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

To cite this version:

HAL Id: hal-01451853
https://hal.archives-ouvertes.fr/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 Thuiller
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

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
Anteposition NOUN Postposition
short long
frequent simple morphologically complex
• A dozen homonyms: in a pair, each homonym tends to occur in one position
• Effect: une agréable soirée (anteposed) une soirée agréable (postposed) ‘a nice evening'

Syntactic aspects
• The presence of a postadjectival dependent makes postposition obligatory
• 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 anteposition in order to equilibrate the elements around the noun
• Definite determiners favor anteposition (Forsgren, 1978)

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

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
General observations
• 6612 occurrences of attributive adjectives modifying a nominal head
• 170 lemmas

Statistical modeling
Mixed-effect logistic regression
• Selection of the probability of anteposition
• 11 fixed-effects (cf. table)
• 1 random effect ⇒ adjectival lemma
• 10 interactions tested: modality with each other fixed-effects
• Removal of non-significant predictors and interactions

Accuracy of the model (10-fold cross-validation)
• mean accuracy = 0.882
• mean concordance probability = 0.947

Results and conclusion
Anteposition is favored by:
• Demonstrative, possessive and definite determiners
• Other dependents of the noun (relative, PP and postposed adjectives)
• 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
⇒ Anteposition 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