

La naissance du style : auteur vs genre aux XVII^e et XIX^e s.

Florian Cafiero¹, Jean-Baptiste Camps², Simon Gabay³ et Marie Puren⁴

¹GEMASS, CNRS / Université Paris-Sorbonne

²Centre Jean-Mabillon, École nationale des chartes, Université Paris Sciences et Lettres

³Université de Neuchâtel

⁴LARHRA, CNRS

Janvier 2020

Dans une procédure d’attribution, le choix des caractères à analyser, comme la méthode de sélection (*culling*), est une étape cruciale [1], afin d’éliminer le bruit pouvant brouiller la signal de l’auteur dans les mesures et garantir la fiabilité des résultats [2]. Parmi les mots-outils, il convient de retirer, par exemple, les pronoms personnels qui introduisent dans les résultats un fort biais de sexe [3] ou de genre [4]. À rebours, en renonçant sciemment à cette phase de sélection, il devient possible de faire ressortir justement ces signaux alternatifs, et ainsi de les comparer avec celui de l’auteur pour mesurer l’importance relative de chacun, voire l’évolution du rapport qu’ils entretiennent dans le temps.

Reprenant le dossier du théâtre français, nous avons décidé de comparer deux époques radicalement différentes : d’une part celle du théâtre classique, avec ses règles strictes scrupuleusement suivies par les dramaturges [5], et d’autre part celle du théâtre du XIX^e s., marquée par l’émancipation de l’auteur romantique d’avec les règles classiques [6]. En comparant les résultats de deux corpus similaires, composés d’une multitude d’auteurs ayant écrit au moins deux pièces qui peuvent être d’un genre différent, il devrait ainsi être possible de déterminer qui, du genre ou de l’auteur, s’impose le plus nettement et ainsi de retrouver quantitativement ce *sacre de l’écrivain* de P. Benichou [7].

Un premier test a été réalisé par la réalisation d’une classification ascendante hiérarchique, utilisant la méthode de Ward [8] et la distance cosinus [9], sur les 100 mots les plus fréquents, sans tri supplémentaire, à l’aide du module `Stylo` pour R [10]. Elle montre une nette inversion entre le XVII^e s. et le XIX^e s.

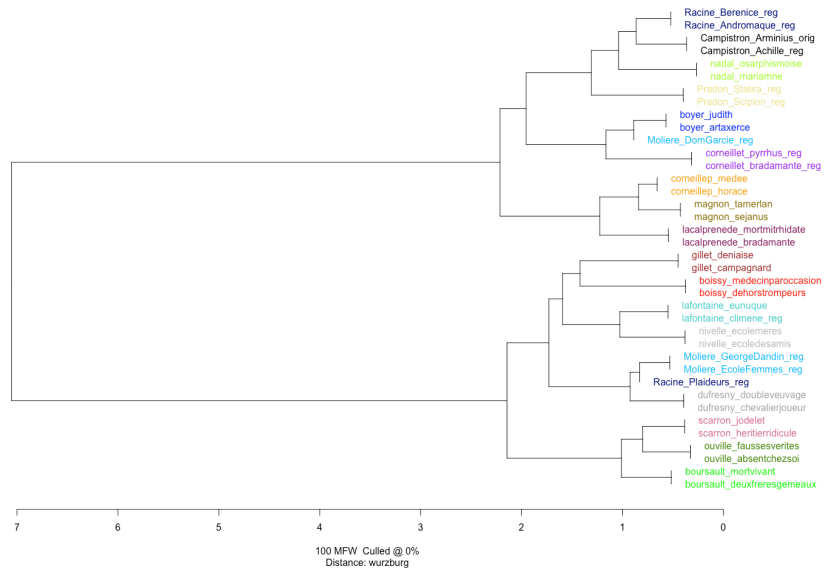


FIGURE 1 – Théâtre français du XVII^e s. : genre vs auteur

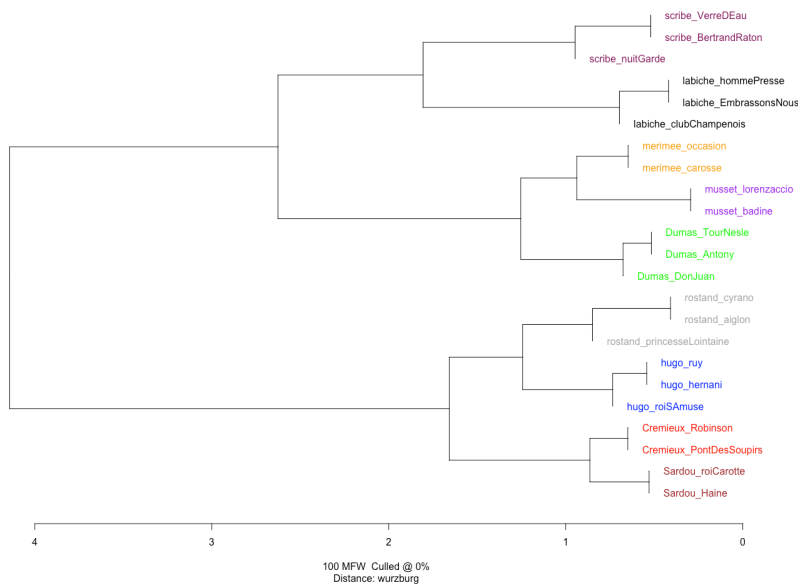


FIGURE 2 – Théâtre français du XIX^e s. : genre vs auteur

Pour la période classique (fig. 1) deux grands groupes organisent la production avec les pièces tragiques (en haut du dendrogramme) et les pièces comiques

(en bas du dendogramme). Le classement par genre prévaut ainsi nettement sur celui par auteur (cf. *Dom Garcie de Navarre* vs les autres pièces de Molière et *Les Plaideurs* vs les autres pièces de Racine). Pour le XIX^e s. (fig. 2), le classement par auteur est de loin le plus important : un opéra-bouffe (*Le Roi Carotte*) et un drame (*La Haine*) de Victorien Sardou ou des Proverbes (*On ne badine pas avec l'amour*) et un drame (*Lorenzaccio*) de Musset se retrouvent ainsi regroupés.

Lors du colloque, nous proposons de présenter les méthodes utilisées pour obtenir ces résultats, ainsi que les étapes nécessaires à la réalisation de cette étude - de la constitution des corpus à leur analyse stylo-métrique. Nous prévoyons également de tester ces hypothèses sur un corpus plus large, de façon à affiner et approfondir ces conclusions préliminaires.

Références

- [1] Mike Kestemont. Function words in authorship attribution. from black magic to theory? In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, pages 59–66, 2014.
- [2] Florian Cafiero and Jean-Baptiste Camps. Why molière most likely did write his plays. *Science Advances*, 5(11) :eaax5489, 2019.
- [3] Matthew L Newman, Carla J Groom, Lori D Handelman, and James W Pennebaker. Gender differences in language use : An analysis of 14,000 text samples. *Discourse Processes*, 45(3) :211–236, 2008.
- [4] Susan C Herring and John C Paolillo. Gender and genre variation in weblogs. *Journal of Sociolinguistics*, 10(4) :439–459, 2006.
- [5] Jacques Scherer. *La Dramaturgie classique en France*. Librairie Nizet, 1973.
- [6] Anne Ubersfeld. *Le Drame romantique*. Belin, 1993.
- [7] Paul Benichou. *Le Sacre de l'Écrivain, 1750-1780. Essai sur l'avènement d'un pouvoir spirituel laïque dans la France moderne*. Joseph Corti, 1973.
- [8] Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301) :236–244, 1963.
- [9] Stefan Evert, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. Understanding and explaining delta measures for authorship attribution. *Digital Scholarship in the Humanities*, 32(suppl_2) :ii4–ii16, 2017.
- [10] Maciej Eder, Jan Rybicki, and Mike Kestemont. Stylometry with R : a package for computational text analysis. *R Journal*, 8(1) :107–121, 2016.