



**HAL**  
open science

## Caractériser les logiques disciplinaires et institutionnelles dans les rapports d'autoévaluation

Virginie Lethier, Emilie Née, Hugo Dumoulin

► **To cite this version:**

Virginie Lethier, Emilie Née, Hugo Dumoulin. Caractériser les logiques disciplinaires et institutionnelles dans les rapports d'autoévaluation. Michelangelo Misuraca; Germana Scepi; Maria Spano. JADT 2022: Proceedings of the 16th international conference on statistical analysis of textual data, 2, Vadistat Press, pp.544-551, 2022. hal-03889991

**HAL Id: hal-03889991**

**<https://hal.science/hal-03889991>**

Submitted on 8 Dec 2022

**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.

# Caractériser les logiques disciplinaires et institutionnelles dans les rapports d'autoévaluation

Virginie Lethier<sup>1</sup>, Emilie Née<sup>2</sup>, Hugo Dumoulin<sup>3</sup>

<sup>1</sup>Université de Bourgogne Franche-Comté –  
virginie.lethier@univ-fcomte.fr

<sup>2</sup>Université Paris Est Créteil – emilie.nee@u-pec.fr

<sup>3</sup>Université Paris Nanterre – hdumoulin@parisnanterre.fr

## Abstract

This paper studies self-evaluation reports of research laboratories in France. We aim to test the hypothesis of discipline and institutional position (status and size) as factors of discursive variation. For this purpose, a principal component analysis (PCA) is deployed on a corpus of 38 reports, articulating 60 descriptors of morphosyntactic order.

**Keywords:** Discours analysis; academic discourse; reports; genre; parts of speech

## Résumé

Cette contribution a pour objet d'étude les rapports d'autoévaluation des laboratoires de recherche en France. Il s'agit de tester l'hypothèse de la discipline et de la position institutionnelle (statut et taille) comme facteurs de variation discursive. Pour cela, est déployée sur un corpus de 38 rapports une Analyse en Composantes Principales (ACP), articulante 60 descripteurs d'ordre morphosyntaxiques.

**Mots clés :** Analyse du discours ; discours académique ; rapport ; genre ; morphosyntaxe

## 1. Introduction

C'est à l'une des formes de la littérature grise académique qui témoignent des mutations de la sphère de l'activité scientifique (Sargeac, 2018) que s'intéresse notre contribution : le rapport d'autoévaluation des laboratoires de recherche universitaires français. Instrument d'évaluation externe de l'activité scientifique, le rapport d'autoévaluation correspond à un écrit standardisé dont la production est fortement encadrée par le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres). Cette standardisation visée par l'Hcéres est-elle néanmoins effective ?

Il paraît légitime de supposer que le rapport d'autoévaluation, parce qu'il est un écrit hybride entre discours scientifique et discours managérial, est susceptible

d'être soumis à des variations (énonciatives, micro-textuelles ou stylistiques, argumentatives) imputables aux logiques disciplinaires et institutionnelles qui traversent la sphère d'activité scientifique. Plus précisément, la discipline dont relève l'unité de recherche qui produit le rapport, sa position institutionnelle et sa taille nous semblent constituer des facteurs de variation de premier plan.

Après avoir présenté brièvement les caractéristiques du rapport d'autoévaluation et les problématiques qu'il soulève, nous rendrons compte de nos choix méthodologiques destinés à émerger, dans une perspective exploratoire, les principes et les lieux textuels de variations de ce genre. Dans un dernier mouvement, nous présenterons les pistes ouvertes par une analyse en Composantes Principales.

## **2. Contexte de la recherche**

### ***2.1. Un écrit hybride, au caractère obligatoire et standardisé***

Témoignant du caractère contractualisé de la recherche scientifique en France, le rapport d'autoévaluation correspond à un écrit de commande produit dans le cadre du processus d'évaluation de l'activité scientifique organisé par l'Hcéres. L'agence évaluatrice contraint fortement la production de ces rapports, dont la comparaison serait facilitée par leur standardisation. En amont, le rapport est un formulaire structuré et encadré par des référentiels ; en aval, il est un imprimé à remplir (Barats, Née, 2020). L'organisation en parties et sous-parties, les intitulés des (sous-)titres, les séquences textuelles qui les composent et la façon dont celles-ci s'articulent sont prédéfinies. Loin de n'affecter que le niveau macrotextuel, la trame imposée contraint tous les versants de la textualité : elle fixe les thématiques à traiter, les outils cognitifs à manipuler, le lexique et les formules à imprimer. Peu propices à la réflexivité cette logique formulaire et le format bref imposé (Rastier, 2013) rentrent en tension avec l'objectif affiché du rapport (s'auto-évaluer). Par ailleurs, si le rapport affiche une visée informative, puisqu'il s'agit de « rendre compte » d'une activité scientifique, il est tout autant caractérisé par une visée argumentative : le rapport est un espace d'autolégitimation et de valorisation de l'unité de recherche dans un contexte concurrentiel que l'exercice participe à alimenter.

Le rapport d'autoévaluation se situe à l'interface de deux types de sphères d'activité, et par suite, de deux types de discours : le discours scientifique d'une part, le discours managérial de l'autre. Appelant à s'acculturer à un lexique, à un « discours (et une rhétorique) de projet », mêlant technicité et idéalité (Sargeac, 2018 : 64), cet écrit est à rédiger sur la base d'outils cognitifs issus de l'entreprise comme la matrice « SWOT ». On notera d'ailleurs que ce sont les codes sémiotiques de la brochure d'entreprise qui cadrent la mise en page de certains rapports d'autoévaluation. Pour autant, en raison de l'objet même du rapport (l'activité scientifique), de la culture socio-professionnelle des scripteurs

et de ses modalités concrètes de production (recyclage et montage de documents, écriture collective), le rapport peut être considéré comme l'un des genres du discours scientifique. Le rapport d'autoévaluation figure ainsi un entre-deux discursif, dans l'espace textuel duquel l'acculturation et/ou l'adhésion aux formats et aux formes du discours managérial sont susceptibles de fluctuer.

## ***2.2. Hypothèses et questions de recherche***

Une abondante littérature s'est attachée à problématiser et décrire les caractéristiques des pratiques diversifiées et des genres pluriels (articles, communications orales) que recouvre la catégorie de « discours scientifique ». Différentes études empiriques (Grossmann, 2010 ; Rinck 2010 ; Fløttum et Vold, 2010, Poudat et Follette 2012) ont ainsi démontré l'existence de logiques disciplinaires ou sectorielles (grands ensembles de disciplines) à l'origine d'importantes variations, affectant notamment les marques de l'énonciation. En nous inspirant de cet état de l'art, nous émettons à notre tour la première hypothèse selon laquelle le rapport d'autoévaluation, en tant que genre, est soumis à un principe de variation d'ordre disciplinaire : la discipline dont relèverait l'unité de recherche productrice du rapport déterminerait les dimensions plurielles de la textualité.

Par ailleurs, en France, les unités de recherche ne sont nullement des institutions aux caractéristiques intrinsèques homogènes. Elles peuvent correspondre statutairement à deux types de structure : les équipes d'accueil (EA) d'une part, les unités mixtes de recherche (UMR), labellisées par le Centre National de la Recherche Scientifique (CNRS). Qu'elles soient des EA ou des UMR, les unités de recherche peuvent être mono- ou inter-disciplinaire. Leur taille, c'est-à-dire leur dotation en personnels statutaires, varie enfin sensiblement. Ces variables constituent, selon nous, des sources potentielles de variation des caractéristiques des rapports d'autoévaluation.

Parmi l'ensemble de ces variables externes (discipline, statut, taille, consistance mono- ou inter-disciplinaire), lesquelles contribuent-elles prioritairement à expliquer les variations observables dans les rapports d'autoévaluation ?

## **3. Méthodologie**

### ***3.1. Corpus***

Notre corpus contient 38 rapports d'autoévaluation – soit 921989 occurrences. Homogène en termes de genre, ce corpus présente plusieurs facteurs de variation. Il intègre ainsi 12 rapports d'unités UMR (labellisation signalée dans le nom du CODE renvoyant à l'UR), contre 26 rapports d'équipes d'accueil. Par ailleurs, ces unités se subdivisent en deux types : monodisciplinaire ou interdisciplinaire. Notre corpus compte ainsi 24 rapports d'unités monodisciplinaires, couvrant les principales disciplines des Sciences humaines :

Discipline	Codes
Droit (n=7)	DROIT1, DROIT2, DROIT3, DROIT4, DROIT5, DROIT6, DROIT7UMR
Langues (n=3)	LANG1, LANG2, LANG3
Histoire (n=2)	HISTI1, HIST2
Littérature (n=2)	LITT1, LITT2
Archéologie (n=1)	ARCH1UMR
Economie (n=1)	ECO1UMR
Gestion (n=1)	GESTION
Info-com (n=1)	INFOCOM
Linguistique (n=1)	LINGUMR
Sciences po. (n=1)	POL1UMR
Psychologie (n=1)	PSYCHO
Thermique (n=1)	THERM1

Et intègre 14 rapports d'unités interdisciplinaires, articulant les disciplines listées ci-dessous :

Codes	Disciplines
INTERA	Physique, mathématique, mécanique, thermique
INTERarUMR	Archéologie, ethnologie
INTER1UMR	Histoire, sociologie
INTER2	Histoire, philosophie
INTER3	Sociologie, philosophie
INTER4UMR	Sciences politiques, sociologie
INTER5	Sociologie, sciences du sport, Médecine
INTER6UMR	Ethnologie, sociologie
INTER7UMR	Géographie, sociologie
INTER8UMR	Architecture, urbanisme, sociologie, anthropologie, géographie
INTER9UMR	Histoire, anthropologie
INTERpsy	Psychologie, psychanalyse
INTERcog	Psychologie, Sciences du sport
INTERlang	Etudes slaves, germaniques, histoire du droit

Enfin, sur la base de leurs ressources humaines fixes, 22 des unités de recherche considérées correspondent à des unités de petite taille<sup>1</sup> (en noir ci-dessus et dans le graphique), 10 à des unités de taille moyenne (en bleu), 6 à des unités importantes (en vert).

### 3.2. Jeu de données

Notre jeu de données articule 60 descripteurs d'ordre morphosyntaxiques, niveau moins sensible aux variations disciplinaires que le vocabulaire. 32 de ces

<sup>1</sup> La masse salariale annuelle M sépare les laboratoires en groupes relativement distincts selon leur petite, moyenne ou grande taille ( $M < 2\,250\,000\text{€}$  ;  $2\,250\,000\text{€} < M < 3\,250\,000\text{€}$  ;  $M > 3\,350\,000\text{€}$ ).

descripteurs sont relatifs à la ponctuation et aux catégories grammaticales, telles que traditionnellement reconnues et étiquetées automatiquement via TreeTagger<sup>2</sup>, appelé par TXM (Heiden et al., 2010), logiciel employé pour segmenter et indexer notre corpus. Ces premières indications ont été complétées par un jeu de 22 descripteurs issus de l'analyseur Cordial<sup>3</sup>, qui renseignent notamment le type de propositions et le type de phrase (verbale ou non) du corpus, et différentes caractéristiques formelles telles que le nombre de lettres par mots. Certains des descripteurs ont été sélectionnés en vue d'affiner la lisibilité des catégories grammaticales traditionnelles, à l'instar des cinq sous-types d'adverbes (lieu, temps, manière, doute, négation) que l'analyseur se propose de repérer, ou du caractère abstrait/concret des noms communs. Six descripteurs spécifiques aux caractéristiques du genre du rapport d'évaluation et aux objectifs de notre recherche ont enfin été intégré : ils se rapportent à l'emploi des principaux pronoms personnels utilisés dans l'article (*il impersonnel*<sup>4</sup>, *on*, *nous*). Concernant le pronom nous, sa place syntaxique a été renseignée (*sujet*, *objet*) au terme d'une procédure d'annotation manuelle.

#### 4. De l'analyse en composantes principales

En vue de mettre au jour les interrelations entre ces descripteurs et de repérer, parmi les rapports d'autoévaluation, des groupes homogènes ou atypiques, nous nous sommes appuyés sur une Analyse en Composantes Principales (ACP). Cette méthode factorielle présente l'intérêt de réduire la complexité de la matrice initiale en un faisceau réduit de variables artificielles.

##### 4.1. Vue sur l'ACP

L'axe 1, qui représente 20.06% de la variance, marque une opposition principalement relative au degré de complexité des phrases. Dans son versant gauche, se concentrent les propositions indépendantes et les indices d'un style à consistance nominale (*noms communs*, *adjectifs*, *noms concrets*), tourné vers la précision et l'énumération (*parenthèses et deux points*) et faisant la place belle aux données chiffrées (*numéral*, *chiffres*, *adjectifs cardinaux*). À droite de l'axe 1 se donne à lire le profil d'une écriture plus complexe (*propositions subordonnées et relatives*, *pronoms relatifs et démonstratifs*, *virgule*), à la consistance plutôt verbale (*phrases verbales*, *adverbe*, *verbes*) recourant à l'ensemble des pronoms personnels.

<sup>2</sup> Institute for Computational Linguistics, University of Stuttgart.

<sup>3</sup> Synapse Développement.

<sup>4</sup> Seuls les cas du « il » dans une construction impersonnelle ont été comptabilisés grâce à une recherche exploitant les possibilités du Corpus Query Language sur corpus étiquetés, dans TXM

L'axe 2, qui extrait 9.04 % du nuage de points, contribue à singulariser dans sa partie haute une énonciation subjectivée et embrayée, centrée sur le « nous » (PRO1PP, verbe1PP, PRO1PP\_sujet). Si l'axe est négativement corrélé aux verbes conjugués à la troisième personne du singulier, ses points extrêmes correspondent à deux marques de ponctuation (*guillemets*, *tirets*) dont la pluralité des fonctions en discours entrave toute interprétative sans retour au texte.

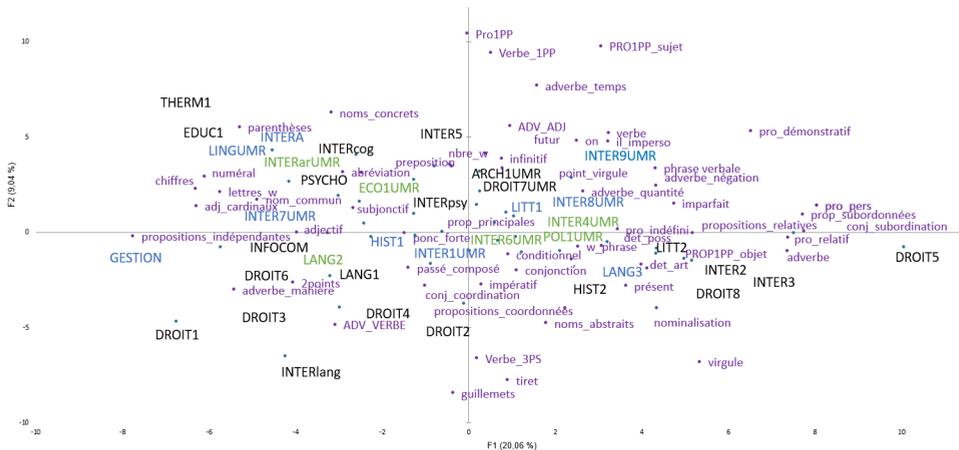


Figure 1. Visualisation des individus et des variables actives

#### 4.2. Éléments d'interprétation

Si l'on considère la position des unités de recherche sur ce graphique, on repérera que le premier facteur semble prioritairement refléter des logiques sectorielles, affiniées (ou contrariées) sur le deuxième facteur par le statut du laboratoire et sa taille. Le facteur 2 tend en effet à attirer, dans sa zone supérieure, les UMR et/ou de taille moyenne et forte, pour reléguer la plupart des équipes d'accueil dans sa zone basse. Le caractère mono-ou interdisciplinaire n'aurait, pour sa part, qu'une très faible valeur explicative. Sous l'effet conjoint des deux premiers facteurs, quatre regroupements notables sont à repérer.

Dans le quart supérieur ouest du graphique, s'entremêlent les sciences physiques (THERM1, INTERA) et certaines sciences humaines dites formelles et/ou outillées : l'économie (ECO1UMR), les sciences de la cognition et de l'expérimentation (PSYCHO, INTERcog, INTERPSY), la linguistique (LINGUMR). C'est dans cette zone que s'inscrivent l'UMR d'urbanisme (INTER7UMR), l'unité interdisciplinaire tissant archéologie et ethnologie (INTERRarUMR), ainsi que les sciences de l'éducation (EDUC1).

Le quart supérieur est du graphique tend, pour sa part, à concentrer des unités UMR ou de taille moyenne tournées vers les sciences sociales, à l'instar des unités interdisciplinaires INTER9UMR (histoire, anthropologie) et

INTER4UMR (sciences politiques, sociologie). Il semblerait que l'inscription de l'unité liée au Droit dans cette zone soit principalement liée à son statut d'UMR. C'est en effet dans le quart inférieur ouest du graphique que se concentrent la plupart des autres unités de droit et de langues (LANG1, LANG2), la gestion et les sciences de l'information-communication, ainsi qu'un laboratoire d'histoire de taille moyenne.

Dans le quart inférieur est du graphique, on retrouvera principalement de petites équipes d'accueil, s'inscrivant dans des disciplines principalement déjà représentées dans la partie ouest de la zone basse du graphique (le droit, les langues, l'histoire), ainsi que deux unités interdisciplinaires partageant de prendre appui sur la philosophie (INTER2 ; INTER3).

## 5. Synthèse

Dans cette contribution, une analyse en composantes principales a été menée en vue de mieux cerner la façon dont des variables extra-discursives se distinguaient et s'articulaient pour expliquer les différences d'écriture des rapports d'autoévaluation. Cette première classe d'analyse, qui doit bien entendu être corroborée par une méthode de classification hiérarchique (CAH), confirme que les rapports d'autoévaluation diffèrent en fonction des disciplines et du statut institutionnel des unités de recherche, et plus discrètement, de sa taille.

D'autre part, nous attendions de cette analyse exploratoire qu'elle oriente nos regards d'analystes du discours vers les lieux morphosyntaxiques matérialisant ces variations. En l'occurrence, la complexité syntaxique des phrases, les marques de l'énonciation, la séquentialité (description prenant la forme de listes ou de développements plus élaborés), l'usage ou l'évitement des embrayeurs, la temporalité (verbes et adverbes), enfin le lexique abstrait et les nominalisations sont autant d'observatoires à prendre en considération pour des recherches ultérieures.

## Bibliographie

Barats C., Née É. (2020). « Structurer, encadrer et lisser l'évaluation : les écrits de l'AERES/HCERES ». *Communication et langage* 203 : 5-25. DOI 10.3917/com1a1.203.0005.

Fløttum K., Vold E.T. (2010). « L'éthos auto-attribué d'auteurs-doctorants dans le discours scientifique ». *Lidil* 41 : 41-58. DOI : 10.4000/lidil.3006.

Grossmann F. (2010). « L'Auteur scientifique. Des rhétoriques aux épistémologies ». *Revue d'anthropologie des connaissances*, vol. 4/3 : 410-426. DOI 10.3917/rac.011.0410.

Heiden S., Magué J.-P., Pincemin B. (2010). « TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement ». *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010, Rome*, pp. 1021-1032.

- Née E., Oger C., Sitri F. (2017). « Le rapport : opérativité d'un genre hétérogène ». *Mots* 114 : 9-24. DOI 10.4000/mots.22752
- Rastier F. (2013). *Apprendre pour transmettre. L'éducation contre l'idéologie managériale*. PUF.
- Sargeac M. (2018). *Le paradoxe de l'isomorphisme dans la concurrence ? Une analyse des discours par et sur les établissements de l'enseignement supérieur français*. Mémoire de master 2, Paris, EHESS/ENS.
- Poudat C., Follette P. (2012). "Corpora and academic writing: A contrastive analysis of research articles in biology and linguistics". In Boulton, A., Rowley-Jolivet, E., Carter-Thomas, S., *Corpus-Informed Research and Learning in ESP: Issues and Applications*. John Benjamins.
- Tutin A. (2010). « Dans cet article, nous souhaitons montrer que... Lexique verbal et positionnement de l'auteur dans les articles en sciences humaines ». *Lidil* 41 : 15-40. DOI : 10.4000/lidil.3040.