
MotiveR : un programme pour la stylistique

Dominique Legallois^{*†1} and Antoine Silvestre De Sacy^{*‡}

¹sorbonne Nouvelle – CNRS – France

Résumé

L'objectif de la communication est de présenter au public un outil informatique écrit sous la forme d'un script R : MotiveR[1]. MotiveR est un programme permettant d'identifier par méthode non supervisée des patrons syntaxiques (les motifs) sur-représentés dans un genre ou chez un auteur, par comparaison avec d'autres genres ou avec d'autres auteurs. Ces motifs sont des unités plus abstraites (ou schématiques) que des ngrammes de tokens, tout en étant plus spécifiques que des ngrammes de Part of Speech. L'équilibre entre schématicité et spécificité des unités est ainsi recherchée.

Certains motifs d'un texte peuvent être statistiquement sur-représentés (par rapport à d'autres textes) ; ils peuvent alors être considérés comme des unités stylistiques caractéristiques.

Nous donnons ici quelques exemples issus d'une analyse dans laquelle nous avons comparé cinq auteurs : Balzac, Dumas, Hugo, Sand et Zola. Le corpus était composé de 10 romans de chacun de ces auteurs (huit pour Hugo). Chez G. Sand, on peut relever, parmi des dizaines d'autres, les deux motifs sur-employés suivants :

1- ADJ et ADJ comme DETPOSS NC

Jeannie était mince et petit comme sa mère, dont il avait toute la retirance. (*François le Champi*)

Elle ne le haïssait point d'être calculateur et positif comme son siècle. (*Indiana*)

2- si ADJ et si ADJ que

Néanmoins Mme Aldini était si gracieuse et si bienveillante, que mon brave homme de père, [...] ne sut que répondre à ses douces paroles et à ses généreuses promesses (*La Dernière Aldini*).

j'avais pour elle un attachement si légitime et si profond, que je ne pensais pas faire un serment téméraire (*La Dernière Aldini*)

Chez Zola :

3- ADJ ainsi que un NC

^{*}Intervenant

[†]Auteur correspondant: dominique.legallois@sorbonne-nouvelle.fr

[‡]Auteur correspondant: antoinedesacy@gmail.com

Un moulin, avec ses ailes, demeura seul, ainsi qu'une épave (*La Terre*)

Par moments, des rues transversales qui dévalaient, des trouées brusques montraient l'immensité de Paris, profonde et large ainsi qu'une mer (*L'œuvre*)

4- ADJ, avec DETPOSS ADJ NC

un instant, il était resté surpris et plein de gêne, devant cette fille déjà savante, avec ses grands yeux candides (*La Joie de vivre*)

La jeune fille qui écoutait, souriante, avec son clair regard si froid et si décidé, eut une brusque affirmation du menton (*L'Argent*).

MotiveR est également programmé pour détecter des motifs sur des corpus en langue anglaise. Ainsi, en comparant les deux traductions anglaises du *Ventre de Paris* (Zola) de Kuransky et de Nelson (publiées toutes les deux en 2009), on peut mettre en évidence grâce à l'outil, ce motif (encore un fois, parmi des dizaines d'autres) spécifique chez Kuransky (c'est-à-dire sur-représentés par rapport à la traduction de Nelson) :

5- , slightly PASTPART by the NOUN

The mat that covered the floor, the soft yellow wallpaper , the imitation oak oilcloth, all gave a coolness to the room, slightly softened by the shine of a brass lamp that hung from the ceiling and sprawled above the table with its large transparent porcelain shade.

The tall brown-haired clerk, with flashing eyes in her calm face , slightly reddened by the cold, sat on a high wooden chair , peacefully writing, apparently undisturbed by the commotion of the hunchback, who seemed to ripple the edges of her skirts.

Outre l'identification et le calcul de motifs (par calcul des spécificités), MotiveR propose également : une représentation en WordCloud, un calcul Tf-idf, une analyse par AFC, une représentation de la densité des motifs dans une œuvre (permettant de voir les endroits dans le texte où un motif est particulièrement employé), un calcul de l'évolution temporelle (ou dans une œuvre) des motifs, un concordancier. Le concordancier permet à partir des calculs statistiques réalisés en amont, d'étudier les figures stylistiques caractéristiques d'un auteur en lecture proche. Chez Louis-Ferdinand Céline, par exemple, sa propension à la dislocation avec rappel par un pronom peut ainsi être modélisée :

6 – le NC, il me VIMP

Les passants, ils me remarquaient (*Mort à crédit*)

La môme, elle me refaisait des gestes (*Mort à crédit*)

La fatma, elle me fait signe de venir (*Mort à crédit*)

La force du programme réside dans sa capacité à fournir à l'utilisateur un ensemble de fonctions exécutables allant des textes bruts, aux analyses statistiques et à des visualisations permettant d'exposer ses résultats, tout en gardant une grande généralité lui permettant d'être utilisé dans des disciplines et problématiques très diverses (stylistique littéraire, attribution d'auteurs, linguistique des genres textuels, etc.). Le format Tidy utilisé favorise l'utilisation des résultats par d'autres outils R.

Le script est facilement utilisable par des utilisateurs non informaticiens. Lors du colloque, nous proposons donc de présenter les différentes fonctions de MotiveR illustrées par plusieurs cas d'analyse.

Références

Legallois, Charnois et Larjavaara, " The balance between quantitative and qualitative literary stylistics : how the method of 'motifs' can help " in Legallois, Charnois et Larjavaara, *The Grammar of Genres and Styles : From Discrete to Non-Discrete Units*, De Gruyter Mouton, 2018, p.168-193

Quiniou, Cellier, Charnois et Legallois, " What About Sequential Data Mining Techniques to Identify Linguistic Patterns for Stylistics ? " in *Lecture Notes in Computer Science*, Springer, 2012.

MotiveR fera par la suite l'objet d'un package R.

Mots-Clés: analyse stylistique, motifs, script R, patrons syntaxiques