



**HAL**  
open science

# Quantitative contribution to the study of the syntax of spoken vs. written language

Juliette Thuilier

► **To cite this version:**

Juliette Thuilier. Quantitative contribution to the study of the syntax of spoken vs. written language. Quantitative Investigations in Theoretical Linguistics (QITL), Sep 2013, Leuven, Belgium. . hal-01451853

**HAL Id: hal-01451853**

**<https://hal.science/hal-01451853>**

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.



# Quantitative contribution to the study of the syntax of spoken vs. written language



Juliette Thuilier

Université Paris-Sorbonne & ALPAGE (INRIA & Paris Diderot)

## Research question

To what extent is the syntax of speech different from the syntax of writing?  
 ⇒ Focus on an alternation phenomenon: the position of the attributive adjective in French

*une agréable soirée* (anteposed)  
*une soirée agréable* (postposed)  
 'a nice evening'

## Hypothesis 1

The choice of the position of attributive adjectives is a non-categorical phenomenon, various factors interacting in a complex way

**Lexical aspects**

- Most of the adjectives have a strong preference for one position
- Bundle of formal lexical properties converging towards each position
- A dozen homonyms: in a pair, each homonym tends to occur in one position  
*un ancien coffre* 'a former chest' / *un coffre ancien* 'an ancient chest'

### Collocational effects

- Some Adjective - Noun pairs are strongly collocational: the position should be influenced by collocational effects as in *à juste titre* 'understandably'

### Syntactic aspects

- The presence of a postadjectival dependent makes postposition obligatory  
*une musique agréable à écouter* / \**une agréable à écouter musique* 'a nice music to hear'
- Pre-modified adjectives and coordinated adjectives occur in both position and are quantitatively more postposed (⇒ *short before long*)
- Presence of other dependents of the noun (PP, relative clause, other adjective) favor anteponement in order to equilibrate the elements around the noun
- Definite determiners favor anteponement (Forsgren, 1978)

## Hypothesis 2

There are subtle differences between the syntax of speech and writing, such as differences between the varieties of a same language

### Dative alternation in varieties of English

- Different effect of animacy of recipient in New Zealand and American English (Bresnan & Hay, 2008)
- Different effect of relative syntactic complexity in Australian and American English (Bresnan & Ford, 2010)

### Aims

- To model the choice between anteponement and postposition
- To characterize the differences between the syntax of speech and writing

### Tools

- Statistical modeling of corpus data : mixed-effect logistic regression
- Focus on interactions between modality and syntactic constraints

## Data

### Extracted from

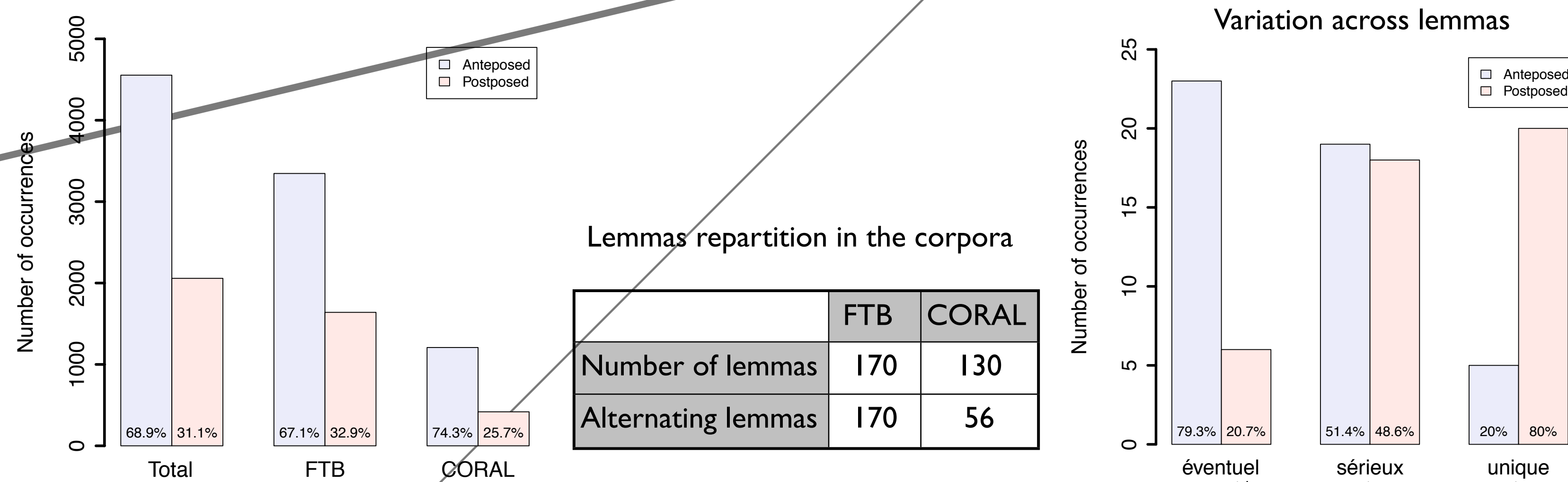
- FTB: French TreeBank, newspaper corpus ⇒ WRITING
- CORAL: CORAL-ROM, spontaneous speech corpus ⇒ SPEECH

### Selection

- Removal of adjectives with postadjectival dependents
- Differentiation of two lemmas for 9 homonyms (ancien, pur, seul, simple, sacré, commun, pauvre, propre, cher)

### General observations

- 6612 occurrences of attributive adjectives modifying a nominal head
- 170 lemmas

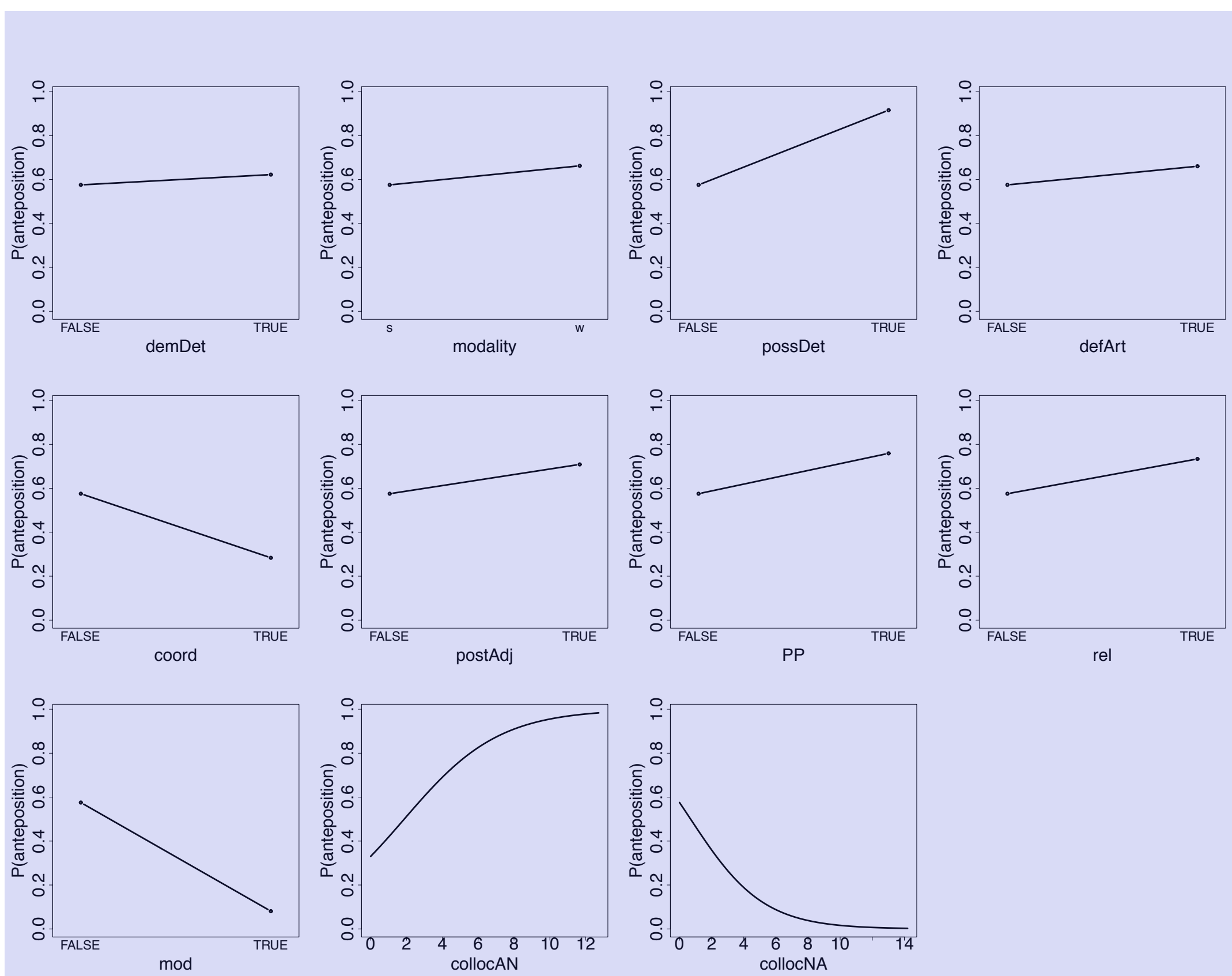


## Annotation

VARIABLES	TYPE	DESCRIPTION
coord	bool	the adjective is coordinated or not
mod	bool	the adjective is pre-modified or not
demDet	bool	the NP is introduced by a demonstrative determiner or not
possDet	bool	the NP is introduced by a possessive determiner or not
defArt	bool	the NP is introduced by a definite article or not
modality	bool	the modality is speech (s) or writing (w)
PP	bool	the NP contains a PP or not
rel	bool	the NP contains a relative clause or not
postAdj	bool	the NP contains a postposed adjective or not
collocAN	real	collocation score for A+N bigram ( $\log(\chi^2)$ )
collocNA	real	collocation score for N+A bigram ( $\log(\chi^2)$ )

Low collinearity : condition number  $\kappa = 8.15$

### Partial effects of fixed-effects



## Statistical modeling

### Mixed-effect logistic regression

- estimation of the probability of anteponement
- 11 fixed-effects (cf. table)
- 1 random effect ⇒ adjectival lemma
- 10 interactions tested: modality with each other fixed-effect
- removal of non-significant predictors and interactions

### Accuracy of the model (10-fold cross-validation)

- mean accuracy = 0.882
- mean concordance probability = 0.947

FIXED EFFECTS				
	ESTIMATE	STD. ERROR	Z-VALUE	PR(> Z )
(Intercept)	-0.70679	0.18881	-3.743	0.000182
demDet = true	0.19508	0.49344	0.395	0.692588
possDet = true	2.07981	0.49308	4.218	2.46e-05
defArt = true	0.36049	0.10687	3.373	0.000743
modality == w	0.36915	0.13451	2.744	0.006063
coord = true	-1.23054	0.26661	-4.616	3.92e-06
postAdj = true	0.58555	0.15354	3.814	0.000137
PP = true	0.84350	0.10448	8.073	6.84e-16
rel = true	0.70982	0.21166	3.354	0.000798
mod = true	-2.73507	0.35934	-7.611	2.71e-14
collocAN	0.37713	0.01849	20.396	< 2e-16
collocNA	-0.44103	0.02000	-22.055	< 2e-16
demDet:modality	1.29997	0.55833	2.328	0.019894
possDet:modality	-1.11834	0.52371	-2.135	0.032727
mod:modality	0.99029	0.39445	2.511	0.012055
RANDOM EFFECTS				
GROUPS	NAME	STD. DEV.		
adjectival lemma	(Intercept)	1.5532		

## Results and conclusion

### Anteponement is favored by:

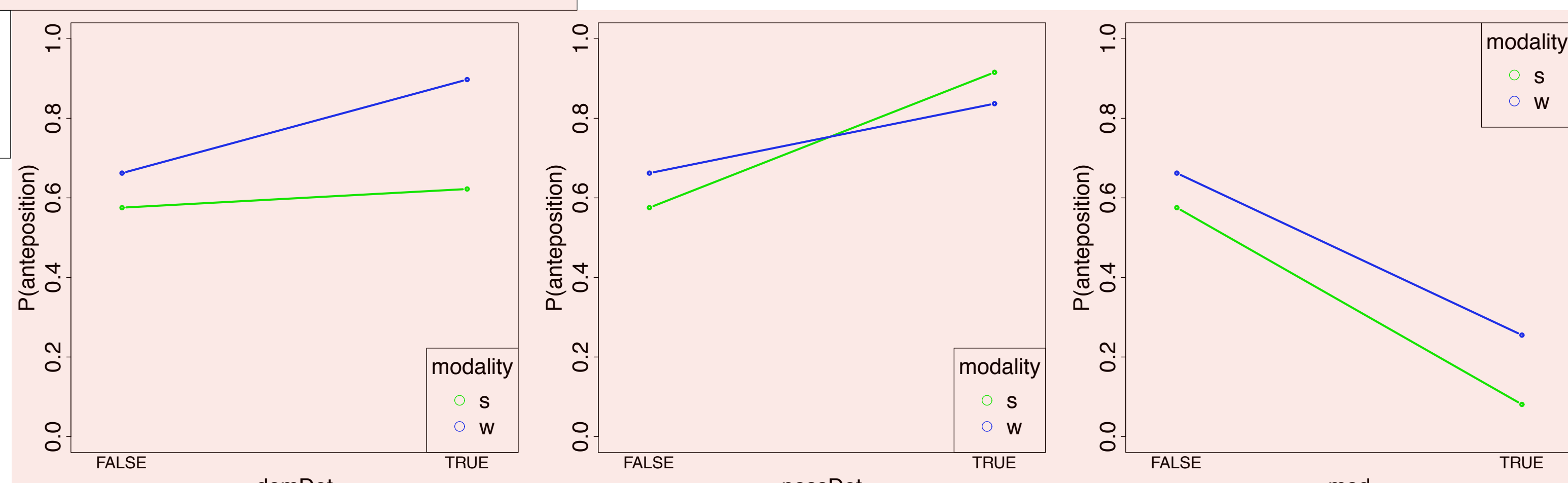
- Demonstrative, possessive and definite determiners
- Other dependents of the noun (relative, PP and postposed adjective)
- High score for AN collocation

### Postposition is favored by:

- Heavy AP (coordinated or modified adjective)
- High score for NA collocation

### Modality :

- Effect of **modifiers** and **possessives** stronger in the syntax of speech ⇒ The stronger effect of modifiers may be due to the fact that in unplanned discourse, speakers tend to massively follow general principles like 'short before long'
- Effect of **demonstratives** stronger in the syntax of writing ⇒ Anteponement with demonstrative determiners is a feature of the written language (monitored language)
- Number of alternating adjectives higher in the written data ⇒ In unplanned discourse, speakers may tend to comply more often with lexical preferences



Selected references

- Abellié and Godard. 1999. La position de l'adjectif épithète en français : le poids des mots. *Recherches Linguistiques de Vincennes* 28: 9-32.
- Agresti. 2007. *An Introduction to Categorical Data Analysis*. Wiley.
- Behaghel. 1909. Von deutscher wortstellung. *Insogermanische Forschungen* 25.
- Blinkenberg. 1933. *L'ordre des mots en français moderne. Deuxième partie*. Copenhagen : Levin & Munksgaard.
- Bresnan, Cueni, Nikitina, and Baayen. 2007. Predicting the dative alternation. In: Boume, Kraemer, and Zwarts (Eds.), *Cognitive Foundations of Interpretation*. Amsterdam : Royal Netherlands Academy of Science.
- Bresnan and Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86 (1).
- Bresnan and Hay. 2008. Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua*, 118(2), 245-259.

- Forsgren. 1978. *La place de l'adjectif épithète en français contemporain, étude quantitative et sémantique*. Stockholm: Almqvist & Wiksell.
- Gelman and Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Hawkins. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Manning and Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: The MIT Press.
- Thuilier. 2012. *Contraintes préférentielles et ordre des mots en français*. PhD Thesis. Université Paris Diderot.
- Wilmet. 1981. La place de l'épithète qualificative en français contemporain : étude grammaticale et stylistique. *Revue de linguistique romane* 45: 17-73.