



**HAL**  
open science

## Libre choix de mode ?

Olaf Mikkelsen

► **To cite this version:**

Olaf Mikkelsen. Libre choix de mode?: Vers une analyse multifactorielle de l'alternance indicatif/subjonctif en français contemporain. 2017. hal-01550566

**HAL Id: hal-01550566**

**<https://hal.science/hal-01550566>**

Preprint submitted on 29 Jun 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Libre choix de mode ?**

### **Vers une analyse multifactorielle de l'alternance indicatif/subjonctif en français contemporain**

*Olaf Mikkelsen*

Université Paris VII Diderot

#### RÉSUMÉ

En suivant Amsili & Guida (2014), ce travail prétend explorer des différents facteurs responsables de l'alternance indicatif/subjonctif en français contemporain. Nous avons choisi d'étudier quatre prédicats épistémiques (*croire que, penser que, être certain que, avoir l'impression que*) en contexte négatif, qui semble être propice à cette alternance. 617 exemples de ces prédicats ont été retirés de Frantext et annotés pour une vingtaine de traits. Sur ce corpus, nous avons effectué une analyse de régression afin d'évaluer le poids relatif de chaque facteur dans le choix de mode.

*« Quant au subjonctif, comment en ignorerait-elle les règles alors qu'il exprime les peurs et les souhaits ? Premier subjonctif (trois ans) : Je ne crois pas que la salle de jeu soit fermée. » (Philippe Forest, L'enfant éternel, 1997)*

L'homme est capable d'imaginer non seulement ce qui est mais également ce qui n'est pas et ainsi franchir le précipice discret existant entre le réel et l'irréel. Cette dichotomie est très souvent évoquée lorsqu'un locuteur essaye d'expliquer ce qu'est le subjonctif, par exemple en français, où ce mode est bien existant. Notre locuteur dirait alors que c'est le mode de l'irréel. Un apprenant du français risquerait de prendre ce constat au pied de la lettre et serait surpris par toute une panoplie d'usages où, même si la situation décrite est bien réelle, le subjonctif est utilisé. Il s'agit de situations incertaines du point de vue de celui qui parle. Malgré cette précision, l'apprenant confronté avec des phrases de type « ça me désole qu'il se soit fâché » s'interrogerait : La personne en question n'est-elle donc pas réellement fâchée ? Ou bien est-ce que, ignorant à propos de la véracité de la colère, on est tout de même désolé ? Le locuteur natif interviendrait alors en ajoutant que le subjonctif est aussi le mode permettant d'exprimer une valeur subjective ou peut-être, si notre locuteur est connaisseur du latin, que c'est le mode de la dépendance (du latin SUBJUNGERE signifiant « mettre sous la dépendance de »), c'est-à-dire de la subordination. Le pauvre apprenant commence à désespérer : n'y a-t-il donc pas un critère transparent pour comprendre l'usage du subjonctif ?

Cette étude n'a pas pour but de rendre compte de la complexité du mode subjonctif en français, mais il est nécessaire de reconnaître la difficulté de notre objet d'étude avant de choisir la manière dont on le traite. Le subjonctif n'est pas le mode de l'irréel, le mode subjectif ou de la dépendance, mais il pourrait englober les trois usages simultanément. Cela suggère que son usage est non-catégorique et sujet à des tendances. Dans ce cas, il est, à notre avis, judicieux d'employer des méthodes quantitatives, qui permettent précisément d'entrevoir des régularités inaccessibles à l'introspection. Peut-être un jour ces mêmes méthodes pourront-elles nous aider à mieux comprendre le subjonctif. Pour le moment nous allons nous limiter à l'étude de l'alternance des deux modes, indicatif et subjonctif, en français contemporain dans certains contextes bien précis.

## 1. L'alternance indicatif/subjonctif

Le point de départ de notre travail est le mémoire de Floriane Guida (Guida 2013), qui décrit le phénomène en grand détail, ainsi qu'un article publié par Pascal Amsili (Amsili & Guida 2014) où les auteurs ébauchent une approche très similaire à la notre, et nous adopterons par la suite leur terminologie et leur conclusions. Nous entendons par alternance indicatif/subjonctif « des contextes linguistiques dans lesquels les locuteurs semblent placés devant un choix entre le mode subjonctif et le mode indicatif » (Amsili & Guida 2014:1) en excluant « les nombreux contextes où en français contemporain le mode [...] est fixé sans ambiguïté par les constructions » (*ibid.*), c'est-à-dire les contextes où seul l'indicatif ou seul le subjonctif est possible. On doit également préciser qu'il s'agit des « cas où les modes indicatif et subjonctif peuvent tous deux être employés sans qu'il y ait de différenciation sémantique claire » (Guida 2013:11) et que « la phrase matrice doit (pouvoir) être entièrement la même » (*idem* p.16). Ainsi, les exemples suivants ne seront pas considérés comme des cas d'alternance (exemples de Guida) :

- (1) Marie croit que Claude **a**<sub>IND</sub> rêvé.
- (2) Marie ne croit pas que Claude **ait**<sub>SBJ</sub> rêvé.
- (3) Il dit qu'il **est**<sub>IND</sub> chinois.
- (4) Salomon, l'agent, m'a dit que je te **fasse**<sub>SBJ</sub> signe.

En (1)-(2) la forme de la phrase matrice est différente : positive dans (1) et négative dans (2). En (3)-(4) le sens du verbe recteur change : déclaratif dans (3) et injonctif dans (4). Ayant identifié ce que nous entendons par alternance indicatif/subjonctif, nous pouvons délimiter quel cas d'alternance étudier. Guida s'intéresse à trois cas :

- I) L'alternance dans les complétives après des prédicats épistémiques de sens positif en contexte négatif ou interrogatif (5) ;
- II) L'alternance dans les conjonctives disloquées à gauche et les conjonctives sujets (6) ;
- III) L'alternance dans les complétives venant après le nom *fait* (7).

- (5) Jean ne croit pas que Marie {**est**<sub>IND</sub>/**soit**<sub>SBJ</sub>} étudiante.
- (6) Que Marie {**est**<sub>IND</sub>/**soit**<sub>SBJ</sub>} étudiant, Jean le croit.
- (7) Le fait que Marie {**est**<sub>IND</sub>/**soit**<sub>SBJ</sub>} étudiante l'étonne.

De ces trois cas, nous nous intéresserons par la suite au premier, et nous avons choisi les quatre prédicats suivants pour le cas d'étude, dans deux formes (négatif et interrogatif), élevant le nombre de contextes à huit :

<i>Croire que</i> + contexte négatif / <i>Croire que</i> + contexte interrogatif <i>Penser que</i> + contexte négatif / <i>Penser que</i> + contexte interrogatif <i>Être certain que</i> + contexte négatif / <i>Être certain que</i> + contexte interrogatif <i>Avoir l'impression que</i> + contexte négatif / <i>Avoir l'impression que</i> + contexte interrogatif
--

Avant de présenter la méthodologie adoptée, nous allons dire deux mots sur les causes possibles de cette alternance. Les causes pourraient bien être historiques, car nous savons que le subjonctif a beaucoup évolué depuis l'ancien français, et il est possible que l'évolution perdure encore aujourd'hui. Danièle Godard (2012) explique les phénomènes d'alternance en tant que manifestations de différents degrés de grammaticalisation syntaxique : en se fondant sur Becker (2010), qui a montré que le changement de mode (indicatif>subjonctif) pour certains prédicats au XVII<sup>e</sup> siècle correspond à des critères sémantiques sous-jacents, Godard propose que, dans les cas où il y a une alternance, le processus de grammaticalisation ne s'est achevé que partiellement. Les causes pourraient également être de nature extralinguistique, en particulier liées à la variation diastratique. Selon Guida « on pourrait imaginer qu'il y ait une distribution socio-stylistique (les deux modes font partie de deux registres différents, deux usages) » (Guida 2013:165). Une troisième possibilité serait d'envisager de multiples facteurs grammaticaux qui opèrent sur le choix de mode : le temps grammatical utilisé dans la forme verbale, la personne du sujet du verbe de la matrice, le type de la phrase, le temps du verbe de la matrice, le type de négation ainsi que le verbe conjugué et sa morphologie (*idem* p.166-169). Nous inclinons pour cette dernière possibilité et notre travail sera précisément d'étudier ces facteurs et d'analyser les interactions possibles entre eux. Or, cela ne veut pas dire que nous rejetons toute cause historique ou extralinguistique, car il est bien possible que le subjonctif soit en voie de disparition dans les générations futures et dans certaines classes sociales. Simplement nous hésitons à croire que ces causes conduisent à une alternance arbitraire entre les modes. Nous répétons volontiers la question posée par Guida : « est-ce parce qu'on est dans un cas d'alternance qu'il y a un libre choix du mode ? » (*idem* p.11).

## 2. Méthodologie

Nous avons utilisé la base de données de textes Frantext pour extraire un corpus (voir annexes ci-dessous) de 1158 exemples des contextes où apparaissent le subjonctif et l'indicatif : 617 exemples dans la forme négative, 541 dans la forme interrogative. Ces exemples ont ensuite été annotés dans un tableur (voir annexes ci-dessous) qui permettra de manipuler les données en utilisant des méthodes statistiques pour tester notre hypothèse de base, à savoir que le choix de mode n'est pas arbitraire, mais dépend de l'interaction de plusieurs facteurs simultanément. Ces méthodes statistiques s'appliquent depuis fort longtemps dans le domaine des sciences sociales, y compris la sociolinguistique, tandis que le basculement vers l'usage de données quantitatives est récent en linguistique théorique : avec l'apparition de l'article de Bresnan et al (2007) sur l'alternance du datif en anglais et la publication subséquente de l'ouvrage de référence (premier dans son genre en linguistique théorique) *Analyzing linguistic data* (Baayen, 2008). Cette mise en œuvre tardive est due, d'une part, à la méfiance générale de la linguistique chomskyenne envers les méthodes quantitatives, et d'autre part à certains problèmes méthodologiques inhérents à l'usage des corpus, notamment le fait que les corrélations dans les données peuvent conduire à des interprétations réductrices en analysant une seule variable (Bresnan *et al* 2007). Avec des données annotées, ce problème peut être détourné en utilisant la régression logistique. Cette technique a commencé à être utilisée en épidémiologie avant de se répandre à d'autres domaines et c'est aujourd'hui la méthode dominante dans les études où la variable dépendante, l'objet d'étude, a une valeur binaire (Speelman in Glynn 2014). Cela est bien le cas de notre étude, car notre variable dépendante est le mode : subjonctif ou indicatif. Il reste toutefois à établir les variables indépendantes, c'est-à-dire les facteurs qui pourraient déterminer le choix de mode dans les deux contextes que nous avons choisi : négatif et interrogatif.

## 3. Guide d'annotation

Nous souhaitons maintenant présenter une liste des 20 traits choisis dans le travail d'annotation. Dans les contextes négatif et interrogatif, le mode dominant est le subjonctif (417/617 exemples), et notre travail est donc de découvrir les facteurs qui favorisent l'usage de l'indicatif (200/617 exemples).

### 3.1 Date de publication (année)

Nous soupçonnions que les textes plus récents présenteraient un taux plus élevé de l'indicatif. Nous avons fait une annotation d'année de la première publication selon les informations que nous avons trouvées dans Frantext. Deux problèmes se présentent : en premier lieu, une granularité à l'année est trop faible par rapport au volume des données, en deuxième lieu, la date de publication ne correspond pas forcément au moment d'écriture, voir les exemples suivants :

- (8) Je ne crois pas que l'on **puisse**<sub>SBJ</sub> citer un seul témoignage qui démente cette affirmation (Charles de Gaulle, *Discours et messages. 1. Pendant la guerre. 1940-1946*, 1970, p. 141).
- (9) Pour moi, je ne crois pas que ce **soit**<sub>SBJ</sub> un de mes plus mauvais livres (Simone de Beauvoir Jean-Paul Sartre, *Entretiens avec Jean-Paul Sartre : août-septembre 1974*, 1981, p. 390).

Le premier problème pourrait se résoudre en réduisant cette granularité à la décennie. Le deuxième n'a pas de solution a priori, mais on pourrait penser que tous les textes auront subi une mise en forme au moment de la publication.

### 3.2 Auteur

Un principe en linguistique de corpus est de ne prendre qu'un exemple par locuteur/ auteur. La raison est le risque que des exemples multiples d'un même sujet entraîne un possible variable parasite, à savoir l'idiolecte du sujet. Nous n'avons pas suivi ce principe, et nous avons en conséquence annoté le trait d' « auteur » afin de pouvoir contrôler son effet.

### 3.3 Registre social (soutenu/courant/relâché)

Comme le souligne Guida « la principale différence entre indicatif et subjonctif dans les contextes qui nous intéressent pourrait bien être une différence qui ressortit à des facteurs socioculturels » (Guida 2013:158). Nous avons choisi une tripartition en nous fondant sur des critères comme les suivants :

Relâché (usage d'expressions populaires, argotiques ou vulgaires, syntaxe simplifiée, usage d'abréviations et vocabulaire familier, interrogatives directes, suppression de *ne* dans la négation, pronom sujet *on* à la place de *nous*). Courant (usage « normal » de la syntaxe et le vocabulaire). Soutenu (phrases très longues, vocabulaire rare, usage de formes verbales inusitées (imparfait et plus-que-parfait du subjonctif), interrogative directe inversée).

Ainsi, des cas prototypiques des trois niveaux pourraient se présenter sous la forme des exemples suivants :

- (10) Ce n'était pas à moi à blâmer les mésalliances discrètes ; je ne croyais pas que le Roi le **pût**<sub>SBJ</sub> faire davantage (Françoise Chandernagor, *L'Allée du Roi*, 1981, p. 512).
- (11) Je vous répète que je resterai avec vous jusqu'à la fin de ce livre, le dernier concernant mon autre vie. Ne croyez pas que je l'**écris**<sub>IND</sub> par plaisir, mais je n'avais pas d'autre possibilité (Patrick Modiano, *Les Boulevards de ceinture*, 1972, p. 149).
- (12) Le comble: le voilà qui m'insulte! - Tu crois pas que je **vais**<sub>IND</sub> bander pour ta gueule, espèce de vieux pédé!... (Bertrand Blier, *Les Valseuses*, 1972, p. 109).

Plusieurs problèmes se présentent. D'abord le problème linguistique, car l'annotation a été effectuée par un francophone non-natif, après il y a le risque non négligeable de faire du casier « courant » un choix par défaut dans des nombreux cas douteux : l'exemple (13) est-il courant ou soutenu ? L'exemple (14) est-il courant ou relâché ? Un troisième problème, encore plus grave, c'est que l'emploi de l'indicatif dans les contextes qui nous intéressent peut lui-même être interprété comme un indice du registre relâché et cela pourrait amener la personne qui fait l'annotation à un biais de sélection.

- (13) Mais la présence de Nestor donnait à toutes choses un air de magie d'une irrésistible emprise. Je ne crois pas que les enfants **aient**<sub>SBJ</sub> un sens esthétique très développé (Michel Tournier, *Le Roi des Aulnes*, 1970, p. 98).
- (14) A chaque bruit, maman me réveille en disant : « Tu ne crois pas que **c'est**<sub>IND</sub> une alerte ? » Des officiers et des soldats anglais passent de temps en temps (Micheline Bood, *Les Années doubles : Journal d'une lycéenne sous l'Occupation*, 1974, p. 23).

#### 3.4 Phrase négative (Oui/Non)

La phrase peut être à la forme négative (Oui) ou interrogative (Non). Ce trait a été éliminé dans l'analyse préliminaire (voir 4 ci-dessous) où nous prenons en compte uniquement les exemples négatifs.

#### 3.5 Type de négation (Descriptive/Métalinguistique)

Selon Guida (2013) une différence entre emploi de l'indicatif et du subjonctif « découle de l'usage de la négation dans la matrice. Ainsi, quand la négation

est métalinguistique, l'indicatif semble plus naturel et préférable au subjonctif » (Guida 2013:161). La négation métalinguistique est une négation qui rejette soit un énoncé antérieur, soit un présupposé (Muller 1991). Elle se caractérise par plusieurs propriétés, dont une prosodie marquée, un effet *garden path*, une dimension échoïque, une clause de rectification indispensable et une interprétation non métalinguistique contradictoire. L'alternative est la négation descriptive, pouvant être définie comme un opérateur « servant à rejeter un énoncé réel ou présupposé » (Muller 1991:43). Vu l'absence presque totale de négations métalinguistiques dans le corpus, ce trait a été éliminé dans l'analyse préliminaire. Nous pensons que ce trait aura un poids plus important dans un éventuel corpus plus large qui intègre des données des corpus oraux (voir ouvertures ci-dessous).

### 3.6 Items à polarité négative (Oui/Non)

Selon Guida (2013) les items à polarité négative seront « nettement admis dans une subordonnée au subjonctif, beaucoup moins dans une subordonnée à l'indicatif » (Guida 2013:159). Nous avons identifié les items à polarité négative suivants dans le corpus : *ni, de, jamais, personne, moindre, grand-chose, quiconque, quoi que ce soit*. Voir les exemples suivants :

- (15) - Christine, je ne crois pas que tu **aies**<sub>SBJ</sub> jamais rencontré Lucien... (Anne-Marie Garat, *Pense à demain*, 2010, p. 382).
- (16) je ne pense pas que quiconque ici **ait**<sub>SBJ</sub> l'intention de laisser indéfiniment entre les mains françaises la police de la circulation (Pierre Mendès-France, *Œuvres complètes. 3. Gouverner, c'est choisir. 1954-1955*, 1986, p. 545).

Cette question de la cooccurrence des items à polarité négative avec l'indicatif, comme le remarquent Amsili et Guida (2014) « mériterait des études plus approfondies, en lien avec le phénomène de *neg-raising* ». La difficulté d'établir la nuance entre interprétation *neg-raising* et interprétation in-situ est cependant bien connue pour les prédicats épistémiques comme (17), car l'interprétation avec *neg-raising* dépend de si on estime le sujet compétent de sa croyance.

- (17) Jean ne croit pas qu'il viendra  
 =  $\neg B_{s,p}$  [S  $\neg$ compétent = in-situ]  
 =  $B_s \neg p$  [S compétent = *neg-raising*]

### 3.7 Phrase interrogative (Oui/Non)

La phrase peut être à la forme interrogative (Oui), interro-négative (Oui) ou négative sans interrogation (Non). Nous avons comme intuition que la forme interro-négative favorise fortement l'indicatif, car une phrase de type « Ne croyez-vous que x ? » semble impliquer que le locuteur tient x pour vrai et cherche une affirmation de son interlocuteur. Selon Guida (2013) il n'y aurait pas de montée de la négation, mais sur le verbe épistémique uniquement. En voici un exemple :

- (18) Vous ne croyez pas que la couleur **a**<sub>IND</sub> changé quand l'Europe a inventé l'ombre ? (André Malraux, *La Corde et les souris. V : [La Tête d'obisdiene]*, 1976, p. 760).  
= Ce n'est pas le cas que vous croyez que la couleur a changé quand l'Europe a inventé l'ombre  
≠ Vous croyez que la couleur n'a pas changé quand l'Europe a inventé l'ombre

### 3.8 Contexte conversationnel (Oui/Non)

Nous comprenons ici pour « contexte conversationnel » des dialogues (entre guillemets dans le tableur) dans les textes littéraires et nous prenons ce trait comme un indice d'un degré plus marqué d' « oralité » dans le discours. Nous avons comme intuition que ce contexte favorise l'indicatif :

- (19) ou un pressentiment plus... - ... plus mortel ? - Si vous voulez.  
Je vis avec un pressentiment obscur de fin du monde et comme je ne crois pas que le monde **va**<sub>IND</sub> finir... (Romain Gary, *Au-delà de cette limite votre ticket n'est plus valable*, 1975, p. 112).  
(20) - Les opinions, oncle Abel, tu ne penses pas que ça **vaut**<sub>IND</sub> les croyances ? (Hervé Bazin, *L'école des pères*, 1991, p. 330).

### 3.9 Verbe de la matrice (Croire, Penser, Être certain, Avoir l'impression)

Nous avons choisi les quatre prédicats mentionnés supra : *croire que, penser que, (être) certain que, (avoir) l'impression que*.

### 3.10 Mode du verbe de la matrice

Nous avons inclus le mode du verbe de la matrice. Les occurrences extrêmement rares du gérondif (21) ou du conditionnel (22-24) dans notre corpus nous ont obligé à éliminer ces exemples.

- (21) et elle, la paysanne des Cards d'une autre façon faillible et ne croyant pas que chose lui **fût**<sub>SBJ</sub> due (Pierre Michon, *Vies minuscules*, 1984, p. 237).
- (22) Et elle: «On ne croirait pas que tu **es**<sub>IND</sub> mon cousin, sale gosse!» (Robert Sabatier, *Les Fillettes chantantes*, 1980, p. 29)
- (23) Je ne croirais pas que **c'est**<sub>IND</sub> une teinte (Jean-Loup Trassart, *Des cours d'eau peu considérables*, 1981, p. 96).
- (24) on ne croirait pas que le train **va**<sub>IND</sub> vers le front (Simone de Beauvoir, *Journal de guerre : septembre 1939-janvier 1941*, 1990, p. 115).

Nous pensons d'ailleurs que si le mode de la matrice est le subjonctif, cela favorise l'indicatif dans la complétive :

- (25) Que l'on ne croie pas que **c'était**<sub>IND</sub> de la susceptibilité malade (Zoé Oldenburg, *Visages d'un autoportrait*, 1977, p. 53).

### 3.11 Personne, nombre et temps du verbe de la matrice

Selon Boysen (1971) et d'autres études « la première personne favoriserait plus le subjonctif que l'indicatif, ou bien la première personne favoriserait plus le subjonctif que les autres personnes » (Guida 2013:166). Voir l'exemple suivant :

- (26) Pourtant, je ne crois pas que les Allemands **aient**<sub>SBJ</sub> vraiment atteint leur objectif (Frédérique Moret, *Journal d'une mauvaise Française*, 1973, p. 44).

Quant au temps verbal « c'est un facteur également mentionnée par Boysen (1971), qui considère que le subjonctif apparaît plus lorsque le verbe de la matrice est au présent » (*idem* p. 168). Il est d'ailleurs logique que l'indicatif soit favorisé lorsque le verbe de la matrice est à l'imparfait (27), car « si le subjonctif imparfait était encore courant au XIX<sup>e</sup> siècle, ce n'est plus guère le cas aujourd'hui » (Amsili & Guida 2014:3).

- (27) Il ne croyait pas que **j'étais**<sub>IND</sub> ce que j'étais dans l'opinion (Pierre Goldman, *Souvenirs obscurs d'un juif polonais né en France*, 1975, p. 142).

Nous avons éliminé les exemples au futur (28) et au passé simple (29) dans l'analyse préliminaire.

- (28) quant aux soldats, tant qu'ils n'auront pas été engagés, ils n'auront pas l'impression que la France **a**<sub>IND</sub> utilisé toutes ses réserves (Pierre Mendès-France, *Œuvres complètes. 1. S'engager. 1922-1943*, 1984, p. 379).
- (29) Elle accepta avec chaleur la proposition de Monsieur Scarron, mais ne crut pas que, jeune et un peu romanesque, je **pusse**<sub>SBJ</sub> choisir d'épouser un homme d'aspect si monstrueux (Françoise Chandernagor, *L'Allée du Roi*, 1981, p. 85).

### 3.12 Matrice subordonnée (Oui/Non)

Nous avons l'intuition qu'une subordonnée introduite par une phrase qui est, elle aussi, une subordonnée, favoriserait l'emploi de l'indicatif :

- (30) Oui, qu'il demeure avec toi, même si tu ne crois pas que tu **crois**<sub>IND</sub> (Claude Mauriac, *Le Rire des pères dans les yeux des enfants*, 1981, p. 48).

### 3.13 Verbe de la complétive et type (Aux/Lex) de verbe

Nous avons inclus le verbe de la complétive, afin de pouvoir contrôler pour l'effet de certains verbes, et ensuite divisé ces verbes en deux groupes : verbes auxiliaires et lexicaux, afin de comparer une éventuelle différence entre p. ex. 'être' lexical et 'être' auxiliaire.

### 3.14 Supplétion (Oui/Non)

Nous avons divisé les verbes en supplétifs (être, aller) et non supplétifs (les autres). Nous aurions pu faire une distinction plus nette entre formes plus ou moins saillantes du point de vue morphologique, car comme souligne Guida (2013:169) « on pourrait envisager que les formes plutôt proches peuvent se confondre ». Toutefois, en travaillant avec des données catégoriques, il est difficile d'opérationnaliser une différence graduelle.

### 3.15 Personne, nombre et temps du verbe de la complétive

Nous avons inclus la morphologie du verbe conjugué lui-même.

### 3.16 Les traits considérés

Nous pouvons finalement regrouper les 20 traits en quatre niveaux :

- I) Niveau socio-historique :  
Date de publication 1.  
Auteur 2.  
Registre 3.
- II) Niveau lexico-morphologique :  
Type et morphologie des verbes 9. 10. 11. 12. 13. 15. 16. 17. 18.  
19. 20.
- III) Niveau syntaxico-pragmatique :  
Type de phrase 4. 7.  
Type de négation 5. 6.  
Subordination 14.
- IV) Niveau discursif :  
Contexte conversationnel 8.

### 3.17 Attentes

Nous nous attendrions à ce que les facteurs suivants favorisent l'emploi de l'indicatif :

Registre : relâché.

Morphologie du verbe : 2<sup>e</sup> et 3<sup>e</sup> personne.

Type de phrase : interrogative, interro-négative.

Présence d'items à polarité négative.

Matrice subordonnée.

Contexte conversationnel.

En même temps, les facteurs suivants favoriseraient l'emploi du subjonctif :

Registre : soutenu/courant.

Morphologie du verbes : 1<sup>er</sup> personne.

Matrice non-subordonnée.

Absence de contexte conversationnel.

Finalement, les facteurs suivants sont problématiques par les raisons invoquées supra :

Date

Auteur

Type de négation

Supplétion.

## 4. Analyse

Après avoir éliminé les exemples qui présentaient des traits extrêmement rares dans notre corpus, nous avons testé certains traits afin d'établir si leur présence dans les exemples sont des indices des structures linguistiques dans les données ou simplement arbitraires. Nous utiliserons le logiciel RStudio pour Mac (Version 0.99.893) pour faire les modèles statistiques et déterminer la probabilité que nos résultats soient dus au hasard. Par la suite, nous analyserons les traits Registre, Type de phrase, Contexte conversationnel, Items à polarité négative, Subordination de la matrice, Personne grammaticale de la matrice ainsi que Type du verbe de la complétive.

### 4.1 Registre

En commençant par le premier, nous avons l'intuition que le registre relâché favorise l'emploi de l'indicatif. La distribution des trois niveaux est la suivante :

<b>Tableau 1</b>	<b>Registre</b>	Relâché	Courant	Soutenu
<b>Mode</b>				
	IND	52	117	24
	SBJ	13	229	170

Il semblerait effectivement que l'indicatif est favorisé par le registre relâché, mais cette distribution est-elle significative ? En utilisant un simple test  $\chi^2$  nous pouvons déterminer la distribution normale attendue :

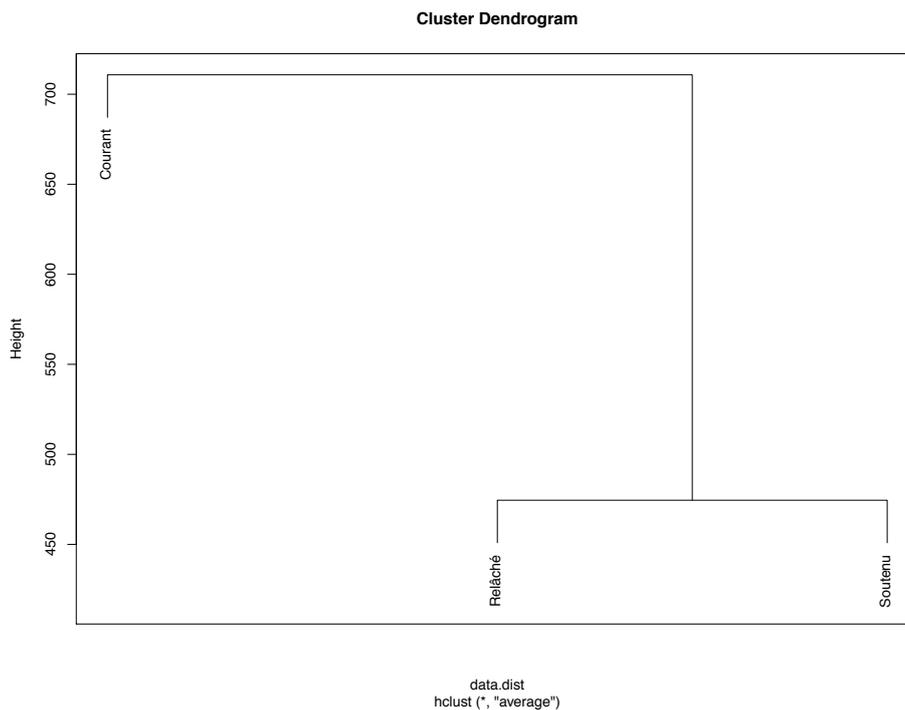
<b>Tableau 2</b>	<b>Registre</b>	Relâché	Courant	Soutenu
<b>Mode</b>				
	IND	20.73554	110.3769	61.8876
	SBJ	44.26446	235.6231	132.1124

La corrélation entre registre et choix de mode est significative (valeur  $p < 2.2e-16$ ), mais pour savoir lequel des facteurs est plus responsable et de quelle façon, nous devons examiner les résidus :

<b>Tableau 3</b>	<b>Registre</b>	Relâché	Courant	Soutenu
<b>Mode</b>				
	IND	6.8658344	0.6304127	-4.8160978
	SBJ	-4.6991940	-0.4314744	3.2962895

Cela nous permet d'affirmer que le registre « relâché » est un indicateur clair de l'emploi de l'indicatif. Or, ce résultat était attendu, et de plus nous avons exprimé des doutes concernant l'annotation du trait de registre. Une façon de contrôler les données serait d'effectuer une analyse de cluster, qui regroupe les facteurs selon une mesure de dissimilarité et les présentent dans un dendrogramme.

Fig. 1 Dendrogramme Registre



En fig. 1 ci-dessus, on constate que le modèle sépare « relâché » et « soutenu » de « courant », ce qui indique qu'il y a des problèmes dans l'annotation. Deux solutions se présentent : soit on refait l'annotation en utilisant la même tripartition, soit on simplifie en deux groupes (p. ex. « formel » et « informel »).

#### 4.2 Type de phrase

Nous avons l'intuition que la phrase interro-négative favorise fortement l'emploi de l'indicatif. Cela semble être le cas (voir tableau 4 ci-dessous).

<b>Tableau 4</b>	<b>Type de phrase</b>	Négative	Interro-négative
<b>Mode</b>			
IND	Occurrences	141	52
	Distribution	172.9025	20.09752
	Résidus	-2.426185	7.116283
SBJ	Occurrences	401	11
	Distribution	369.0975	42.90248
	Résidus	1.660557	-4.870609

Les résidus montrent qu'il y a une forte corrélation indicatif/phrase interro-négative (7.116283), et cette corrélation est significative (valeur  $p < 2.2e-16$ ).

#### 4.3 Contexte conversationnel

Nous attendions à ce qu'il favorise l'emploi du mode indicatif, mais les résultats sont moins conclusifs que dans le cas de l'interrogation :

<b>Tableau 5</b>	<b>Type de phrase</b>	Négative	Interro-négative
<b>Mode</b>			
IND	Occurrences	110	83
	Distribution	130.4744	62.52562
	Résidus	-1.792455	2.589296
SBJ	Occurrences	299	113
	Distribution	278.5256	133.47438
	Résidus	1.226812	-1.772196

Les valeurs ne s'écartent pas beaucoup des valeurs attendues. Toutefois, il y a une corrélation indicatif/contexte conversationnel (2.589296), et cette corrélation est significative (valeur  $p < 0.0001355$ ).

#### 4.4 Polarité

<b>Tableau 6</b>	<b>Items à polarité négative</b>	Non	Oui
<b>Mode</b>			
IND	Occurrences	191	2
	Distribution	174.4975	18.50248
	Résidus	1.2492650	-3.8364914
SBJ	Occurrences	356	56
	Distribution	372.5025	39.49752
	Résidus	-0.8550364	2.6258159

Comme montre tableau 6, la prédiction de Guida (voir 3.6 ci-dessus) se vérifie, car la corrélation entre les items à polarité négative et le subjonctif est significative (valeur  $p < 1.012e-06$ ).

#### 4.5 Subordination de la matrice

Ce facteur ne joue aucun rôle dans le choix de mode : sa distribution est similaire à l'attendue (11/9.88) et la corrélation très faible (0.35) n'est pas significative (valeur  $p < 0.6604$ ).

<b>Tableau 7</b>	<b>Matrice</b>	Non-subordonnée	Subordonnée
<b>Mode</b>			
IND	Occurrences	182	11
	Distribution	183.1107	9.889256
	Résidus	-0.08208371	0.35320926
SBJ	Occurrences	392	20
	Distribution	390.8893	21.110744
	Résidus	0.05618069	-0.24174758

Nous espérons toutefois que ce trait aura un poids plus important dans le contexte interrogatif.

#### 4.6 Personne grammaticale

Nous attendions (voir 3.11 ci-dessus) une corrélation entre première personne du verbe de la matrice et le mode subjonctif.

<b>Tableau 8</b>	<b>Personne</b>	1 <sup>re</sup>	2 <sup>e</sup>	3 <sup>e</sup>
<b>Mode</b>				
IND	Occurrences	87	74	32
	Distribution	130.7934	40.51405	21.69256
	Résidus	-3.829267	5.260897	2.213071
SBJ	Occurrences	323	53	36
	Distribution	279.2066	86.48595	46.30744
	Résidus	2.620871	-3.600724	-1.514696

Cette prédiction se vérifie et nous constatons une corrélation 1<sup>re</sup> personne/subjonctif ainsi qu'une anti corrélation 1<sup>re</sup> personne/indicatif. Et de plus on trouve une forte corrélation 2<sup>e</sup> personne/indicatif. Ces corrélations sont significatives (valeur  $p < 8.655e-16$ ).

#### 4.7 Mode de la matrice

Il y a une corrélation significative (valeur  $p < 0.002032$ ) entre le mode subjonctif de la matrice et l'emploi du mode indicatif dans la complétive.

<b>Tableau 9</b>	<b>Mode</b>	Impératif	Indicatif	Subjonctif
<b>Mode</b>				
IND	Occurrences	19	164	10
	Distribution	19.4595	169.3934	4.147107
	Résidus	-0.1041654	-0.4143940	2.8740738
SBJ	Occurrences	42	367	3
	Distribution	41.5405	361.6066	8.852893
	Résidus	0.0712941	0.2836244	-1.9671069

#### 4.8 Type du verbe

Finalement on ne peut montrer qu'une corrélation très faible et non-significative (valeur  $p < 0.2296$ ) entre le type de verbe et le choix de mode.

<b>Tableau 10</b>	<b>Type du verbe</b>	Auxiliaire	Lexical
<b>Mode</b>			
IND	Occurrences	67	126
	Distribution	73.69091	119.3091
	Résidus	-0.7794319	0.6125596
SBJ	Occurrences	164	248
	Distribution	157.30909	254.6909
	Résidus	0.5334678	-0.4192552

#### 4.9 Analyse de régression logistique

Afin de vérifier nos résultats et de regarder les effets des variables entre elles, nous avons créé deux modèles de régression logistique différents pour prédire le choix de mode. Nous appliquerons GLM avec le package MASS et LRM avec le package rms. Dans le premier modèle nous utiliserons les variables « Phrase interrogative », « Registre » et « Contexte conversationnel », dans le deuxième « Personne grammaticale de la matrice », « Items à polarité négative » et « Mode de la matrice ». Nous présenterons les résultats des deux modèles en tableau 11 (ci-dessous). Dans un souci de concision, seuls les coefficients ainsi que quelques statistiques seront inclus, notamment le degré de liberté (ddl), la fonction de vraisemblance ( $G^2$ ), la statistique de concordance (C), Nagelkerke pseudo  $R^2$  (N.  $R^2$ ) et une estimation inférentielle (bootstrapped) de N.  $R^2$  (Boot.  $R^2$ ). Les coefficients apparaissent avec les niveaux  $\alpha$  ( $^0 p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ).

<b>Tableau 11</b>	Modèle 1	Modèle 2
<u>Coefficients</u>		
Phrase interro-négative	-2.2972***	–
Registre : Relâché	-1.7113***	–
Registre : Soutenu	1.2704***	–
Contexte conversationnel	0.2343 <sup>0</sup>	–
Personne : 1 <sup>re</sup>	–	1.0133***
Personne : 2 <sup>e</sup>	–	-1.7098***
Mode matrice : Indicatif	–	-2.2674***
Mode matrice : Subjonctif	–	-3.1492***
Items à polarité négative : Oui	–	2.2868**
<u>Statistiques du modèle</u>		
ddl	4	5
G <sup>2</sup>	146.24	126.19
C	0.758	0.727
N. R <sup>2</sup>	0.301	0.264
Boot. R <sup>2</sup>	0.2863	0.2406

La qualité classificatoire de nos modèles est bonne, avec des scores C de 0.758 et 0.727 respectivement, leur complexité n'est pas trop élevée (4 et 5 degrés de liberté) et les scores R<sup>2</sup> sont acceptables. Pour ce qui est des variables du premier modèle, on constate que le contexte interro-négatif ainsi que le registre relâché sont des bons prédicteurs de l'emploi de l'indicatif, comme registre soutenu l'est pour l'emploi du subjonctif. Ces corrélations sont toutes significatives ( $p < 0.001$  dans les trois cas). Or, la corrélation contexte conversationnel-subjonctif (non pas de l'indicatif comme nous le pensions) est faible et non significative (valeur  $p < 0.33$ ). Dans le deuxième modèle toutes les variables sont significatives. L'emploi d'items à polarité négative et le verbe de la matrice en 1<sup>re</sup> personne étant des bons prédicteurs pour le subjonctif, le verbe de la matrice en 2<sup>e</sup> personne pour l'indicatif. Finalement, les modes indicatif et subjonctif dans la matrice semblent tous les deux être des prédicteurs de l'emploi de l'indicatif dans la complétive, ce qui est un indice que nous avons un problème dans le modèle. Il vaut mieux laisser d'un côté cette variable pour l'instant. En tableau 12 (ci-dessous) nous présentons un troisième modèle dans lequel sont inclus seulement les variables « Phrase interrogative », « Registre », « Personne grammaticale de la matrice » et « Items à polarité négative ».

<b>Tableau 12</b>	<b>Modèle 3</b>	
<u>Coefficients</u>		
Phrase interro-négative	-1.9835***	
Registre : Relâché	-1.5097***	
Registre : Soutenu	1.3217***	
Personne : 1 <sup>re</sup>	1.2941***	
Personne : 2 <sup>e</sup>	0.7614*	
Items à polarité négative : Oui	2.2742**	
<u>Statistiques du modèle</u>		
ddl	6	
G <sup>2</sup>	186.02	
C	0.804	
N. R <sup>2</sup>	0.371	
Boot. R <sup>2</sup>	0.3552	

La valeur prédictive de ce modèle est très élevée et les corrélations entre toutes les variables sont significatives. Nous constatons que le contexte interro-négatif ainsi que le registre relâché favorisent l'emploi de l'indicatif, les items à polarité négative, le registre soutenu ainsi que la première personne grammaticale celui du subjonctif. La dernière variable, à savoir la deuxième personne grammaticale, est un faible prédicteur et sa valeur est changée de négative (indicatif) à positive (subjonctif). Nous préférons donc pour le moment de ne pas tenir compte de cette variable.

### **5. Discussion**

Dans cette étude préliminaire, nous ne prenons en compte que le contexte négatif. De plus, certains problèmes se présentent, soit dans le travail d'annotation (voir discussion sur le registre en 3.3 ci-dessus), soit dans les modèles eux-mêmes (voir discussion sur les modes dans la matrice en 4.9 ci-dessus). Compte tenu de ces difficultés, que pouvons-nous apprendre ? Les résultats de notre étude vont dans le sens que nous l'attendions : le choix de mode dans les contextes qui permettent une alternance libre (en apparence) entre subjonctif et indicatif est influencé par la cooccurrence de, au moins, cinq facteurs, inclus dans le tableau 13 ci-dessous :

<b>Tableau 13</b>		
<b>Facteurs favorisant le subjonctif</b>	<b>Facteurs favorisant l'indicatif</b>	
Items à polarité négative	Contexte interro-négatif	
Registre soutenu	Registre relâché	
Verbe de la matrice en 1 <sup>re</sup> personne		

Beaucoup de travail reste à faire : la chose la plus urgente est de finir l'annotation des 541 exemples en contexte interrogatif, deuxièmement faudrait-il faire réviser la colonne de « registre » par des francophones natifs et finalement, mais non pour cela moins important, une révision complète des traits contenues dans la guide d'annotation sera nécessaire. Nous avons constaté que certains traits (notamment « Contexte conversationnel » et « Type de verbe ») ne jouent pas un rôle dans le choix de mode dans notre corpus, mais avant de les éliminer il serait judicieux de refaire notre analyse avec plus des données. Nous attendons que un trait comme « Subordination de la matrice » aura un poids plus important dans le contexte interrogatif. Pour terminer, nous envisageons de comparer des données écrites (Frantext) avec des données orales ([Orfeo](#)) en ajoutant un trait « type de texte » (écrit/oral).

## Corpus

Base textuelle FRANTEXT, ATILF - CNRS & Université de Lorraine. Site internet : <http://www.frantext.fr>.

## Annexes

Fichier word contenant 1158 exemples de Frantext.

Fichier excel contenant 617 exemples annotés en contexte négatif

## Références

- Amsili, Pascal & Floriane Guida. 2014. 'Vers une analyse factorielle de l'alternance indicatif/subjonctif'. In *SHS Web of Conferences*, volume 8, p. 2313-2331 : EDP Sciences.
- Baayen, Harald. 2008. *Analyzing linguistic data*. Cambridge: Cambridge University Press.
- Boysen, Gerhard. 1969. 'Le mode de la proposition complétive préposée dans les langues romanes'. In *Revue romane*, 4(1), 10-19.
- Bresnan, Joan & Anna Cueni & Tatiana Nikitina & Harald Baayen. 2007. 'Predicting the dative alternation'. In *Cognitive foundations of interpretation*, p. 69-94.

- Farkas, Donka. 2003. 'Assertion, Belief and Mood Choice'. Manuscrit présenté dans le *Workshop on Conditional and Unconditional Modality*, Vienna.
- Glynn, Dylan & Justyna Robinson (Eds.). 2014. *Corpus Methods for Semantics. Quantitative studies in polysemy and synonymy*. Amsterdam: John Benjamins.
- Godard, Danièle. 2012. 'Indicative and subjunctive mood in complement clauses: from formal semantics to grammar writing'. In *Empirical Issues in Syntax and Semantics* 9, ed. Christopher Piñón, pp. 129-148.
- Godard, Danièle & Barbara Hemforth & Jean-Marie Marandin. 2012. 'Negation and mood alternation in French. A preliminary experimental investigation of polarity mood'. Manuscrit proposé en hommage à Ivan Sag.
- Gries, Stefan Th. 2009. *Quantitative corpus linguistics with R*. London: Routledge.
- Gries, Stefan Th. 2013. *Statistics for linguistics with R*, 2<sup>nd</sup> edition. London: Routledge.
- Guida, Floriane. 2013. *Deux modes pour un contexte. L'alternance des modes indicatif et subjunctif en français contemporain*. Mémoire de master 2 recherche. Université Paris-Diderot.
- Huot, Hélène. 1986. 'Le subjunctif dans les complétives: subjectivité et modalisation'. In M. Ronat & D. Couquaux, Eds., *La grammaire modulaire*, p. 81–101. Paris: Éditions de Minuit.
- Muller, Claude. 1991. *La négation en français*. Genève: Droz.
- Poplack, Shana. 1990. 'Prescription, intuition et usage: le subjunctif français et la variabilité inhérente'. In *Langage et société*, 54(1), 5–33.
- Schlenker, Philippe. 2005. 'The lazy Frenchman's approach to the subjunctive: Speculations on reference to worlds and semantic defaults in the analysis of mood'. In *Proceedings of Going Romance XVII*.