



**HAL**  
open science

# An Experimental Approach to French Attributive Adjective Syntax

Juliette Thuilier

► **To cite this version:**

Juliette Thuilier. An Experimental Approach to French Attributive Adjective Syntax. Christopher Piñón. Empirical Issues in Syntax and Semantics, 10, , pp.287-304, 2014, Experimental Syntax and Semantics, 1769-7158. hal-01451697

**HAL Id: hal-01451697**

**<https://hal.science/hal-01451697>**

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.

# An Experimental Approach to French Attributive Adjective Syntax

*Juliette Thuilier*

In this paper, we study the alternation of adjective position in the noun phrase. We postulate that this phenomenon is influenced by various factors interacting in a complex way and favoring one position over the other. Thus we use an experimental approach in order to determine which factors are indeed involved in the choice and how they interact. Our approach is based on a corpus data modeling and a questionnaire experiment.

*Keywords:* word order, corpus, questionnaire, statistical modeling

## 1 Introduction

In French, as well as in other Romance languages, attributive adjectives can appear both before or after the noun, as shown in example (1).

- (1) a. une agréable soirée (prenominal position)  
a nice evening  
b. une soirée agréable (postnominal position)  
a evening nice

The postnominal position is generally considered the canonical position because (i) adjectives appear more frequently in this position, both in terms of lemmas and tokens (Forsgren 1978, Wilmet 1981, Thuilier et al. 2010, among others),<sup>1</sup> (ii) most of the new adjectives created in the language appear in postnominal position (Noailly 1999). However, although it is not as frequent as the postnominal position, the prenominal position appears to be the preferred position for short and frequently occurring adjectives. Moreover, it will be shown in this paper that the adjectives that appear in both positions in corpus data seem to occur more frequently before the noun.

In this paper, we focus on the lexical and syntactic aspects of adjective position alternation. We postulate that this phenomenon is influenced by various factors interacting in a complex way and favoring one position over the other. Thus we need an experimental approach in order to determine which factors are indeed involved in the choice and how they interact. Our approach is based on experiences using corpus data and questionnaires. It has been inspired by the work by Bresnan et al. (2007), Bresnan (2007) and Bresnan and Ford (2010) on the dative alternation in English. It also follows up on previous works by Thuilier et al. (2010) and Thuilier et al. (2012), which are corpus studies based on written data extracted from a newspaper corpus and comparing the effect of several factors on adjective position by using statistical modeling. In comparison to these previous works, the present paper relies on (i) speech data

I thank the following people for their help and/or comments at different stages of this work: Anne Abeillé, Christophe Benzitoun, Benoît Crabbé, Laurence Danlos, Sarra El Ayari, Gwendoline Fox, Margaret Grant, Delphine Tribut. I also thank the anonymous reviewer as well as the editor of the present volume, Christopher Piñón.

<sup>1</sup>For instance, in the corpus study by Thuilier et al. (2010), 71.9% of the adjective occurrences appear after the noun, and 84.5% of the adjectival lemmas are only found in postnominal position.

in addition to written data; (ii) more accurately annotated data, in particular concerning the potentially homonymous adjectives (cf. section 3.1); (iii) a comprehension experiment investigating the link between the results of the corpus data modeling and the metalinguistic choices of speakers.

The paper is organized as follows: section 2 is dedicated to describing the semantic, syntactic, and lexical aspects of the alternation phenomenon; in section 3, we will present the corpus data and the multifactorial statistical modeling; section 4 will discuss a questionnaire experiment showing that the results of the statistical modeling correlate with speaker preferences about adjective position.

## 2 The Phenomenon

The alternation of noun-adjective ordering is a long-debated issue in French linguistics and has generated a huge literature (Blinkenberg 1933, Reiner 1968, Waugh 1977, Forsgren 1978, Wilmet 1981, Delbecque 1990, Bouchard 1998, Abeillé and Godard 1999, Noailly 1999, Thuilier et al. 2012, among others). Without reviewing all of the literature, we will give the main factors that have been mentioned and that we will study on the basis of corpus data.

### 2.1 Semantic Aspects

The semantics of adjectives as well as the semantics of noun-adjective combinations is a complex problem, as shown by the literature: Kamp (1975), McNally and Kennedy (2008), among many others. In French, this semantic problem interacts with the two possible positions of the adjective, which adds complexity to the problem. Given that establishing an exhaustive review of this problem is beyond the scope of this article, we will give a brief overview of the links between position and semantics.

The general idea is that preposed adjectives tend to be subjective, as *petite* in (2), or intensional, as *vrai* in (3), whereas the postposed ones are inclined to be intersective (or predicative), as *fragile* in (4).

- (2) une petite souris  
a small mouse
- (3) un vrai complot  
a true plot
- (4) un vase fragile  
a vase fragile

Some linguists postulated that the alternation of position is a purely semantic phenomenon. In broad outline, Waugh (1977) and Bouchard (1998) considered that preposed adjectives modify internal components of the noun, whereas postposed ones assign the noun referent a property that cannot be assigned to a sub-component of the noun. This approach leads to postulating that there is a systematic difference of meaning between the preposed and the postposed version of the same adjective. However, this generalization appears to be false. First, as pointed out by Abeillé and Godard (1999), there are noun-adjective sequences with the same meaning regardless of the position of the adjective. In (5), both NPs mean ‘a charming boy’, without any possible variation in interpretation.

- (5) a. un charmant garçon  
       a charming boy  
       b. un garçon charmant  
       a boy charming

Second, we observe semantic effects linked to the position of the adjective in the case of some specific adjective-noun combinations. For instance, the adjective *gros* 'fat' can acquire an intensifying value when it is preposed to agentive nouns, such as *fumeur* 'smoker' in (6), but this value is not present with other agentive nouns, such as *coiffeur* 'hairdresser' in (7).

- (6) a. un gros fumeur  
       a fat smoker  
       'a heavy smoker'  
       b. un fumeur gros  
       a smoker fat  
       'a fat smoker'
- (7) a. un gros coiffeur  
       a fat hairdresser  
       b. un coiffeur gros  
       a hairdresser fat

Moreover, Abeillé and Godard (1999) pointed out that in the case of *un gros fumeur*, the prenominal position is compatible with both interpretations: a person who smokes a lot or a person who is a fat smoker. This means that this is not the position that requires a specific interpretation, but the adjective-noun combination itself.

Thus, following Abeillé and Godard (1999), we consider that there is no semantic property categorically associated with one position. We assume that the semantics does not account for the entire phenomenon (contra Bouchard 1998 and Waugh 1977) and that the choice of the position is mainly driven by lexical properties and syntactic constraints.

## 2.2 Lexical Aspects

Adjectives show individual preferences which are shaped by formal properties: length, frequency, and morphological properties.

The length of words and constituents plays a role in word order and alternation phenomena (Hawkins 1994). SVO languages as French tend to prefer the short-before-long order. In the case of adjectives, Wilmet (1981), Forsgren (1978), and Thuilier (2012) noticed that what matters is the length of the adjective itself,<sup>2</sup> with the following tendency: *short adjectives first, long adjectives last*. These corpus studies showed that most of the monosyllabic adjectives are preposed, while adjectives containing more than two syllables are more frequently postposed.

Since Zipf's (1932) work, we know that there is a strong correlation between length and frequency, such as the more frequent the word, the shorter it tends to be. Given the above mentioned *short first* and *long last* preference, corpus data display the expected tendency: frequent adjectives are inclined to be preposed, whereas rare ones tend to be postposed (Wilmet 1981,

<sup>2</sup>In corpus data (Forsgren 1978, Thuilier 2012), the relative length of the noun and the adjective does not appear to be as relevant as the length of the adjective itself. Thuilier (2012:142-145) showed that the slight effect of the relative length observed in corpus data can be understood as the result of the effect of adjective length.

**Table 1**

Lexical properties and adjective position

Prenominal position	Postnominal position
short	long
frequent	rare
morphologically simple	morphologically complex

Thuilier 2012). Besides the relation between length and frequency, the effect of frequency on adjective position may be explained by an hypothesis on the diachronic evolution of adjective syntax. Following Bybee (2006), we consider that highly frequent words and word sequences are strengthened in their morphosyntactic structure and are resistant to change. In Old French, the most frequent order was ‘adjective noun’, and some contexts allowed the adjective to be postposed (Buridant 2000). We can hypothesize that the postposing default rule developed in Modern-French did not affect highly frequent adjectives, because these were resistant to change.

The morphological complexity of the adjective seems to affect its position: derived adjectives tend to be postposed. Apart from adjectives derived by conversion, complex adjectives are generally longer than simple ones, which favors their postposition. Despite this length effect, other properties have been identified as playing a role in their preference for the postnominal position. In particular, part of the deverbals and denominals can be substituted by relative clauses, as shown in (8) and (9). The ability of derived adjectives to be replaced by syntactically more complex and obligatorily postposed sequences correlates with a significant proportion of occurrences of derived adjectives in postnominal position.

- (8) Deverbal
- a. une décision contestable  
a decision questionable
  - b. une décision que l’on peut contester  
a decision that one can contest
- (9) Denominal
- a. les résultats semestriels  
the results semiannual
  - b. les résultats du semestre  
the results of-the semester

In sum, previous works on corpus data showed that a bundle of formal lexical properties converges at each position, as summarized in Table 1.

### 2.3 Syntactic Aspects

The alternation of position is affected by the internal structure of the adjective phrase (AP) and the noun phrase (NP). We will present five syntactic factors based on the following elements: post-adjectival dependent, pre-modified adjective, coordination of adjectives, other noun dependent in the NP, and type of determiners introducing the NP.

First, adjectives followed by a dependent must be postposed to the noun (Thuilier 2012, Abeillé and Godard 1999, Blinkenberg 1933), as shown in (10). This is the only categorical constraint. The other syntactic constraints do not impose, but rather favor one position over the

other.

- (10) a. une musique agréable à écouter  
 a music nice to hear  
 b. \*une agréable à écouter musique  
 a nice to hear music

Pre-modified adjectives can be both preposed and postposed to the noun, as in (11).

- (11) a. une très agréable soirée  
 a very nice evening  
 b. une soirée très agréable  
 an evening very nice

However, the presence of a modifier makes the AP longer, thereby favoring its postposition. Forsgren (1978: 159) observed that among 559 pre-modified adjectives in his corpus data, 73.4% are postposed, whereas only 66% of single adjectives are in this position. This suggests that in addition to the length of the adjective, the length of the AP also plays a role in the alternation.

Furthermore, if an adjective with a very strong preference for one position is pre-modified, its preference becomes less strong by means of the modifier (Wilmet 1981, Abeillé and Godard 1999). For example, the adjective *bon* 'good' strongly prefers the prenominal position (the NP in (12b) sounds odd), but can easily be postposed to the noun when it is pre-modified, as in (12c).

- (12) a. un bon poulet  
 a good chicken  
 b. ?un poulet bon  
 a chicken good  
 c. un très bon poulet / un poulet très bon  
 a very good chicken / a chicken very good

Likewise, the adjective *familial* 'family' has a strong preference for postnominal position, as in (13a-b), but can be preposed if it is pre-modified, as in (13c).

- (13) a. une berline familiale  
 a sedan family  
 b. ?une familiale berline  
 a family sedan  
 c. une berline très familiale / une très familiale berline  
 a sedan very family / a very family sedan

Both the prenominal position and the postnominal position are also possible for coordinated adjectives, as shown in (14). As has been observed for pre-adjectival modifiers, coordination tends to favor the postnominal position because of the length of the AP. Forsgren (1978) found around 73% of coordinated adjectives, and around 67% of noncoordinated adjectives in postnominal position.

- (14) a. un petit et confortable canapé  
 a small and comfortable sofa

- b. un canapé petit et confortable  
 a sofa small and comfortable

Moreover, coordination is comparable to pre-modification insofar as it allows adjectives with strong lexical preferences to have more flexibility. For instance, *grand* 'big' and *calme* 'quiet' sound better, respectively, in prenominal position and in postnominal position, as in (15a-b). Once coordinated, these adjectives can be naturally either preposed or postposed to the noun, as in (15c).

- (15) a. un grand appartement  
 a big apartment  
 b. un appartement calme  
 an apartment quiet  
 c. un grand et calme appartement / un appartement grand et calme  
 a big and quiet apartment / an apartment big and quiet

The effect of coordination is also observable when both adjectives have a strong preference for the same position. For example, Abeillé and Godard (1999) draw attention to the case of two intensional adjectives, *vrai* 'true' and *false* 'false', which sound very odd when they are postposed, as in (16b) and (17b). However, the coordination of these adjectives can occur either before or after the noun, as in (18).

- (16) a. des vrais coupables  
 some true culprits  
 b. ?des coupables vrais  
 some culprits true
- (17) a. des faux coupables  
 some false culprits  
 b. ?des coupables faux  
 some culprits false
- (18) a. des vrais ou faux coupables  
 some true or false culprits  
 b. des coupables vrais ou faux  
 some culprits true or false

Grevisse and Goosse (2007) mentioned a tendency to produce, in planned and written discourse, "balanced NPs," with material before and after the head noun in order to avoid the accumulation of postnominal dependents. For example, when the NP contains a prepositional phrase (PP), which cannot be preposed to the noun, placing the adjective before the noun avoids separating the noun from its complement, as shown in (19c).

- (19) a. un recueil [de textes grecs]<sub>PP</sub>  
 a collection of texts Greek  
 b. un recueil récent [de textes grecs]<sub>PP</sub>  
 a collection recent of texts Greek  
 c. un récent recueil [de textes grecs]<sub>PP</sub>  
 a recent collection of texts Greek

More generally, the presence of dependents postposed to the noun, as relative clauses, PPs, or other adjectives, tends to favor the prenominal position of adjectives:

- (20) a. l'air habituel [que Paul joue]<sub>RC</sub>  
 the-tune usual that Paul plays  
 b. l'habituel air [que Paul joue]<sub>RC</sub>  
 the-usual tune that Paul plays
- (21) a. un animal étrange [indomptable]<sub>A</sub>  
 a animal strange untameable  
 b. un étrange animal [indomptable]<sub>A</sub>  
 a strange animal untameable

According to Forsgren's (1978) corpus study, the nature of the determiner introducing the NP influences the position of the adjective. This author observed that definite determiners, for example, demonstratives, possessives or definite articles, favor the prenominal position. For each NP, the indefinite counterpart in (b) sounds less natural.

- (22) a. cet éblouissant spectacle (demonstrative)  
 this dazzling show  
 b. un éblouissant spectacle  
 a dazzling show
- (23) a. son habituel refrain (possessive)  
 her usual record  
 b. un habituel refrain  
 a usual record
- (24) a. le traditionnel thé (definite article)  
 the traditional tea  
 b. un traditionnel thé  
 a traditional tea

#### 2.4 Specific Combinations of Nouns and Adjectives

Given that we are interested in the factors affecting the placement of attributive adjectives with respect to the noun, it is important to mention that the noun itself plays a role in a number of cases.

First, some adjective-noun pairs are strongly collocational in the sense that the choice of the adjective depends on the noun. For instance, the noun *hommage* 'tribute' is generally associated with the adjective *vibrant* 'vibrant' in order to idiomatically refer to a big or intense tribute. Not only does the collocational effect affect the selection of the adjective with respect to the noun, but it also affects its position. Indeed, the adjective *vibrant* is inclined to be postposed to the noun, as in (25), partly due to the fact that it is a derived adjective (cf. section 2.2). Nevertheless, the noun *hommage* strongly favors its placement in prenominal position, as in (26).

- (25) a. une voix vibrante / ?une vibrante voix  
 a voice vibrant / a vibrant voice



- b. un ton vibrant / ?un vibrant ton  
 a tone vibrant / a vibrant tone
- (26) un vibrant hommage  
 a vibrant tribute

Second, as mentioned in section 2.1, we observe that some adjective-noun combinations convey a particular meaning when the adjective is preposed. (6) and (7) above show that in a number of cases, the noun selects the adjective and its position. Thus not only is the position of the adjective determined by its lexical properties, but it is also affected by the particular noun the adjective combined with.

### 3 Corpus Data Modeling

By looking over the factors playing a role in adjective position alternation, we observed that a variety of constraints influences the choice for one position. In order to better understand their effects and to capture their relative importance, we conducted a corpus study. Using statistical modeling, we tested most of the factors mentioned in the previous section with attested data excerpted from speech and written corpora. We assume that, with statistical tools, we are able to free ourselves from variations due to the sampling of the corpora.

#### 3.1 Building the Database

The data were excerpted from two corpora:

- the *French TreeBank* (henceforth, FTB), which comprises 20,000 sentences (400,000 tokens) from the newspaper *Le Monde* fully annotated and manually validated for syntax purposes (Abeillé et al. 2003, Abeillé and Barrier 2004).
- the French part of the spoken corpus C-ORAL-ROM (henceforth, CORAL), which comprises about 300,000 tokens (Cresti and Moneglia 2005)

We must make an initial observation concerning the adjective position alternation in the FTB data. In this corpus, there are 1,750 adjectival lemmas in attributive position. These include 1,488 only-postposed adjectives and 92 only-preposed ones. These only-preposed and only-postposed lemmas represent around 64% of the 13,399 adjectival occurrences. Thus, only 170 lemmas occur in both positions. These alternating adjectives represent 4,486 occurrences and thus are the most frequent lemmas on average. These observations are summarized in Table 2.

So, even though we assume that alternation is possible for the entire adjective category, (i) for a number of adjectives the alternation is very rare and the probability that we observe it in a corpus is low; (ii) more than two fifths of the adjectives (747) in attributive position appear only once in the corpus, thus making it impossible to regard alternation for them.

**Table 2**

Attributive adjectives in the FTB corpus

	Number of lemmas	Number of occurrences
Only-preposed adjectives	92 (5.3%)	462 (3.3%)
Only-postposed adjectives	1,488 (85%)	8,485 (60.9%)
Adjectives in both positions	170 (9.7%)	4,986 (35.8%)
Total	1,750 (100%)	13,933 (100%)

The fact that the proportion of nonalternating adjectives is so high means that for a large part of the data, the identity of the adjective is enough to categorically determine its position in the dataset. This kind of data distribution leads to convergence problems of estimation algorithms with the statistical tools used here (see section 3.4).

Given that we are interested in the factors explaining the alternation and that including nonalternating adjectives raises an issue of statistical soundness, we focus on adjectives that do alternate in the FTB. In a sense, this methodological choice limits the scope of the present corpus study because we don't have a comprehensive picture of the entire category position alternation. However, we made sure that the data and the statistical modeling are reliable.

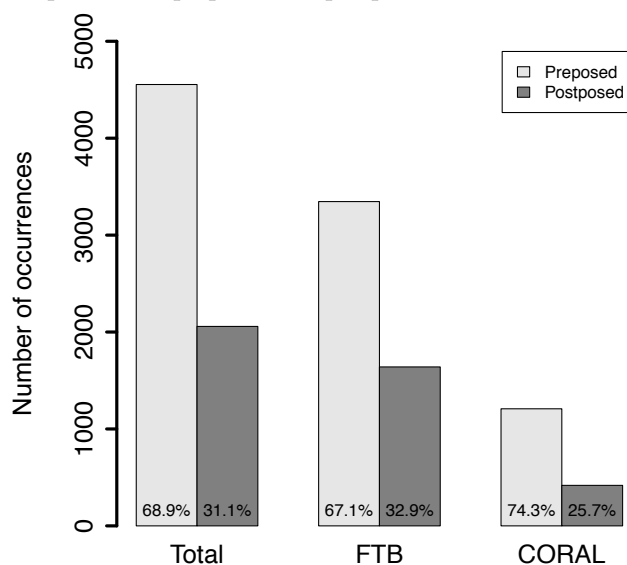
To build our database, we first excerpted the attributive adjectives that appeared in both positions in the FTB. Given that the presence of post-adjectival dependents categorically determines the position of the AP (cf. section 2.3), we left these adjectives aside. Then we excerpted the same adjectives from CORAL.

We set apart two lemmas for each of the following potentially homonymous adjectives: *ancien* 'ancient/former', *propre* 'own/clean', *pur* 'pure', *seul* 'alone/single', *simple* 'simple/modest'. For each adjective, both meanings are illustrated in examples (27)–(31).

- (27) a. un coffre ancien  
a chest old  
'an ancient chest'  
b. un ancien coffre  
a old chest  
'a former chest'
- (28) a. son propre pantalon  
her own pants  
b. son pantalon propre  
her pants clean
- (29) a. un pur produit  
a pure product  
'an archetypal product'  
b. un produit pur  
a product pure  
'a pure product' (not mixed)
- (30) a. un seul homme  
a alone man  
'a single man'  
b. un homme seul  
a man alone  
'a lonely man'
- (31) a. une simple phrase  
a simple sentence  
'a mere sentence'  
b. une phrase simple  
a sentence simple  
'a simple sentence'

**Figure 1**

Proportions of preposed and postposed adjectives



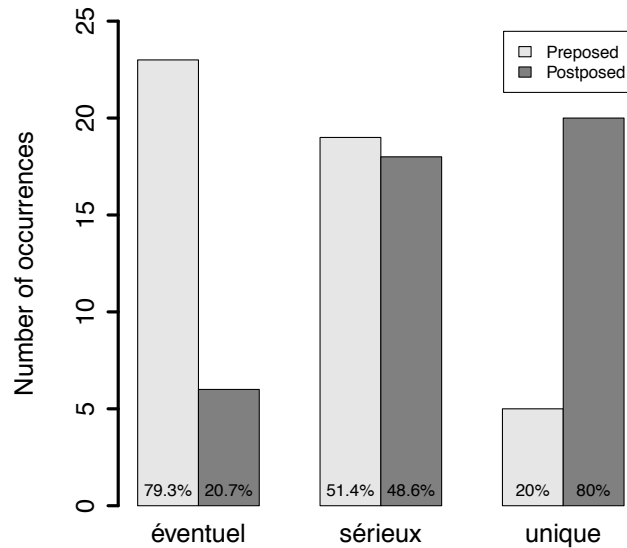
As shown in the examples, the meaning of these adjectives is canonically associated with a position, but one can find occurrences of each meaning in both positions (see Thuilier 2012 for examples). We observed alternation in the data for 5 out of the 10 disambiguated lemmas. These alternating adjectives, whose meaning is presented in (27a), (28a), (29a), (30a), and (31a), were thus included in the database. For instance, the examples in (32) show the adjective *ancien* with the ‘old’ meaning in both positions; the examples in (33) display both positions for the ‘own’ meaning of the adjective *propre* (these examples are extracted from CORAL corpus).

- (32) a. c’est plus de la variété plus des trucs anciens  
 it-is more some commercial-music more some things old  
 ‘the more it’s commercial music, the more it’s old stuff’
- b. j’avais acheté beaucoup de livres déjà quand j’avais les anciens appareils  
 I-had bought a-lot of books already when I-had the old devices  
 ‘I had already bought a lot of books when I had the old devices’
- (33) a. chacune des communautés faisait passer son intérêt propre avant  
 each of-the communities made pass its interest own before  
 l’intérêt national  
 the-interest national  
 ‘each community put its own interest before the national interest’
- b. on les attache sur nos propres maillots  
 we them tie on our own jerseys  
 ‘we tie them on our own jerseys’

### 3.2 Alternating Adjectives in Two French Corpora

The database contains 6,612 occurrences of attributive adjectives: 4,986 from FTB, 1,626 from CORAL. As shown in Figure 1, there are 68.9% of the adjectives occurring in prenominal position in the dataset, which means that adjectives that do alternate in the corpus data tend to

**Figure 2**  
Variation across lemmas



be preposed. Moreover, this proportion is higher in speech than in writing. Given that in our data, writing corpus corresponds to planned discourse and speech data comprise more spontaneous and unplanned discourse, it may be the case that unplanned discourse slightly favors prenominal position.

As for the number of lemmas, FTB data contain 170 alternating adjectives, whereas there are only 130 lemmas in CORAL and 43% of them appear in both positions. These observations are presented in Table 3. Thus, there is less alternation in speech than in writing. This seems to reveal that in spoken French, the adjectives tend to have a more fixed behavior than in the written variant.

Section 2.2 mentions that lexical properties have an effect on the adjective position, leading to particular behavior for each adjective. This can be observed in the database through variation according to the lemmas. For instance, as shown in Figure 2, the adjective *unique* ‘unique’ is preposed in 20.7% of the cases, whereas *sérieux* ‘serious’ appears in this position in 51.4% and *petit* ‘small’ in 98.6%.

These observations go against the idea that the default position of an adjective is after the noun and argue for considering that there is not a canonical position for the adjective category as a whole but rather a canonical position for each lemma.

### 3.3 Annotation of the Data

In order to capture the constraints described in section 2, the data were annotated for the 11 variables presented in Table 4. The first eight variables are binary variables capturing syntactic

**Table 3**  
Repertition of the lemmas

	FTB	CORAL
Number of lemmas	170	130
Alternating lemmas	170	56
	100%	43%

**Table 4**  
Variables annotated in the database

	Variables	Description
1	COORD	the adjective is coordinated or not
2	MOD	the adjective is pre-modified or not
3	DEMDet	the NP is introduced by a demonstrative determiner or not
4	POSSDET	the NP is introduced by a possessive determiner or not
5	DEFART	the NP is introduced by a definite article or not
6	PP	the NP contains a PP or not
7	REL	the NP contains a relative clause or not
8	POSTADJ	the NP contains a postposed or not
9	COLLOCAN	collocation score for A+N bigram ( $\log(\chi^2)$ )
10	COLLOCNA	collocation score for N+A bigram ( $\log(\chi^2)$ )
11	MODALITY	the modality is speech (s) or writing (w)

constraints mentioned in the literature. Variables 9 and 10 (COLLOCAN and COLLOCNA) were designed in order to take into account the influence of the noun combined with the adjective (cf. section 2.4). Their values correspond to  $\chi^2$  scores (Manning and Schütze 1999)<sup>3</sup> calculated with data from the *Est-Républicain* corpus<sup>4</sup> and they estimate the strength of the association of the noun and the adjective in a given position. Finally, in order to know whether the way data were produced affects the adjective position, we included the MODALITY variable (variable number 11 in the table).

### 3.4 Multifactorial Statistical Modeling

The statistical modeling of adjective position alternation was done using mixed-effects logistic regression (Agresti 2007, Gelman and Hill 2006). This statistical tool allows one to model the behaviour of a binary variable. More precisely, in our case, it estimates the probability that the adjective will be preposed to the noun as a function of the predictive variables presented in Table 4. One advantage of the mixed-effects logistic regression model 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 prenominal position and postnominal position on new data. This way, we can assess how well the model generalizes from the training set.

The construction of the model consists in estimating the coefficients that are associated with each variable. Each coefficient can be interpreted as the preference of its variable: in the case of a variable having only positive values, a positive coefficient indicates a preference for

<sup>3</sup>Using contingency tables (2-by-2 tables) such as the ones presented below, “[t]he [ $\chi^2$ ] statistic sums the differences between observed and expected values in all squares of the table, scaled by the magnitude of the expected values” (Manning and Schütze 1999:169).

	Noun = <i>hommage</i>	Noun ≠ <i>hommage</i>		Noun = <i>hommage</i>	Noun ≠ <i>hommage</i>
Pre-adj = <i>vibrant</i>	152	8607	Post-adj = <i>vibrant</i>	10	8749
Pre-adj ≠ <i>vibrant</i>	238	2797624	Post-adj ≠ <i>vibrant</i>	380	4578757

In other words, the  $\chi^2$  statistic is an estimation of the distance between the observed frequencies and the expected frequencies for independent variables. So, the more greater the distance, the higher the  $\chi^2$  and the stronger the association of the noun and the adjective.

<sup>4</sup>It is a newspaper corpus comprising 148 million words and downloadable from <http://www.cnrtl.fr/corpus/estrepublikain/>.

**Table 5**

Corpus model

*Random effects*

	Variance	Std.Dev.
ADJECTIVE	2.3938	1.5472

Number of obs: 6621, groups: ADJECTIVE, 170

*Fixed effects*

	Estimate	Std. Error	z value	P(>  z )
Intercept	-0.782	0.182	-4.304	<.001
DEMDET=1	1.226	0.246	4.99	<.001
POSSDET=1	1.185	0.235	5.04	<.001
DEFART=1	0.370	0.107	3.47	<.001
POSTADJ=1	0.587	0.154	3.82	<.001
PP=1	0.840	0.104	8.04	<.001
REL=1	0.714	0.210	3.40	<.001
COLLOCAN	0.378	0.018	20.52	<.001
MOD=1	-1.957	0.174	-11.26	<.001
COORD=1	-1.266	0.266	-4.76	<.001
COLLOCNA	-0.443	0.020	-22.12	<.001
MODALITY = w	0.458	0.121	3.78	<.001

prenominal position, and negative one a preference for postnominal position. Besides the predictive variables, also called fixed effects, mixed-effect models are able to take into account the variation in the data by means of random effects.

In our case, the adjectival lemma is the random effect in order to model the adjectival idiosyncrasies. Each lemma constitutes a group in the data, which is assigned a randomly varying normally distributed effect in the model. Thus, associating each value of the random effect with a specific coefficient accounts for the different behaviors according to adjectives (cf. previous subsection).

Using our database, we built a model with 11 fixed effects and 1 random effect. All the effects are significant and thus participate in predicting the position of the adjective.

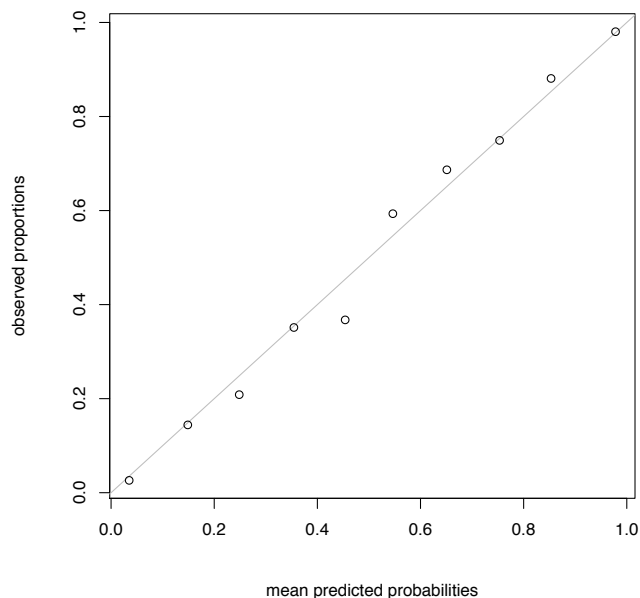
The corpus model is presented in Table 5. For each random effect, the standard deviation of the normal distribution is given. For the fixed effect, the estimated coefficient (Estimate) indicates the way each variable affects the adjective position. The *p*-values testify that the coefficients associated with the variables are significantly different from 0 (i.e. the variable has a significant effect). The model has a mean accuracy of 0.88 (10-fold cross-validation) and the mean concordance probability is  $C = 0.947$  (10-fold cross-validation). These numbers indicate that the model's predictions are very accurate. The goodness of fit can also be evaluated by means of the graph presented in Figure 3. The plot compares the grouped mean probabilities with the observed proportions of prenominal position. A perfect fit would correspond to the straight line. The distribution of the data points suggests that the model fits the data very well.

### 3.5 Results

Each coefficient associated with fixed effects can be interpreted as the preference for a position: a positive coefficient indicates a preference for prenominal position and a negative one for postnominal position. For example, the model shows that the nature of the determiner has an effect on the position: demonstrative, possessive determiners and definite articles favor pre-

**Figure 3**

Observed proportions of prenominal position and the corresponding mean predicted probabilities for the corpus model



posed adjectives. This result can be identified as providing evidence that in anaphoric contexts, when the relation between the referent of the noun and the property denoted by the adjective is established, the adjective can be more easily preposed (Waugh 1977).

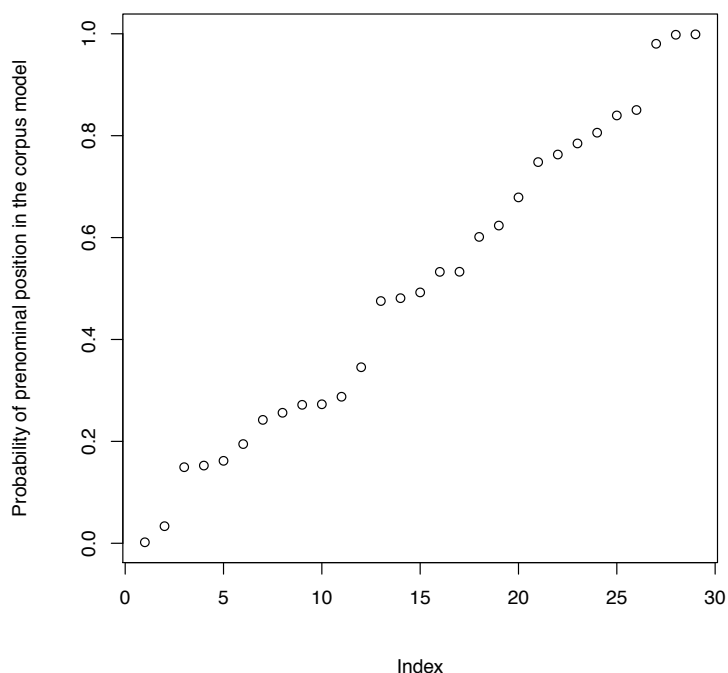
As expected, the presence of a relative clause, a PP, or another adjective after the noun also favors the prenominal position. Given that it is argued that in planned and written discourse, adjectives are inclined to be preposed when they occur with other dependents postposed to the noun, we checked the significance of the interaction between MODALITY and the three variables: PP, REL, and POSTADJ. The fact that these interactions were not significant ( $p > .1$ ) tends to show that the tendency to produce “balanced NPs” applies in both speech and writing production.<sup>5</sup>

Moreover, APs containing coordinated adjectives or pre-adjectival modifiers tend to be postposed. This can be analyzed as a clear effect of heaviness: long and complex APs are inclined to be postposed. This is in accordance with the generalization that in SVO languages, heavy constituents tend to appear last.

Concerning lexical preferences and noun-adjective combination, each adjective has a more or less strong preference for one position, which is captured by the random effect. The noun the adjective is combined with also affects the choice: the more the adjective and the noun tend to be a collocation in a given order, the more the sequence is inclined to occur in the given order. Moreover, the model shows that the noun appearing with the adjective can favor the non-preferred position. For instance, the adjective *fort* ‘strong’ has a slight preference for prenominal position, but when it is combined with the noun *point* ‘point’, the postnominal position is strongly preferred because the sequence *point fort* is much more likely to be produced.

<sup>5</sup>Thuilier and Grant (2014) found complementary evidence. They studied the effect of postnominal PPs on the position of adjectives using a sentence-recall experiment in order to test whether the presence of NP dependents affects sentence production in real time. Preliminary results show a clear effect of postnominal PPs on the position of adjectives in sentence production, which means that the tendency to produce “balanced NPs” seems to be also at play in speech production.

**Figure 4**  
Corpus model probability of questionnaire sentences



We observed that there is less alternation in the speech data. One explanation could be that in unplanned discourse, speakers tend to comply more often with lexical preferences because they have less time to produce the NP. Finally, contrary to what we observed in section 3.2, the model shows that the prenominal position is more likely in written data than in speech. This means that the apparent preference for preposed adjectives in speech data reflects the effect of a set of lexical idiosyncrasies and syntactic constraints that the statistical modeling allows us to neutralize.

#### 4 Speaker Preferences

The corpus model estimates the probability of prenominal position of each adjectival occurrence given the syntactic environment and taking into account the specificity of each lemma (random effect). A questionnaire experiment was then conducted to test whether these probabilities are related to the judgments of native speakers.

##### 4.1 Methodology

Our hypothesis is that, for many speakers, the frequency of choice for prenominal position will correspond to the probability of prenominal position estimated in the corpus model. Thus, we hypothesize that the factors favoring one position over the other will favor the choice of the speakers for the same position during a metalinguistic task.

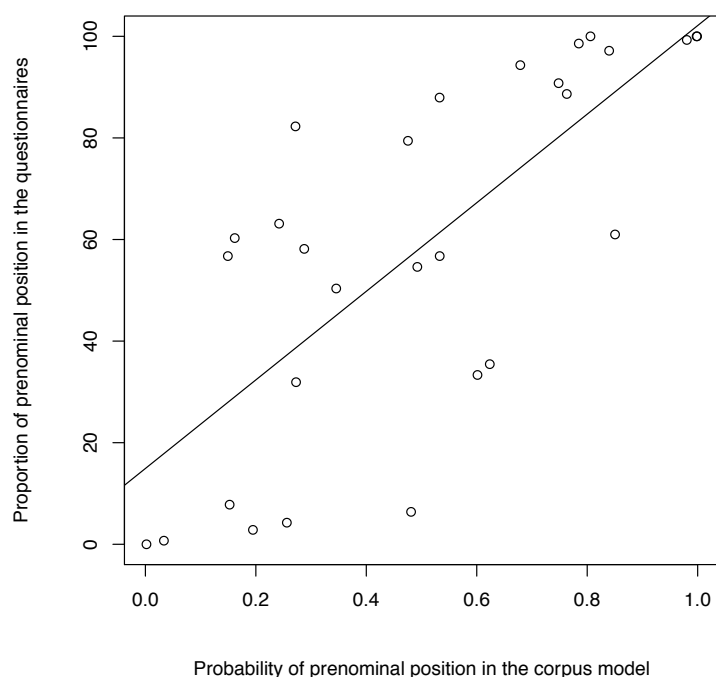
The questionnaire is made up of 29 sentences picked out from the database (the FTB part) and selected according to their probability in order to have a sample containing the range of possible probabilities (from 0 to 1). The probabilities of prenominal position for the sentences are represented in Figure 4.

Each sentence is part of a pair of sentences containing the original sentence and a modified



**Figure 5**

Correlation between corpus model probability and proportion of prenominal position



version with the adjective-noun sequence in the opposite order. As shown in (34a), in both versions of the sentence, the NP is in bold and colored letters in order to help the participant to notice the difference in the pair. The pairs and the sentences within the pairs are randomly ordered in each questionnaire.

- (34) a. Henri Guitton a joué **un rôle important** dans la modernisation de  
 Henri Guitton has played a role important in the modernization of  
 l'enseignement de l'économie en France.  
 the-teaching of the-economics in France
- b. Henri Guitton a joué **un important rôle** dans la modernisation de  
 Henri Guitton has played a important role in the modernization of  
 l'enseignement de l'économie en France.  
 the-teaching of the-economics in France

The participants were contacted via social networks and scientific mailing lists. 141 participants completed the questionnaire online. During the experiment, they saw both versions of the sentence on the screen and were asked to select their preferred version by means of a check box.

#### 4.2 Results

As predicted, the proportion of choice for preposed adjectives significantly correlates with the probability of prenominal position estimated in the corpus model: 0.74 ( $p < .0001$ ). As shown in Figure 5, the correlation is not perfect but there is a clear relation between the probability of prenominal position and the preferences of the speakers.

This result suggests that language users are sensitive to the factors used in the corpus

model when they make metalinguistic choices. More precisely, if the context strongly favors one position, the speakers tend to mostly choose this position, whereas when the context is not clearly in favor of one position, a part of the speakers selects one position and the others choose the other position. This result is in accordance with Bresnan's (2007) experimental work on the dative alternation in English. Her experiment (Experiment 1 in the paper) indicated that subjects' intuitions are affected by the same constraints as those that have an effect on the probability of dative PP realization calculated in a corpus model.

Finally, this experiment is an argument in favor of the idea that the statistical modeling proposed on the basis of usage data is an appropriate way of describing and accounting for a rather complex syntactic phenomenon such as the alternation of attributive adjective position.

## 5 Conclusion

This paper presented an experimental approach to the alternation of adjective position in the NP, combining the modeling of corpus data and a questionnaire experiment.

From the linguistic point of view, the results suggest that there are three levels of organization involved in the phenomenon. The first level is related to the lexicon insofar as each adjective has a more or less strong preference for one position. In the model, this is captured via the random effect. The second level concerns the combination of two lexical items: the noun can strongly affect the position of the adjective as the collocation variables show in the model. The third level is related to syntax and corresponds to the constraints concerning the structure of the AP and the NP.

We have offered a very accurate modeling of the phenomenon, based on corpus data and providing the probability of having a preposed adjective in a given context. The result of the comprehension experiment showed that the probabilities estimated in the corpus model seem to partly reflect the speaker preferences. This is a further argument in favor of the idea that what corpus data tell us is in accordance with a form of linguistic knowledge of language users.

## References

- Abeillé, Anne, and Nicolas Barrier. 2004. Enriching a French treebank. In *Proceedings of Language Resources and Evaluation Conference*, 2233–2236. Lisbon.
- Abeillé, Anne, Lionel Clément, and François Toussenet. 2003. Building a treebank for French. In *Treebanks: Building and Using Parsed Corpora*, 165–187. Dordrecht: Kluwer Academic Publishers.
- Abeillé, Anne, and Danièle Godard. 1999. La position de l'adjectif épithète en français : le poids des mots. *Recherches Linguistiques de Vincennes* 28:9–32.
- Agresti, Alan. 2007. *An introduction to categorical data analysis*. Wiley.
- Blinkenberg, Andreas. 1933. *L'ordre des mots en français moderne. Deuxième partie*. Copenhagen: Levin & Munksgaard.
- Bouchard, Denis. 1998. The distribution and interpretation of adjectives in French: A consequence of bare phrase structure. *Probus* 10:139–183.
- Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In *Roots: linguistics in search of its evidential base*, ed. Sam Featherston and Wolfgang Sternefeld, 77–96. Berlin: Mouton de Gruyter.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen. 2007. Predicting the dative alternation. In *Cognitive foundations of interpretation*, ed. G. Boume, I. Kraemer, and J. Zwarts, 69–94. Amsterdam: Royal Netherlands Academy of Science.
- Bresnan, Joan, and Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86:186–213.

- Buridant, Claude. 2000. *Grammaire nouvelle de l'ancien français*. Sedes.
- Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82:711–733.
- Cresti, Emanuela, and Massimo Moneglia, ed. 2005. *C-ORAL-ROM Integrated reference corpora for spoken Romance languages*. Amsterdam, USA: John Benjamins.
- Delbecq, Nicole. 1990. Word order as a reflexion of alternate conceptual construals in French and Spanish. Similarities and divergences in adjective position. *Cognitive Linguistics* 1:349–416.
- Forsgren, Mat. 1978. *La place de l'adjectif épithète en français contemporain, étude quantitative et sémantique*. Stockholm: Almqvist & Wiksell.
- Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Grevisse, Maurice, and André Goosse. 2007. *Le bon usage*. 14ème édition: De Boeck Université.
- Hawkins, John. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Kamp, Hans. 1975. Two theories about adjectives. In *Formal semantics of natural language*, ed. Edward Keenan, 123–155. Cambridge: Cambridge University Press.
- Manning, Christopher D., and Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. Cambridge: MIT Press.
- McNally, Louise, and Christopher Kennedy, ed. 2008. *Adjectives and adverbs: Syntax, semantics and discourse*. Oxford University Press.
- Noailly, Michèle. 1999. *L'adjectif en français*. Paris: Ophrys.
- Reiner, Erwin. 1968. *La place de l'adjectif épithète en français : théories traditionnelles et essai de solution*. Vienna et Stuttgart: W. Braumüller.
- Thuilier, Juliette. 2012. *Contraintes préférentielles et ordre des mots en français*. Doctoral Dissertation, Université Paris 7.
- Thuilier, Juliette, Gwendoline Fox, and Benoît Crabbé. 2010. Fréquence, longueur et préférences lexicales dans le choix de la position de l'adjectif épithète en français. In *Actes du 2ème Congrès Mondial de Linguistique Française*, ed. Franck Neveu, Valelia Muni Toke, Thomas Klingler, Jacques Durand, Lorenz Mondada, and Sophie Prévost, 2197–2210. Nouvelle-Orléans.
- Thuilier, Juliette, Gwendoline Fox, and Benoît Crabbé. 2012. Prédire la position des adjectifs épithètes en français. *Linguisticae Investigationes* 35:28–75.
- Thuilier, Juliette, and Margaret Grant. 2014. Influence of noun dependents on French adjective placement in sentence production. Poster presented at *International Workshop on Language Production* in Geneva.
- Waugh, Linda R. 1977. *A semantic analysis of word order: Position of the adjective in French*. Leiden: E. J. Brill.
- Wilmet, Marc. 1981. La place de l'épithète qualificative en français contemporain : étude grammaticale et stylistique. *Revue de linguistique romane* 45:17–73.
- Zipf, George Kingsley. 1932. *Selected studies of the principle of relative frequency in language*. Cambridge, USA: Harvard University Press.

Université Rennes 2 and Laboratoire de Linguistique Formelle (CNRS/Paris 7)  
 juliette.thuilier@univ-rennes2.fr