



**HAL**  
open science

# Rise and Fall of Theatrical Genres in Early Modern France

Florian Cafiero, Simon Gabay

► **To cite this version:**

Florian Cafiero, Simon Gabay. Rise and Fall of Theatrical Genres in Early Modern France: A Centroid-Based Approach. DH Benelux, May 2023, Bruxelles, Belgium. hal-04093598

**HAL Id: hal-04093598**

**<https://hal.science/hal-04093598>**

Submitted on 10 May 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.



Distributed under a Creative Commons Attribution 4.0 International License

# Rise and Fall of Theatrical Genres in Early Modern France: a Centroid-Based Approach

**Florian Cafiero**

Sciences Po Paris

florian.cafiero@sciencespo.fr

**Simon Gabay**

Université de Genève

simon.gabay@unige.ch

## Abstract

Literature is not stable : it changes. By taking the example of two genres experiencing different destinies, tragedy and comedy, we demonstrate that this change results from the pressure exerted by two opposing forces : that of respecting the rules of the genre on the one hand, and that of the personal expression of the author of the other. The study of these two forces is done by means of the distance of the plays with the centroids of their respective authors or genres, a new method in stylometry.

The 17<sup>th</sup> c. is the century of the great French tragedy. Corneille and Racine, to name a few, still are among the authors whose notoriety goes beyond the borders of the French-speaking world. Who could have predicted in 1691, when *Athalie* was published, that this major genre of French literature would be, to use Christian Biet’s words, “dead” (Biet, 2010, p. 167) a century later? Indeed, at the end of the 18<sup>th</sup> c., tragedy gave way to bourgeois drama with Diderot and Beaumarchais, which would experience its hours of glory in the 19<sup>th</sup> c. with Hugo (*Hernani*, *Ruy Blas*), Vigny (*Chatterton*) or Musset (*Lorenzaccio*). Conversely, comedy knows another fate : it survives the 18<sup>th</sup> c. and remains, throughout the 19<sup>th</sup> c., one of the favourite genres of playwrights.

Many academics have tried to understand this disappearance of tragedy (Allain (1965); Biet (2010). . .) : the rejection of ancient themes by the romantic movement, the incompatibility of tragic values with those carried by the Enlightenment... In contrast, the persistence of the comic genre has not been explained by research. We believe that distant approaches can explain this continuity by revealing a profound internal mutation, partly driven by a more marked presence of the author to the detriment of the rules of the genre.

## 1 Set up

Studying computationally comedy and tragedy is not trivial. Identifying contents from a specific genre and understanding the rhythm and determinants of these genres’ evolution (Moretti, 2005) are much-debated tasks inside the field of Computational Literary Studies. Recent papers on “gothic novels” (Underwood, 2016) or “detective novels” (Jockers, 2013) demonstrate the complexity of such research.

Depending on context, the detection of the genre of a text relies on different features, be they syntactic (Gianitsos et al., 2019), lexical (Al-Yahya, 2018; Lex et al., 2010), topic-related (Schöch, 2017), or on character’s interaction (Hettinger et al., 2015). However, the French Early Modern theatre raises specific problems : because of its often identical sources of inspiration, or its strict compositional rules (be they implicit or explicit), it is characterised by a very strong stylistic and thematic homogeneity compared to the production of the following centuries – a fact that inevitably shows up in computational analyses (Cafiero and Camps, 2019; Cafiero et al., 2020; Cafiero and Camps, 2021; Gabay, 2021). To maximise the linguistic information we get from our texts, we rely on character 3-grams (Kestemont, 2014; Sapkota et al., 2015). We choose to keep here the 2,000 most frequent ones as the others were not frequent enough to yield significant effects – but our results remain stable with other values<sup>1</sup>. We compute the z-score for all the 3-grams and perform a vector-length Euclidean normalisation (Evert et al., 2017).

To evaluate the heterogeneity within a given set of data compared to a reference situation, we compute centroids for each author and genre. The centroid is defined by the arithmetic mean position of a list of  $n$  points ( $C = \frac{1}{n} \sum_{i=1}^n x_i$ ), i.e. in our case

---

1. Our code is distributed with this paper for controlling purposes.

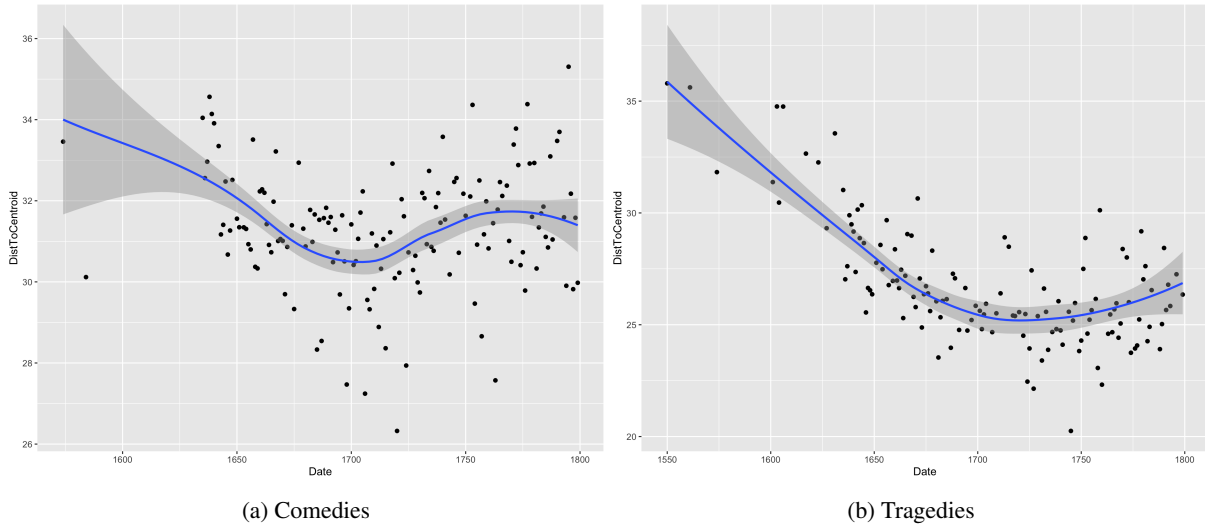


FIGURE 1 – Evolution of the distance from the centroid over time : locally estimated scatterplot smoothing (LOESS). Each point represents a play, the abscissa is that of the date of writing, the ordinate is that of the distance from the centroid.

the normalised frequency of all the 3-grams of an author or a genre.

Our data come from the *Théâtre classique* corpus (Fièvre, 2007; Camps et al., 2021), which transcriptions have all been linguistically normalised (i.e. aligned with contemporary spelling), which simplifies the comparison of plays written at different times, and therefore in different states of the French language. Only the plays belonging to the tragic and comic genres are kept, excluding hybrid genres such as heroic comedy or tragicomedy because they do not offer a clear signal but a mixed one (Schöch, 2017). For the calculation of the centroids of the authors, only the authors who have written more than three plays of the same genre are kept. Our *Terminus post quem* is 1550 and our *terminus ad quem* is 1800 (the number of tragedies decreasing drastically thereafter).

## 2 Experiments

### 2.1 Experiment A : plays vs the centroid of their genre

By comparing all the plays with the centroid of their respective genre and placing the measured distances of the plays on a chronological axis, it is possible to trace the evolution of comedy and tragedy over time. On fig. 1, the lower the point on the y-axis, the shorter the distance to the centroid : the linguistic practice of the play is more in line with the average of the genre. Conversely, the higher a point, the greater the distance from the centroid : the observed data diverge from the average for the

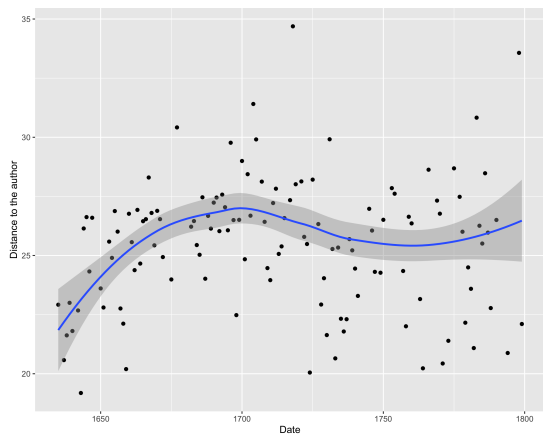
genre. These phenomena can be translated into literary terms : a short distance displays an increased search by the author to conform to the genre, possibly a confinement in it, while a longer distance shows a taking of freedom with the codes.

If we compare the distance of comedies (cf. fig. 1a) and of tragedies (cf. fig. 1b) from their respective centroids over time, we obtain two very different curves. For tragedies, we observe an uninterrupted drop, testifying to ever more similar practices among authors who conform to a common mold. Conversely, for comedies, we observe first an identical trend, and then a clear inflection at the beginning of the 18<sup>th</sup> c., which reflects a diversification in writing practices.

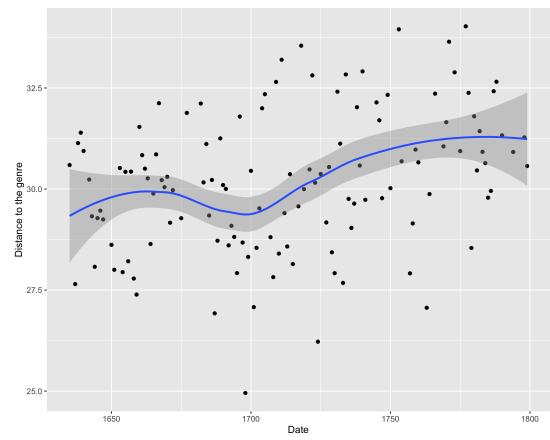
### 2.2 Experiment B : auctorial signal vs generic signal

A possible explanation for this change in writing could be found in the clearer affirmation of the personal style of writers, or at least of certain writers, which would temper the attraction exercised by the genre on their work. In order to test this hypothesis, we need to

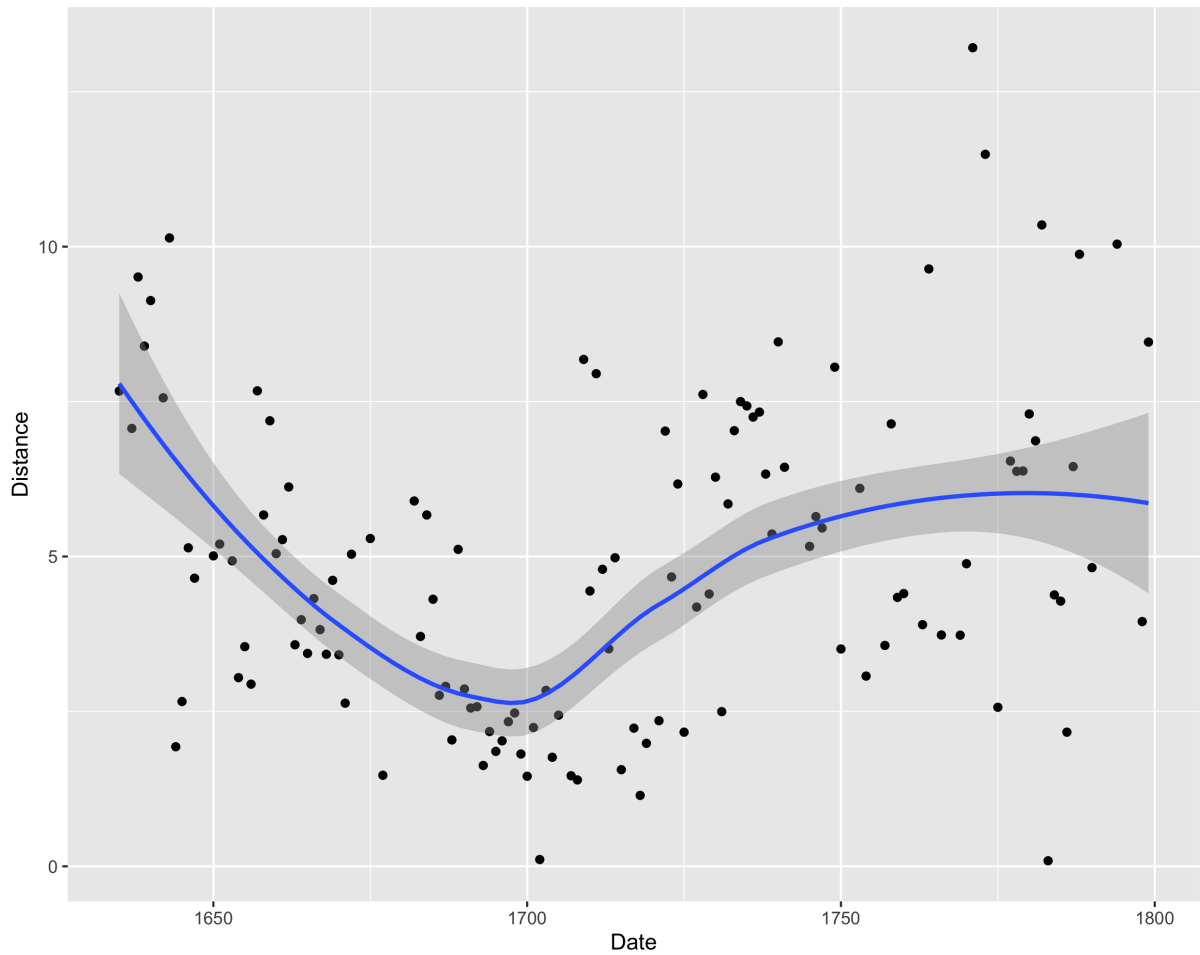
1. dynamically recalculate the centroid of the genre by using a subset of comedies written 30 years before and after the publication of each play. This prevents comedies that are too old or too recent from exerting too great an influence on the position of the centroid ;
2. calculate the distance between each comedy and the centroid of all other comedies written



(a) Distance between a play and the centroid of other comedies written by the same author.



(b) Distance between a play and the centroid of all comedies written 30 years before and after



(c) Euclidean distance between the two previous distances.

FIGURE 2 – Separation of the authorial and generic signal for the evolution of the comic genre.

by the same author, by selecting only authors who have written at least two other comedies.

The comparison of the two curves computed with a local regression confirms that in 1700 the distance with the dynamic centroid of the comedies stops falling to increase again (cf. fig. 2b), but it also shows that at the exact same time the distance with the centroid of the comedies of the same author follows an opposite movement : instead of continuing to increase, it begins to stagnate, even to regress slightly (cf. fig. 2a). The calculation of the euclidean distance between the distances with the centroid of comedies and the centroid of comedies written by the same author clearly highlight this divergence in trend (cf. fig. 2c).

### 3 Discussion

It is possible to distinguish two phases in the history of comedy during early modern France, with a turning point between the 17<sup>th</sup> and 18<sup>th</sup> c. Once the codes of the literary genre have been established, comedy manages, unlike tragedy, to halt its fossilisation : by stopping to use common linguistic features, authors reverse the movement of convergence. This liberation of the practice could be done in favour of more personal practices, as evidenced by the curve formed by the distance of each play with the other comedies by the same playwright, which regresses slightly. In other words, insofar as the author is not bound by external generic rules, he can somehow follow his own.

It is important to note that this movement is not clear : we do not find a massive trend, followed uniformly by all the authors. They act in dispersed order : only a non-parametric regression makes it possible to highlight what appears to be a general trend. This growing dispersion of authorial practices is logical : comedy finds itself torn between an old code, still active, and a multiform renewal which is set up in opposition to the old code.

As shown by Jacques Scherer, the genius of 17th century French theater was, behind seemingly strict rules, to leave writers a margin of freedom (Scherer, 2014). This freedom was swallowed up by tragedy, which died of it, and regained by comedy, which continued to exist.

### References

Maha Al-Yahya. 2018. [Stylometric analysis of classical arabic texts for genre detection](#). *The Electronic Library*, 36(5) :842–855.

Mathé Allain. 1965. [Voltaire et la fin de la tragédie classique française](#). *The French Review*, 39(3) :384–393.

Christian Biet. 2010. *La Tragédie*, 2e edition. Coursus. Armand Colin.

Florian Cafiero and Jean-Baptiste Camps. 2019. [Why Molière most likely did write his plays](#). *Science advances*, 5(11) :eaax5489.

Florian Cafiero and Jean-Baptiste Camps. 2021. [‘psyché’ as a rosetta stone? assessing collaborative authorship in the french 17th century theatre](#). *Proceedings CEUR-WS*, 1613 :0073.

Florian Cafiero, Jean-Baptiste Camps, Simon Gabay, and Marie Puren. 2020. [La naissance du style: auteur vs genre aux xviii et xixe siècles](#). In *Colloque Humanistica 2020 – Archives du colloque*.

Jean-Baptiste Camps, Simon Gabay, Paul Fièvre, Thibault Clérice, and Florian Cafiero. 2021. [Corpus and models for lemmatisation and pos-tagging of classical french theatre](#). *Journal of Data Mining and Digital Humanities*.

Stefan Evert, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. 2017. [Understanding and Explaining Delta Measures for Authorship Attribution](#). *Digital Scholarship in the Humanities*, 32(suppl 2) :ii4–ii16.

Paul Fièvre. 2007. [Théâtre classique](#). Online corpus of French plays.

Simon Gabay. 2021. [Beyond idiolectometry? on racine’s stylometric signature](#). In *Conference on Computational Humanities Research 2021*, pages 359–376.

Efthimios Gianitsos, Thomas Bolt, Pramit Chaudhuri, and Joseph P. Dexter. 2019. [Stylometric classification of ancient greek literary texts by genre](#). In *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 52–60.

Lena Hettinger, Martin Becker, Isabella Reger, Fotis Jannidis, and Andreas Hotho. 2015. [Genre classification on german novels](#). In *26th International Workshop on Database and Expert Systems Applications (DEXA)*, pages 249–253.

Matthew L Jockers. 2013. *Macroanalysis: Digital methods and literary history*. University of Illinois Press.

Mike Kestemont. 2014. [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.

Elisabeth Lex, Andreas Juffinger, and Michael Granitzer. 2010. [A comparison of stylometric and lexical features for web genre classification and emotion classification in blogs](#). In *2010 Workshops on Database and Expert Systems Applications*, pages 10–14. IEEE.

Franco Moretti. 2005. *Graphs, Maps, Trees : Abstract Models for a Literary History*. Verso.

- Upendra Sapkota, Steven Bethard, Manuel Montes, and Thamar Solorio. 2015. **Not all Character N-Grams are Created Equal: A Study in Authorship Attribution.** In *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics : Human language technologies*, pages 93–102.
- Jacques Scherer. 2014. *La Dramaturgie classique*. Hors collection. Armand Colin.
- Christof Schöch. 2017. **Topic modeling genre: An exploration of french classical and enlightenment drama.** *Digital Humanities Quarterly*, 11(2).
- Ted Underwood. 2016. **The life cycles of genres.** *Cultural Analytics*, 2.