
			devoir	104	
donc			falloir	653	éventuellement, potentiellement
			pouvoir	255	
			devoir	506	
de ce fait					
puisque	pouvoir	34	pouvoir	68	visiblement
			devoir	14	
			falloir	20	
car	pouvoir	88	pouvoir	125	probablement, peut-être, vraiment, évidemment
	devoir	48	devoir	28	
	falloir	73	falloir	87	
tandis que					
alors que			devoir	145	
cependant			devoir	30	
			pouvoir	78	
mais	pouvoir	153	pouvoir	703	assurément, certainement, évidemment, forcément, franchement, nécessairement, probablement, sûrement, vraiment
			devoir	195	
			falloir	1077	
pourtant					
afin que/de	devoir	79	pouvoir	132	
pour que	falloir	1109	pouvoir	630	
	devoir	88			

Tableau 5 : Propriétés des connecteurs relevées dans ER

Connecteurs testés dans ER	Accompagnateurs spécifiques révélant la propriété de hiérarchisation				Accompagnateurs spécifiques révélant la propriété énonciative
	Verbe modal avant	LL	Verbe modal après	LL	
puis					
ensuite	pouvoir	61	pouvoir	340	
	falloir	17	falloir	109	
			devoir	21	
donc	pouvoir	26	pouvoir	557	forcément, vraiment
	falloir	18	falloir	866	
	devoir	13	devoir	569	
de ce fait			devoir	31	
			pouvoir	11	
puisque	pouvoir	87	pouvoir	25	apparemment, évidemment, visiblement, forcément, manifestement
	falloir	34	falloir	18	
	devoir	20	devoir	53	
car	pouvoir	173	pouvoir	171	vraiment, forcément, simplement, certainement, peut-être
	falloir	376	falloir	358	
	devoir	122	devoir	97	
alors que	pouvoir	81	pouvoir	116	véritablement
			devoir	417	
tandis que					
cependant			pouvoir	54	
			falloir	66	
			devoir	90	
mais	pouvoir	383	pouvoir	1042	peut-être, évidemment, forcément, sûrement, certainement, apparemment, personnellement, véritablement, sérieusement, nécessairement
	falloir	26	falloir	1945	
	devoir	74	devoir	502	
pourtant			devoir	16	
			falloir	29	
afin que/de	devoir	91	pouvoir	628	
	falloir	15			
pour que	devoir	123	pouvoir	1043	réellement, simplement, véritablement
	falloir	1077			

B. AFC de cooccurrences

Pour lire les AFC, les étiquettes des connecteurs (encadrées en bleu) sont les suivantes : ENSU pour *ensuite*, PUIS, DCFT pour *de ce fait*, PUIQ pour *puisque*, CAR, DONC, CPDT pour *cependant*, MAIS, PRTT pour *pourtant*, TAND pour *tandis*, ALSQ pour *alors que*, AFIN pour *afin de/que* et PRQU pour *pour que*. Les adverbess sont en rouge.

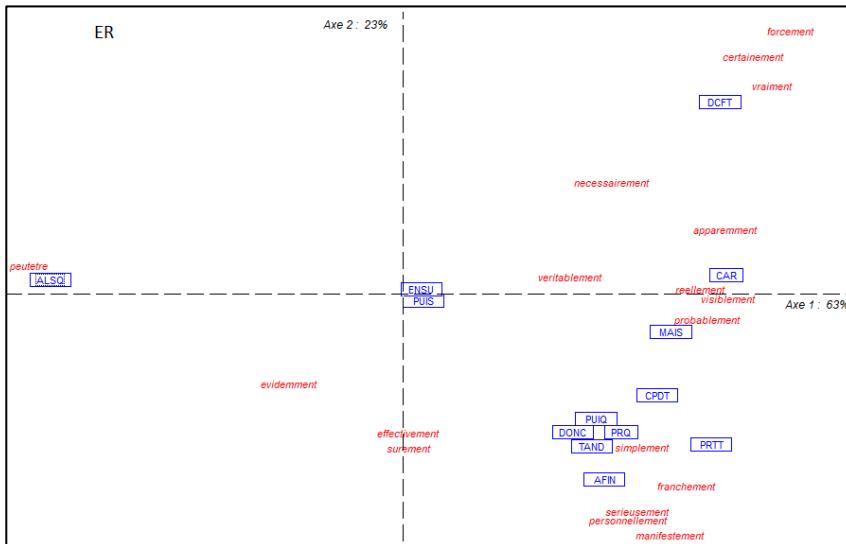


Figure 1. AFC de cooccurrences réalisée dans l'ER

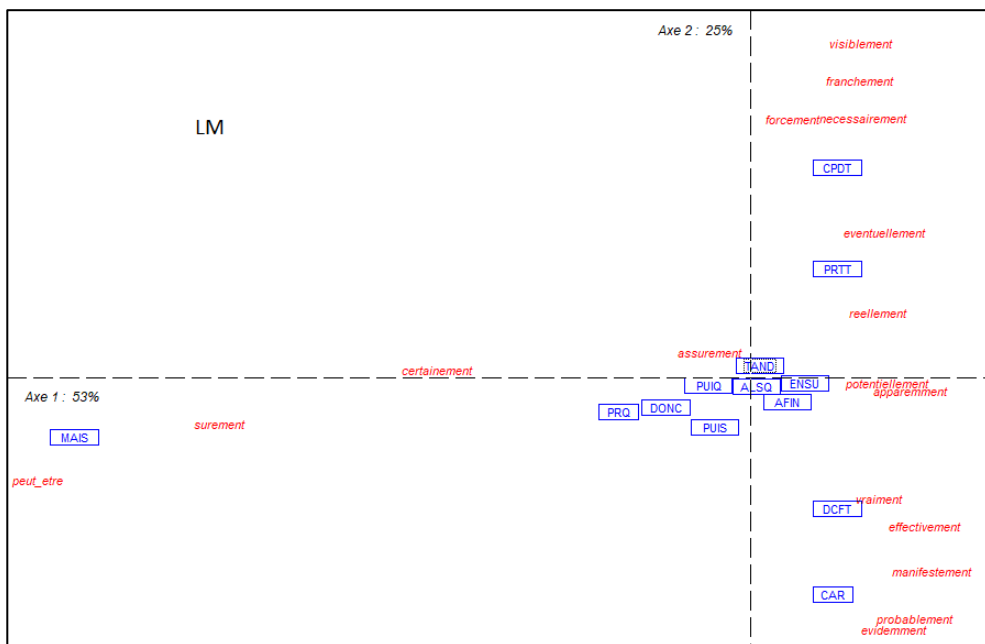


Figure 2. AFC de cooccurrences réalisée pour LM

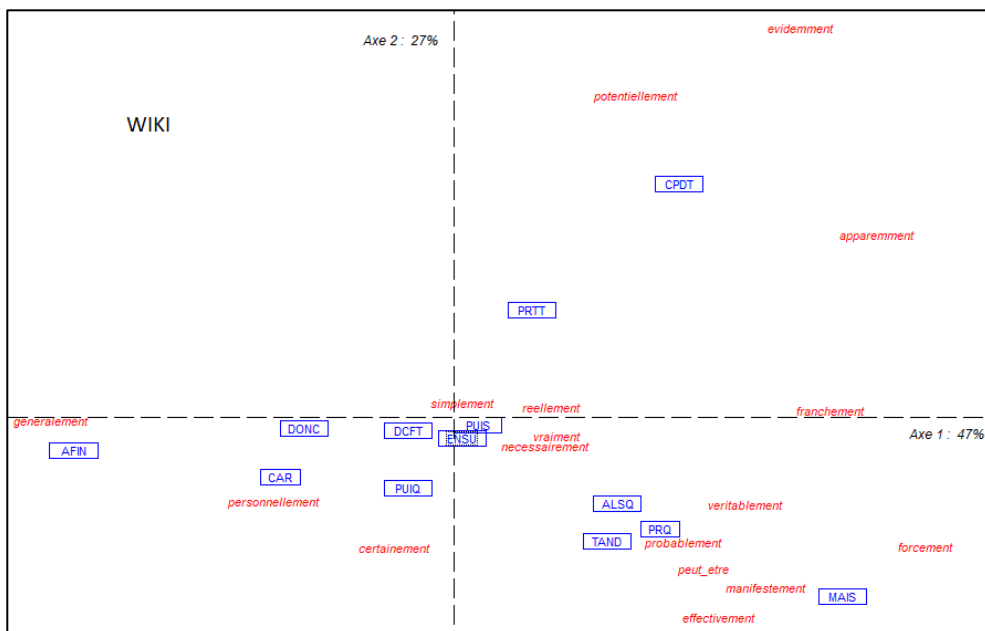


Figure 3 : AFC de cooccurrences réalisée dans WIKI

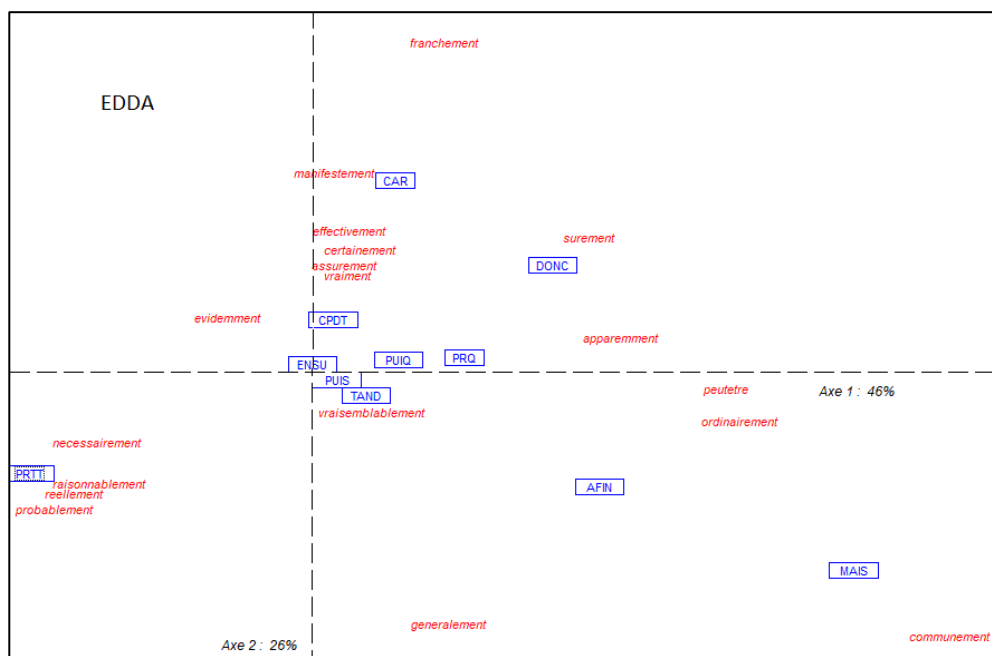


Figure 4 : AFC de cooccurrences réalisée dans EDDA

Références bibliographiques

- Abeillé, A. & Godard, D. (dir.) (2021). *La Grande Grammaire du Français*. Paris : Actes Sud.
- Anscombe, J.-C. (2002). *Mais/pourtant* dans la contre-argumentation directe : raisonnement, généricité, et lexique, *Linx*, 46, 115-131.
- Anscombe, J.-C. & Ducrot, O. (1977). Deux *mais* en français ?, *Lingua*, 43, 23-40.
- Authier-Revuz, J. (1984). Hétérogénéité(s) énonciative(s). *Langages*, 73, 98-111.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics. Investigating Language Structure and Use*. New York : Cambridge University Press.
- Brunet, E. (2012). Nouveau traitement des cooccurrences dans Hyperbase. *Corpus [en ligne]*, 11. Mis en ligne le 18 juin 2013, consulté le 8 septembre 2020. URL : <http://journals.openedition.org/corpus/2275>
- Ducrot, O. et al. (1980). *Les mots du discours*. Paris : Minit.
- Firth, J. (1957). A Synopsis of Linguistic Theory 1930-1955. *Studies in Linguistic Analysis*, Special Volume of the Philological Society, pp. 1-32.
- Guimier, C. (2000). Non-congruence et congruence : alors que vs tandis que. *Syntaxe et Sémantique*, 1, 80-112.
- Jayez, J. & Rossari, C. (2004). Parentheticals as Conventional Implicatures. *Handbook of French Semantics*. Stanford : CSLI, 211-229.
- Mellet, S. (2009). Dialogisme, parcours et altérité notionnelle : pour une intégration en langue du dialogisme ?. *Langue française*, 163, 157-173.
- Molinier, Ch. & Levrier, F. (2000). *Grammaire des adverbes. Description des formes en -ment*. Genève/Paris : Droz.
- Rossari, C., Ricci, C. & Dolamic, L. (2018). Le conditionnel appliqué à *devoir/dovere* et son potentiel argumentatif. *Langue française*, 200, 105-120.
- Rossari, C., Montrichard, C., Pamuksaç, A., Ricci, C. & Sanvido, L. (à paraître). Structures argumentatives et analyses statistiques : étude des combinatoires entre connecteurs et indications énonciatives. *Humanidades & Inovação*.
- Roulet, E., Auchlin, A., Moeschler, J., Rubattel, C. & Schelling, M. (1987). *L'articulation du discours en français contemporain*. Berne : Peter Lang

- Sitri, F. (2015). *Parcours en analyse du discours : enjeux et méthode. Autour d'écrits professionnels*. Linguistique. Université Sorbonne nouvelle-Paris3. Thèse pour une habilitation à diriger des recherches.
- Sperber, D. & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford : Blackwell.
- Rocci, A. (2017). *Modality in argumentation*. Oxford : Springer.
- Toulmin, S. E. (2003) [1958]. *The uses of argument*. Cambridge : Cambridge University Press.
- Vigier, D., Moncla, L., Brenon, A. McDonough, K & Joliveau, T. (2020). Classification des entités nommées dans l'Encyclopédie ou dictionnaire raisonné des sciences des arts et des métiers par une société des gens de lettres (1751-1772). *Congrès Mondial de Linguistique Française - CMLF 2020*. URL : <https://hal.archives-ouvertes.fr/hal-02578029/document>

ⁱ Cet article s'inscrit dans le cadre du financement d'un projet FNS intitulé « L'ancrage argumentatif des formes modales. Etude sur corpus avec un éclairage comparatif entre français et italien », projet n. 100001F_192247.

ⁱⁱ Notre choix de prendre en compte des textes à visée informative est motivé par le fait de documenter l'usage des connecteurs dans la langue ordinaire, qui n'est pas esthétisante comme l'est le discours littéraire.

ⁱⁱⁱ *Le Monde* 2010 compte 17'895'009 tokens, *L'Est Républicain* 25'825949, *Wikipédia* 2019 18'715455 et *EDDA* 11'759825. Nous remercions l'équipe du projet DISCO (CollEx-Persée) pour le partage du corpus EDDA annoté (Vigier *et al.* 2020).

^{iv} Les connecteurs retenus ne présentent pas d'ambiguïté faussant les résultats en raison d'emplois où la même forme ne fonctionne pas comme un connecteur. Seuls *de ce fait* et *alors que* peuvent avoir un emploi très différent (*cela est issu de ce fait* ou *c'est alors qu'il est arrivé*), mais après vérification ces emplois sont marginaux et donc ne faussent pas les résultats statistiques ; en revanche ils peuvent prendre différentes valeurs que nous n'avons pas codées ; nos résultats valent pour toutes les valeurs confondues d'un connecteur donné, considérant que leur sémantique globale est intrinsèquement liée aux propriétés testées.

^v A partir de la valeur 10.83, la rencontre entre deux items a 99.9 % de chances de ne pas être due à la distribution aléatoire de chacun des items du corpus consulté. Les valeurs négatives sont fondées sur le même principe, à savoir que la 'non rencontre' entre les deux items au-delà du seuil de -10.83 a 99.9 % de chances de ne pas être due au hasard. Ces indications servent de base pour interpréter les valeurs au-dessus de ce seuil comme des attractions et en dessous comme des dés-attractions.

^{vi} Nous utilisons les trois catégories d'adverbes mobilisées, car, comme montré par les indices de spécificité, les connecteurs y sont sensibles et ce sont des catégories suffisamment vastes pour prétendre à ce type de méthode.

^{vii} Pour réaliser ces AFC, nous avons mobilisé le logiciel Hyperbase, conçu et développé par Étienne Brunet (laboratoire Base Corpus Langage de l'université de Nice Sophia Antipolis), qui permet un tel calcul (Brunet 2012).

^{viii} Nous avons pris la liberté de prendre en considération des adverbes qui ne sont pas catégorisés comme disjonctifs de style ou disjonctifs d'attitude quand ces derniers nous semblent avoir des emplois proches de l'une ou l'autre de ces catégories dans nos corpus : *raisonnablement* qui est souvent conforme aux disjonctifs de style dans EDDA et *potentiellement* qui paraît proche des modaux, notamment dans son utilisation fréquente avec *pouvoir*.