



HAL
open science

Syntaxe du français parlé vs. écrit : le cas de la position de l'adjectif épithète par rapport au nom

Juliette Thuilier

► **To cite this version:**

Juliette Thuilier. Syntaxe du français parlé vs. écrit : le cas de la position de l'adjectif épithète par rapport au nom. Travaux Interdisciplinaires sur la Parole et le Langage, 2013, Le français parlé, 29. hal-01451579

HAL Id: hal-01451579

<https://hal.science/hal-01451579>

Submitted on 1 Feb 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.

TIPA. Travaux interdisciplinaires sur la parole et le langage

29 (2013)
Le français parlé

Juliette Thuilier

Syntaxe du français parlé vs. écrit : le cas de la position de l'adjectif épithète par rapport au nom

Avertissement

Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur.

Les œuvres figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage soit personnel, soit scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document.

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.

revues.org

Revues.org est un portail de revues en sciences humaines et sociales développé par le Cléo, Centre pour l'édition électronique ouverte (CNRS, EHESS, UP, UAPV).

Référence électronique

Juliette Thuilier, « Syntaxe du français parlé vs. écrit : le cas de la position de l'adjectif épithète par rapport au nom », *TIPA. Travaux interdisciplinaires sur la parole et le langage* [En ligne], 29 | 2013, mis en ligne le 18 décembre 2013, consulté le 22 mars 2014. URL : <http://tipa.revues.org/1066>

Éditeur : Laboratoire parole et langage

<http://tipa.revues.org>

<http://www.revues.org>

Document accessible en ligne sur :

<http://tipa.revues.org/1066>

Document généré automatiquement le 22 mars 2014.

© Travaux interdisciplinaires sur la parole et le langage

Juliette Thuilier

Syntaxe du français parlé vs. écrit : le cas de la position de l'adjectif épithète par rapport au nom

1. Introduction

1 En français, l'adjectif peut être antéposé (1-a) ou postposé au nom (1-b).

(1) a. un magnifique tableau

b. un tableau magnifique

2 Dans cet article, nous comparons la syntaxe du français parlé (FP) avec celle du français écrit (FE) à partir de ce phénomène d'alternance de position de l'adjectif épithète dans le SN. Notre objectif est double : premièrement, il s'agit de proposer un modèle statistique permettant de rendre compte de l'alternance de position dans des données attestées à partir d'un ensemble de facteurs linguistiques ; deuxièmement, il s'agit d'étudier ce phénomène à travers la perspective de la comparaison de la syntaxe du FP et du FE.

3 La position de l'adjectif par rapport au nom a été très largement étudiée en linguistique française. De nombreux facteurs ont été identifiés comme participant au choix du placement de l'adjectif. Cependant, notre travail présente l'originalité de confronter les facteurs postulés dans la littérature avec des données attestées extraites de corpus annotés en syntaxe et d'adopter un point de vue nouveau sur ce phénomène, à savoir la comparaison entre FP et FE, ce qui, à notre connaissance n'a jamais été fait auparavant.

4 Nous proposons une approche quantitative pour aborder ce problème, considérant que dans le cas de phénomènes régis par des contraintes non-catégoriques, ce type d'approche permet d'avoir une vue d'ensemble, de repérer et de modéliser les contraintes générales qui interviennent, comme l'ont par exemple montré Bresnan *et al.* (2007) et Bresnan & Ford (2010) dans leur travaux sur l'alternance dative en anglais. Notre approche est complémentaire d'une approche plus qualitative telle que celle menée par C. Benzitoun (ce volume).

5 Dans cet article, nous reviendrons sur un ensemble de facteurs qui ont été trouvés dans la littérature sur le sujet, puis nous présenterons les données en fournissant une description du comportement des facteurs étudiés en fonction de la position de l'adjectif et du mode de production des données (FP vs. FE). Nous exposerons ensuite les outils statistiques utilisés puis la modélisation du phénomène lui-même. Enfin, nous ferons la synthèse des observations et des résultats, avant de conclure.

2. Le problème de la position de l'adjectif

6 Le phénomène de l'alternance de position a été largement étudié en linguistique française selon des approches diverses, comme le montre par exemple l'article de Delomier (1980). Sans passer en revue l'intégralité des facteurs et approches proposés dans la littérature, nous reprenons une liste de facteurs que nous allons étudier à partir de données de corpus.

2.1. Aspects lexicaux

7 Différentes dimensions concernant l'adjectif sont susceptibles d'avoir une influence sur le choix de la position. Nous mentionnerons ici la longueur, la fréquence et la morphologie de l'adjectif. La longueur est une dimension qui intervient de façon générale dans les phénomènes d'ordre des mots et d'alternance de constructions (Hawkins, 1994) et dans les langues SVO comme le français, la tendance est la suivante : *court avant long*. Dans le cas de la position de l'adjectif, les études de Wilmet (1981) et de Forsgren (1978) montrent que c'est la longueur de l'adjectif pris isolément qui est pertinente¹. Ces auteurs constatent dans leurs études de corpus qu'une majorité d'adjectifs monosyllabiques apparaît en antéposition et qu'à partir de

trois syllabes, les adjectifs sont plus fréquemment postposés. Wilmet (1981) observe que plus les adjectifs de son corpus sont fréquents, plus ils ont tendance à être monosyllabiques : 9 monosyllabiques sur les 10 adjectifs les plus fréquents, 38 monosyllabiques sur les 50 plus fréquents et 58 monosyllabiques sur les 100 plus fréquents. Ainsi, la relation entre position et longueur peut être interprétée comme un effet de fréquence, ce que propose d'ailleurs Wilmet (1981).

- 8 Il existe une corrélation forte entre fréquence et position : les adjectifs les plus fréquents ont tendance à être antéposés et les adjectifs moins fréquents sont plutôt postposés. Par exemple, dans les données de Wilmet, les trois lemmes adjectivaux les plus fréquents sont *grand*, *petit* et *bon*. Ils se présentent en antéposition respectivement à 96.8%, 98.7% et 97.5%. Par contraste, les lemmes qui n'apparaissent qu'une fois chez Wilmet (1980), comme *coléreux*, *invalide* ou *ocre*, sont, dans une large majorité, postposés. Cette tendance n'est pas sans exception. Les adjectifs de couleur, *blanc*, *bleu*, *noir* et *rouge*, font partie des 15 lemmes adjectivaux les plus fréquents dans le corpus de Wilmet et apparaissent pourtant à 97.4% en postposition. De même, l'adjectif *plein* fait partie des vingt adjectifs les plus fréquents de son corpus et il n'a pas de préférence pour une position ou l'autre (49% de postposition). Si ce sont bien les lemmes les plus fréquents qui apparaissent en antéposition, le nombre de lemmes se présentant dans cette position est très inférieur au nombre de ceux qui se présentent en postposition. Wilmet (1981) observe que seuls 175 lemmes adjectivaux sont antéposés dans ses données, soit 4.6% des lemmes. Ainsi, les lemmes rencontrés en antéposition sont très fréquents mais peu nombreux. À l'inverse, les adjectifs postposés sont moins fréquents en moyenne, mais on observe une grande variété de lemmes. Une fois le rôle de la fréquence établi, il reste une question : en quoi la fréquence de l'adjectif peut-elle influencer sa position par rapport au nom ? Un élément de réponse peut être apporté par les travaux sur l'effet de la fréquence en diachronie. D'un point de vue diachronique, l'un des effets de la fréquence relevée par Bybee (2009) est le suivant : les mots et les séquences très fréquents sont renforcés dans leur structure morpho-syntaxique et résistent donc aux changements généraux de règles liés à l'évolution de la langue. En ancien français, l'ordre le plus répandu était Adjectif - Nom pour l'ensemble des adjectifs. Cependant, la possibilité de postposer les adjectifs existait déjà. Buridant (2000 : p. 211) explique que la postposition pouvait résulter d'une mise en relief ou de considérations stylistiques (assonance, rime). Sachant que l'ordre prépondérant en ancien français était Adjectif - Nom, on peut émettre l'hypothèse selon laquelle la règle générale, développée en français moderne, qui consiste à postposer l'adjectif au nom, a peu affecté les lemmes adjectivaux très fréquents, ces derniers étant résistants aux changements en raison de leur fréquence élevée. Cette idée est en accord avec le constat de Glatigny (1967) : « *Les adjectifs antéposés appartiennent en très grande majorité au fonds ancien de la langue* » (Glatigny, 1967 : p. 209).
- 9 La morphologie de l'adjectif a également un impact sur sa position. De façon générale, les adjectifs morphologiquement construits ont tendance à préférer la postposition. Hormis les convert², les adjectifs construits sont généralement plus longs que les adjectifs simples et leur longueur favorise donc leur postposition. En plus de ce simple effet de longueur, d'autres caractéristiques des adjectifs construits ont été identifiées comme intervenant dans leur préférence pour la postposition. On note notamment que les déverbaux et dénominaux ainsi que les participes présents et passés peuvent, pour certains, être substitués par un groupe plus complexe ayant une valeur sémantique équivalente. C'est le cas du déverbal *contestable* qui peut être substitué par la relative *que l'on peut contester* (ou *qui peut être contesté*), ainsi que du dénominal semestriel remplaçable par le SP *du semestre*. De même, les participes *alarmant* et *attendu* peuvent être substitués respectivement par *qui alarme* et *qui est attendu*. La possibilité que présente un certain nombre d'adjectifs construits de commuter avec une séquence syntaxiquement plus complexe et obligatoirement postposée coïncide avec une importante tendance à la postposition.
- 10 Nous considérons que l'ensemble des caractéristiques lexicales façonne le comportement de chaque adjectif et aboutit à des préférences plus ou moins marquées pour une position ou pour l'autre.

2.2. Aspects syntaxiques

11 Nous présentons cinq éléments relatifs à la syntaxe du syntagme adjectival (SA) et à celle du syntagme nominal (SN) : la présence d'un dépendant post-adjectival, la présence d'un modifieur pré-adjectival, la présence d'une coordination d'adjectifs, la présence d'autres dépendants du nom et le type de déterminant introduisant le SN.

12 L'unique contrainte catégorique qui impose une position à l'adjectif est une contrainte syntaxique : la présence d'un dépendant post-adjectival. Lorsqu'un adjectif épithète est complété (2) ou modifié (3) par un constituant apparaissant à sa droite, ce dernier doit être obligatoirement postposé (Abeillé & Godard, 1999 ; Blinkenberg, 1933).

- (2) a. un homme fier de son fils
 b. *un fier de son fils homme
 (3) a. une femme belle à croquer
 b. *une belle à croquer femme

13 La présence d'un modifieur pré-adjectival est possible en antéposition comme en postposition.

- (4) a. une très agréable soirée
 b. une soirée très agréable

14 Lorsqu'un adverbe accompagne l'adjectif, la longueur du SA est plus importante. Or, si jusqu'ici nous n'avons considéré que la longueur de l'adjectif, il semble pertinent d'observer la longueur du SA dans les cas où l'adjectif est modifié. Forsgren (1978 : p. 159) observe que, parmi les 559 adjectifs modifiés par un adverbe, 73.4% sont en postposition, alors que seuls 66.1% des adjectifs formant à eux seuls le SA apparaissent en postposition. Ces chiffres suggèrent que la présence d'un adverbe favorise la postposition. Forsgren explique ce constat en partie par l'application du principe *court avant long* au couple Nom / SA. Par ailleurs, les adjectifs présentant une forte préférence lexicale pour une position, voient cette préférence assouplie par la présence d'un adverbe. C'est le cas de l'adjectif *bon* qui présente une forte préférence pour l'antéposition quand il est seul (5-a), mais accepte facilement la postposition une fois modifié (5-b). De même, l'adjectif *familial* préfère très largement la postposition (6-a), mais peut être antéposé quand il est accompagné d'un adverbe (6-b). La présence d'un adverbe tel que *très* est donc un facteur facilitant la mobilité des adjectifs ayant des préférences lexicales très marquées.

- (5) a. un bon poulet / un poulet bon
 b. un très bon poulet / un poulet très bon
 (6) a. la berline familiale / la familiale berline
 b. la berline très familiale / la très familiale berline

15 La coordination d'adjectifs est possible en antéposition comme en postposition.

- (7) a. une table belle et longue
 b. une belle et longue table

16 De la même façon que le modifieur pré-adjectival, la coordination simple d'adjectifs assouplit les préférences lexicales. Dans les SN présentés en (8), on observe que l'adjectif *grand* est plus naturel en antéposition, tandis que, pour l'adjectif *calme*, la postposition est plus naturelle. Une fois les deux adjectifs coordonnés, il apparaît que les deux positions sont tout à fait acceptables.

- (8) a. un grand appartement / un appartement grand
 b. un calme appartement / un appartement calme

c. un grand et calme appartement / un appartement grand et calme

- 17 Abeillé & Godard (1999) signalent le cas des intensionnels *vrai* et *faux*, pour lesquels la postposition est difficile quand l'adjectif est seul (9-a,b). L'exemple (9-c) montre que la coordination de deux adjectifs ayant une très forte préférence lexicale pour l'antéposition peut apparaître en postposition.

(9) a. des faux coupables / des coupables faux
 b. des vrais coupables / des coupables vrais
 c. des vrais ou faux coupables / des coupables vrais ou faux

- 18 La coordination est donc un moyen de neutraliser les préférences lexicales. En ce qui concerne les tendances d'un point de vue quantitatif, on peut poser l'hypothèse que la coordination favorise la postposition en raison de l'importante longueur du SA. Forsgren (1978) observe 72.9% de postposition dans le cas des adjectifs coordonnés. Cela indique une légère préférence pour la postposition, étant donné que la proportion d'adjectifs postposés est de 67.2% dans l'intégralité de son corpus.
- 19 Grevisse & Goosse (2007) évoquent l'existence d'une tendance à équilibrer les éléments à l'intérieur du SN, notamment dans la langue écrite. Par exemple, lorsque le nom est accompagné d'un complément prépositionnel, l'antéposition de l'adjectif permet de ne pas séparer le nom de son complément, comme en (10-c).

(10) a. un recueil récent / un récent recueil
 b. un recueil récent de thèmes grecs
 c. un récent recueil de thèmes grecs

- 20 Plus généralement, la présence d'autres éléments postposés au nom peut favoriser l'antéposition de l'adjectif. Les éléments susceptibles d'apparaître dans le SN sont des relatives (11), des SP (12) ou des adjectifs (13). Dans les SN donnés en (11-c) et (12-c), la présence de la relative ou du SP n'impose pas de position, mais semble rendre l'antéposition de l'adjectif *habituel* un peu plus naturelle.

(11) a. l'air habituel / l'habituel air
 b. l'air habituel que joue Paul
 c. l'habituel air que joue Paul
 (12) a. sa tête habituelle / son habituelle tête
 b. sa tête à coucher dehors habituelle
 c. son habituelle tête à coucher dehors

- 21 Lorsque deux adjectifs sont modificateurs du même nom, comme en (13), la possibilité d'antéposer l'un des deux adjectifs permet d'obtenir un SN plus équilibré.

(13) a. un animal étrange indomptable
 b. un étrange animal indomptable

- 22 Nous posons l'hypothèse que la présence d'autres dépendants du nom favorise l'antéposition de l'adjectif. Aucun des travaux sur corpus que nous avons consultés ne fait d'observation concernant ce point.
- 23 Forsgren (1978) émet l'hypothèse selon laquelle la nature du déterminant est un élément formel permettant de mieux expliquer la position de l'adjectif. Dans ses données, cet auteur observe que l'adjectif a tendance à être postposé quand le déterminant est indéfini (*un, de, des, du*). Il estime que l'antéposition observée avec les indéfinis est principalement due aux adjectifs ayant une valeur évaluative ou quantificatrice. Lorsque le déterminant est défini (*le,*

les, ce, ces, son, ses), il observe que l'antéposition est légèrement favorisée. La proportion d'antéposition est particulièrement élevée pour les déterminants possessifs.

2.3. Aspects sémantiques

24 Une large part de la littérature relative à la position de l'adjectif épithète est consacrée à la dimension sémantique de cette alternance. De façon générale, la sémantique de l'adjectif ainsi que la sémantique de la combinaison de l'adjectif avec le nom est un problème complexe, comme en témoigne la littérature théorique sur le sujet : Kamp (1975), McNally & Kennedy (2008) parmi beaucoup d'autres. En français, ce problème sémantique entre en interaction avec la position variable de l'adjectif, ce qui donne lieu à un problème d'autant plus complexe qu'il existe bien des liens entre position et sens, mais « *il n'y a pas de propriété sémantique générale qui soit liée de manière parfaitement régulière à l'ordre relatif du N et du A* » (Abeillé & Godard, 1999 : p. 12).

25 Il existe un ensemble d'adjectifs pour lequel on admet généralement l'existence de deux homonymes qui apparaissent chacun dans une position déterminée. Dans ces cas, la position de l'adjectif serait déterminée par sa sémantique. Voici une liste d'adjectifs homonymes telle qu'on la trouve fréquemment : *pauvre, pur, propre, simple, brave, cher, faux, sale, seul, vrai, sacré, ancien, commun*. Les exemples suivants illustrent la relation entre sens et position pour certains de ces adjectifs homonymes.

- (14) ce pauvre garçon vs. ce garçon pauvre
- (15) un pur produit vs. un produit pur
- (16) son propre pantalon vs. son pantalon propre
- (17) une simple phrase vs. une phrase simple
- (18) un brave garçon vs. un garçon brave
- (19) une sacrée histoire vs. une histoire sacrée
- (20) un ancien coffre vs. un coffre ancien

26 Pour chaque paire d'exemples, le sens constaté en antéposition est différent de celui observé en postposition d'un point de vue vériconditionnel³ : un *pauvre garçon* peut être riche ; un *pur produit* peut être mélangé ; son *propre pantalon* peut être sale ; une *simple phrase* peut être grammaticalement complexe ; un *brave garçon* peut être un lâche. Cependant, la relation entre sens et position n'est pas aussi tranchée que ces exemples peuvent le laisser paraître. En effet, on peut trouver une interprétation normalement réservée à la postposition en antéposition, comme le montrent les exemples suivants.

- (20) Ils habitent tous Avenue Q, le **plus pauvre quartier** de New-York, à l'opposé des riches Avenues A, B et C.⁴
- (21) Michel enfila sa plus belle chemise de trappeur à gros carreaux rose et vert et son **plus propre pantalon** de velours côtelé bleu cyanosé.⁵
- (22) Devant un public amoureux d'**anciennes chansons**, Patrick a interprété un répertoire très traditionnel.⁶
- (23) Le fondement juridique est discutable, car il s'attaque au **très sacré droit** de propriété.⁷

27 L'inverse, à savoir l'interprétation normalement associée à l'antéposition en postposition, est plus rare, mais il est possible d'en trouver des exemples. Nous donnons trois exemples, pour les adjectifs *ancien, pur* et *brave*.

- (24) L'émergence du mot « folie » correspond à l'inquiétude d'un poète qui cherche encore les moyens de restaurer l'**unité ancienne**.⁸
- (25) Peut-on dire que le sarkozysme est le **produit le plus pur** de la Ve République ?⁹

(26) Le personnage de John C. Reilly, un **type brave et mou** qui se métamorphose au milieu de la scène en gros con macho.¹⁰

28 Nous considérons donc qu'il existe des adjectifs homonymes qui se présentent de façon majoritaire dans une position plutôt que l'autre. Pour ces homonymes, il faut distinguer deux entrées lexicales. Cependant, nous estimons que le sens de chaque homonyme n'est pas catégoriquement corrélé à une position.

29 Certains linguistes ont réduit le problème de la position de l'adjectif à la sémantique liée à la position. Waugh (1977) et Bouchard (1998) ont chacun élaboré une théorie visant à rendre compte de la distribution des adjectifs épithètes en s'appuyant sur le rapport entretenu entre le nom et l'adjectif selon la position. Waugh (1977) adopte une approche structuraliste. Elle considère que chaque adjectif a un sens invariant et que c'est la position ainsi que le nom que l'adjectif modifie qui déterminent l'interprétation de l'adjectif en contexte. Dans le cas de la postposition, la combinaison du nom et de l'adjectif est conçue comme l'intersection de deux parties du discours différentes, alors que l'antéposition implique que le sens lexical du nom est présupposé. Bouchard (1998) adopte un point de vue similaire, estimant, qu'en position pré-nominale, l'adjectif modifie des composants internes au nom, alors que postposé, l'adjectif modifie le nom pris comme un tout et lui assigne une propriété qui ne peut pas être attribuée à un sous-composant du nom. Ce type d'approche, visant à expliquer la position de l'adjectif par une différence de relation entre le nom et l'adjectif, oblige à postuler l'existence d'une différence sémantique systématique selon la position de l'adjectif. Or, une telle généralisation semble connaître des contre-exemples. D'abord, il existe des séquences Nom Adjectif qui ont un sens identique quelle que soit la position de l'adjectif, comme dans les SN présentés en (27) (tirés de Abeillé & Godard, 1999 : p. 28) où il n'y a pas de nuance sémantique permettant de distinguer une interprétation pour le syntagme avec adjectif antéposé ou postposé.

(27) a. un charmant jeune homme
b. un jeune homme charmant

30 De plus, dans certains cas, les différences sémantiques observées pour un couple Nom Adjectif ne sont pas valables pour n'importe quel couple Nom Adjectif. L'exemple (28) fait apparaître que l'interprétation intensificatrice associée à l'antéposition de *gros* avec un nom comme *fumeur*, n'est pas valable en combinaison avec le nom *coiffeur* : *un gros coiffeur* renvoie à un individu qui est coiffeur et qui a la propriété d'être gros. De plus, comme le notent Abeillé & Godard (1999 : p. 13) pour la séquence *un gros fumeur*, « l'antéposition est compatible avec les deux interprétations », à savoir qu'il s'agit soit d'un individu qui fume beaucoup, soit d'un individu qui est fumeur et gros. Cela indique que ce n'est pas la position qui impose une interprétation à la combinaison Nom Adjectif, mais la combinaison elle-même.

(28) a. un gros fumeur
b. un fumeur gros
c. un gros coiffeur
d. un coiffeur gros

31 Les différences sémantiques observées selon la position de l'adjectif ne sont donc pas systématiques et varient très fortement en fonction des items lexicaux combinés.

2.4. Effets de figement

32 L'importance de la combinaison Nom Adjectif se retrouve dans les cas où l'adjectif épithète est impliqué dans une expression figée. Ces cas sont nombreux comme en témoigne le livre de Gross (1996). Nous reproduisons, dans les exemples (29) à (31), des expressions figées extraites de cet ouvrage.

(29) noms

- a. canard boiteux
- b. bête noire
- c. tête brûlée
- d. poids lourd

(30) locutions adjectivales

- a. dans le plus simple appareil
- b. d'âge avancé, d'accès difficile, d'un seul tenant
- c. en chute libre, en bonne posture
- d. sur la corde raide
- e. sous haute surveillance

(31) locutions adverbiales

- a. à bras raccourcis
- b. de longue date
- c. de guerre lasse
- d. tambour battant

- 33 De façon générale, la position de l'adjectif ne peut varier, dans la mesure où elle est fixée par l'expression dans laquelle apparaît l'adjectif en question. Deux remarques s'imposent en ce qui concerne ces effets de figements. Premièrement, la position de l'adjectif dans ces expressions semble respecter une position courante pour chacun des adjectifs. Ainsi, *long* et *bon* sont antéposés, tandis que *battant* ou *las* apparaissent en postposition. Deuxièmement, malgré leur caractère figé, certaines expressions offrent une souplesse quant à la position de l'adjectif. Parmi celles que nous avons énumérées précédemment, il semble que la postposition du SA soit possible dans l'exemple (30-a) comme cela est montré en (32). De même, l'exemple (33) atteste du fait que l'adjectif de (30-b) peut être antéposé, même s'il semble que la version précédemment citée soit préférée.

(32) dans l'appareil le plus simple

(33) de difficile accès¹¹

- 34 Nous observons donc que la position de l'adjectif peut être dépendante d'effets de figement. Dans la plupart des cas, la position est imposée par l'expression utilisée, mais cette position est celle qui est généralement attendue, étant donné les préférences lexicales de chaque adjectif.

2.5. Français parlé (FP) vs. français écrit (FE)

- 35 Comme nous l'avons précisé dans l'introduction, le phénomène d'alternance de position de l'adjectif n'a jamais réellement été envisagé dans une perspective comparatiste entre FP et FE. Les études sur corpus (Wilmet, 1981 ; Forsgren, 1978) portent sur des données écrites. Il est donc nécessaire de s'interroger sur les effets des facteurs listés en FP.
- 36 En premier lieu, dans la mesure où le FP relève en grande partie d'énoncés spontanés et non préparés, on peut supposer que les effets lexicaux vont être plus forts et notamment que la position « préférée » d'un adjectif va être choisie plus fréquemment. Cela reviendrait à dire qu'un temps de production des énoncés plus limité favoriserait le choix de la solution la plus fréquente. En suivant cette idée, on arrive à l'hypothèse selon laquelle la possibilité d'outrepasser les préférences lexicales et de placer l'adjectif dans la position non préférée serait moins importante à l'oral et serait déclenchée par la présence de facteurs favorisant fortement la position non-préférée.
- 37 Dans le même ordre d'idées, les séquences privilégiées Nom Adjectif pourraient être d'autant plus respectées en FP que la construction du SN se fait rapidement et ne peut pas être « travaillée ». Ainsi, on peut penser que les séquences fréquentes sont privilégiées en FP, notamment si l'on adopte le point de vue des grammaires basées sur l'usage (Bybee, 1998 ;

Goldberg, 2006 ; parmi d'autres) selon lequel nous stockons des séquences de plusieurs mots fréquemment rencontrées en plus des mots seuls dans notre lexique mental.

38 Au niveau syntaxique, on peut émettre l'hypothèse selon laquelle l'effet de la longueur du SA, due à la présence de modificateurs ou de coordinations, est plus important en FP. En effet, la tendance à organiser les éléments selon le principe *court avant long* est générale en français¹². Ainsi, comme la parole orale est en général non préparée et spontanée, on peut supposer que les locuteurs suivent des « automatismes » et tendent donc à appliquer ce type de contraintes plus massivement. Si cette hypothèse est vraie, on s'attend à ce que la présence d'un modificateur ou d'une coordination favorise largement la postposition en FP. En ce qui concerne la présence d'autres éléments dans le SN, comme Grevisse & Goosse (2007) le notent, la tendance à équilibrer les éléments de part et d'autre du nom tête semble être plus caractéristique d'une langue écrite plus travaillée. On s'attend donc à ce que ce type de facteurs ait un effet plus fort en FE qu'en FP. Enfin, en ce qui concerne les déterminants, nous ne voyons pas de raisons *a priori* pour que les effets soient différenciés en FE et en FP.

3. Les données de corpus

39 Nous proposons une étude du phénomène à partir de données de FE et de FP. Pour cela nous avons utilisé deux corpus :

40 - French Treebank (FTB), corpus composé d'articles du journal *Le Monde* datant de 1989 à 1993. Il contient environ 870 000 mots annotés pour les parties du discours, la flexion, les lemmes et la constituance (Abeillé *et al.* 2003). Dans le cadre de cette étude, nous utilisons la sous-partie du FTB annotée en fonctions grammaticales (Abeillé & Barrier, 2004) qui contient environ 385 000 mots. Deux courts extraits de ce corpus se trouvent en annexe.

41 - CORAL-ROM (CORAL), corpus de parole spontanée transcrite, distribué par ELRA. Il contient environ 300 000 mots annotés en parties du discours et en lemmes (Cresti & Moneglia, 2005). Deux courts extraits de ce corpus se trouvent en annexe.

42 Etant donné que le FTB est annoté en constituants, nous avons pu extraire automatiquement l'intégralité des adjectifs modifiant une tête nominale. Dans la mesure où nous nous intéressons à la possibilité d'alternance, nous avons écarté les adjectifs présentant un dépendant post-adjectival pour lesquels la postposition est obligatoire. En procédant ainsi, nous avons obtenu près de 14 000 occurrences d'adjectifs épithètes qui représentent 1750 lemmes adjectivaux. Cependant, parmi l'ensemble de ces lemmes, seuls 170 sont attestés dans les deux positions. Etant donné que nous nous intéressons à l'alternance de position, nous avons centré notre travail sur ces 170 adjectifs. Dans le FTB, ils représentent 4986 occurrences, avec 67,1 % d'antéposition. En ce qui concerne CORAL, nous avons extrait automatiquement l'ensemble des occurrences des 170 lemmes, puis nous avons trié manuellement les occurrences correspondant à des adjectifs modifiant une tête nominale. Seuls 130 adjectifs parmi les 170 ont des occurrences remplissant la fonction épithète. Nous avons ainsi obtenu 1626 occurrences, avec 74,3 % d'antéposition. Nos données comptent donc 6612 occurrences d'adjectifs épithètes, avec un taux général d'antéposition de 68,9 %. Ces informations sont récapitulées dans le tableau 1.

Tableau 1. Proportions d'antéposition et de postposition dans les données étudiées

	Total		FTB		CORAL	
Occurrences	6612	100 %	4986	100 %	1626	100 %
Antéposition	4554	68,9 %	3346	67,1 %	1208	74,3 %
Postposition	2058	31,1 %	1640	32,9 %	418	25,7 %
Nombre de lemmes	170		170		130	

43 Il existe une importante variation de la position selon l'adjectif considéré : par exemple, *unique* est antéposé dans 20% des cas, alors que *sérieux* l'est dans 51,4% et *petit* dans 98,6%. De plus, les données de CORAL présentent moins d'alternances de position que celles de FTB : tandis que les 170 adjectifs apparaissent dans les 2 positions dans nos données d'écrit, seuls 56 (43,1% des 130 lemmes) d'entre eux présentent l'alternance à l'oral. Deux éléments

peuvent permettre de comprendre cette différence. Premièrement, étant donné que l'échantillon de données d'oral est trois fois plus petit que celui de données écrites, la probabilité d'avoir des occurrences dans la position non préférée des adjectifs peu enclins à l'alternance est moins élevée. Deuxièmement, cette observation va dans le sens de l'hypothèse que nous avons formulée dans la section 2.5, selon laquelle les adjectifs présentant une préférence pour une position sont plus difficilement rencontrés dans la position non-attendue en FP. Cela signifie que le choix pour la position préférée de l'adjectif constituerait une contrainte plus forte à l'oral qu'à l'écrit.

3.1. Annotations

44 Les données ainsi extraites ont été annotées pour un certain nombre de variables reprenant en grande partie les facteurs détaillés dans la section 2. Nous avons en premier lieu distingué les adjectifs extraits de FP et de FE, en utilisant la variable *MODEDEPROD* qui peut prendre deux valeurs : FE et FP.

3.1.1. Niveau lexical

45 Comme nous l'avons précisé dans la section 2.3, il existe des adjectifs homonymes pour lesquels il est nécessaire de distinguer deux entrées lexicales. Parmi, les 170 adjectifs de notre base de données, nous avons donc différencié deux lemmes pour les adjectifs suivants : *seul*, *simple*, *pur*, *propre* et *ancien*. Nous avons observé que seuls cinq lemmes sur les dix annotés sont attestés dans les deux positions. Nous présentons dans les exemples (33) à (37) des cas illustrant l'alternance de position pour ces cinq lemmes. Nous n'avons retenu que ces cinq lemmes dans nos données.

(33) a. ouais d'un autre orchestre en fait c'est là c'est plus de la variété plus des **trucs anciens** (CORAL)

b. et puis j'avais acheté beaucoup de livres déjà quand j'avais les **anciens appareils** le Elgi (CORAL)

(34) a. on s'est très vite rendu compte qu'il y avait aucune volonté nationale pour maintenir le pays ensemble c'est-à-dire que chacun des chacune des communautés faisait passer son **intérêt propre** avant l'intérêt national (CORAL)

b. on les attache sur nos **propres maillots** ou soit des maillots qui ont été donnés pour pour les foulées de Poitiers (CORAL)

(35) a. c'est la folie de la **folie pure** et simple (CORAL)

b. ça a été un système coopératif de **pure autogestion** qui a complètement fonctionné (CORAL)

(36) a. pour que ce travail ne soit pas une **simple mesure** supplémentaire ils rendront compte de leurs investigations aux élus du conseil général tous les six mois (CORAL)

b. c'est la folie de la **folie pure** et **simple** (CORAL)

(37) a. Pour les **voitures** particulières **seules**, la situation est un peu meilleure (FTB)

b. [...] la création d'une franchise générale pour les deux premières ruptures de travail pourrait ne pas être limitée aux **seules** très petites **entreprises**. (FTB)

46 Nous avons ensuite spécifié pour chaque adjectif sa fréquence, sa longueur et son statut morphologique. Afin d'estimer la longueur et la fréquence nous avons utilisé la base de données

Lexique 3.72 (New, 2006; New *et al.*, 2001). La fréquence correspond à la fréquence par million de lemmes dans les corpus Livres et Sous-Titres de films de Lexique 3.72. La longueur est exprimée en nombre de syllabes. En ce qui concerne la morphologie, nous avons d'abord repéré automatiquement les adjectifs dérivés à l'aide de l'analyseur morphologique DERIF (Namer, 2002), puis procédé à une vérification manuelle. Nous avons classé les 170 lemmes en deux catégories : simple ou construit.

47 En ce qui concerne la fréquence, nous observons les tendances attendues à savoir que les lemmes les plus fréquents apparaissent plus souvent en antéposition que les lemmes moins fréquents. Afin d'illustrer cela, nous avons créé dans les données quatre groupes de taille à peu près équivalente présentés dans le tableau 2.

Tableau 2. Position des adjectifs en fonction de leur fréquence

	Antéposés		Postposés		Totaux		Nombre de lemmes
F >= 524.8	1490	98,7 %	19	1,3 %	1509	100 %	6
FP	632	98,6 %	9	1,4 %	641	100 %	6
FE	858	98,8 %	10	1,2 %	868	100 %	6
524.8 > F >= 197.5	1508	83,5 %	297	16,5 %	1805	100 %	11
FP	328	87 %	49	13 %	377	100 %	11
FE	1180	82,6 %	248	17,4 %	1428	100 %	11
197.5 > F >= 46.2	986	57,8 %	721	42,2 %	1707	100 %	46
FP	191	50,1 %	190	49,9 %	381	100 %	43
FE	795	60 %	531	40 %	1326	100 %	46
46.2 > F	570	35,8 %	1021	64,2 %	1570	100 %	107
FP	57	25,1 %	170	74,9 %	227	100 %	70
FE	513	37,6 %	851	62,4 %	1364	100 %	107

48 De façon logique, les lemmes les plus fréquents sont les moins nombreux. En ce qui concerne le FP, le tableau montre que pour les deux groupes contenant les adjectifs les plus fréquents l'antéposition est largement favorisée, de façon équivalente ou supérieure au FE. En revanche, dans les deux groupes les moins fréquents, la postposition est significativement plus observée en FP qu'en FE. Enfin, ce tableau permet d'observer que les lemmes qui n'apparaissent pas dans le corpus de FP sont les lemmes les moins fréquents, ce qui est tout à fait attendu : 37 des 40 lemmes manquants appartiennent à la tranche de fréquence la plus basse. Ainsi, la forte représentation de l'antéposition dans les données de FP est probablement due à la sous-représentation d'adjectifs susceptibles d'être postposés.

49 Les tendances observées pour la longueur correspondent également à ce qui est attendu (cf. tableau 3) : les adjectifs d'une syllabe sont à plus de 80 % en antéposition, alors que ceux de deux y sont attestés dans près de 70 % des cas et que la proportion tombe à environ 40 % pour les adjectifs de plus de deux syllabes.

Tableau 3. Position des adjectifs en fonction de leur longueur

	Antéposés		Postposés		Totaux		Nombre de lemmes
1 syllabe	1778	82,9 %	367	17,1 %	2145	100 %	44
FP	513	85,6 %	86	14,3 %	599	100 %	36
FE	1265	81,8 %	281	18,2 %	1546	100 %	44
2 syllabes	2322	69,1 %	1037	30,9 %	3359	100 %	65
FP	640	77,2 %	189	22,8 %	829	100 %	55
FE	1682	66,5 %	848	33,5 %	2530	100 %	65
Syllabes > 2	454	41 %	654	59 %	1108	100 %	61
FP	55	27,8 %	143	72,2 %	198	100 %	39
FE	399	43,8 %	511	56,2 %	910	100 %	61

50 Il est intéressant de remarquer que ces tendances sont encore plus marquées pour le FP avec seulement 27 % d'antéposition pour les adjectifs de 3 syllabes et plus. Ainsi, il semble que les caractéristiques de fréquence et de longueur aient un effet plus marqué en FP qu'en FE, notamment en ce qui concerne les adjectifs relativement longs et peu fréquents, pour laquelle la postposition est plus systématique.

51 Enfin, nos données comptent 59 adjectifs morphologiquement construits qui représentent 1035 occurrences. Parmi ces adjectifs, on observe 43,6 % d'antéposition, alors que les adjectifs simples sont antéposés dans 73,6 % des cas. Ainsi les adjectifs construits tendent à apparaître en postposition, ce qui va dans le sens des observations concernant la longueur dans la mesure où la complexité morphologique tend à être positivement corrélée à la longueur des mots. De nouveau, il semble que l'effet observé soit plus net pour les données de FP : les adjectifs complexes (199 occurrences) sont antéposés dans seulement 33,2 % des cas.

52 L'ensemble des observations concernant les caractéristiques lexicales en FP semble montrer que les effets de la fréquence, de la longueur et de la nature morphologique sont plus forts en FP. On remarque notamment que les adjectifs longs, peu fréquents et construits ont tendance à être plus massivement postposés en FP qu'en FE. Cela va dans le sens de l'hypothèse selon laquelle, étant donné le temps réduit dont on dispose pour construire un énoncé en parole spontanée, les locuteurs tendent à suivre plus fréquemment des tendances générales.

3.1.2. Niveau syntaxique

53 La récolte d'informations syntaxiques s'est faite de façon automatique dans le corpus arboré FTB et de façon manuelle dans le corpus CORAL. Nous nous intéressons ici à neuf variables :

- COORD : l'adjectif considéré est coordonné (oui = 1, non=0) ;
- MOD : l'adjectif considéré est modifié par un élément en position pré-adjectival (oui = 1, non=0) ;
- ADJANT : le SN contient au moins un adjectif antéposé en dehors de l'adjectif considéré (oui = 1, non=0) ;
- ADJPOST : le SN contient au moins un adjectif postposé en dehors de l'adjectif considéré (oui = 1, non=0) ;
- SP : le SN contient au moins un syntagme prépositionnel (oui = 1, non=0) ;
- REL : le SN contient au moins une relative (oui = 1, non=0) ;
- DEM : le SN est introduit par un déterminant démonstratif (oui = 1, non=0) ;
- POSS : le SN est introduit par un déterminant possessif (oui = 1, non=0) ;
- DEF : le SN est introduit pas un article défini (oui = 1, non=0) ;

54 En ce qui concerne la configuration du SA, on observe que la présence d'une coordination d'adjectifs favorise nettement la postposition comme cela est attendu : on observe 66,7 % de postposition en présence d'une coordination, alors que les adjectifs postposés ne représentent que 30 % pour le reste des données. De même, la présence d'un modifieur pré-adjectival favorise la postposition avec 57 % des adjectifs modifiés dans cette position, contre seulement 28,9 % des adjectifs non-modifiés. En revanche, ces deux facteurs n'ont pas le même comportement vis-à-vis de la distinction entre FP et FE : tandis que l'effet de la coordination ne semble pas être sensible au mode de production des données, la présence d'un modifieur semble favoriser plus nettement la postposition en FP. En effet, 68,1 % des adjectifs modifiés sont postposés en FP alors qu'en FE, on n'en observe que 52,1 % dans cette position.

55 De façon générale, la présence d'autres éléments que l'adjectif considéré dans le SN semble favoriser l'antéposition de l'adjectif. En ce qui concerne les variables ADJPOST, SP et REL, la tendance observée correspond aux attentes si l'on estime que l'on tend à équilibrer le nombre d'éléments de part et d'autre du nom tête. Plus précisément, on observe 78,6 % d'antéposition en présence d'un SP, 74,3 % en présence d'une relative et 78,6 % quand il y a déjà un adjectif en postposition. Ces observations sont valables en FP comme en FE. Cependant, l'effet de SP et ADJPOST semble plus fort en FE, alors que l'effet de REL apparaît plus important en FP.

56 Contrairement aux trois variables que nous venons de commenter, l'effet dû à la présence d'un adjectif en antéposition est inverse à celui attendu si l'on se réfère à l'idée d'équilibre dans le SN. En effet, la variable ADJANT favorise légèrement l'antéposition avec 72,5 % d'adjectifs

antéposés, contre 68,7 % dans les cas où il n'y a pas d'adjectif antéposé. Cette tendance générale ne semble valable que dans le cas du FE : en FP, on observe une préférence plutôt légère pour la postposition en présence d'un adjectif antéposé (env. 29 % de postposition pour ADJANT = 1, contre env. 25 % pour ADJANT = 0).

57 En ce qui concerne la nature du déterminant en début de SN, les données étudiées montrent que la présence d'un démonstratif ou d'un possessif semble favoriser fortement l'antéposition de l'adjectif : 80,4 % et 79,8 % d'antéposés respectivement pour les démonstratifs et les possessifs. Cependant, l'influence de ces déterminants semble varier selon le mode de production des données : la tendance à antéposer l'adjectif en présence d'un possessif est particulièrement importante en FP avec 88,6 % d'antéposition, tandis que la présence du démonstratif n'a un effet qu'en FE. Enfin, la présence d'un article défini ne semble pas avoir d'effet d'après les proportions observées dans les données (68,4 % d'antéposés avec article défini et 69,2% sans).

3.1.3. Combinaison du nom et de l'adjectif

58 Le choix de la position d'un adjectif peut être influencé par le nom avec lequel il se combine, soit pour des raisons sémantiques, soit parce que l'ordre choisi correspond à une façon conventionnelle de dire les choses. Nous avons donc cherché à prendre en compte le couple Adjectif Nom et leur position réciproque.

59 Pour obtenir une estimation statistique de la relation d'association entre l'item adjectival et l'item nominal, nous avons employé une métrique couramment utilisée pour identifier les collocations : le #² (Manning & Schütze, 1999). Etant donné que le calcul des collocations nécessite un gros volume de données, nous avons eu recours au corpus de l'Est-Républicain (corpus journalistique de 148 millions de mots, disponible sur le site du CNRTL¹³). Nous avons procédé à l'extraction des bigrammes de lemmes Adjectif - Nom et Nom - Adjectif et créé deux listes, l'une servant à identifier les collocations avec adjectif antéposé, l'autre pour les collocations avec adjectif postposé. À partir de ces listes, nous avons calculé la valeur de #² de chaque bigramme. Pour chaque couple Nom - Adjectif présent dans nos données, nous avons spécifié les scores d'association #² pour chaque ordre possible : nom suivi d'adjectif, adjectif suivi de nom. Nous avons ainsi obtenu deux variables :

60 - COLLOCAN : score de #² pour la séquence ordonnée Adjectif - Nom (échelle logarithmique) ;
 61 - COLLOCNA : score de #² pour la séquence ordonnée Nom - Adjectif (échelle logarithmique).
 62 Notons que dans le cas des couples Nom - Adjectif qui n'ont pas été rencontrés dans le corpus Est-Républicain dans un ordre ou dans l'autre, nous avons mis la valeur zéro pour la variable concernée.

62 De plus, afin de réduire l'étendue des valeurs des deux variables et ainsi réduire l'effet des valeurs extrêmes, nous avons effectué une transformation logarithmique. Plus précisément, les variables COLLOCAN et COLLOCNA ont pour valeur $\log(\#^2 + 1)$, afin d'éviter d'avoir des valeurs infinies ($= \log(0)$) dans le cas des couples Nom - Adjectif ayant un #² égal à zéro. Pour un couple ordonné, plus la valeur de COLLOCAN (respectivement COLLOCNA) est élevée, plus la force d'association du nom et de l'adjectif en antéposition (respectivement postposition) est importante.

63 Pour évaluer l'effet de la combinaison du nom et de l'adjectif, nous observons les proportions d'antéposition et de postposition selon que les variables sont égales à zéro ou supérieures à zéro.

64 Les chiffres du tableau 4 montrent qu'un score d'association supérieur à zéro favorise l'antéposition dans le cas de COLLOCAN et la postposition dans le cas de COLLOCNA. En repérant les couples Nom - Adjectif avec un fort degré d'association, ces deux variables rendent compte de l'importance du nom qui est combiné avec l'adjectif.

Tableau 4. Position des adjectifs en fonction des scores d'association COLLOCAN et COLLOCNA

	Antéposés		Postposés		Totaux	
COLLOCAN = 0	555	32,7 %	1141	67,3 %	1696	100 %
FP	117	34,8 %	219	65,2 %	336	100 %
FE	438	67,8 %	922	32,2 %	1360	100 %
COLLOCAN > 0	3999	81,3 %	917	18,7 %	4916	100 %

	FP	1091	84,6 %	199	15,4 %	1290	100 %
	FE	2908	80,2 %	718	19,8 %	3626	100 %
	COLLOCNA = 0	2748	84 %	523	16 %	3272	100 %
	FP	745	86,9 %	112	13,1 %	857	100 %
	FE	2004	83 %	411	17 %	2415	100 %
	COLLOCNA > 0	1805	54 %	1535	46 %	3340	100 %
	FP	463	60,2 %	306	39,8 %	769	100 %
	FE	1342	52,2 %	1229	47,8 %	2571	100 %

65 Les observations faites à partir de pourcentage dans le corpus permettent de détecter de grandes tendances. Cependant, l'utilisation de ces mesures présente un certain nombre de limites. Nous avons vu que les aspects lexicaux semblent avoir une grande importance dans le choix de l'adjectif. Comment s'assurer alors que l'effet d'une contrainte syntaxique n'est pas un artefact lié à un ou deux adjectifs ayant un comportement particulier ? Par exemple, l'adjectif *propre* est dans la très grande majorité des cas antéposé dans nos données et il est majoritairement attesté dans des séquences du type *son propre X*. Or nous avons constaté que le déterminant possessif favorise fortement l'antéposition. Cet effet ne serait-il pas simplement dû au fait que la plupart des possessifs s'observe avec l'adjectif *propre* ? Dans ce cas-là, ce ne serait pas le déterminant possessif qui favoriserait l'antéposition mais la séquence entière et l'adjectif concerné. De même, on pourrait se demander si les effets de certaines contraintes syntaxiques ne sont pas la conséquence d'autres contraintes syntaxiques.

66 C'est une des raisons pour lesquelles nous utilisons une modélisation statistique multifactorielle qui permet de prendre en compte simultanément l'effet de l'ensemble des contraintes. Une autre raison, en lien avec la première, est que la modélisation permet de quantifier le rôle de chaque contrainte et de modéliser leurs interactions, ce qui va bien au-delà de ce que permettent de simples pourcentages. Enfin, l'utilisation de méthodes de statistique inférentielle permet de négliger les variations liées aux contingences des corpus utilisés et donc de produire des résultats valables au-delà des données étudiées.

4. Modélisation statistique multifactorielle

67 L'approche proposée est quantitative et repose sur la modélisation statistique d'une importante collection de données attestées. Pour cela, nous utilisons la régression logistique à effets-mixtes (Agresti, 2007 ; Gelman & Hill, 2006).

4.1. Aspects techniques

68 La régression logistique à effets-mixtes¹⁴ permet de modéliser le comportement d'une variable binaire. Dans notre cas, cela permet d'estimer la probabilité que l'antéposition soit choisie en fonction des variables prédictives présentées dans la section précédente. Ainsi, nous allons estimer pour chaque occurrence d'adjectif la probabilité d'antéposition en fonction des variables telles que déterminant possessif, déterminant démonstratif, présence d'un modificateur pré-adjectival etc.

69 Un des avantages de ce type de modèle statistique est qu'il est prédictif, dans le sens où l'on peut construire un modèle sur un ensemble de données et utiliser ce modèle pour prédire le choix de la position de l'adjectif sur des données nouvelles. De cette façon, nous pouvons évaluer à quel point le modèle généralise à partir de l'ensemble de données d'entraînement.

70 La construction du modèle consiste à estimer les coefficients qui sont associés à chaque variable prédictrice. Chaque coefficient peut être interprété comme la préférence de la variable qu'il accompagne : dans le cas d'une variable ayant uniquement des valeurs positives (ce qui est le cas de l'ensemble de nos variables prédictrices), un coefficient positif indique une préférence pour l'antéposition et un coefficient négatif une préférence pour la postposition. De plus, nous exploitons la possibilité de tester la significativité des interactions entre plusieurs variables au sein du modèle afin d'évaluer quels facteurs présentent un comportement différent selon le mode de production des données.

71 Etant donné que chaque adjectif présente un comportement spécifique (cf. section 3), nous considérons qu'il existe une variation au niveau de la position de l'adjectif selon le lemme

utilisé. Cette variation est due aux caractéristiques de l'adjectif, telles que sa longueur, sa fréquence, sa morphologie, sa sémantique etc. Au-delà des variables prédictives que nous venons de décrire (aussi appelées effets fixes), les modèles à effets-mixtes permettent de prendre en compte la variation dans les données grâce aux effets aléatoires. Nous utilisons donc ces effets aléatoires pour capter la variation liée aux lemmes adjectivaux. Cela signifie que la variable indiquant le lemme adjectival (LEM) est un effet aléatoire dans notre modèle afin de rendre compte des idiosyncrasies relatives à chaque adjectif. Chaque valeur de LEM, c'est-à-dire chaque lemme, constitue un groupe dans les données, auquel est assigné, dans le modèle, un coefficient distribué normalement et variant aléatoirement. Ainsi, en associant chaque valeur de l'effet aléatoire avec un coefficient spécifique, on rend compte des différents comportements selon l'adjectif utilisé. De cette façon, les variations inter-adjectifs sont prises en compte et nous pouvons observer les effets des variables syntaxiques au-delà des spécificités lexicales.

72 A partir des données décrites dans la section 3, nous avons construit un modèle contenant douze effets fixes (MODEDEPROD, MOD, COORD, SP, REL, ADJANT, ADJPOST, DEF, POSS, DEM, COLLOCAN et COLLOCNA), un effet aléatoire (LEM) et les interactions entre la variable MODEDEPROD et chacune des autres variables prédictives, soit onze interactions. Afin de rendre le modèle le plus compact possible, nous avons éliminé les variables et les interactions qui n'avaient pas un effet significatif sur la probabilité d'antéposition¹⁵. Ainsi, l'ensemble des variables prédictives ainsi que quatre interactions ont été conservées dans le modèle (interactions entre MODEDEPROD et DEM, MODEDEPROD et POSS, MODEDEPROD et ADJANT, MODEDEPROD et MOD.)

Tableau 5 : Modèle statistique obtenu à partir des données annotées

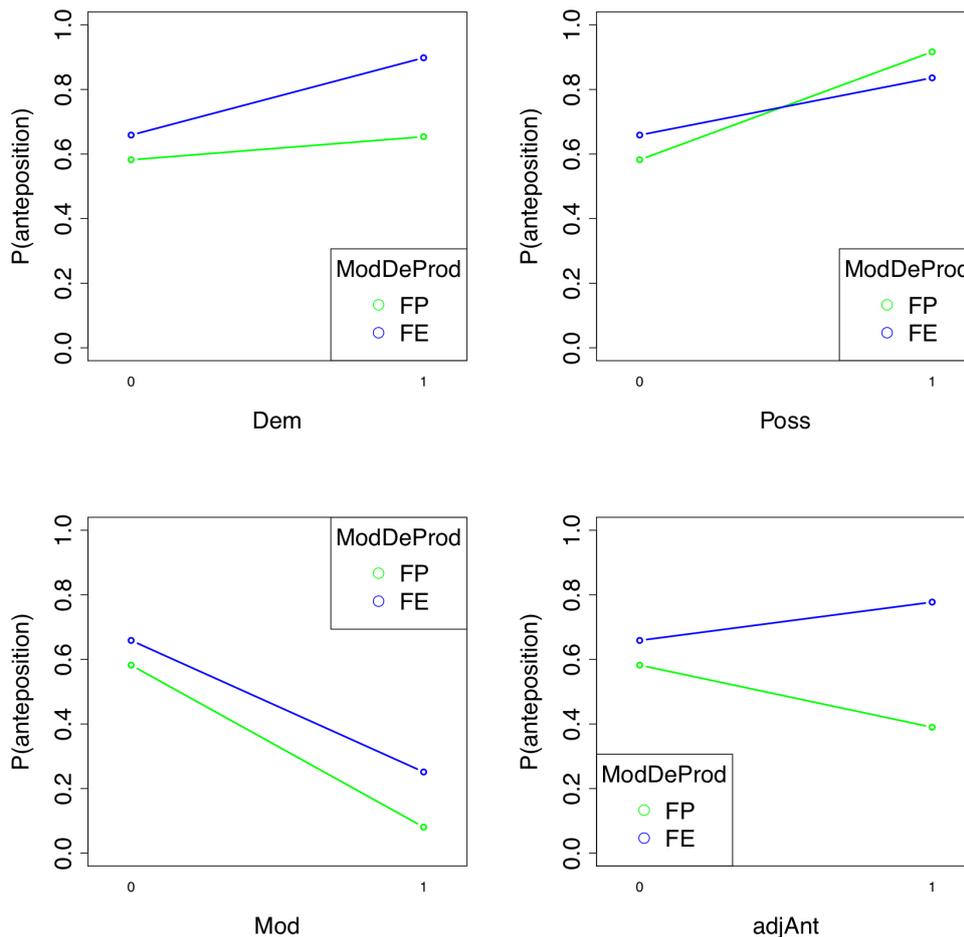
Effets aléatoires :				
	Groupes	Nom	Variance	Ecart-type
	LEM	(Intercept)	2.4156	1.5542
Nombre d'obs. : 6612 ; groupes : LEM, 170				
Effets fixes :				
		Estimation	p.value	
	(Intercept)	-0.68009	0.000336	
	MODEDEPROD=FE	0.32568	0.016677	
	DEM = 1	0.30327	0.545139	
	POSS = 1	2.06531	3.82e-5	
	DEF = 1	0.36736	0.000608	
	COORD = 1	-1.22670	4.22e-6	
	MOD = 1	-2.76641	1.50e-14	
	ADJANT = 1	-0.77964	0.207560	
	ADJPOST = 1	0.58592	0.000140	
	SP = 1	0.84944	4.54e-16	
	REL = 1	0.70661	0.000864	
	COLLOCAN	0.37753	<2e-16	
	COLLOCNA	-0.44393	<2e-16	
	ADJANTADJANT=1:MODEDEPROD=FE	1.37298	0.041050	
	DEM=1:MODEDEPROD=FE	1.21472	0.031642	
	POSS=1:MODEDEPROD=FE	-1.09463	0.036655	
	MOD=1:MODEDEPROD=FE	1.01762	0.00997	

4.2. Le modèle

73 Le modèle statistique obtenu est présenté dans le tableau 5. Pour l'effet aléatoire, nous donnons l'écart-type et la variance de la distribution normale ; pour les effets fixes, nous présentons le coefficient estimé et la *p-value* témoignant du fait que les coefficients associés aux variables sont significativement différents de zéro, autrement dit que la variable a un effet significatif. L'interaction de variables 1 et 2 est indiquée en utilisant la notation « Variable 1 : Variable 2 ».

- 74 Comme nous l'avons précisé, le signe du coefficient permet d'interpréter les préférences des variables prédictrices. Ainsi, si l'on considère les variables non impliquées dans des interactions, la postposition est favorisée par la présence d'une coordination dans le SA. De plus, plus le score d'association du nom et de l'adjectif est élevé, plus la postposition a tendance à être choisie pour COLLOCNA et plus l'antéposition a tendance à être choisie dans le cas de COLLOCAN. L'antéposition est également favorisée par la présence d'autres dépendants dans le SN (relative, SP ou adjectif postposé) et la présence d'un déterminant défini.
- 75 Afin d'interpréter les coefficients associés aux variables impliquées dans des interactions, il est nécessaire de prendre en compte simultanément les coefficients associés à l'interaction et aux deux variables prenant part à cette interaction. Nous présentons sous forme graphique les effets partiels des interactions (figure 1). Chaque graphique représente l'effet d'une interaction sur la probabilité d'antéposition de l'adjectif. Ainsi, on observe que les déterminants démonstratifs et possessifs favorisent fortement l'antéposition. Cependant l'effet du démonstratif est significativement moins fort en FP qu'en FE, et inversement, l'effet du possessif est plus fort en FP qu'en FE. En présence d'un modifieur pré-adjectival, la postposition est privilégiée et ce, de façon plus marquée en FP qu'en FE. Enfin, en ce qui concerne la présence d'un adjectif antéposé, les résultats de la modélisation vont dans le même sens que les observations faites sur la base de proportions : ADJANT favorise l'antéposition en FE et la postposition en FP.
- 76 Ainsi, les tendances observées sur la base de proportions se retrouvent pour la plupart dans la modélisation. Cependant, le type d'approche proposée permet de faire la distinction entre les cas où le mode de production a une réelle influence (cf. les quatre variables impliquées dans les interactions) et ceux où les proportions ne font état que d'observations non généralisables (cf. section 3, effet du FE sur SP et ADJPOST). Cette approche a également permis de faire émerger l'effet de l'article défini qui ne se manifestait pas en termes de proportions.
- 77 Le modèle statistique obtenu permet d'estimer les préférences de chaque adjectif. En effet, chaque valeur de la variable LEM, effet aléatoire dans le modèle, se voit attribuer un coefficient (appelé intercept aléatoire) qui s'interprète de la même façon que les coefficients associés aux effets fixes : un coefficient positif marque une préférence pour l'antéposition et un coefficient négatif pour la postposition. De plus, ces intercepts aléatoires peuvent se comparer entre eux : plus un intercept est élevé, plus la préférence pour l'antéposition est forte, et inversement, plus un intercept est bas, plus la postposition est favorisée par le lemme. L'estimation des « préférences lexicales » dans le cadre de cette modélisation statistique est plus précise que de simples proportions dans la mesure où les contraintes syntaxiques sont prises en compte simultanément. Nous présentons dans la figure 2 les coefficients estimés pour les 28 adjectifs les plus fréquents dans nos données (sur la figure, l'adjectif est accompagné de sa fréquence dans les données). On observe par exemple que les adjectifs *bon* et *grand* ont une très forte préférence pour l'antéposition, alors que *britannique* et *total* préfèrent massivement la postposition. Les adjectifs *important* et *différent* ont une légère préférence pour respectivement l'antéposition et la postposition. Enfin, l'adjectif *long* ne présente pas de préférence particulière.

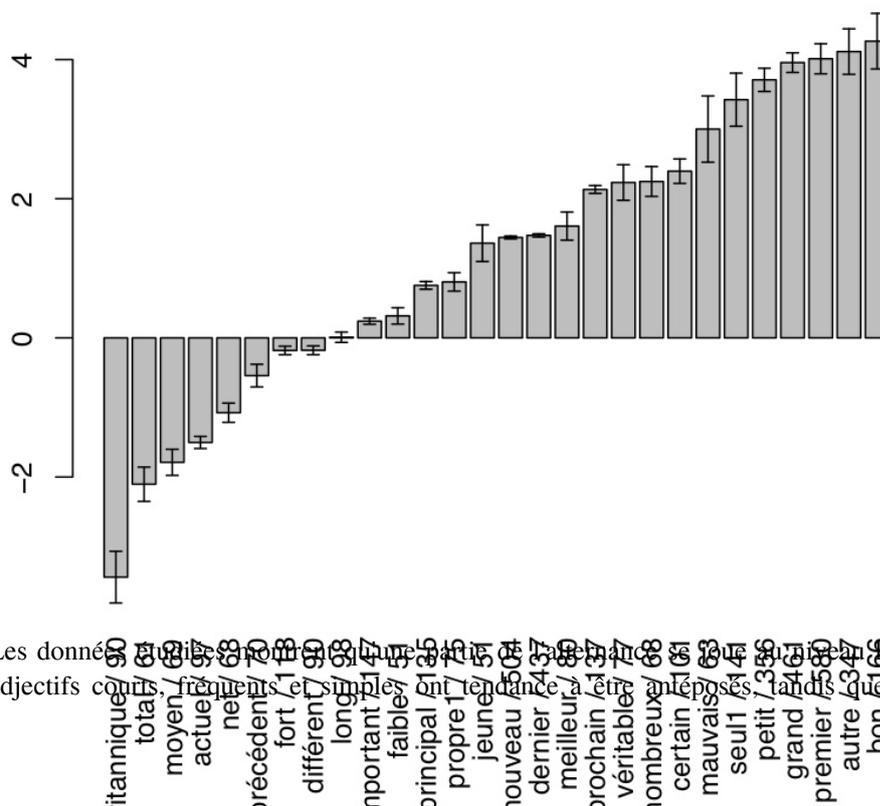
Figure 1. Effets partiels des interactions dans le modèle



Figure

1. Effets partiels des interactions dans le modèle

Figure 2 : Intercepts aléatoires associés aux lemmes les plus fréquents dans la modélisation



78 Les données sont issues de la base de données de la Banque de Français (Banque de Français, 2008) et ont été filtrées pour ne retenir que les adjectifs courts (moins de 10 lettres) et les longs (plus de 10 lettres).

5. Synthèse

rare et construits tendent à être postposés. On observe également qu'en FP ces tendances sont renforcées et notamment que la postposition est plus massivement choisie dans le cas d'adjectifs qui présentent les caractéristiques favorisant cette position. Pour expliquer ces résultats, nous avons émis l'hypothèse selon laquelle dans le contexte de la parole spontanée, les locuteurs tendent à suivre plus fréquemment des tendances générales.

79 Au niveau syntaxique, nous observons que, de façon générale, la nature du déterminant influence la position : les déterminants démonstratifs, possessifs et les articles définis favorisent l'antéposition. La présence de propositions relatives, de SP ou d'un autre adjectif après le nom favorise également l'antéposition. De plus, les SA contenant des adjectifs coordonnés ou des modificateurs adverbiaux ont tendance à être postposés, ce qui va dans le sens de la tendance à placer les constituants plus complexes après le nom. Enfin, on observe que plus le nom et l'adjectif ont tendance à former une collocation, plus l'ordre de cette collocation tend à être respecté. Cela signifie qu'en plus des spécificités lexicales, il est nécessaire de prendre en compte le nom pour rendre compte de la position de l'adjectif. Il est important de noter que l'alternance de position de l'adjectif semble être un phénomène unifié au niveau syntaxique, dans la mesure où l'ensemble de ces facteurs ont des effets communs en FE et FP. Seule la présence d'un adjectif en antéposition a des conséquences différentes : en FP, elle favorise l'antéposition, tandis qu'en FE, elle favorise la postposition. Trois autres interactions avec le mode de production des énoncés apparaissent significatives dans le modèle : les déterminants possessifs et démonstratifs ainsi que la présence d'un modificateur pré-verbal. Alors qu'en FE, le démonstratif favorise très nettement l'antéposition, ce facteur ne joue qu'un rôle très faible en FP. On peut remarquer que l'antéposition de l'adjectif en présence d'un démonstratif est, dans certains cas, caractéristique d'une langue travaillée, plus représentée à l'écrit qu'à l'oral, comme le montrent les deux phrases suivantes extraites du FTB.

(38) Au centre de **ce gigantesque bouleversement**, l'enjeu primordial restera cependant l'emploi.

(39) A telle enseigne que c'est pratiquement toute l'industrie des biens d'équipement qui se trouve, peu ou prou, engagée dans **ce périlleux processus** : pour attirer une clientèle qui se dérobe, les vendeurs sont amenés à consentir des rabais de plus en plus importants.

80 Par ailleurs, les tendances générales observées pour le possessif et pour la présence du modificateur sont renforcées en FP : l'antéposition est fortement favorisée lorsque le SN est introduit par un possessif et la présence d'un modificateur déclenche la postposition du SA dans la grande majorité des cas.

81 Le fait que l'effet de la présence d'un modificateur adverbial soit plus important en FP va dans le sens de l'hypothèse, déjà formulée, selon laquelle, comme la production orale est le plus souvent non préparée, les locuteurs ont tendance à suivre plus massivement des tendances générales telles que *court avant long*. Cependant, si les locuteurs se soumettaient de façon plus massive à la contrainte de longueur en FP, on devrait observer la même tendance pour la coordination. Or, ce n'est pas le cas. Notons tout de même qu'il se peut que l'effet attendu ne soit pas observé en raison du nombre limité de données relatives à la coordination en FP (23 seulement)¹⁶. En ce qui concerne l'effet du possessif, nous ne voyons pas pour le moment ce qui pourrait expliquer une préférence plus marquée en FP. Enfin, il est important de remarquer que l'hypothèse selon laquelle la présence d'autres éléments dans le SN serait plus susceptible d'influencer la position de l'adjectif en FE, n'est pas vérifiée dans les données. Nos résultats montrent au contraire que l'influence des dépendants du nom est active en FP comme en FE.

6. Conclusion

82 Ce travail apporte des données et des résultats sur un aspect peu abordé de la question de la place de l'adjectif, à savoir la comparaison entre FP et FE. Les résultats obtenus permettent d'établir que les effets de la plupart des facteurs syntaxiques étudiés vont dans le même sens en FP et en FE. Cependant, la syntaxe du FP semble différer de celle du FE, dans la mesure où une partie des facteurs présente une intensité différente selon le mode de production du discours.

- 83 Pour obtenir les résultats présentés, nous avons procédé à une analyse quantitative de la position de l'adjectif en FE et en FP. La méthodologie utilisée permet de prendre en compte simultanément les variables syntaxiques et les spécificités lexicales. Plus précisément, l'un des intérêts de l'utilisation de méthodes statistiques telles que les modèles à effets-mixtes dans le cadre d'un travail de modélisation de phénomène d'ordre des mots tient à la possibilité de capter deux dimensions en tension : la dimension relative au lexique et la dimension relative aux contraintes générales du système de la langue. En associant des effets aléatoires aux adjectifs, le modèle à effets mixtes tient compte de la structuration des données autour de chaque lemme, et traduit ainsi la dimension locale et spécifique apportée par chaque item lexical. Simultanément, ce même modèle rend compte, grâce aux effets fixes, de tendances générales qui traversent les données. Les modèles à effets mixtes semblent donc être adéquats pour capter la complexité des données de la langue et l'imbrication entre le niveau lexical et le niveau systémique. De plus, en utilisant les interactions de variables, il est possible de comparer l'effet des contraintes générales dans différentes « variantes » d'une même langue : anglais américain et anglais australien, comme dans Bresnan & Ford (2010) ; français parlé et français écrit, comme dans notre étude.
- 84 Le travail proposé ouvre un certain nombre de prolongements. Premièrement, nous avons limité notre étude à 170 adjectifs. Il faudrait donc étendre le travail à d'autres adjectifs. Deuxièmement, étant donné que le nombre de données orales est trois fois inférieur à celui des données écrites, il est nécessaire de réunir et d'annoter un nombre beaucoup plus important de données afin de s'assurer de la validité des résultats et hypothèses concernant le FP. Enfin, nous avons pris en compte le nom avec lequel est combiné l'adjectif. Cependant, comme le montrent bien Benzitoun *et al.* (2010) avec l'exemple de l'adjectif *prochain*, il existe un certain nombre de cas dans lesquels l'alternance est dirigée par un contexte plus précis, spécifique à un adjectif. Il serait donc intéressant de trouver un moyen d'introduire ce type de résultats au niveau de la modélisation, afin de mesurer comment interagissent les différents niveaux : lexique, combinaisons nom-adjectif, constructions spécifiques à un adjectif, contraintes syntaxiques générales.

Bibliographie

- Abeillé, A., L. Clément & F. Toussanel (2003) Building a treebank for French. In Abeillé, A. (ed), *Treebanks*, Dordrecht : Kluwer.
- Abeillé, A. & D. Godard (2004) De la légèreté en syntaxe, *Bulletin de la Société de Linguistique de Paris*, XCIX(1), p. 69-106.
- Abeillé A. & Barrier N. 2004. Enriching a French treebank. In Proceedings of Language Resources and Evaluation Conference (LREC), Lisbonne.
- Agresti, A. (2007) *An Introduction to Categorical Data Analysis*, Wiley.
- Benzitoun C., S. Bresson, L. Budzinski, J.-M. Debaisieux et K. Holzheimer (2010) Quand un corpus rencontre un adjectif du troisième type. Etude distributionnelle de *prochain*, *Corpus*, 9, p. 245-264.
- Bresnan, J., A. Cueni, T. Nikitina, and H. Baayen. 2007. Predicting the dative alternation. In Boume, Kraemer, and Zwarts (eds.), *Cognitive Foundations of Interpretation*, Amsterdam: Royal Netherlands Academy of Science.
- Bresnan J. & M. Ford (2010) Predicting syntax: Processing dative constructions in American and Australian varieties of English, *Language*, 86(1), p. 186-213.
- Cresti E. & M. Moneglia (2005) *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam: J. Benjamins.
- Blinkenberg A. (1928) *L'ordre des mots en français moderne. Première partie*. Copenhague : Host & Son.
- Blinkenberg A. (1933) *L'ordre des mots en français moderne. Deuxième partie*, Copenhague : Levin & Munksgaard.
- Bouchard D. (1998) The distribution and interpretation of adjectives in French: A consequence of bare phrase structure, *Probus*, 10(2), p. 139-183.

- Buridant C. (2000) *Grammaire nouvelle de l'ancien français*, Sedes.
- Bybee J. (2009) Language universals and usage based theory, In M. Christiansen, C. Collins & S. Edelman (eds), *Language Universals*, Oxford: Oxford University Press, p. 17-40.
- Bybee, J (1998) The emergent lexicon. In Gruber, M.C., Higgins, D., Olson, K.S., Wycsocki, T. (eds), *CLS 34: The panels*, University of Chicago, Chicago Linguistic Society
- Delomier D. (1980) La place de l'adjectif en français, bilan des points de vue et théories du XXe siècle, *Cahiers de lexicologie*, 37, p. 5-34.
- Forsgren M. (1978) La place de l'adjectif épithète en français contemporain, étude quantitative et sémantique. Stockholm: Almqvist & Wiksell.
- Gelman, A. & J. Hill (2006) *Data Analysis Using Regression and Multilevel/ Hierarchical Models*, Cambridge: Cambridge University Press
- Glatigny M. (1967) La place des adjectifs épithètes dans deux oeuvres de Nerval, *Le français moderne*, 35(1).
- Goldberg, A. (2006) *Constructions at Work: the nature of generalization in language*, Oxford University Press.
- Grevisse M. & A. Goosse, (2007) *Le bon usage*, 14ème édition : De Boeck Université.
- Hawkins J. (1994) *A Performance Theory of Order and Constituency*, Cambridge: Cambridge University Press.
- Howell D. C. (1998) *Méthodes statistiques en sciences humaines*. Série Internationale. Paris: De Boeck.
- Kamp H. (1975) Two theories about adjectives, In E. Keenan (ed), *Formal Semantics of Natural Language*, Cambridge : Cambridge University Press, p. 123-155.
- Gross G. (1996) *Les expressions figées en français. Noms composés et autres locutions*, Paris : Ophrys.
- Manning, C. D. & H. Schütze (1999) *Foundations of Statistical Natural Language Processing*, Cambridge, MA : The MIT Press.
- McNally L. & C. Kennedy (2008) *Adjectives and Adverbs: Syntax, Semantics and Discourse*, Oxford University Press.
- New B. (2006) Lexique 3 : une nouvelle base de données lexicales. In P. Mertens, C. Fairon, A. Dister & P. Watrin, (eds), Actes de la conférence sur le Traitement Automatique des Langues Naturelles, Universitaires de Louvain.
- New B., C. Pallier, L. Ferrand & R. Matos (2001) Une base de données lexicales du français contemporain sur internet : LEXIQUE, *L'année psychologique*, 101, p. 447-462. <http://www.lexique.org>
- Thuilier, J. (2012) *Contraintes préférentielles et ordre des mots en français*, Thèse de doctorat, Université Paris Diderot. Disponible sur : <http://tel.archives-ouvertes.fr/tel-00781228/>
- Thuilier J., G. Fox & B. Crabbé (2010) Fréquence, longueur et préférences lexicales dans le choix de la position de l'adjectif épithète en français, in F. Neveu, V. M. Toke, T. Klingler, J. Durand, L. Mondada & S. Prévost (eds), Actes du 2ème Congrès Mondial de Linguistique Française 2010 (CMLF 2010), Nouvelle-Orléans.
- Tribout D. (2010) Les conversions de nom à verbe et de verbe à nom en français. Thèse de doctorat, Université Paris Diderot. Disponible sur : <http://www.llf.cnrs.fr/Gens/Tribout/publis/these-tribout-2010.pdf>
- Waugh L. R. (1977) *A semantic Analysis of Word Order: Position of the Adjective in French*. Leuven: E. J. Brill.
- Wilmet M. (1980) Antéposition et postposition de l'épithète qualificative en français contemporain : matériaux, *Travaux de linguistique*, 7, p. 179-201.
- Wilmet M. (1981) La place de l'épithète qualificative en français contemporain : étude grammaticale et stylistique, *Revue de linguistique romane*, 45, p. 17-73.

Annexe

Extraits du corpus French Treebank

Extrait 1

Un mois plus tard , après avoir admis de payer lui_-même la coûteuse amende , qui aurait dévoré les fonds_propres de cellier , le holding faisait même état d' une recapitalisation de sa filiale , pour un montant de 30 millions de francs , et de l' obtention d' importants crédits à court terme (125 millions de francs) . Seuls ces crédits semblent avoir été obtenus , en_raison_de la multiplication des démarches auprès_des banques et des pouvoirs_publics .

Les autres mesures n' ont pu être appliquées , tant la direction a semblé paralysée : l' ensemble du personnel , cadres_supérieurs compris , a même observé huit jours de grève , du 9 au 17 décembre , pour demander - et obtenir - le départ du PDG , jugé personnellement responsable de tous ces maux !

En 1968 , la société n' avait connu qu' une heure d' arrêt_de_travail .

Extrait 2

Dans des cantons où tout le monde se connaît , il n' est pas question un instant de ne_pas être là .

" La coordination analyse les choses autrement .

" Nous n' avons pas les chefs , mais nous avons les troupes " , dit , en parlant du syndicalisme officiel , M. Jacques Laigneau , l' un de ses pères fondateurs .

Quant_à l' opération " blocus de Paris " , il refuse le terme d' échec : " l' agressivité des forces policières était telle qu' il y aurait eu des morts .

Les gars sur les tracteurs parlaient d' aller chercher les fusils .

Nous avons voulu éviter le pire .

" quant_à M. Philippe Arnaud , l' autre responsable national de la coordination , il précise : " Nous sommes plus déterminés que jamais " .

leur solidarité et ils montent à Paris voter les décisions qui nous condamnent .

Extraits du corpus CORAL-Rom

Extrait 1

en &plu [/] il faut tournasser / c' est-à-dire faire le pied d' un bol par exemple // # \$ ensuite il faut &euh # &euh attendre que ça sèche // \$ ensuite il faut cuire / ensuite il faut poser l' émail / # \$ et cuire une deuxième fois / sachant que la cuisson dure &euh six heures / et &euh ensuite six heures de refroidissement // # \$ &euh # donc le tournage / en fait / c' est qu' une toute petite partie de [/] du métier // # \$ il y a aussi pétrir la terre / enfin il y a des tas de choses à côté / &euh # que les gens ne s' imaginent pas // # \$ et puis &euh les gens s' imaginent aussi que c' est [/] ah ouais la poterie comme dans les années soixante-dix / où ça a commencé à [/] # à devenir &euh # \$ fréquent de voir des gens qui partaient &euh dans le sud / &euh # &euh faire de la poterie / que c' était facile / &euh que c' était &euh # &euh [/] et [/] et aussi ils ont une image aussi de [/] de [/] de poterie / # \$ &euh # rustique / traditionnelle / &euh alors qu' on peut faire &euh des milliers d' autres choses / &euh # qui n' ont rien à voir avec la poterie &euh des années soixante-dix / qu' on [/] qu' on s' imagine très &v [/] très vite dès qu' on parle de poterie / # \$ &euh les grès au sel / enfin les # &euh [/] les tasses &euh # \$ enfin bref # hhh c' est pas facile à décrire / mais # \$ voilà quoi //

Extrait 2

*EST: / le problème c' est que pour moi / c' est quand ? bè &mercre [/] mercredi c' était férié pareil / ça [/] a bloqué mon mail tu vois // j' attendais des trucs <importants> // \$

*DEL: <oui ben c' était ça ouais> // \$ c' était mercredi dernier ouais // # \$ *EST: donc &euh à mon avis / ou ils arrivent pas à nettoyer bien / ou il faudra trouver / parce que ça [/] # ça va recrasher pendant les cinq jours où ce sera férié là // # \$ *DEL: <ouais> // \$

*EST: / <mercredi> \$ jeudi vendredi samedi dimanche / mh // # \$ *DEL: mais moi je pense qu' il y a des gens tu sais / nous [/] enfin moi je [/] je [/] je laisse mon message // \$ avant je laissais plus les messages sur le serveur // # \$ *EST: <ouais> // \$

*DEL: <mais> \$ je les laisse sur le serveur // \$ et je les nettoie au fur et à mesure // # \$ et je demande à ce qu' ils soient supprimés au bout de dix jours / # \$

Notes

1 La longueur relative entre le nom et l'adjectif a également été envisagée (Forsgren, 1978 ; Thuilier *et al.* 2011), mais elle semble n' être qu' une conséquence de la tendance observée selon la longueur de l'adjectif seul (Thuilier, 2012 : p. 142-145).

2 Les convertis sont les mots construits grâce au procédé de conversion, procédé morphologique de formation d'unités lexicales qui se caractérise par un changement de catégorie lexicale et par une identité phonologique (Tribout, 2010).

3 Les conditions de vérité dans lesquelles le syntagme nominal contenant l'adjectif antéposé est vrai sont différentes des conditions de vérité dans lesquelles le syntagme nominal contenant l'adjectif postposé est vrai.

4 <http://www.evene.fr/theatre/actualite/interview-bruno-gaccio-avenue-q-spectacle-musical-broadway-bobin-780159.php>, page consultée le 15 février 2012.

5 <http://busterk.blog.lemonde.fr/2010/09/09/la-carte-de-sejour-et-le-territoire-national/>, page consultée le 1er février 2012.

6 Phrase extrait du corpus Est-Républicain : <http://www.cnrtl.fr/corpus/estrepublikain/>

7 <http://www.jeuxvideo.com/forums/1-27-7666091-1-0-1-0-0-0.htm>, page consultée le 15 février 2012.

8 Phrase extraite de l'ouvrage de D. Ménager (1979) : *Ronsard*, Travaux d'humanisme et Renaissance, Droz.

9 http://www.mediapart.fr/journal/culture-idees/170910/pierre-rosanvallon-lechec-du-sarkozysme-la-panne-de-la-gauche?page_article=3, page consultée le 15 février 2012.

10 <http://tedsifflera3fois.com/2011/12/14/carnage-critique/>, page consultée le 21 mars 2012.

11 http://www.fabula.org/actualites/h-cixous-peinetures-ecrits-sur-l-art_41766.php, page consultée le 14 septembre 2012 : « *Ces textes ont été publiés dans des catalogues et des revues d'art de difficile accès, ou parfois seulement en traduction* ».

12 Cette tendance se retrouve par exemple dans l'organisation des constituants après le verbe (Blinkenberg, 1928 ; Abeillé & Godard, 2004, Thuilier 2012).

13 Site du corpus Est-Républicain : <http://www.cnrtl.fr/corpus/estrepublikain/>.

14 Pour plus de détails à propos de la régression logistique à effets-mixtes, se reporter à Thuilier (2012) : chapitre 2, pour une présentation générale ; chapitre 4, pour son application au problème de la position de l'adjectif épithète.

15 Pour éliminer les effets fixes non significatifs, nous avons utilisé le test du rapport de vraisemblance (*Likelihood Ratio Test*). Cette méthode est décrite par exemple dans Howell (2009), p. 156-157.

16 C. Benzitoun (communication personnelle) suggère que le fait que la coordination d'adjectifs favorise l'antéposition à l'oral est attendu, dans la mesure où une coordination d'adjectifs est le témoin d'une langue travaillée dans laquelle les locuteurs tendent à enfreindre les règles spontanées et massives telles que *court avant long*.

Pour citer cet article

Référence électronique

Juliette Thuilier, « Syntaxe du français parlé vs. écrit : le cas de la position de l'adjectif épithète par rapport au nom », *TIPA. Travaux interdisciplinaires sur la parole et le langage* [En ligne], 29 | 2013, mis en ligne le 18 décembre 2013, consulté le 22 mars 2014. URL : <http://tipa.revues.org/1066>

À propos de l'auteur

Juliette Thuilier

Université Paris-Sorbonne & Alpage (INRIA – Paris Diderot)
juliette.thuilier@paris-sorbonne.fr

Droits d'auteur

© Travaux interdisciplinaires sur la parole et le langage

Résumés

Grâce à une approche quantitative sur des données de corpus, nous étudions les contraintes syntaxiques et lexicales intervenant dans le choix de la position de l'adjectif épithète, en français parlé et en français écrit. La modélisation statistique de 6612 occurrences d'adjectif représentant 170 lemmes adjectivaux permet de mettre en lumière le rôle de onze contraintes syntaxiques tout en prenant en compte les spécificités lexicales. Les déterminants démonstratifs, possessifs et les articles définis favorisent l'antéposition. De plus, les SA contenant des adjectifs coordonnés ou des modificateurs adverbiaux ont tendance à être postposés. La présence de propositions relatives, de SP ou d'un autre adjectif après le nom favorise également la postposition. Enfin, on observe que plus le nom et l'adjectif ont tendance à former une collocation, plus l'ordre de cette collocation tend à être respecté. Nous montrons également que quatre contraintes syntaxiques interagissent avec le mode de production des données (oral ou écrit) ce qui signifie que ces contraintes ont un comportement différent selon le mode de production des données. Nous émettons des hypothèses pour tenter de comprendre ces différences entre français parlé et français écrit.

The syntax of speech and the syntax of writing: the case of the position of attributive adjectives in French

In French, attributive adjectives (A) can appear both before or after the noun (N):

(1) a. *une agréable soirée* (anteposed)

b. *une soirée agréable* (postposed) « a nice evening »

We compare the difference between the syntax of spoken French (SF) and written French (WF) on the basis of this alternation phenomenon. We aim to determine in which cases the syntax of this phenomenon is different in SF and WF, and to quantify these differences. The methodology is inspired by the work by Bresnan et al. (2007) and Bresnan and Ford (2010) on dative alternation in English. Using statistical modeling on data extracted from written and spoken corpora, we test syntactic factors found in the literature (Abeillé and Godard, 1999; Wilmet, 1981; Forsgren, 1978; Blinkenberg, 1933 a. o.)

We assume that, with statistical tools (logistic regression – Agresti, 2007 – and mixed-effect models – Gelman and Hill, 2006), we are able to free ourselves from variations due to the sampling of the corpora. Moreover, one advantage of the mixed-effect logistic regression is that it is predictive, in the sense that one can build a model on a set of data and use this model to predict the choice between ante- and postposition on unseen data. This way, we can evaluate how well the model generalizes from the training set. Lastly, we make use of the possibility of testing the significance of interaction between different factors in order to evaluate which syntactic factors have a different behavior according to the medium used (spoken vs. written). To build our database, we first extracted the attributive As that appeared in both positions in the syntactically annotated newspaper corpus *French Treebank* (FTB, Abeillé and Clément 2004), leaving aside As with post-adjectival dependents. We then extracted the same As from the spoken corpus C-ORAL-ROM (CORAL, Cresti and Moneglia 2005). Besides the variable capturing the medium used (SF vs. WF), these data were annotated for 11 variables concerning the syntactic environment of each A in context: (1) the A is coordinated, (2) the A is modified by an adverbial element; the NP contains (3) an other A in postposition, (4) an other A in ante- or postposition, (5) a relative clause, (6) a PP; the determiner of the NP is (7) demonstrative, (8) possessive, (9) definite; a measure of collocation for (10) the ordered sequence A+N and (11) the ordered sequence N+A (collocations estimated with χ^2 , Manning and Schütze, 1999). We also differentiated two lemmas in context for 5 As: *ancien* 'ancient/former', *pur* 'pure', *seul* 'alone/single', *simple* 'simple/modest', *propre* 'own/clean'.

The database contains 6612 occurrences of attributive As (4986 in FTB, 1626 in CORAL) representing 170 lemmas, with 68.9% of ante- or postposition (67.1% in FTB, 74.3% in CORAL). There is variation according to the lemmas: for instance, the A *unique* 'unique' is ante- or post-

20% of the cases, whereas *sérieux* 'serious' appears in this position in 51.4% and *petit* 'small' in 98.6%. Moreover, there is less alternation in spoken data than in written ones: the 170 lemmas appear in both positions in FTB, while only 56 (72,8% of the 130 lemmas attested in CORAL) of them are really alternating in CORAL. This seems to reveal that in spoken French, the As tend to have a more fixed behavior than in the written variant. One can hypothesize that the more the speech is spontaneous, the more the A occurs in its preferred position, that is the more frequent position.

We used mixed-effects logistic regression to estimate the probability that the anteposition will be chosen as a function of 12 predictive variables (the 11 syntactic variables and the medium used: WF or SP). The construction of the model consists in estimating the coefficients that are associated with each variable. Besides the predictive variables, also called fixed effects, mixed-effects models are able to take into account the variation in the data by means of random-effects. In our case, the adjectival lemmas are the random effects in order to model the adjectival idiosyncrasies. We built a model with 12 fixed-effects, 1 random-effect and 11 interaction between the medium and the 11 syntactic variables. We tested all the interactions between the medium and the 10 syntactic variables interactions. We removed predictors and interactions that were non-significant at the 0.05 level step by step, but keeping in the model non-significant fixed-effects for predictors that participated in significant interactions. All the fixed-effects as well as 4 interaction were significant and thus participate in predicting the position of the As.

Each coefficient associated with fixed-effects can be interpreted as the preference for a position: a positive coefficient indicates a preference for anteposition and a negative one for postposition. Thus the model shows that the nature of the determiner influences the position: demonstrative, possessive and definite determiners favor the anteposition. Moreover, APs containing coordinated As or adverbial modifiers tend to be postposed, which confirms that speakers tend to put « heavy » APs after the N. The occurrence of a relative clause, a PP or another A after the N also favors the anteposition. Finally, the N the A is combined with affects the choice: the more the A and the N tend to be a collocation in a given order, as in *à juste titre* 'understandably', the more the sequence tend to occur in the given order. There is also a significant effect of the medium: SF favors postposition compared to WF. Insofar as each of these syntactic variables favor the same position in WF as well as in SF, we consider that the phenomenon is syntactically unified in both variants. There is only one factor that do not have the same effect: the presence of an anteposed adjective. It favors anteposition in SF and postposition in WF. The three other interactions with the medium show that the observed effect is strengthened or weakened in SF. First, demonstratives strongly favor anteposition in WF, whereas in SF it has a weak effect. Second, the possessive determiner and the adverbial modifier tend to be more strongly associated with anteposition in SF.

Entrées d'index

Mots-clés : linguistique de corpus, approche quantitative, modélisation statistique, French TreeBank, C-ORAL-ROM

Keywords : corpus linguistics, quantitative approach, statistical modeling, French TreeBank, C-ORAL-ROM