



HAL
open science

A contribution of discourse analysis to the morphology of nominalizations. Study of the use of nominalizations in the genre of the scientific activity report using a corpus linguistics approach

Hugo Dumoulin

► **To cite this version:**

Hugo Dumoulin. A contribution of discourse analysis to the morphology of nominalizations. Study of the use of nominalizations in the genre of the scientific activity report using a corpus linguistics approach. International Symposium of Morphology (ISM0) 2023, Sep 2023, Nancy, France. hal-04287318

HAL Id: hal-04287318

<https://hal.science/hal-04287318v1>

Submitted on 15 Nov 2023

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.

A contribution of discourse analysis to the morphology of nominalizations. Study of the use of nominalizations in the genre of the scientific activity report using a corpus linguistics approach

Hugo Dumoulin

Laboratoire MoDyCo UMR7114 & Université Paris Nanterre

Introduction



Figure 1. Archive boxes as they appear in Nanterre university archive room

- Framework of ArchivU project (Labex *Les passés dans le présent*)
- Scientific activity reports in french universities (1964-today)
- Study of discourse genre (Bakhtine 1980) : describe the link between a social activity sphere and linguistic observables
- Exploring here an interface with morphology : nominalizations and the affix rivalry discussion (Missud & Villoing 2020)

Corpus and methodology

Corpus and databases

- RAPPORTS : 38 HCERES reports in 2018 (Lethier, Née, Dumoulin 2021)
- VCEUX : annual greetings of the presidents of the French republic since the 1970s (Leblanc 2016)
- ADMIN : activity reports from two French administrations (ADMR and Ucanss) in 2018
- SCIENTEXT : corpus of scientific articles (Tutin & Grossman 2014)
- LEXIQUE 3 : Frantext contemporary novels & subtitles corpus (New, Pallier, Brysbaert & Ferrand 2004)

Methodology

- ↓ TXT files to XML format
- ↓ Tokenization & lemmatisation with TreeTagger
- ↓ Filtering nominalizations with Demonette database (Hathout & Namer 2014)
- ↓ R (Ihaka & Gentleman 1983) for Correspondence analysis and TXM textometry software (Heiden et al. 2010) for cooccurrence visualization

Results

Nominalization in general and the administrative and scientific discourse genres

	Corpus size	Verbal nominalizations			
		Type	Token	Token frequency (%)	Type/Token Ratio (sample : 40 000)
RAPPORTS	921 898	1 028	26 552	2,8801	0,232
VCEUX	61 197	263	975	1,5932	0,355
SCIENTEXT	4 185 487	1564	119 396	2,8526	0,208
ADMIN	43 145	226	1266	2,9343	0,179
Lexique3_romans	~14 700 000	2497		0,36089	
Lexique3_films	~50 000 000	8226		0,20973	

Figure 2. Distribution of verbal nominalization among discourse genres

- Nominalizations appear to be **more frequent** in scientific and administrative discourse genre
- Nominalizations show **less lexical diversity** in those genres (lowest TTR)

Results

A statistical attraction for distinct suffixes among genres

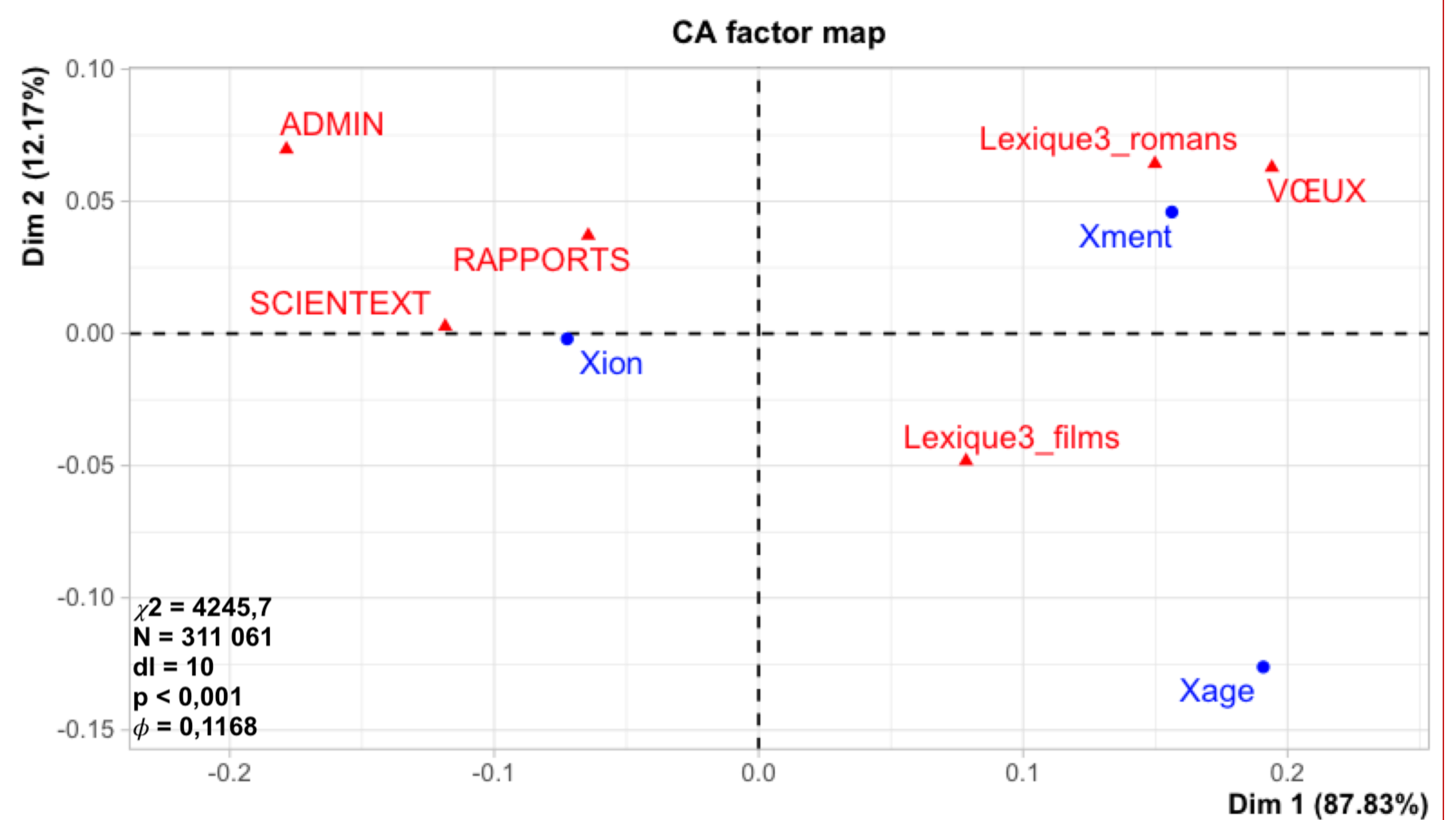


Figure 3. Correspondence analysis (Benzécri 1971) representing the distribution of derivational suffixes among the corpora

- A major opposition between Xion and the rest (87,83% of variance)
- The attraction of Xion for scientific and administrative genres : an abstract suffix (Martin 2008)
- Xage confirms its « concrete nature » (Missud & Villoing 2021) : representing oral interaction
- Xment appears to be more formal and literary

Nominalization as a part of a routinization process

Requête [demonext="Xion"][frlemma="de|du"]

Cooccurent	Fréquence	CoFréquence	Indice	Distance moyenne
NOM unité	1912	254	54	3.2
NOM produit	260	84	48	1.2
NOM connaissance	337	82	36	1.4
VER:pres rechercher	56	35	32	3.0
NOM savoir	332	76	32	2.0
NOM donnée	709	111	30	1.9
NOM écosystème	82	39	30	2.0
NOM recherche	5080	403	29	3.0
NAM Introduction	28	23	26	3.0
NOM colloque	1067	123	21	2.3
NOM activité	1364	143	20	3.3

Figure 4. Specific co-occurents in the complement position of -ion nominalizations (distance to the right : 5 occurrences).

- CQP query to find the cooccurents in the object position of nominalizations
- High Lafon indice (1981) for lexemes related to the lexical field of scientific activity

>> The nominalization takes part in a discourse routine (Sitri & Véniard 2017)

Requête [demonext="Xment"][frlemma="de|du"]

Cooccurent	Fréquence	CoFréquence	Indice	Distance moyenne
NOM recherche	5080	157	21	2.9
NOM outil	325	24	10	2.2
NOM chercheur	1200	48	10	2.3
NOM donnée	709	35	10	1.9
ADJ nouveau	1440	53	10	2.5
NOM intérêt	214	19	10	1.6
NOM fond	41	10	9	1.0
DET:ART un	13124	253	8	2.9
NAM CTEM	8	5	7	3.2
NOM signal	15	6	7	1.0
NOM doctorants	922	35	7	2.4
NOM professeur	351	20	7	2.5

Figure 5. Specific co-occurents in the complement position of -ment nominalizations (distance to the right : 5 occurrences).

- Most of the time the agent argument is **not** represented in a syntactic position

>> That routine constraints discours semantics in the corpus : events that affects scientific activity seem to occur « by themselves »

>> A bureaucratic discourse genre ? (See Dumoulin 2022)

Conclusion

- Nominalizations contribute to the discursive semantics of discourse genres, especially with **routinization in the administrative and scientific report genre**
- Speech genres contribute to **distinguish rival affixes with a discursive criteria**