La correspondance de motifs, un outil pour l'analyse du discours ?

Correspondence Patterns, a Tool for Discourse Analysis?

Marine Riguet et Mohamed Amine Boukhaled

Édition électronique

URL : <https://journals.openedition.org/revuehn/312>

DOI : 10.4000/revuehn.312

ISSN : 2736-2337

Éditeur

Humanistica

Référence électronique

Marine Riguet et Mohamed Amine Boukhaled, « La correspondance de motifs, un outil pour l'analyse du discours ? », *Humanités numériques* [En ligne], 1 | 2020, mis en ligne le 01 janvier 2020, consulté le 15 juillet 2021. URL : <http://journals.openedition.org/revuehn/312> ; DOI : <https://doi.org/10.4000/revuehn.312>

Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la Licence Creative Commons Attribution 4.0 International.

La correspondance de motifs, un outil pour l'analyse du discours ?

Correspondence Patterns, a Tool for Discourse Analysis?

Marine Riguet et Mohamed Amine Boukhaled

Résumés

Dans la seconde moitié du XIX^e siècle, une partie de la critique littéraire française revendique le caractère scientifique de son discours, au point de prendre pour modèles certains essais de biologie, de médecine, d'anthropologie, de psychologie ou encore de sociologie. Cet article se propose de mener une enquête préliminaire afin de repérer un éventuel point commun discursif entre la critique littéraire et les sciences de la fin du XIX^e siècle. Pour ce faire, notre étude s'appuie sur une méthode computationnelle procédant par extraction des motifs syntaxiques et analyse factorielle des correspondances (AFC). Plutôt qu'une analyse stylistique, nous entreprenons ici une exploration à l'échelle du discours, conduite par la découverte de motifs surreprésentés dans un texte par rapport à un corpus donné.

In the second half of the 19th century, some French literary critics proclaimed the scientific nature of their discourse and regarded scientific essays (in biology, medicine, anthropology, psychology or sociology) as models. This article proposes to conduct a preliminary study to identify a possible discursive common ground between literary criticism and the sciences at the end of the 19th century. To do so, our study is based on a computational method, using extraction and correspondence analysis (CA) for syntactic patterns. Rather than a stylistic analysis, we undertake here a transdiscursive exploration, led by the discovery of relevant patterns in a text within a given corpus.

Entrées d'index

MOTS-CLÉS : littérature, analyse de discours, analyse des données, critique littéraire, interdisciplinarité, motif syntaxique, traitement automatique des langues

KEYWORDS: literature, discourse analysis, data analysis, literary criticism, interdisciplinarity, syntactic pattern, natural language processing

Introduction

¹ Dans une fin du XIX^e siècle où émergent bon nombre de nouveaux discours scientifiques (psychologique, sociologique, biologique, etc.), la critique littéraire construit elle-même la « positivité » de son discours, au sens où l'entend Michel Foucault dans *L'Archéologie du savoir*, c'est-à-dire en définissant des domaines de *validité*, de *normativité*, et d'*actualité* à partir desquels elle pourra s'unifier et s'autonomiser (Foucault 2008, 85). Aussi est-ce la question d'une formation discursive qui se pose en premier. Malgré les formules tonnantes d'une critique devenue « science de la littérature » pour les uns, « science des faits humains » pour les autres, l'enjeu ne nous semble pas tant résider dans la délimitation d'un champ disciplinaire que dans la modulation d'un discours pensé à partir d'une matrice de type scientifique. C'est dire que la critique littéraire entend moins collecter les savoirs que se donner la légitimité d'en produire de nouveaux. Cette re-production, à partir de modèles, nous semble donc passer d'abord par un processus plus ou moins conscient d'imitation, de reprise, avant de générer à son tour un nouveau discours. Gérard Genette le pose clairement, « l'imitation n'est plus une simple reproduction, mais bien une production nouvelle : celle d'un autre texte dans le même style, d'un autre message dans le même code » (Genette 1992, 110). S'il souligne la fécondité de l'imitation intertextuelle, c'est qu'elle ne porte pas sur un texte lui-même, mais sur un style ou un genre, par le biais d'une « matrice de compétence » commune entre imité et imitatif.

Car imiter, en littérature comme ailleurs, suppose toujours [...] la constitution préalable (consciente et volontaire ou non : les imitations juvéniles sont le plus souvent des contagions subies) d'un modèle de compétence dont chaque acte d'imitation sera une performance singulière, le propre de la compétence étant, ici comme ailleurs, de pouvoir engendrer un nombre illimité de performances correctes. (Genette 1992, 108)

² Ainsi l'imitation se réalise-t-elle indirectement par la reprise d'une même matrice plutôt que de son produit. Par là seul, le texte imitatif peut pleinement rivaliser avec le texte imité, en se posant comme nouvelle « performance correcte », c'est-à-dire en prétendant, tout en étant autre, aux mêmes compétences. Cela nous paraît tout aussi valable à l'échelle du discours. Pour ériger la science en modèle-rival¹, la critique

imite moins un texte scientifique en particulier qu'elle ne tente de se positionner sur une matrice de compétence commune – celle d'un discours de la connaissance, apte à générer des productions nouvelles.

3 Afin d'identifier les éventuels emprunts ou imitations de la critique littéraire sur le plan macrostructurel, il faudrait adopter pour postulat que la science se constitue au niveau du discours ; qu'elle a un langage, un genre, qui lui sont propres : qu'elle existe dans sa façon de *produire* un savoir. Mais sans nier pour autant la pluralité de ses formes d'expression, de ses écritures et de sa rhétorique, soit, en somme, sa propre littérarité ; sans non plus chercher à définir son discours en réduisant ses manifestations à un corps préconstruit dont il ne pourrait que dépendre. Afin de contourner la dichotomie, nécessairement caricaturale, entre un discours littéraire et un discours scientifique, afin de penser en termes de matrice commune des performances singulières, il nous a semblé plus juste de recourir, avant toute chose, à une enquête discursive computationnelle et transversale. Contrairement à une approche par classification, qui préétablirait des caractéristiques linguistiques d'un discours (scientifique) pour en chercher la trace dans un autre (littéraire), nous privilégions à ce dessein une approche plus heuristique. L'emploi d'une méthode computationnelle nous offre ici la possibilité d'analyser automatiquement des textes sans connaissance préalable, sans catégories discursives posées *a priori*. Plus spécifiquement, elle permet une première exploration transdiscursive, conduite par la découverte de motifs linguistiques surreprésentés dans un texte par rapport à un corpus donné.

Méthode

4 Nous choisissons un outil de la stylistique computationnelle inédit, qui procède par extraction de motifs à l'aide de méthodes statistiques et automatiques, afin de caractériser un style ou un type de textes particulier (Mahlberg 2014, Quiniou *et al.* 2012, Boukhaled 2016). Pour autant, confessons d'entrée de jeu que notre étude n'a pas de prétention stylistique : à rebours des approches traditionnelles, notre souci n'est pas de caractériser la singularité d'un auteur ou d'une œuvre qui se manifesterait dans l'écart, mais d'observer si se dessinent éventuellement des types de discours à partir de traits linguistiques communs entre des groupes de textes. Nous éprouvons d'ailleurs une réticence à parler de style dans ce cadre. Malgré les récents travaux de stylistique computationnelle, il nous semble en effet que l'extraction de motifs récurrents, ou *patterns*², si elle fournit bien des indices de forme, se situe sur un plan flou où sont en corrélation sémantique et syntaxe, individuel et collectif, et dans lequel la notion de style sonne creux. Le motif a déjà donné lieu à de nombreuses définitions, parmi lesquelles nous relevons celle de Dominique Longrée et de Sylvie Mellet, qui met bien en exergue son double aspect multidimensionnel et récurrent :

La notion de motif est conçue comme un moyen de conceptualiser la multidimensionnalité (ou le caractère multiniveaux) de certaines formes récurrentes qui sollicitent à la fois le lexique, les catégories grammaticales et la syntaxe, éventuellement la prosodie, la métrique. (Longrée et Mellet 2013)

5 Notre analyse présuppose en effet que lexicale, catégories grammaticales et syntaxe se conditionnent mutuellement, ce que le motif permet précisément de représenter.

6 Cela posé, nous concentrons le repérage computationnel sur le motif morphosyntaxique, en suivant la méthode employée par Francesca Frontini et Mohamed Amine Boukhaled au cours de leurs travaux (Boukhaled, Frontini, et Ganascia 2015 ; Frontini, Boukhaled, et Ganascia 2015). L'analyse morphosyntaxique nous permet en effet d'approcher l'agencement de certains segments de discours. Elle part bien sûr de la langue qui, si l'on reprend la définition posée par Roland Barthes dans *Le Degré zéro de l'écriture*, correspond à cette grammaire collective, ce « corps de prescriptions et d'habitudes, commun à tous les écrivains d'une époque » (Barthes 2009, 15). Mais elle permet aussi de sortir de ce commun de la langue en identifiant des segments qui se désolidarisent, qui se singularisent, et qui approchent alors des formes d'écriture idiosyncratiques, « en dehors de l'installation des normes de la grammaire et des constantes du style » (Barthes 2009, 18). C'est la raison pour laquelle nous abordons cette extraction de motifs morphosyntaxiques dans une visée comparatiste : non pas pour définir des formes d'écriture, puisqu'une telle méthode nous semblerait bien insuffisante pour cerner véritablement cette singularité qui fait œuvre³ – ce qui nous éloigne des intentions de la stylistique computationnelle ; mais pour repérer des segments discursifs communs à certains auteurs plutôt qu'à d'autres, témoignant d'une intention, d'une pensée et, peut-être, d'une construction discursive communes. De ce fait, nous détournons l'outil computationnel de son usage premier. Concrètement, l'extraction des motifs morphosyntaxiques dans une perspective comparatiste nous permet de tester empiriquement les questions suivantes : trouve-t-on des marques d'un discours scientifique par rapport à un discours littéraire sur le plan syntagmatique ? Peut-on réellement tracer une distinction domaniale à partir de motifs récurrents communs ? Et est-ce à ce niveau que se joue l'imitation d'un « discours scientifique » par la critique littéraire ?

7 Pour mener cette expérience, nous entendons le texte comme un ensemble de séquences syntaxiques, segmentées et étiquetées à l'aide de l'outil TreeTagger⁴. Dans ce contexte, on appelle motif un segment syntagmatique séquentiel présent dans une séquence syntaxique. Tels sont par exemple des motifs morphosyntaxiques particulièrement récurrents dans un discours de langue française : <DET><NOM><ADJ> ; ou <NOM><VER><ADV> ; ou encore <PRP><DET><NOM>. Un algorithme d'extraction de motifs séquentiels⁵ nous permet alors d'extraire les motifs d'un texte avec leurs fréquences d'apparition (relatives au nombre de phrases totales dans le texte et atteignant un seuil minimum de 1 %⁶).

8 Cette méthode automatique propose une première entrée pour saisir, sur l'axe syntagmatique, la cohésion du texte comme discours (Frontini, Boukhaled, et Ganascia 2015). Mais elle nous intéresse surtout en raison de la possibilité de comparer, regrouper ou espacer des textes selon la fréquence de représentation des motifs d'un texte par rapport aux autres ; autrement dit, de confronter des textes de nos différents corpus de critique littéraire, de sciences exactes et de sciences humaines, afin d'observer d'éventuelles corrélations entre les domaines dont ils sont is-

sus et la surreprésentation de certains motifs morphosyntaxiques. Pour rendre cette expérience plus significative, nous décidons de la mener sur trois corpus transdisciplinaires restreints et formés par genre :

- un corpus « Étude », composé de huit études critiques et scientifiques de portée analytique
- un corpus « Théorie », composé de neuf essais littéraires et scientifiques à la visée plus théorique
- un corpus « Histoire », composé de huit ouvrages d'histoire, d'histoire littéraire et d'histoire naturelle

⁹ Les textes ont été numérisés au préalable dans le cadre de notre thèse de doctorat (Riguet 2018), avec l'aide du Labex Obvil (Observatoire de la vie littéraire), dans le cadre d'un projet d'édition numérique des ouvrages de critique littéraire française publiés entre le xvii^e et le xx^e siècles : ils ont été ocrés à partir de la première édition définitive et structurés en XML-TEI. La majorité d'entre eux est consultable en ligne⁷.

¹⁰ Après l'extraction automatique des motifs morphosyntaxiques, une analyse factorielle des correspondances (AFC) est menée sur chacun des corpus (Benzécri 1982)⁸. De cette façon, nous traitons uniquement la représentation des motifs d'un texte relativement à leur fréquence d'apparition dans les autres textes du corpus⁹. Nous l'avons dit, c'est moins la caractérisation d'un écart que l'exploration empirique d'éventuelles constructions discursives communes que nous espérons obtenir par ce rapport.

Corpus « Étude »

¹¹ Afin d'éprouver les éventuels types de discours, nous avons cherché à former un corpus le plus homogène possible, constitué autour d'une même intention d'écriture, quoique pour des objets d'études distincts. Dans cette perspective, nous avons sélectionné parmi notre masse de textes ceux qui nous semblaient partager une même visée analytique et explicative, ceux dont le ton pouvait le mieux coïncider, que ce soit en critique littéraire, ou en sciences exactes ou humaines ; c'est-à-dire en nous concentrant sur des études et en écartant tout traité, manifeste théorique ou histoire. Pour rendre les résultats de l'expérience lisibles, il nous a toutefois fallu limiter le nombre de textes ; d'autres titres auraient tout à fait pu répondre à ces critères, et il y a donc nécessairement une part d'arbitraire dans l'établissement des corpus à prendre en compte, ensuite, dans l'interprétation de nos résultats. Mais dans notre volonté d'identifier un éventuel commun discursif, nous avons veillé à construire un corpus à la fois le plus cohérent possible dans l'intention des auteurs, et le plus diversifié possible concernant leurs positionnements esthétiques et les objets traités. Nous avons également tâché d'exclure les ouvrages qui regorgeaient de citations textuelles et de discours rapporté (en particulier en critique littéraire), et de choisir des textes de tailles équivalentes en retranchant les notes de bas de page, de façon à rendre l'analyse de fréquence des motifs signifiante. Ainsi avons-nous sélectionné, par ordre chronologique, les *Leçons d'anatomie comparée* de Georges Cuvier (1805), le tome I des *Leçons de physiologie expérimentale appliquée à la médecine* de Claude Bernard (1855), le tome I des *Portraits littéraires* de Charles Augustin Sainte-Beuve (1862), les *Leçons cliniques sur les maladies des vieillards et les maladies chroniques* de Jean-Martin Charcot (1874), le tome I des *Contemporains* de Jules Lemaître (1886), les *Études littéraires : dix-neuvième siècle* d'Émile Faguet (1887), *Quelques écrivains français* d'Émile Hennequin (1890) et le tome I des *Études et Portraits* de Paul Bourget (1905).

¹² La figure 1, générée à partir du module FactoMineR de R (Husson, Josse, et Lê 2008), rend visible l'analyse de correspondances de ces textes, c'est-à-dire qu'elle les répartit dans un espace bidimensionnel selon la fréquence de représentation relative de leurs motifs les uns par rapport aux autres.

13 Dans cette représentation, les textes sont spatialisés selon leurs motifs morphosyntaxiques. Le point de rencontre (0 ; 0) de l'abscisse et de l'ordonnée marque la situation d'un motif qui aurait obtenu la même fréquence relative dans l'ensemble des textes du corpus. Plus un texte est proche d'un autre, plus la correspondance de motifs entre les deux est élevée.

14 Cette figure est d'emblée très éloquente. À l'exception du texte d'É. Faguet, qui se trouve en situation marginale par rapport au reste du corpus, nous obtenons deux agglomérats distincts autour de l'axe des abscisses, l'un au-dessus rapprochant nos auteurs scientifiques (Cl. Bernard, G. Cuvier, J.-M. Charcot), l'autre en dessous concentrant nos auteurs littéraires (P. Bourget, J. Lemaître, Ch. Sainte-Beuve, É. Hennequin). Il y aurait donc, dans ce corpus, des motifs morphosyntaxiques plus ou moins propres aux textes scientifiques, et d'autres plus ou moins propres à la critique littéraire. Mais allons plus loin. Sur la figure 2 sont représentés les 10 motifs à la force de contribution la plus grande ; soit, pour le dire autrement, les 10 motifs les plus discriminants, ceux qui saillent à la lumière de notre méthode comparatiste et qui, par leur sous- ou leur surreprésentation, causent cette répartition spatiale des textes.

15

Commençons peut-être par remarquer que É. Faguet obtient cette position médiane pour avoir très peu de motifs discriminants relativement aux autres auteurs. Les motifs qui lui sont propres, qui expliquent son écart à droite mais qui restent néanmoins trop peu représentés pour jouer un rôle fort dans la contribution totale, sont majoritairement dus aux majuscules des débuts de vers poétiques qu'il cite dans ses études. En outre, É. Faguet recourt souvent à une même construction emphatique de la phrase, en débutant par le motif <PRO:DEM><VER> : « C'est un divertissement malin et poli », « C'est la note dominante d'*Émaux et Camées* », « C'est une variété, mais très distincte, de la littérature descriptive », « C'est du moins, assurément, le plus difficile à atteindre », etc. Cette emphase assertive, servant l'art de la formule efficace, ne ressort pas dans les autres textes et explique en partie, sur la figure, l'exclusion de É. Faguet vers la droite. Mais laissons de côté ce pan de l'étude qui, nous l'avons dit, n'est pas l'objet de notre expérience.

16

En ce qui concerne un commun discursif, des motifs de forte contribution apparaissent en revanche très nettement du côté des textes scientifiques d'une part, et littéraires de l'autre. Les motifs 1665, 1696 et 1673, dans le coin supérieur gauche de la figure, ont une fréquence relative particulièrement forte chez Cl. Bernard ; or, tous les trois sont les déclinaisons d'un même segment séquentiel :

- Motif 1696 : <PRP:"du"><NOM><PRP:"dans">
- Motif 1673 : <PRP:"du"><NOM><PRP:"dans"> <DET:ART>

- Motif 1665 : <PRP:"du"><NOM><PRP:"dans"><DET:ART><NOM>

17

Ils mettent en avant une itération importante qui caractérise, chez Cl. Bernard, ses analyses physiologiques : « du sucre dans l'organisme », « du sucre dans les urines », « du diabète dans le sang », « du sucre dans la fonction », etc. Ces motifs, par l'enchaînement de l'article partitif « du » et de la préposition « dans », se rapportent clairement à l'étude de la composition de l'intériorité organique. Or cette idée se retrouve dans le motif 1588 (<PRP:"dans"><DET:ART><NOM><SENT>), particulièrement discriminant pour G. Cuvier et Cl. Bernard, et emblématique de ce type de fin de phrase : « dans la respiration », « dans le sang », « dans la mâchoire », « dans les organes », etc. Déclinaison de ce dernier, le motif 1804 (<PRP:"dans"><DET:ART><NOM>) est lui aussi un des motifs les plus surreprésentés du corpus, présent dans tous les textes, mais avec une force de contribution beaucoup plus grande chez J.-M. Charcot, G. Cuvier et Cl. Bernard, ainsi que le laisse voir la figure 2. Ce sont par conséquent les compléments circonstanciels de lieu qui se trouvent être en quelque sorte marqueurs, dans ce corpus, d'un discours plus scientifique que littéraire ; mais d'un lieu intérieur, organique, dont les savants cherchent à analyser et expliquer le fonctionnement.

18

De ce fait, le sujet d'étude paraît déterminant dans l'organisation syntaxique du discours. Car les critiques littéraires, eux, s'attachent finalement moins à étudier la composition intérieure de quelque chose qu'à caractériser ce qui se rapporte à quelqu'un. Le motif 444 (<NOM><PRP:"de"><NAM>), qui a dans ce corpus la plus grande force de contribution, exprime bien cette tendance avec une nette surreprésentation chez l'ensemble des critiques littéraires (É. Hennequin et Ch. Sainte-Beuve en tête), et une sous-représentation chez les trois savants. Or il est assez clair qu'il se fait l'archétype d'une étude portant soit sur des personnes réelles, soit sur des personnages, comme en témoignent les exemples suivants : « actes de Rougon », « volte-face d'Hernani », « maison de Condé », « élève de La Bruyère », « héritage de Molière », « désenchantement de René », « poétique de Chateaubriand », « frère de Milton », etc. Plus intéressant encore, le motif 1585 (<PRP:"de"><DET:POS><NOM>) prolonge cette construction d'un discours critique centré sur les singularités – principalement biographiques ou littéraires – d'un homme : « de son talent », « de son œuvre », « de son existence », « de ses sentiments », « de son cœur », « de ses lectures », « de ses penchants », « de son art », « de sa maîtresse », etc. De façon générale, on remarque que les déterminants possessifs apparaissent dans la majorité des motifs qui se trouvent discriminants pour les textes de critique littéraire par rapport aux textes scientifiques. Parmi ceux que partagent P. Bourget, É. Faguet, É. Hennequin, J. Lemaître et Ch. Sainte-Beuve, les motifs 1262 (<NOM><PRP:"de"><DET:POS><NOM>) et 2198 (<NOM><ADJ><PRP:"de"><DET:POS>) contribuent fortement à cette répartition disciplinaire, tels qu'on les retrouve dans les exemples suivants : « pièces nombreuses de sa décadence », « défauts essentiels de son œuvre », « caractéristique de son génie », « naïfs de son temps », etc. Enfin, le motif 1482 (<ADJ><KON:"et"><ADJ>), visible dans la figure 2, fait quant à lui apparaître un discours critique privilégiant la qualification binaire : « maladif et taciturne », « rudes et amers », « sagace et pénétrante »,

« noble et poli », « élégant et grave », « naïf et vrai », etc. ; l'image littéraire naît ainsi, plurivoque, dans le binôme adjectival, tandis que la description scientifique lui préfère la précision de l'adjectif unique.

19

Ainsi, l'exploration de ce premier corpus « Étude » tendrait empiriquement à dessiner une distinction de nature entre deux discours, indissociable du domaine traité, penchant chez les uns vers l'analyse de composition intérieure, et chez les autres vers l'explication de caractéristiques extérieures ; plus encore, l'aménagement syntagmatique du discours exprime une démarcation entre, du côté scientifique, une portée générale du propos (articles indéfinis et noms communs), et, du côté de la critique, une recherche du particulier (noms propres et adjectifs possessifs).

Corpus « Théorie »

20

Dédié à des textes plus théoriques, notre second corpus réunit par ordre chronologique la *Physiologie des écrivains et des artistes* d'Émile Deschanel (1864), l'*Introduction à l'étude de la médecine expérimentale* de Cl. Bernard (1865), *Système de la logique déductive et inductive* de John Stuart Mill traduit par Louis Peisse à partir de la 6^e édition (Mill 1866), le tome I *De l'intelligence* d'Hippolyte Taine (1892), *La Critique scientifique* d'Émile Hennequin (1888), *L'Avenir de la science* d'Ernest Renan (1890), *La Méthode scientifique de l'histoire littéraire* de Georges Renard (1900), *Le Problème du style* de Remy de Gourmont (1902), et *La Théorie physique, son objet, sa structure* de Pierre Duhem (1906). Comme précédemment, nous tirons de l'analyse de correspondances des motifs morphosyntaxiques la spatialisation suivante (figure 3).

21 Deux constats s'imposent immédiatement au vu de cette figure. C'est d'abord une configuration complètement différente de notre première expérience qui se dessine, avec une répartition plutôt homogène des textes, quel que soit le domaine dont ils sont issus : É. Deschanel, G. Renard, P. Duhem, R. de Gourmont, E. Renan, J. Stuart Mill et Cl. Bernard obtiennent une grande proximité dans la représentation de leurs motifs. Mais, par opposition, force est de noter l'écart étonnant de H. Taine, et celui encore plus flagrant de É. Hennequin, qui, pour le second au moins, nous semblent d'emblée relever d'un « cas aberrant », c'est-à-dire relatif à un défaut de l'expérience elle-même.

22 Concernant É. Hennequin, les motifs morphosyntaxiques qui caractérisent son texte avec la plus grande force de contribution sont de deux ordres. Dans les motifs 2145 (<PRP:"d"><DET:ART><NOM><PRP:"d">), 2138 (<PRP:"d"><DET:ART><NOM><PRP:"d"><NOM>) ou 2177 (<NOM><PRP:"d"><NOM><VER>) – tels qu'on les retrouve par exemple dans les expressions suivantes : « d'une œuvre d'art », « d'un état d'âme », « d'une méthode d'investigation », « d'une rupture d'équilibre » –, c'est le « d' » qui, par le hasard des voyelles, se trouve discriminant. Quant aux motifs 2181 (<NOM><PUN:";"><PRP:"de"><ART>), 2141 (<DET:ART><NOM><PUN:";"><PRP:"par">), ou encore 2132 (<NOM><ADJ><PUN:";">), ils mettent en avant la récurrence de la virgule, que É. Hennequin emploie effectivement en abondance dès lors qu'il multiplie les propositions, comme dans le cas suivant :

L'*Hamlet*, la *Divine Comédie*, la symphonie en ut dièse mineur, une cathédrale gothique, le *Bon Samaritain* de Rembrandt, sont des œuvres excitantes à un haut degré parce qu'elles sont tristes, et dénuées cependant de tristesse, parce qu'elles n'ont de la douleur que le choc et non la blessure. Les mots « sensation du beau » sembleront donc désigner cette situation d'esprit : excitation intense d'un ou plusieurs sentiments ordinaires ; absence des images positivement c'est-à-dire personnellement douloureuses, qui accompagnent et timbrent d'habitude cette excitation intense ; en d'autres termes, le transport, le heurt de la douleur, sans son amertume ou sa terreur. Et la douleur entière, la vraie, le désir de l'éviter, étant les derniers mobiles de toute l'activité animale, humaine et sociale, nous comprenons maintenant pourquoi les suprêmes émotions esthétiques sont improductives d'actes, comme nous l'avons dit au commencement de ce chapitre ; ces émotions comprennent toutes les souffrances harcelantes de l'existence, mais sans les aiguillons des périls, des angoisses, des menaces, des maux prévus ou ressentis. (Hennequin 1888, 39)

23

En ce sens, notre méthode computationnelle fait bien ressortir certains traits d'écriture propres à É. Hennequin dans ce corpus. Il nous faut toutefois reconnaître que le degré de précision des étiquettes syntaxiques choisies perd ici de sa pertinence : en identifiant le type de préposition (« d' ») ou de signe de ponctuation (la virgule), l'étiquetage réduit les correspondances de motifs à la composition syntaxique semblable. Telles sont donc les premières limites de notre approche. Le cas de H. Taine peut se justifier d'une façon à peu près similaire. Les quatre motifs responsables de son isolement sur la droite (figure 3) contiennent tous l'étiquette <NUM>, dont ils révèlent la récurrence ; et en effet, il n'est pas rare de trouver dans *De l'intelligence* des numériques : « 451 billions par seconde », « vingt sensations », « deux mémoires », « deux états », « deux systèmes », « quatre langues », etc., alors que ces mêmes motifs figurent pour les autres textes dans moins de 1 % des phrases, y compris chez les auteurs scientifiques. Bien évidemment, cette caractéristique stylistique n'est pas sans intérêt, et met au jour, chez H. Taine, l'effort d'une analyse scientifique qui passe quasi systématiquement par un recensement précis. Mais, dans notre visée exploratoire, ces résultats très singuliers parasitent quelque peu les phénomènes plus globaux que nous nous efforçons d'observer.

24

Pour les autres textes du corpus, en revanche, on peut noter une certaine homogénéité des principaux motifs morphosyntaxiques (figure 4).

25

Le deuxième motif qui possède, pour l'ensemble, la force de contribution la plus grande révèle chez tous nos auteurs, et tout particulièrement chez Cl. Bernard et E. Renan, le recours très fréquent au segment syntagmatique suivant : <ADV><VER><ADV>, propre aux formules verbales négatives : « n'est point », « ne pourrait cependant », « n'ont pas été certainement », « ne conçoit nullement », etc. Il est complété par le motif 751 (<PRO:PER><ADV><VER><ADV>), qui se retrouve dans les expressions emphatiques suivantes : « je ne doute pas que », « il n'est pas étonnant que », « on ne conçoit pas que », etc. Cet usage très clairement argumentatif de la négative retient notre attention dans la mesure où il témoigne d'un effort rhétorique comparable des critiques littéraires et des savants. Or c'est bien cette même intention du discours que pointent les motifs récurrents de ce corpus théorique. Parmi les moins discriminants, c'est-à-dire parmi ceux qui sont le mieux partagés par nos auteurs (avec un taux de fréquence relative presque égal d'un texte à l'autre, et une force de contribution faible), se rencontrent des constructions syntaxiques tout à fait typiques du genre argumentatif. Le motif (<PRO:PER><VER><KON:"que">), par exemple, s'impose dans tous les textes de critique comme de science, et participe de la progression logique du discours : « on sait que », « il semble que », « nous croyons que », « on voit que », « il résulte que », « nous découvrons que », « on prouve que », etc. Il permet d'établir un lien causal entre les énoncés. À ce raisonnement vient bien sûr s'ajouter une rhétorique de la persuasion, qui perçoit notamment au travers des modalités déontiques, telles que dans le

motif (<PRO:PER><VER><VER:infi>) : « il faut éviter », « il faut reconnaître », « il faut marcher », « il faut conclure », « il faut maintenir », etc. Nous ne tenterons pas d'établir une liste exhaustive de ces motifs communs, qui deviendrait aussi vite fastidieuse qu'inutile, tant nous croisons ici les ressorts traditionnels de la rhétorique française. En revanche, il ne nous semblait pas superflu d'en relever rapidement l'application transdisciplinaire.

26

Cette cohésion discursive entre les domaines, nous la constatons peut-être mieux encore par le biais des motifs morphosyntaxiques les plus discriminants (avec, donc, une grande force de contribution). Preuve en est que le motif 803 (<PRO:PER><PRO:PER><VER>), segment par excellence de l'expression modale, est surtout caractéristique des textes de Cl. Bernard et É. Deschanel, et relativement plus éloigné des textes de P. Duhem et de É. Hennequin. En formant un syntagme verbal de ce type – « il me semble », « je vous jure », « je vous accorde », « il me suffit », « on me reprochera » – il sert la mise en scène d'une oralité qui implique doublement l'auteur et le lecteur, sur le modèle de l'art oratoire qui s'accorde ici à une visée professorale. Force est de rappeler pourtant que les textes de ce corpus « Théorie », contrairement à ceux du premier, ne sont pas des compilations de leçons ou de discours qui auraient été d'abord prononcés. Il s'agit bien d'un effort rhétorique partagé par certains auteurs critiques et scientifiques, façonnant un *ethos* dans le but de convaincre. De même, le motif 787 (<VER:infi><DET:ART><NOM>) participe à asseoir un style prescriptif ; plus propre aux textes de G. Renard, J. Stuart Mill, Cl. Bernard et E. Renan, il vient soit compléter un complément circonstanciel de cause (précédé des prépositions « pour » ou « afin de »), soit poursuivre un verbe de modalité (de l'ordre du devoir ou du vouloir) : « considérer les faits », « ordonner les événements », « construire l'histoire », « assigner un rôle », « former un système », « trouver la formule », « défendre l'instruction », « cultiver le peuple », « chercher la raison », « connaître leurs usages », etc. Outre la double visée didactique et normative qui apparaît, nous soulignons surtout le recours de tout discours théorique à une rhétorique traditionnelle, contrairement à l'idée reçue d'une science a-rhétorique qui pourrait être opposée au genre argumentatif de la critique littéraire.

27

Plus encore, le motif 802 (<PRO:PER><VER><ADV>), particulièrement discriminant pour Cl. Bernard, R. de Gourmont et E. Renan, révèle quelques nuances du discours dès lors qu'on l'observe dans le détail. Tandis que les deux critiques littéraires, R. de Gourmont et E. Renan, font quasi exclusivement appel à des connecteurs logiques qui appuient la validité déductive de leur raisonnement – « il sied donc », « il est ainsi », « il faut cependant », « on peut donc », « on dira encore », « nous verrons aussi », etc. –, le même segment renvoie chez Cl. Bernard à une récurrence d'adverbes d'intensité qui accentuent la modalisation du discours, comme suit : « il résulte nécessairement », « ils sont évidemment », « il déduit logiquement », « ils observent réellement », etc. N'en déplaise à la scientificité du propos, l'art de persuader est en marche. En ce sens, l'exploration de ce corpus nous invite à penser l'existence d'une rhétorique scientifique ; une rhétorique que, pour autant, nous ne pouvons pour l'instant considérer avec les socio-constructivistes comme propre à un « discours scientifique » que les théoriciens littéraires em-

prunteraient, mais peut-être à l'inverse, comme un agencement discursif passant nécessairement par certaines règles du genre argumentatif qui, en littérature, possède déjà ses codes.

Corpus « Histoire »

28

Le dernier volet de notre approche herméneutique porte sur un corpus plus historique, comprenant chronologiquement la *Philosophie zoologique* de Jean-Baptiste Lamarck (1809), le tome I du *Tableau de la littérature au Moyen Âge* d'Abel-François Villemain (1830), le tome I de l'*Histoire de la littérature française* de Désiré Nisard (1883), *De l'origine des espèces* de Charles Darwin dans la première traduction française de Clémence Royer (Darwin 1870)¹⁰, le *Tableau de la France* de Jules Michelet (1875), les tomes I et II des *Origines de la France contemporaine* d'Hippolyte Taine (1901), l'*Histoire de la littérature française* de Gustave Lanson (1895), et le *Manuel de l'histoire de la littérature française* de Ferdinand Brunetière (1898). L'analyse de correspondances des motifs morphosyntaxiques nous en fournit la spatialisation suivante (figure 5).

FIGURE 5. ANALYSE DE CORRESPONDANCES DU CORPUS « HISTOIRE »

29

Là encore, la concentration des textes de critique est remarquable par rapport aux autres. Alors que les courants revendiqués par certains de ces auteurs sont *a priori* très distincts – il suffirait d'opposer les projets esthétiques et philosophiques de D. Nisard et de H. Taine –, ils présentent une très forte homogénéité de motifs par contraste avec les

textes d'histoire et de sciences naturelles. Surtout, D. Nisard à la langue classique se retrouve finalement plus proche de Ch. Darwin/Cl. Royer et J.-B. Lamarck que de H. Taine. Au regard de ces résultats, il paraît donc difficile d'affirmer que certains critiques littéraires dits positivistes calquent davantage un discours scientifique que d'autres. Les phénomènes d'imitation discursifs sont du moins trop fins pour être perçus par notre méthode computationnelle à échelle macroscopique. Il est également intéressant de noter d'entrée de jeu la position des critiques, à peu près à équidistance de J. Michelet et de Ch. Darwin/Cl. Royer. Là encore, c'est l'habituel partage des discours, présupposant entre autres une filiation plus étroite entre des discours d'humanités, qui est remis en cause. La répartition globale des textes par correspondances de motifs se comprend mieux à la lumière de la figure 6 :

FIGURE 6. LES MILLE MOTIFS AYANT LA FRÉQUENCE ABSOLUE LA PLUS FORTE - CORPUS « HISTOIRE »

30

La majorité des motifs esquisse une forme de continuum entre deux pôles plus discriminants, l'un représenté par le texte de J. Michelet, l'autre par J.-B. Lamarck. La situation isolée de A.-F. Villemain (à droite) tient d'abord au fait que les motifs discriminants des autres textes du corpus sont chez lui sous-représentés ; en outre, les quelques motifs qui semblent caractériser son texte correspondent à des sortes d'hapax par rapport au reste du corpus, peu signifiants dans la mesure où ils renvoient à des citations latines mal analysées par notre étiqueteur morphosyntaxique.

31

Une analyse rapide permet de dessiner les grandes lignes de force de cette répartition d'ensemble. Les motifs qui obtiennent chez J. Michelet le plus fort taux de contribution, ceux donc qui tirent les critiques vers le haut de la figure, sont les segments syntaxiques formés par des noms

propres, à l'instar des motifs 964 (<NOM><PRP:"de"><NAM>), 840 (<DET:ART><NAM><PUN:";">), 953 (<PRP:"de"><NAM><PUN:";">) ou 1276 (<DET:ART><NAM><SENT>). Rien d'étonnant, en effet, à ce que l'histoire et l'histoire littéraire partagent des structures semblables dès lors qu'elles abordent le rôle tenu par des personnes réelles. Aux antipodes, J.-B. Lamarck, qui ne partage aucun de ses motifs fréquents avec J. Michelet, arbore une pléthore de constructions logiques marquées par différentes combinaisons d'adverbes et de conjonctions. Nous pouvons citer, parmi d'autres, le motif 3086 (<ADV><PUN:";"><ART><NOM>) exprimé ainsi : « cependant, la durée », « alors, les animaux », « ainsi, les genres », « néanmoins, les limites », etc., que J.-B. Lamarck partage surtout avec Ch. Darwin/Cl. Royer et F. Brunetière ; ou le motif 3096 (<PRO:DEM><ADV><PUN:";">) qui renvoie à l'itération « en outre, », si remarquable chez lui qu'elle semble plutôt de l'ordre du tic de langage. En somme, c'est le raisonnement logique qui ressort de l'organisation des textes scientifiques, par contraste avec ceux de A.-F. Villemain ou J. Michelet. Les motifs discriminants partagés par Ch. Darwin/Cl. Royer et J.-B. Lamarck se rapportent qui plus est aux propositions relatives : le motif 3087 (<PUN:";"><PRO:DEM><PRO:REL>), par exemple, formant les « , ce qui » ; ou encore les motifs 825 (<DET:ART><NOM><PRO:REL><PRO:PER>) – « l'héritage qu'on en tire », « les pensées qui se présentent », etc. – et 654 (<PRO:REL><PRO:PER><PRO:PER>) – « qu'on les pose », « qu'on les discute », « qu'ils nous tirent », etc. Or ces motifs semblent davantage correspondre à des idiolectes. Très présents chez F. Brunetière et D. Nisard, ils le sont beaucoup moins chez G. Lanson et H. Taine, qui préfèrent recourir avec J. Michelet à l'énumération (ce dont témoignent nombre de motifs de ce type : <NOM><ADJ><PUN:";"><ADJ><PUN:";">). En ce sens, la proximité que nous observons entre les critiques F. Brunetière et D. Nisard et les savants Ch. Darwin/Cl. Royer et J.-B. Lamarck tiendrait davantage à des concordances d'écriture qu'à un genre discursif. Donnons pour preuve le motif 3095 (<NOM><PUN:";"><KON:"que">), servant la construction de propositions relatives ; alors qu'il pourrait sembler *a priori* commun dans la langue française, sa représentation dans notre corpus est éloquente, avec 174 occurrences chez J.-B. Lamarck, 133 chez Ch. Darwin/Cl. Royer, 79 chez F. Brunetière, 65 chez D. Nisard, 54 chez A.-F. Villemain, et moins de 1 % d'occurrences chez G. Lanson, H. Taine et J. Michelet. Ces derniers privilégient à l'inverse des formules cumulatives construites par ponctuations ; chez H. Taine, par exemple :

De tous côtés, dans les Mémoires, on aperçoit en raccourci quelqu'une de ces vies seigneuriales. Telle est celle du duc de Gesvres, premier gentilhomme de la chambre, gouverneur de Paris et de l'Île-de-France, ayant en outre les gouvernements particuliers de Laon, de Soissons, de Noyon, de Crépy en Valois, la capitainerie de Mousseaux et vingt mille livres de pension, véritable homme de cour, sorte d'exemplaire en haut relief des gens de sa classe [...] Sa mémoire est étonnante pour les parentés et les généalogies ; il possède à fond la science précieuse de l'étiquette ; à ces deux titres, il est un oracle et très consulté. (Taine 1901, 177-178)

Cette abondance de ponctuation, d'incises, cette préférence de l'asyn-dète sur les conjonctions, marque cette écriture énumérative que F. Brunetière et D. Nisard s'emploient justement à rejeter.

Se dessine en fait, dans notre corpus, une ligne de partage entre une langue d'époque (G. Lanson, H. Taine, J. Michelet) et un héritage classique (D. Nisard, F. Brunetière). C'est en effet cette langue classique que F. Brunetière loue dans son étude sur Blaise Pascal :

Au fond, c'est tout simplement méconnaître ici l'une des beautés de la prose française du XVII^e siècle, je veux dire cet agencement savant, ou, pour donner l'idée de quelque chose de plus vivant, cette savante articulation des parties qui se tiennent si bien toutes ensemble, par le seul jeu des conjonctions, que le secours de la virgule et du point et virgule en devient presque superflu. Je me chargerais de ponctuer Bossuet rien qu'avec des points. (Brunetière 1916, 72)

Apparaît ici formulé ce que nous venons précisément d'observer dans les résultats de notre extraction automatique. F. Brunetière et D. Nisard recourent bien davantage aux constructions par propositions relatives sur le modèle du XVII^e siècle. L'un comme l'autre traduisent dans leur syntaxe cette « nostalgie du simple » (Bury 2004) partagée par une partie des auteurs du XIX^e siècle, qui reconnaissent dans N. Boileau, dans B. Pascal, dans J.-B. Bossuet ou dans R. Descartes un idéal conjuguant clarté et mesure – une langue d'autorité. D. Nisard hisse lui aussi B. Pascal parmi les maîtres de la pensée et de la langue françaises ; et il voit plus encore en R. Descartes la perfection du génie :

L'usage d'une langue étant de rendre universelle la communication des idées, et les hommes ne communiquant point entre eux par leurs différences, mais par leurs ressemblances, dont la principale est la raison, une langue est arrivée à sa perfection quand elle est conforme à ce que nous avons de commun, la raison. Telle est la langue de Descartes. (Nisard 1854, 103)

Cette langue représente pour D. Nisard la quintessence du XVII^e siècle, à savoir une écriture mise au service de la raison et qui, de fait, incarne « la recherche et l'expression de la vérité » (Nisard 1844, 863). Une corrélation se voit ainsi établie entre la simplicité de la langue et la puissance rationnelle, si ce n'est philosophique, du propos. Or, c'est ce même souci de clarté et de mesure que l'on peut déduire des textes scientifiques de notre corpus. Elle permet d'expliquer certaines concordances de syntaxe entre J.-B. Lamarck, Ch. Darwin/Cl. Royer, F. Brunetière et D. Nisard, notamment dans les constructions rigoureuses par conjonctions et propositions subordonnées.

Conclusion

Si l'on peut conclure à l'efficacité d'une telle méthode heuristique pour une première exploration transdiscursive et comparatiste de textes, cette dernière pourrait gagner en efficacité grâce à un nettoyage particulier des textes, comprenant une suppression systématique du discours rapporté, des listes, des tableaux, ou encore des locutions étrangères. Cette préparation poussée du corpus permettrait de réduire le bruit et les cas aberrants relevés dans nos résultats.

Dans le cadre de notre travail, l'analyse de correspondance de motifs, menée sur nos trois corpus et proposée comme enquête préliminaire, permet d'ores et déjà de tirer plusieurs conclusions, à commencer par

l'impossibilité d'opter pour une approche dualiste entre discours littéraire et discours scientifique dans la seconde moitié du XIX^e siècle. Si l'on observe bien sûr des corrélations entre l'agencement syntaxique du discours et le domaine d'étude, il est tout aussi clair que les textes, quelle que soit leur nature, se rencontrent par la mise en place de stratégies discursives communes. On a bien, pour reprendre la terminologie de G. Genette, le recours à un « même code » selon le genre du texte et l'intention de l'auteur, notamment lorsqu'il s'agit d'enseigner et de convaincre.

38

Les motifs discriminants, qui ont pu être observés dans le détail, donnent des indices plus précis sur les connivences entre un auteur et un autre. Dans le champ de la critique littéraire, ils permettent de réorienter l'histoire littéraire traditionnelle qui tend à dessiner, à cette époque, une ligne nette entre une critique d'influence positiviste d'un côté, une critique classique ou impressionniste de l'autre. Or cette répartition binaire des critiques par rapport à un modèle scientifique ne se retrouve pas aussi franchement dans nos analyses. Nous arrivons plutôt au constat d'une interpénétration des types de discours, et de similitudes d'écritures transdomaniales, selon les orientations rhétoriques adoptées par les auteurs. Ainsi notre étude incite-t-elle à une réflexion plus approfondie sur la rhétorique épistémique mise en œuvre conjointement, dans cette fin du XIX^e siècle, par les sciences de la vie, les sciences humaines et la critique littéraire.

Bibliographie

Barthes, Roland. 2009 [1972]. *Le Degré zéro de l'écriture, suivi de nouveaux essais critiques*. Paris : Le Seuil.

Benzécri, Jean-Paul. 1982. *Histoire et préhistoire de l'analyse des données*. Paris : Bordas-Dunod.

Bernard, Claude. 1855. *Leçons de physiologie expérimentale appliquée à la médecine, faites au Collège de France*. Tome I : Cours du semestre d'hiver 1854-1855 ; publié par Henri Lefèvre. Paris : J.-B. Baillière et fils. https://obvil.sorbonne-universite.fr/corpus/critique/bernard_lecons-physio-experimentale.

Bernard, Claude. 1865. *Introduction à l'étude de la médecine expérimentale*. Vol. 1. Paris : J.-B. Baillière et fils. https://obvil.sorbonne-universite.fr/corpus/critique/bernard_introduction-medecine-experimentale.

Boukhaled, Mohamed Amine. 2016. *On Computational Stylistics : Mining Literary Texts for the Extraction of Characterizing Stylistic Patterns*. Thèse de doctorat en informatique, université Pierre-et-Marie-Curie. <https://tel.archives-ouvertes.fr/tel-01535586>.

Boukhaled, Mohamed Amine, Francesca Frontini, et Jean-Gabriel Ganascia. 2015. « Une mesure d'intérêt à base de surreprésentation pour l'extraction des motifs syntaxiques stylistiques ». *Actes de la 22^e conférence sur le traitement automatique des langues naturelles*. <https://hal.sorbonne-universite.fr/hal-01198409/document>.

Bourget, Paul. 1905. *Études et portraits [1889]*, t. I : *Portraits d'écrivains et notes d'esthétique* (éd. définitive). Vol. 1. Paris : Plon-Nourrit. http://obvil.sorbonne-universite.site/corpus/critique/bourget_etudes-01.

Brunetière, Ferdinand. 1898. *Manuel de l'histoire de la littérature française*. Vol. 1. Paris : C. Delagrave. https://obvil.sorbonne-universite.fr/corpus/critique/brunetiere_manuel-histoire.

Brunetière, Ferdinand. 1916. *Études critiques sur l'histoire de la littérature française : 1^{re} série [1880]*. Vol. 1. Paris : Hachette. https://obvil.sorbonne-universite.fr/corpus/critique/brunetiere_etudes-critiques-01.

Bury, Mariane. 2004. *La Nostalgie du simple : essai sur les représentations de la simplicité dans le discours critique au XIX^e siècle*. Paris : Champion.

Charcot, Jean-Martin. 1874. *Leçons cliniques sur les maladies des vieillards et les maladies chroniques*. Paris : A. Delahaye.

Cuvier, Georges. 1805. *Leçons d'anatomie comparée*. Vol. 1. Paris : Crochard.

Darwin, Charles. 1870. *De l'origine des espèces par sélection naturelle, ou Des lois de transformation des êtres organisés* (3^e éd.). Traduit par Clémence Royer. Paris : V. Masson et fils ; Guillaumin et Cie. https://obvil.sorbonne-universite.fr/corpus/critique/darwin_origine-des-especes.

Deschanel, Émile. 1864. *Physiologie des écrivains et des artistes ou Essai de critique naturelle*. Vol. 1. Paris : L. Hachette. https://obvil.sorbonne-universite.fr/corpus/critique/deschanel-emile_physiologie-ecrivains-et-artistes.

Duhem, Pierre. 1906. *La Théorie physique : son objet et sa structure*. Paris : Chevalier et Rivière. http://fr.wikisource.org/wiki/Fichier:Duhem_-_La_Th%C3%A9orie_physique,_1906.djvu.

Faguet, Émile. 1887. *Études littéraires : dix-neuvième siècle*, 34^e éd. Vol. 1. Paris : Société française d'imprimerie et de librairie. https://obvil.sorbonne-universite.fr/corpus/critique/faguet_etudes-litteraires-19e.

Foucault, Michel. 2008 [1^{re} éd. : 1969]. *L'Archéologie du savoir*. Paris : Gallimard.

Fournier-Viger, Philippe, Antonio Gomariz, Ted Gueniche, Azadeh Soltani, Cheng-Wei Wu, et Vincent S. Tseng. 2014. « SPMF : A Java Open-Source Pattern Mining Library ». *The Journal of Machine Learning Research* 15 (1) : 3569-3573. <http://jmlr.org/papers/volume15/fournierviger14a/fournierviger14a.pdf>.

Frontini, Francesca, Mohamed Amine Boukhaled, et Jean-Gabriel Ganascia. 2015. « Linguistic Pattern Extraction and Analysis for Classic French Plays ». *Presentation at the CONSCILA Workshop*. <https://pdfs.semanticscholar.org/219b/58a00a540de3ab8ecd95eb21cedd2ed15134.pdf>.

Genette, Gérard. 1992. *Palimpsestes : la littérature au second degré*. Paris : Le Seuil.

Girard, René, Jean-Michel Oughourlian, et Guy Lefort. 2001 [1^{re} éd. : 1978]. *Des choses cachées depuis la fondation du monde*. Paris : Grasset.

Gourmont, Remy (de). 1902. *Le Problème du style. Questions d'art, de littérature et de grammaire*. Vol. 1. Paris : Mercure de France. https://obvil.sorbonne-universite.fr/corpus/critique/gourmont_probleme-du-style.

Hennequin, Émile. 1888. *La Critique scientifique*. Vol. 1. Paris : Perrin. https://obvil.sorbonne-universite.fr/corpus/critique/hennequin_critique-scientifique.

Hennequin, Émile. 1890. *Quelques écrivains français. Flaubert, Zola, Hugo, Goncourt, Huysmans, etc.* Vol. 1. Paris : Perrin. https://obvil.sorbonne-universite.fr/corpus/critique/hennequin_quelques-ecrivains-francais.

Husson, François, Julie Josse, et Sébastien Lê. 2008. « FactoMineR : An R Package for Multivariate Analysis ». *Journal of Statistical Software* 25 (1).

Lamarck, Jean-Baptiste. 1809. *Philosophie zoologique, ou exposition des considérations relatives à l'histoire naturelle des animaux*. Tome 1. Paris : Dentu.

Lanson, Gustave. 1895. *Histoire de la littérature française* [1894] (3^e éd. rev. et corr.). Paris : Hachette. https://obvil.sorbonne-universite.fr/corpus/critique/lanson_histoire-litterature-francaise.

Lemaître, Jules. 1886. *Les Contemporains. Études et portraits littéraires. Première série*. Vol. 1. Paris : H. Lecène et H. Oudin. http://obvil.sorbonne-universite.site/corpus/critique/le-maitre_contemporains-01.

Longrée, Dominique, et Sylvie Mellet. 2013. « Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours ». *Langages* 189 (1) : 65-79. <https://doi.org/10.3917/lang.189.0065>.

Mahlberg, Michaela. 2014. *Corpus Stylistics and Dickens's Fiction*. New York, Abingdon, Oxfordshire : Routledge.

Michelet, Jules. 1875. *Tableau de la France : géographie physique, politique et morale*. Paris : Lacroix.

Mill, John Stuart. 1866. *Système de la logique déductive et inductive : exposé des principes de la preuve et des méthodes de recherche scientifique*. Traduit par Louis Peisse. Tome I. Paris : Ladrangé. http://classiques.uqac.ca/classiques/Mill_john_stuart/systeme_logique/systeme_logique_tdm.html.

- Moretti, Franco. 2016. « Abstraction, motifs récurrents et forme ». *Revue d'histoire littéraire de la France* 116 (3) : 521-532. <https://doi.org/10.3917/rhlf.163.0521>.
- Nisard, Désiré. 1844. « Descartes et son influence sur la littérature française ». *Revue des deux mondes* 8.
- Nisard, Désiré. 1854. *Histoire de la littérature française*. Tome I (10^e éd.). Paris : Firmin-Didot. http://obvil.sorbonne-universite.site/corpus/critique/nisard_histoire-litterature-01.
- Nisard, Désiré. 1883. *Histoire de la littérature française*. Tome II. Paris : Firmin-Didot.
- Quiniou, Solen, Peggy Cellier, Thierry Charnois, et Dominique Legallois. 2012. « What about Sequential Data Mining Techniques to Identify Linguistic Patterns for Stylistics ? ». Dans *Computational Linguistics and Intelligent Text Processing*, édité par Alexander Gelbukh, 166-177. Berlin, Heidelberg : Springer.
- Renan, Ernest. 1890. *L'Avenir de la science : pensées de 1848*. Paris : C. Lévy. https://obvil.sorbonne-universite.fr/corpus/critique/renan_avenir-science.
- Renard, Georges. 1900. *La Méthode scientifique de l'histoire littéraire*. Vol. 1. Paris : F. Alcan. https://obvil.sorbonne-universite.fr/corpus/critique/renard_methode-scientifique.
- Riguet, Marine. 2018. *La Littérature laboratoire (1850-1914) : quand la critique littéraire défie la science*. Thèse de doctorat en littérature française, Sorbonne-Université.
- Riguet, Marine. 2019. « La critique laboratoire : fabrique du XIX^e siècle ou fantasme éternel ? ». *Romantisme* 183 : 56-65. <https://www.cairn.info/revue-romantisme-2019-1-page-56.html?contenu=article>.
- Sainte-Beuve, Charles-Augustin. 1862. *Portraits littéraires*, nouvelle édition revue et corrigée. T. I : Boileau, Pierre Corneille, La Fontaine, Racine, Jean-Bapt. Rousseau, Le Brun, Mathurin Régnier [et] André Chénier, George Farcy, Diderot, l'abbé Prévost, M. Andrieux, M. Jouffroy, M. Ampère, Bayle, La Bruyère, Millevoye, Charles Nodier. Vol. 1. Paris : Garnier frères. https://obvil.sorbonne-universite.fr/corpus/critique/sainte-beuve_portraits-litteraires-ed-nouv_01.
- Taine, Hippolyte-Adolphe. 1892. *De l'intelligence*, 6^e éd. Tome I. Paris : Hachette. https://obvil.sorbonne-universite.fr/corpus/critique/taine_intelligence-01.
- Taine, Hippolyte-Adolphe. 1901. *Les Origines de la France contemporaine. L'Ancien Régime*, 24^e éd. Tomes I et II. Paris : Hachette. https://obvil.sorbonne-universite.fr/corpus/critique/taine_france.
- Villemain, Abel-François. 1830. *Cours de littérature français : tableau de la littérature au Moyen Âge en France, en Italie, en Espagne et en Angleterre*. Tome 1. Paris : Pichon et Didier. https://obvil.sorbonne-universite.fr/corpus/critique/villemain_tableau-litterature-moyen-age.

Notes

- 1 Nous empruntons cette expression à René Girard (Girard, Oughourlian, et Lefort 2001). Par là, nous entendons que la modélisation de la science par la critique littéraire ne signifie ni assimilation ni soumission passive des sciences, mais mise en scène d'un affrontement au travers duquel cette dernière revendique son identité (Riguet, 2019).
- 2 Jusqu'à présent, il ne nous semble pas que la distinction entre les deux termes ait été établie nettement. Nous gardons tout du moins l'usage qu'en fait Franco Moretti dans son récent article (Moretti 2016).
- 3 Nous rejoignons ici le constat établi par Franco Moretti dans son article « Abstraction, motifs récurrents et forme » (2016) : « [...] la beauté de Dieu est absente des motifs. L'harmonie également. Elles ne peuvent pas y être présentes, car les motifs récurrents se situent à la croisée de deux domaines trop hétérogènes pour être réellement "harmonisés" : le monde souvent chaotique des données empiriques, et les structures conceptuelles bien ordonnées de la théorie littéraire. »
- 4 Parmi les étiquettes syntaxiques proposées par TreeTagger pour la langue française (<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/french-tagset.html>), nous avons éliminé celles dont le détail nous semblait, pour notre étude, superfétatoire (tel que le temps des verbes), et qui auraient risqué de parasiter les résultats en démultipliant la variété de motifs morphosyntaxiques.
- 5 Nous avons eu recours à l'*open source* SPMF (Fournier-Viger *et al.* 2014).

6 Ce seuil de fréquence de 1 % nous paraît essentiel, non seulement pour éviter de nous noyer dans la masse de motifs extraits, mais également pour éliminer les risques de confusion entre la syntaxe du critique et celle d'un auteur qu'il citerait.

7 www.obvil.paris-sorbonne.fr/corpus/critique.

8 Nous renvoyons à la thèse de Mohamed Amine Boukhaled précédemment citée, à partir de laquelle a été réalisée notre étude.

9 De façon à rendre l'analyse possible, nous conservons uniquement les 100 premiers motifs qui ont la plus grande force de contribution pour chaque texte.

10 L'emploi d'une traduction pour ce type d'expérience pose bien sûr question, plus encore pour *De l'origine des espèces* de Ch. Darwin que pour le *Système de logique* de J. Stuart Mill, dans la mesure où l'on connaît les libertés de plume prises par Cl. Royer. Pour autant, puisque nous ne nous attachons pas à déterminer le style particulier d'un auteur mais plutôt les éventuels traits d'un genre ou les influences entre domaines, la traduction d'un ouvrage aussi majeur nous paraît tout aussi intéressante. Nous gardons néanmoins à l'esprit, chaque fois qu'un motif morphosyntaxique s'applique à *De l'origine des espèces*, qu'il se rapporte moins au propos de Ch. Darwin lui-même qu'à l'écriture de Cl. Royer ; aussi renvoyons-nous dans ce paragraphe au couple auctorial Ch. Darwin/Cl. Royer pour éviter toute confusion.

Auteurs

Marine Riguet

Labex Obvil, Sorbonne-Université, Paris, France

Marine Riguet, post-doctorante au Labex Obvil (Observatoire de la vie littéraire), a soutenu en 2018 à Sorbonne-Université une thèse au croisement de la littérature, de l'épistémologie et des humanités numériques, intitulée *La Littérature laboratoire (1850-1914) : quand la critique littéraire défie la science*.

marineriguet@gmail.com

Mohamed Amine Boukhaled

UMR 8094 Lattice, CNRS, Montrouge, France

Mohamed Amine Boukhaled, post-doctorant au laboratoire Lattice (Langues, textes, traitements informatiques, cognition), a soutenu en 2016 à Sorbonne-Université une thèse d'informatique et de traitement automatique des langues, intitulée *On Computational Stylistics : mining Literary Texts for the Extraction of Characterizing Stylistic Patterns*.

mohamed.boukhaled@gmail.com

Droits d'auteur

Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la [Licence Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).