

‘Psyché’ as a Rosetta Stone? Assessing Collaborative Authorship in the French 17th Century Theatre

Florian Cafiero¹, Jean-Baptiste Camps²

¹GEMASS / CNRS / Université Paris-Sorbonne, 59-61 rue Pouchet, 75017 Paris, France

²École nationale des chartes / Université PSL, 65 rue de Richelieu, 75002 Paris, France

Abstract

During the 17th century, a significant number of collaborations emerged between playwrights, among which authors as famous as Pierre Corneille, Thomas Corneille or Molière, as well as Philippe Quinault or Jean Donneau de visé. The actual division of labour between authors can sometimes be deduced from historical documents, but is most of the time uncertain. In this paper, we try to address this question by using the information we got from one specific instance of collaboration: *Psyché* (1671). We first try to assess the accuracy of the notice to the reader of the printed edition of the play, where each author’s involvement is clearly claimed, using machine learning and “rolling stylometry” methodology. We then use the optimal parameters already applied to this play to analyse other collaborative works of the time, in particular cases of potential collaboration between Thomas Corneille and Jean Donneau de Visé in *Circé* and *L’Inconnu*.

Keywords

authorship attribution, French literature, 17th century, rolling stylometry, collaborative authorship

1. Introduction

1.1. Collaborative authorship and stylometry: the challenges of the *Théâtre classique*

Stylometry, and notably ‘rolling stylometry’, has been successfully used to identify the co-authors of a literary work in several cases. With Burrows’ delta, it has for instance been used to assess Ford’s claims about his implications in collaborations with Joseph Conrad [25], to determine the beginning of Vostaert’s intervention on Dutch Arthurian novel Roman van Walewein [5], or to understand Lovecraft’s and Eddy’s implication in *The Loved Dead* [13]. A distance-based approach has also advanced our understanding of the collaboration between Julius Caesar and General Hirtius, and confirmed that pseudo-Caesar texts had been written by an anonymous writer [16]. Principal Components Analysis was used to visualise the importance of Hildegard of Bingen’s last secretary Guibert-Martin de Gembloux in her late production [15]. Using support-vector machines, rolling stylometry more recently helped to confirm John Fletcher and William Shakespeare’s collaboration for *Henry VIII* [22]. Co-authorship between Nobel Prize winner Yasunari Kawabata and one alleged ghostwriter was detected using in parallel various supervised machine-learning settings [29].


CHR 2021: Computational Humanities Research Conference, November 17–19, 2021, Amsterdam, The Netherlands

✉ florian.cafiero@cnrs.fr (F. Cafiero); Jean-Baptiste.Camps@chartes.psl.eu (J. Camps)

🆔 0000-0002-1951-6942 (F. Cafiero); 0000-0003-0385-7037 (J. Camps)

© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

Yet, the task raises a number of difficulties. First, style change detection is not a completely solved problem, and even recent competitions [30] have shown that precisely determining *style breaches*, i.e. places where the authorship switches in a collaborative text, was still a complex task.

The task is all the more complicated for French 17th century plays, known as “*théâtre classique*”. The 17th century is a time when the notion of authorship settles in France [31]. Yet, as stated by Quiel, the need for originality or the fear of being too derivative were not yet a major worry for the authors of the time:

French playwrights do not refrain from imitating or adapting whichever text is likely to be presented on a stage, sometimes without even bothering to make significant additions or to add characteristics from their own literary style [23].¹

Strongly codified, building their plots on the same Spanish, Italian, Greek or Latin models, these plays are often very homogeneous, which makes it more difficult to properly attribute texts. In particular, similarities induced by the literary genre or subgenre can be as strong as similarities induced by the authors’ idiolect [27, 3, 4].

Imitation could even go to the point of piracy or plagiarism, such as attempting to steal another author play, or led to disputes about the true source of a story. The polygraph and literary “entrepreneur” Donneau de Visé [28] – famous later on as the powerful founder and editor of the monthly *Mercure Galant*, a collective literary periodical that published a mix of recycled manuscript or printed pieces, reader contributions (poems, etc.) and his own original material – was familiar of such practices at least in his early career. He more or less started his career by trying to steal *Sganarelle ou le cocu imaginaire* from Molière: before Molière could publish himself his own play, Donneau published in 1660, with the help of the printer Jean Ribou, both a pirate edition of Molière’s *Sganarelle*, in which he added his commentaries and that he went to the point of dedicating to Molière himself (!), and a plagiarised play, *La Cocue imaginaire*, where he reversed masculine and feminine roles [8]. He also had an important dispute with Quinault around the *Mère Coquette*, two plays with this name being published, in 1665 by Quinault and 1666 by Donneau.

A final challenge is related to the diverse potential nature of collaborative writing. According to Pennebaker and Ireland [21], three main hypotheses can be made on the result of collaborative writing: the “Just-like-another-member-of-the-team hypothesis” where collaborative writing can be distributed in portions successively attributable to the idiolect of one of the authors; “The average person hypothesis” where the resulting style is an average of the authors’ idiolects; and finally the “synergy hypothesis” where the contact situation and interactions between different individual idiolects create a resulting singular style, different of each individual one. This last hypothesis tends to be verified in famous cases such as the Lennon and McCartney collaboration or the one between Hamilton and Madison.

As we will see, suspected collaborative writing cases in 17th century French theatre – even though we might suspect them on declarative grounds to fall mainly in the first “Just-like-another-member-of-the-team hypothesis” – also presents clue of a division of the work on different authorial levels, for instance content versus form, narrative versus versification.

To help us address the various collaborative authorship problems raised by the writings of this century, we thus try to work on one of the best documented collaboration of the time: *Psyché*.

¹All translations are our own.

1.2. 'Psyché' and its notice 'to the reader'

Psyché is a tragedy-ballet in five acts, written in free verse, and created in 1671, during the very long festivities following the peace of Aix-la-Chapelle in 1668. It originates in Louis XIV's desire to give a new show in the "Salle des Machines" of the Tuileries Palace. Built in 1660 by renowned architect Louis Le Vau, this theatre took its name from the machinery designed by Gaspare, Carlo and Lodovico Vigarani, allowing for spectacular effects and complex set changes [6]. The acoustics of this theatre were poor. It was thus abandoned, not being used since 1662. But its large capacity, and the existing impressive sets from Cavalli's opera *Ercole Amante*, would have drawn the French king to commission a new play specifically designed for this place.

In 1758, Lagrange-Chancel reported in the preface to his own *Orphée* that several authors would have proposed a project for this occasion.

The late King having resolved to give to all his court one of these great celebrations in which he liked to have a rest from his works, wanted to take advice from Racine, Quinault, and Molière, which, among the the great geniuses of this century, he regarded as the most capable of contributing, by their talent, to the magnificence of his pleasures. To that effect, he asked them to pick a subject for which they could use an excellent decor representing the underworld, kept safe and sound in the furniture storage unit. Racine proposed the subject of Orphée; Quinault, the abduction of Proserpine, which he subsequently turned into one of his most beautiful operas; and Molière, with the help of the great Corneille, championed the subject of Psyché, which prevailed over the two others [17].

Through the correspondence of the Vigarani, we know that the decision to give this *Psyché* at the *Salle des Machines* was made only a few weeks before it was played. In a letter written on December 12th, 1670, Vigarani explained that they were "preparing a great show, to be performed for the Epiphany at the Tuileries theatre" [24], The imminence of the deadline forced everyone to rush to get the work done. As stated in a letter written on December 15th, 1670, "Carlo is very busy because of the show prepared for the Epiphany. He is very tired. He is doing his best to please the King, but he doubts he will be strong enough to continue."

The lack of time apparently had consequences on Molière's ability to finish the play. This led to a singularity: an official account of each author's implication. In a notice from the publisher "*au lecteur*" (to the reader), we find this explanation on how the work is supposed to have been divided:

This work is not written by a single hand. Mr Quinault wrote all the poetry of the parts set to music, except the Italian Complaint. Mr de Molière wrote the outline of the play, set its arrangement - he focused more on the beauties and the pump of the show than on its strict observance of the rules. Regarding versification, he did not get the time to execute it in its entirety. Carnival was approaching, and the insisting demands of the King, who wanted to entertain himself several time before Lent, forced him into accepting some assistance. Thus, only verses from the Prologue, the First Act, the first scene of the Second Act, and the first scene from the third Act are his work. Mr. Corneille used two weeks to versify the rest; and this way, His Majesty's orders were satisfied in time [20].

Does this notice to the reader seem plausible? The situation where the King's urgent demands change the author's agenda was not unprecedented. For instance, when in 1664, Louis XIV commissioned a new play with ballet to Molière for the "*Plaisirs de l'île enchantée*" festivity in Versailles, the latter could not finish in time the versification of his play. As stated in its first edition [19], "an order from the King, who pressed this matter, forced [the author] to finish all the rest in prose." Only thirty percent of *La Princesse d'Elide* is thus in verse, the rest being in prose.

Stylistic studies also seem to confirm the plausibility of this notice. *Psyché* is one of the rare examples of mixed verses by their authors. Before that, Pierre Corneille had only used it once, in *Agésilas* in 1666. The same goes for Molière, who also mixed various types of verses in his *Amphitryon* in 1668. And the way the two authors use mixed verses is different [2]. While Molière did not hesitate to use heptasyllables in *Amphitryon*, Corneille never used a single one of them in *Agésilas*. This distinction is still observable in *Psyché*: in the part attributed to Molière, we find 37 heptasyllables - which fits the proportion observed in *Amphitryon*; in the part attributed to Corneille, we do not find any heptasyllable.

Without taking this notice "to the reader" for granted, we can consider that it gives potentially truthful information regarding the play, that we will first try to disprove or verify.

1.3. Hollywood in the 17th century: Special effects, music and the "*Pièces à machines*"

Psyché raises a few specific concerns because of its genre. This play is indeed a rare instance of *pièce à machines*, a subgenre very much in favour between the 1650s and the 1670s, but of which only 15 plays or so have been composed [33, 32]. Changes of scenery in each act, flying characters, raging seas, thunderous blows... the "*pièces à machines*" make the most of the machinery available at the time, to propose spectacular shows to the audience, including passages set to music. The first model of the genre probably is *Andromède* (1650) by Pierre Corneille [1].

But other authors followed him, sometimes collaborating to produce that kind of plays, like Thomas Corneille (Pierre Corneille's younger brother) and Jean Donneau de Visé. Amongst the most impressive shows of the time, the first collaboration between the two authors, *Circé* (1675), encouraged them to work together for other works. This play was thus quickly followed by another collaboration: *L'Inconnu* (1675). It seems that Donneau de Visé's implication in this play could be very significant. It draws its inspiration from Donneau de Visé's own tenth short story [18] in *Les Nouvelles Galantes, Comiques et Tragiques* (1669). In Thomas Corneille's obituary notice, Donneau de Visé himself claimed he played an important part in the writing:

to make progress, I wrote the whole play in prose, and while I was writing the prose of the second act, he was transforming the prose of the first into verse; and as prose is easier than verse, I had the time to write those of the entertainments, and especially the dialogue of Love and Friendship, which did not displease the public [7].

Yet, despite these claims, the only name cited in the "*Privilège du roi*" of the first printed edition is Thomas Corneille. Why omitting to mention Donneau de Visé? We know from the *Registre de La Grange* that Donneau de Visé was paid fees for the writing of the play, and he received the same amount for it as Thomas Corneille. It is thus historically extremely

unlikely that Donneau de Visé would not have contributed to the play. But did he overstate his implication in the writing of the play? Or were there only “commercial reasons” to avoid mentioning his name - Thomas Corneille being more respected as a playwright than he was?

Thus, using the parameters found optimal in our benchmark phase, and then tested on *Psyché*, we will try to work on *Circé* and *L’Inconnu*, two *pièces à machines* also written in verse and allegedly written collaboratively.

The scarcity of “*pièces à machines*” however makes it challenging to build a stylometric approach only on a subgenre-specific approach. We thus compare two methods in this paper: a genre-specific approach, on a small dataset, and a cross-genre approach, on a considerably larger dataset (see appendix A).

2. Materials and methods

2.1. Choice of plays and dataset

To investigate these difficulties, we built two different analytic setups, with two very different philosophies:

a **genre-specific approach**, in which we built a training corpus including one play from the same subgenre for each of the three involved authors, as well as two control authors – one play by Boyer, and two plays for Quinault, as his plays were significantly shorter (see appendix A.1.1). We then benchmarked different sample lengths, using a leave-one-out approach. We lacked sufficient data to include Thomas Corneille in this setup, because the two available plays that could fit the definition, *Circé* and *L’Inconnu* are suspected to be collaborations with Donneau de Visé that we want to analyse later on.

a **cross-genre approach** in which we took all available single-author verse or mixed plays containing more than 400 verses from each candidate in the *Théâtre classique* corpus [12]. In order to recreate the conditions of the actual analysis on the *Pièces à machines*, that is to evaluate the performance in a cross-genre setup on a given specific genre or sub-genre not represented in the training set, we set apart a subgroup of heroic comedies as an unseen test set on which to benchmark the models (see appendix A.2.1 and A.2.2). For the mixed plays, we retained them only if there were more than 400 verses to be extracted.

For each case studied, the set of candidate authors is known. We thus use an authorship attribution rather than an authorship verification setup.

In terms of features, after suppressing editorial punctuation and lowercasing the texts, we extract character 3-grams, a standard choice in authorship attribution [14, 26].

2.2. Calibration

The chosen size of the sample can be seen as a trade-off between accuracy and granularity: the smaller the samples are, the better the “resolution” of the analysis and the ability to locate precise stylistic breaks or identify limited shifts in hands [10]; the bigger they are, the more statistically reliable is our computation, with optima often observed in the 2500-5000 words range [9]. To find a good compromise, we experimented with a variety of lengths, from 10 to 300 verses, an upper limit close to the size of an act (Table 1). Each time, we normalised the

data by using z-scores for variables and applied Euclidean vector-length normalisation (i.e., L2 normalisation) to texts [11] and trained a linear Support Vector Classifier (SVC) model, using a Python `sklearn` pipeline (see Code section). These results show that a first peak in all metrics is reached at length 150 verses for the smaller genre-specific corpus, with perfect scores, as well as for the larger cross-genre corpus when considering the F1 score (with $F1 = 0.92$).

Table 1

Scores resulting from the SVM training benchmark on samples of length from 10 to 500 verses; precision, recall and F1 score are given, as well as the support (number of test samples). The small same-genre models are tested using a leave-one-out approach, then averaged; the larger cross-genre models are tested on out-of-domain plays.

Sample length (verses)	genre-specific				cross-genre			
	Prec.	Rec.	F1	support	Prec.	Rec.	F1	support
10	0.85	0.85	0.85	1069	0.57	0.47	0.46	1120
20	0.92	0.92	0.92	533	0.68	0.59	0.59	558
30	0.95	0.94	0.94	354	0.75	0.67	0.67	371
40	0.96	0.96	0.96	264	0.76	0.71	0.70	277
50	0.98	0.98	0.98	212	0.81	0.75	0.75	222
100	0.99	0.99	0.99	103	0.87	0.81	0.82	110
150	1.00	1.00	1.00	68	0.93	0.91	0.92	72
200	1.00	1.00	1.00	51	0.94	0.92	0.92	53
250	1.00	1.00	1.00	41	0.93	0.90	0.91	42
300	0.97	0.97	0.97	32	0.96	0.94	0.94	35

Detailed scores for the retained sample length on the cross-genre corpus show that Boyer is the best recognised, while precision for Donneau de Visé and Molière is better (the model is never wrong in attributing plays to them) than recall (the model misses a few samples), while the opposite is true for both Corneille brothers (Table 2).

Once the sample size of 150 verses chosen, we train a final SVM model for each setup, on the complete training set, and then, following rolling stylometry methods, we apply this model to every successive portion of length n , with a step of 1 (and so, an overlap of $n - 1$ between two successive portions, e.g., verses 1-150, 2-151, 3-152...). We then extract the classification, and plot the decision function for each classifier.

Like with many geometrical methods, this type of analysis rests on the representation of texts or samples as points in a high dimensional space, based on the frequency of the selected features. In our case, the frequency of each type of character 3-grams are used as a coordinate on one axis of an high-dimensional space.

A support vector machine computes a hyperplane in this high-dimensional space, in order to achieve the best separation between two sets of dots (i.e., text from author 1, text from all other authors). Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class (called functional margin), since in general the larger the margin the lower the generalisation error of the classifier.

The decision function tells us how close each sample is to the hyperplane separating each class. A negative value means that the sample is outside, a positive, inside. The higher the score, the deeper inside the class is located a dot, which can be interpreted as a strength of the authorial markers or an increase in the confidence of the classifier.

By monitoring the decision function of all candidate authors, we can see when a portion of the text is getting closer to the style of an individual author. When the value of the decision function for one author gets high, while the values for all others remain low, it is easy to

Table 2

Detailed class scores on out-of-domain test for the SVM trained on the cross-genre setup, with sample length 150 verses.

	Prec.	Rec.	F1	support
BOYER	1.00	1.00	1.00	12
CORNEILLEP	0.86	1.00	0.92	12
CORNEILLET	0.82	1.00	0.90	14
DONNEAUDEVISE	1.00	0.91	0.95	11
MOLIERE	1.00	0.75	0.86	12
QUINAULT	0.90	0.82	0.86	11
macro avg	0.93	0.91	0.92	72
weighted avg	0.93	0.92	0.92	72

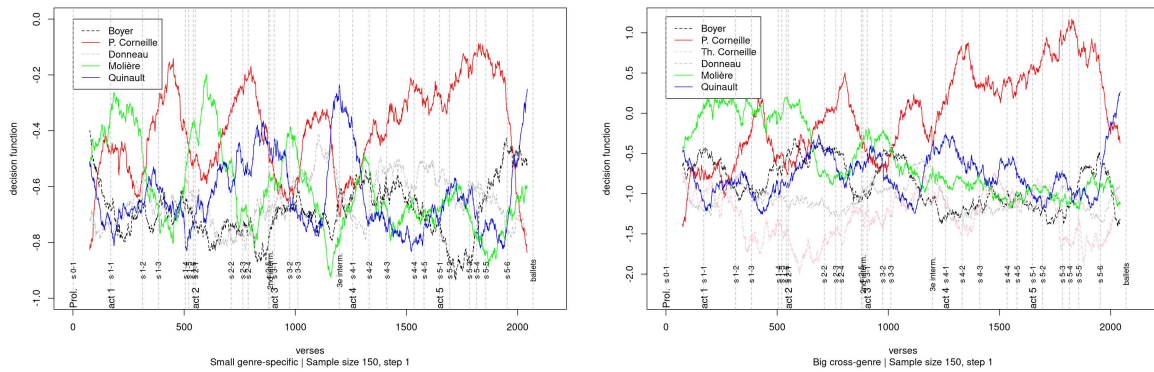


Figure 1: Value of the decision function for each classifier and each successive rolling sample; each sample is placed horizontally at it's median point in verses and vertical grey dashed lines denote scenes. Both setups achieve largely consistent results.

attribute this portion to this candidate. Yet, in the case where all decision function would decrease or remain low simultaneously, the status of the portion is hard to assess, and could be alternatively attributed to a synergy between several candidates, to the intervention of an author outside the set or could be the results of some kind of noise, in particular generic discrepancies between the training set and the portion being assessed.

Each sample's horizontal placement on the graph is determined by its **median** point, in verses (fig. 1). For instance, the score attributed to the sample ranging from the 400th to the 550th verse will be placed at the 475th verse. This implies a small but significant distortion regarding the placement of each score in the timeline of each play. It must be taken into account when reading the graph.

3. Results

3.1. *Psyché* by P. Corneille, Molière and Quinault

Both setups provide globally consistent results. The end of the prologue, most of the first act, and the beginnings of the second and the third acts are attributed to Molière. Molière's imprint then seems to gradually fade away, while Pierre Corneille's contribution appears constantly growing throughout the play. This result seems solid, and even conservative regarding Molière's

share of the work. When using the cross-genre setting, the precision for Molière is 100 % while the recall is lower, at 0.75 % (Table 2). On the opposite, the recall for Pierre Corneille is perfect (100 %), while the precision is lower (86 %). This setting should thus underestimate Molière’s participation in the play. Yet, even this computation confirms that the statements made in the notice from the publisher to the reader regarding him are globally accurate. A small but interesting difference between our analysis and the notice concerns the spike of Corneille’s decision function at the end of the first act. According to it, it could be possible that Corneille had a hand in finishing the first act, under a fashion comparable to that of the other acts (begun by Molière, ended by him) but in much smaller proportions, considering the brevity of scenes 4 to 6 of the first act (forty verses in total).

Performance on Quinault is also satisfying, even if his interventions can be sometimes quite short: his alleged part in the prologue barely exceeds 50 verses, his “*second intermède*” is less than 20 verses long etc. His intervention in the prologue is a bit masked by the 150 verses window. Yet, we can see that the decision function for Quinault is high at the very beginning. The finale (“*Cinquième intermède*”) and the 60-verse long “*troisième intermède*” at the end of the third act are quite visible, and attributed to Quinault, which is consistent with the notice to the reader. Even the very short “*second intermède*” is detected by both methods, and especially clearly in the genre-specific approach.²

3.2. ‘Circé’ by T. Corneille and J. Donneau de Visé

If we consider the results of the cross-genre setup on ‘Circé’ (fig. 2) – the only setup with training material for Thomas –, the major implication of Thomas Corneille in the writing of this play is self-evident. Donneau de Visé seems to rise during a small passage of the prologue, somewhere around the third scene or perhaps the “*Prologue de la musique et de la comédie*” (accounting for the blurriness due to window size), a part of sung dialog, which would be consistent with some of his claims. Yet, no other passage emerges that would make it even seem plausible that he partly versified them.

The spike of Quinault at act 2, scene 7 matches also a sung dialog (“*dialogue de Sylvie et de Tircis, qui se chante*”). Quinault obviously knew how to write passages to be sung and collaborated with Thomas Corneille in other occasions. Did Donneau de Visé have another try at plagiarism, after having experimented with such practices earlier in his career? Or did Donneau or Thomas Corneille ask a small amount of help to a colleague for a specific passage? Without further analysis, it is difficult to answer, yet it is to be noted that the value of the decision function only barely crosses 0 (implying positive class membership) on one single point. This could be an artefact due to generic attractions, because Quinault produced an important number of musical texts, represented in the training material. It is in any case deserving of further investigation before drawing any firm conclusion.

3.3. ‘L’Inconnu’ by T. Corneille and J. Donneau de Visé

When applying the same method on *L’Inconnu* (fig. 3), here again, Donneau de Visé’s implication is hard to assess, while Thomas Corneille’s style seems easily recognised. Small passages for which he claims authorship could well be too difficult to detect which such windows - but appeared in *Psyché*. Here, the decision function seems barely affected. It is to be noted that

²It is to be noted that we removed the “*Premier intermède*”, written in Italian (by Lully), for obvious reasons.

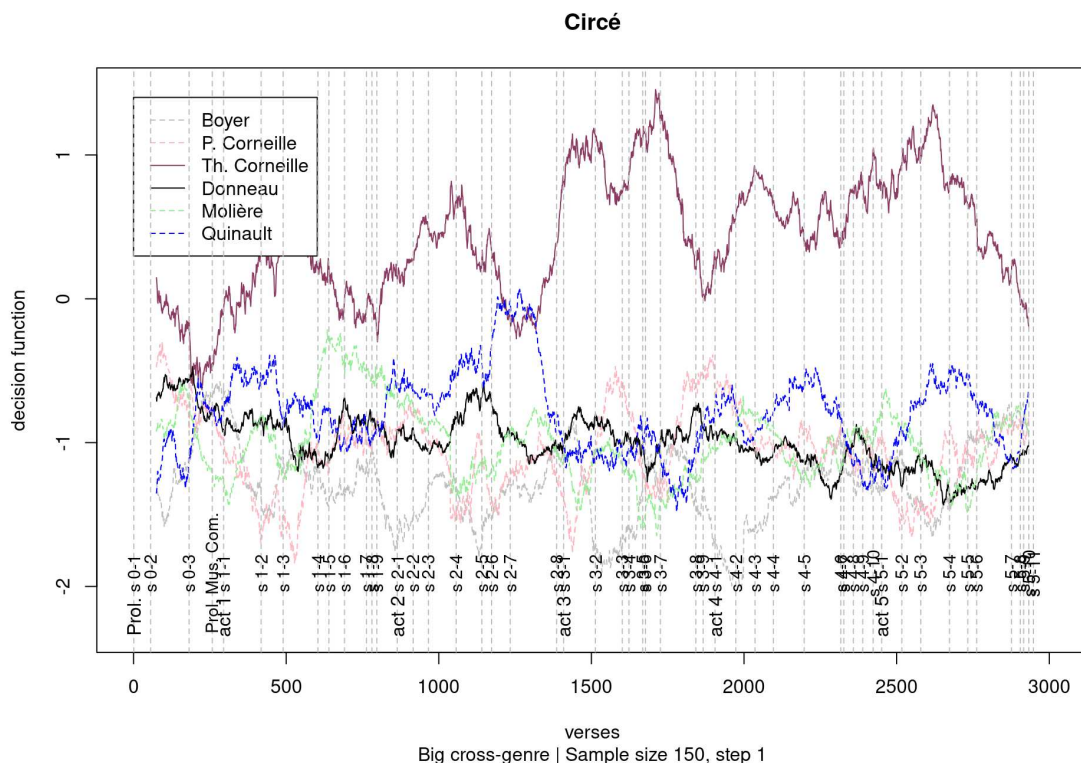


Figure 2: Results for *Circé*, using the cross-genre setup.

the value of the decision function for Thomas collapses below 0 on a few occasions, especially during the long dialogues at the end of act 1 and act 2. This could also receive an interpretation based on generic discrepancies between the training set and this part of the text.

4. Discussion

On *Psyché* the results of the rolling analysis verifies very closely the self-declaration of the notice. In comparison, we seem only to identify a non declared limited intervention of Corneille at the end of the first act.

Our results seem to confirm that Donneau de Visé’s contribution to the final writing of the plays he co-signed with Thomas Corneille was quite scarce. This of course does not mean that his contribution to those plays was non-existent. Stylometric analyses such as the ones performed here mostly detect the style of the person who actually wrote the last version of the sentences. Thomas Corneille for instance versified *Molière*’s famous comedy in prose, *Dom Juan*, after his death and stylometry attributes the play to Thomas Corneille without a blink [3], while Molière arguably contributed quite a bit to the final result... Donneau de Visé could have given a lot of insights about the intrigue, written large passages in prose etc. But for now, his contribution strictly to the verses seem even scarcer that what he claimed after Thomas Corneille’s death.

In broader terms, the clues gathered here seem to point mainly towards the first case of collaborative writing described by Pennebaker and Ireland [21], the “Just-like-another-member-

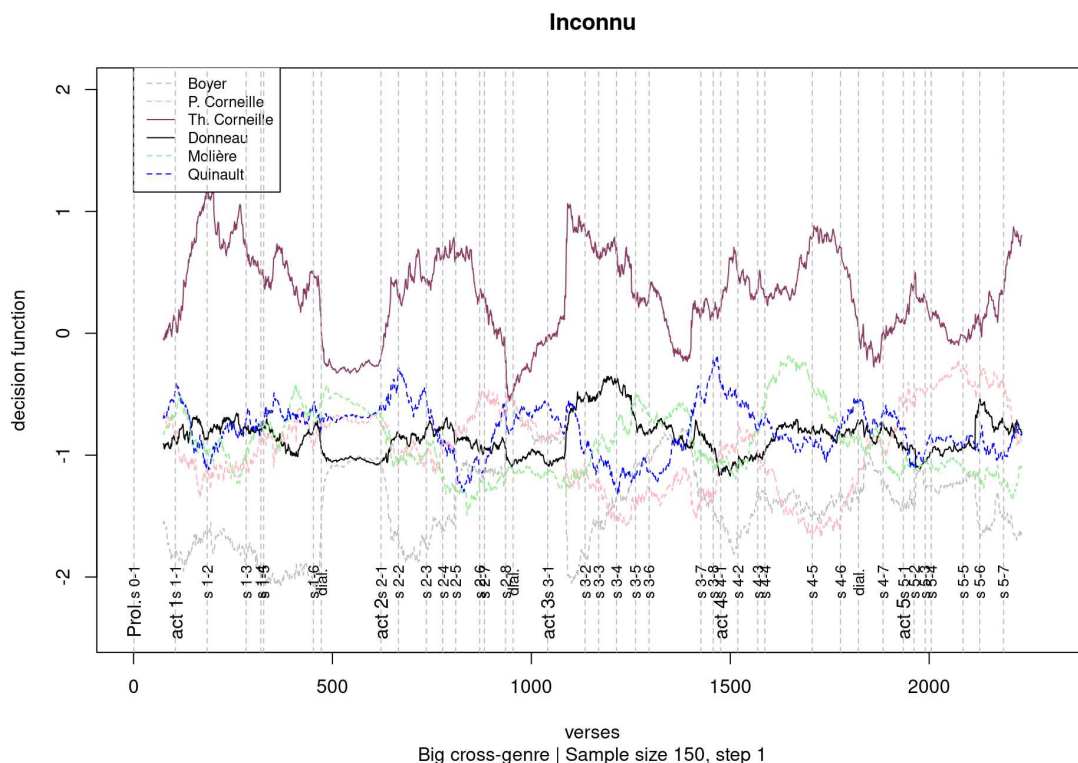


Figure 3: Results for *l’Inconnu*, using the cross-genre setup.

of-the-team hypothesis”: each portion is mostly attributable to the individual style of a single author, i.e., the one responsible for the final form of the text, not for its content or for previous formulations (in particular in the case of –necessarily heavy – transpositions between prose and verse). Yet, this would be deserving of further research to improve still our understanding of authorial collaborations during the *Grand Siècle* and beyond. For now, it remains impossible to say if some points of the texts, where the decision function for all candidate authors collapse, can be attributed to synergies instead of, for instance, generic disturbances (see the case of the *Inconnu*).

5. Further research

Further research are still needed to confirm and extend the results on *Circé* and *L’Inconnu*, and more generally on collaborative writing during this age. Increasing the size of the training set, especially for Donneau de Visé could be a first lead. In particular, we were not able to secure access to usable digital text of plays such as *Les Amours de Vénus et d’Adonis* (1670) or *Amours de Bacchus et d’Ariane* (1672), two *pièces à machine* he wrote alone during the same decade as his collaborations with Thomas Corneille. Running an efficient OCR and post-correction of those texts, already digitised by the *Bibliothèque nationale de France*, should thus be an important next step.

In terms of analysed features, we could extend our work to account for metrical features [22], accounting for instance for verse length in syllables, and the sequence of such lengths in

the “*vers libres*” parts. We could also try to cross stylistic with thematic features, to further investigate contribution to the plot by opposition to contributions to the versification.

The Quinault spike in *Circé* could also make us think that generic attractions are still an issue in our study: having written numerous opera librettos, Quinault could be a designated candidate for whatever looks like a sung passage to our SVM model. We should thus check for possible imbalances in the training corpus. First experiments however seem to show that our results stay even when downsizing Quinault’s sung passage in the training set.

Finally, we could extend our process to collaborations in prose - which were quite numerous in the *théâtre classique* in general, and which also occurred in a *pièce à machines* such as *La Devineresse* by Thomas Corneille and Jean Donneau de Visé.

Code and data availability

Code and datasets are available at [10.5281/zenodo.5517801](https://zenodo.org/record/5517801).

Acknowledgments

We thank Pr. Georges Forestier for his input on *Psyche* and the subgenre of “*pièces à machines*”, Thibault Clérice for fruitful discussions on machine-learning and stylometry, and anonymous reviewers for their careful reading and insightful suggestions. Errors remain our own.

References

- [1] N. Akiyama. “Corneille et ses pièces à machines”. In: *Dix-septième siècle* 3 (2010), pp. 403–417.
- [2] R. Bray. “L’Introduction des vers mêlés sur la scène classique”. In: *Pmla* 66.4 (1951), pp. 456–484.
- [3] F. Cafiero and J.-B. Camps. “Why Molière most likely did write his Plays”. In: *Science advances* 5.11 (2019), eaax5489. DOI: [10.1126/sciadv.aax5489](https://doi.org/10.1126/sciadv.aax5489).
- [4] F. Cafiero, J.-B. Camps, S. Gabay, and M. Puren. “La naissance du style: auteur vs genre aux XVII^e et XIX^e siècles”. In: *Humanistica 2020*. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02577853/>.
- [5] K. van Dalen-Oskam and J. Van Zundert. “Delta for Middle Dutch: Author and Copyist Distinction in Walewein”. In: *Literary and Linguistic Computing* 22.3 (2007), pp. 345–362.
- [6] M. De Pure. *Idée des spectacles anciens et nouveaux*. Minkoff, 1668.
- [7] J. Donneau de Visé. “[Notice nécrologique]”. In: *Mercure galant* (1710), pp. 270–299. URL: <https://gallica.bnf.fr/ark:/12148/bpt6k6351123z/f276>.
- [8] J. Donneau de Visé. *Préface à Sganarelle, ou le Cocu imaginaire*. Ed. by G. Forestier and C. Fournial. Paris: Jean Ribou, 1660. URL: <http://idt.huma-num.fr/notice.php?id=317>.
- [9] M. Eder. “Does Size matter? Authorship Attribution, Small Samples, Big Problem”. In: *Literary and Linguistic Computing* 30.2 (2015), pp. 167–182. DOI: [10.1093/llc/fqt066](https://doi.org/10.1093/llc/fqt066). URL: <https://academic.oup.com/dsh/article/30/2/167/390738>.

- [10] M. Eder. “Rolling Stylometry”. In: *Digital Scholarship in the Humanities* 31.3 (2016), pp. 457–469.
- [11] S. Evert, T. Proisl, F. Jannidis, I. Reger, S. Pielström, C. Schöch, and T. Vitt. “Understanding and Explaining Delta Measures for Authorship Attribution”. In: *Digital Scholarship in the Humanities* 32.suppl_2 (2017), pp. ii4–ii16. DOI: 10.1093/llc/fqx023. URL: <https://academic.oup.com/dsh/article/32/suppl%5C%5F2/ii4/3865676>.
- [12] P. Fièvre. *Théâtre classique*. 2007. URL: <http://www.theatre-classique.fr/>.
- [13] A. A. Gladwin, M. J. Lavin, and D. M. Look. “Stylometry and Collaborative Authorship: Eddy, Lovecraft, and ‘The Loved Dead’”. In: *Digital Scholarship in the Humanities* 32.1 (2017), pp. 123–140.
- [14] M. Kestemont. “Function Words in Authorship Attribution: From Black Magic to Theory?” In: *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*. 2014, pp. 59–66.
- [15] M. Kestemont, S. Moens, and J. Deploige. “Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux”. In: *Digital Scholarship in the Humanities* 30.2 (2015), pp. 199–224.
- [16] M. Kestemont, J. Stover, M. Koppel, F. Karsdorp, and W. Daelemans. “Authenticating the Writings of Julius Caesar”. In: *Expert Systems with Applications* 63 (2016), pp. 86–96.
- [17] F.-J. (Lagrange-Chancel. *Œuvres de Monsieur de Lagrange Chancel revues et corrigées par lui-même*. 1758.
- [18] P. Mélése. *Un homme de lettres au temps du grand roi, Donneau de Visé: fondateur du Mercure galant*. Librairie Droz, 1936.
- [19] Molière. *La Princesse d’Elide*. Robert Ballard/Thomas Jolly/Guillaume de Luynes/Louis Billaine, 1665.
- [20] Molière. *Psyché: tragédie-ballet, par I.B.P. Molière*. 1671. URL: <https://gallica.bnf.fr/ark:/12148/bpt6k70160j/f7.item>.
- [21] J. W. Pennebaker and M. E. Ireland. “Using Literature to Understand Authors: The Case for Computerized Text Analysis”. In: *Scientific Study of Literature* 1.1 (2011), pp. 34–48. DOI: 10.1075/ssol.1.1.04pen. URL: <https://www.jbe-platform.com/content/journals/10.1075/ssol.1.1.04pen>.
- [22] P. Plecháč. “Relative Contributions of Shakespeare and Fletcher in Henry VIII: An Analysis Based on Most Frequent Words and Most Frequent Rhythmic Patterns”. In: *Digital Scholarship in the Humanities* (2019).
- [23] F. G. Quiel. “Comedia palaciega et tragi-comédie française au XVII^e siècle: adresse au lecteur et transfert littéraire dans la Cassandre de l’abbé Boisrobert”. In: *Revista de Lenguas Modernas* (2010).
- [24] G. Rouchès. *Inventaire des lettres et papiers manuscrits de Gaspare, Carlo et Lodovico Vigarani conservés aux Archives d’État de Modène, 1634-1684*. Champion, 1913.
- [25] J. Rybicki, D. Hoover, and M. Kestemont. “Collaborative Authorship: Conrad, Ford and Rolling Delta”. In: *Literary and Linguistic Computing* 29.3 (2014), pp. 422–431.

- [26] U. Sapkota, S. Bethard, M. Montes, and T. Solorio. “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*. 2015, pp. 93–102.
- [27] C. Schöch. “Fine-Tuning our Stylometric Tools: Investigating Authorship, Genre, and Form in French Classical Theater”. In: *Digital Humanities 2013: Conference Abstracts*. 2013, pp. 383–86.
- [28] C. Schuway. *Un entrepreneur des lettres au XVIIIe siècle: Donneau de Visé, de Molière au Mercure galant*. Lire le XVIIIe siècle 69. Paris: Classiques Garnier, 2020. DOI: 10.15122/isbn.978-2-406-09572-9.
- [29] H. Sun and M. Jin. “Collaborative Writing of ‘Otome no minato’”. In: *Structure, Function and Process in Texts* (2018), p. 116.
- [30] M. Tschuggnall, E. Stamatatos, B. Verhoeven, W. Daelemans, G. Specht, B. Stein, and M. Potthast. “Overview of the Author Identification Task at PAN-2017: Style Breach Detection and Author Clustering”. In: *CLEF (Working Notes)*. 2017.
- [31] A. Viala. *Naissance de l’écrivain*. Minuit, 1985.
- [32] H. Visentin. “La tragédie à machines ou l’art d’un théâtre bien ajusté”. In: *Littératures classiques, hors-série, 2002. Mythe et histoire dans le théâtre classique. Hommage à Christian Delmas* (2002). DOI: 10.3406/licla.2002.1825. URL: <https://www.persee.fr/doc/licla%5C%5F0992-5279%5C%5F2002%5C%5Fhos%5C%5F1%5C%5F1825>.
- [33] H. Visentin. “Le théâtre à machines: Succès majeur pour un genre mineur”. In: *Littératures classiques* 51.1 (2004), pp. 205–222.

A. Plays used as training

A.1. Genre-specific setup

A.1.1. Train (with leave-one-out test)

author	title	date	n. words
Boyer, Claude	LES AMOURS DE JUPITER ET DE SÉMÉLÉ, TRAGÉDIE	1666	20554
Corneille, Pierre	LA CONQUÊTE DE LA TOISON D’OR, TRAGÉDIE	1661	22779
Donneau de Visé, Jean	LES AMOURS DU SOLEIL, PASTORALE.	1671	20898
Molière	Amphitryon, Comédie	1668	17603
Quinault, Philippe	THÉSÉE, TRAGÉDIE	1675	9200
Quinault, Philippe	CADMUS et HERMIONE, TRAGÉDIE	1673	6840

A.2. Cross-genre setup

A.2.1. Train

author	title	date	n. words
Boyer, Claude	AGAMEMNON, TRAGÉDIE.	1680	16638
Boyer, Claude	LES AMOURS DE JUPITER ET DE SÉMÉLÉ, TRAGÉDIE	1666	20554
Boyer, Claude	ARISTODÈME	1648	13642
Boyer, Claude	ARTAXERCE, TRAGÉDIE	1683	16777
Boyer, Claude	CLOTILDE, TRAGÉDIE.	1659	20711
Boyer, Claude	JUDITH, TRAGÉDIE	1695	14507
Boyer, Claude	LISIMÈNE OU LA JEUNE BERGÈRE, PASTORALE	1672	18040
Boyer, Claude	LA MORT DES ENFANTS DE BRUTE, TRAGÉDIE.	1648	14634
Boyer, Claude	OROPASTE OU LE FAUX TONAXARE	1663	22328
Boyer, Claude	LA PORCIE ROMAINE	1646	16131
Boyer, Claude	PORUS OU LA GÉNÉROSITÉ D’ALEXANDRE, TRAGÉDIE.	1648	15305
Boyer, Claude	TYRIDATE, TRAGÉDIE	1649	16941
Corneille, Pierre	AGÉSILAS, TRAGÉDIE	1666	20383

author	title	date	n. words
Corneille, Pierre	ANDROMÈDE, TRAGÉDIE.	1651	16439
Corneille, Pierre	ATTILA, ROI DES HUNS, TRAGÉDIE	1668	18913
Corneille, Pierre	LE CID, TRAGI-COMÉDIE	1637	18273
Corneille, Pierre	LE CID, TRAGÉDIE	1682	18226
Corneille, Pierre	CINNA ou LA CLÉMENCE D'AUGUSTE, TRAGÉDIE	1643	18001
Corneille, Pierre	CINNA ou LA CLÉMENCE D'AUGUSTE, TRAGÉDIE	1682	18313
Corneille, Pierre	CLITANDRE, COMÉDIE	1682	16383
Corneille, Pierre	DON SANCHE D'ARAGON, COMÉDIE HÉROÏQUE	1649	19191
Corneille, Pierre	LA GALERIE DU PALAIS ou L'AMIE RIVALE	1637	18411
Corneille, Pierre	HÉRACLIUS, EMPEREUR D'ORIENT, TRAGÉDIE	1647	19614
Corneille, Pierre	HORACE, TRAGÉDIE	1641	18327
Corneille, Pierre	L'ILLUSION COMIQUE, COMÉDIE	1639	17623
Corneille, Pierre	MÉDÉE, TRAGÉDIE	1639	15921
Corneille, Pierre	MÉDÉE, TRAGÉDIE	1682	15897
Corneille, Pierre	MÉLITE OU LES FAUSSES LETTRES, COMÉDIE	1633	20284
Corneille, Pierre	MÉLITE, COMÉDIE	1682	18730
Corneille, Pierre	LE MENTEUR, COMÉDIE	N/A	19189
Corneille, Pierre	LA MORT DE POMPÉE, TRAGÉDIE	1644	18583
Corneille, Pierre	NICOMÈDE, TRAGÉDIE	1651	19229
Corneille, Pierre	OEDIPE, TRAGÉDIE	1659	20532
Corneille, Pierre	OTHON, TRAGÉDIE	1665	19119
Corneille, Pierre	PERTHARITE ROI DES LOMBARDS, TRAGÉDIE	1653	19334
Corneille, Pierre	LA PLACE ROYALE ou L'AMOUREUX EXTRAVAGANT, COMÉDIE	1637	15076
Corneille, Pierre	POLYEUCTE MARTYR, TRAGÉDIE	1643	18642
Corneille, Pierre	RODOGUNE, TRAGÉDIE	1647	19085
Corneille, Pierre	SERTORIUS, TRAGÉDIE	1662	20041
Corneille, Pierre	SOPHONISBE, TRAGÉDIE	1663	18885
Corneille, Pierre	LA SUITE DU MENTEUR, COMÉDIE	1645	20353
Corneille, Pierre	LA SUIVANTE, COMÉDIE	1637	17188
Corneille, Pierre	SURENA GÉNÉRAL DES PARTHES, TRAGÉDIE	1675	18771
Corneille, Pierre	THÉODORE, VIERGE ET MARTYRE, TRAGÉDIE CHRÉTIENNE	1646	19451
Corneille, Pierre	TITE ET BÉRÉNICE, COMÉDIE HEROÏQUE	1671	18803
Corneille, Pierre	LA CONQUÊTE DE LA TOISON D'OR, TRAGÉDIE	1661	22779
Corneille, Pierre	LA VEUVE OU LE TRÂTRE TRAH, COMÉDIE	1634	19310
Corneille, Pierre	LA VEUVE, COMÉDIE	1682	19823
Corneille, Thomas	L'AMOUR À LA MODE, COMÉDIE.	1651	20384
Corneille, Thomas	ARIANE, TRAGÉDIE	1672	18737
Corneille, Thomas	LE BERGER EXTRAVAGANT, PASTORALE BURLESQUE.	1652	19730
Corneille, Thomas	BRADAMANTE, TRAGÉDIE	1695	13758
Corneille, Thomas	CAMMA, REINE DE GALATIE, TRAGÉDIE	1661	20702
Corneille, Thomas	LE CHARME DE LA VOIX, COMÉDIE	1658	19698
Corneille, Thomas	LE COMTE D'ESSEX, TRAGÉDIE	1678	17025
Corneille, Thomas	LA COMTESSE D'ORGUEIL, COMÉDIE	1690	20697
Corneille, Thomas	DARIUS, TRAGÉDIE	1659	20512
Corneille, Thomas	DON CÉSAR D'AVALOS, COMÉDIE.	1661	19482
Corneille, Thomas	LES ENGAGEMENTS DU HASARD, COMÉDIE.	1662	18651
Corneille, Thomas	LE FEINT ASTROLOGUE, COMÉDIE	1651	20223
Corneille, Thomas	LE FESTIN DE PIERRE, COMÉDIE	1677	21978
Corneille, Thomas	LE GALANT DOUBLÉ, COMÉDIE.	1659	20794
Corneille, Thomas	LE GEÔLIER DE SOI-MÊME, COMÉDIE.	1655	19230
Corneille, Thomas	MAXIMIAN, TRAGÉDIE	1662	20419
Corneille, Thomas	MÉDÉE, TRAGÉDIE EN MUSIQUE	1693	9659
Corneille, Thomas	LA MORT D'ANNIBAL, TRAGÉDIE	1669	19903
Corneille, Thomas	LA MORT D'ACHILLE, TRAGÉDIE	1673	18121
Corneille, Thomas	LA MORT DE L'EMPEREUR COMMODORE, TRAGÉDIE	1657	19953
Corneille, Thomas	PERSÉE ET DÉMÉTRIUS, TRAGÉDIE.	1662	21509
Corneille, Thomas	PYRRHUS, ROI D'ÉPIRE, TRAGÉDIE.	1661	21246
Corneille, Thomas	STILICON, TRAGÉDIE	1664	21267
Corneille, Thomas	THÉODAT, TRAGÉDIE	1673	18863
Corneille, Thomas	TIMOCRATE, TRAGÉDIE	1662	19804
Donneau de visé, Jean	LES AMOURS DU SOLEIL, PASTORALE.	1671	20898
Donneau de visé, Jean	LA COCUE IMAGINAIRE, COMÉDIE	1660	6278
Donneau de visé, Jean	L'EMBARRAS DE GODARD, OU L'ACCOUCHÉE, COMÉDIE	1668	8203
Donneau de visé, Jean	LE GENTILHOMME GUESPIN, COMÉDIE	1670	7413
Donneau de visé, Jean	LES INTRIGUES DE LA LOTERIE, COMÉDIE	1670	11179
Donneau de visé, Jean	LA MÈRE COQUETTE, OU LES AMANTS BROUILLÉS, COMÉDIE	1666	11907
Donneau de visé, Jean	LA VEUVE À LA MODE, COMÉDIE	1668	6123
Molière	AMPHITRYON, COMÉDIE	1668	17603
Molière	LE DÉPIT AMOUREUX	1656	19021
Molière	L'ÉCOLE DES FEMMES, COMÉDIE.	1663	19377
Molière	L'ÉCOLE DES MARIS, COMÉDIE	1661	12161
Molière	L'ÉTOURDI ou LES CONTRE-TEMPS, COMÉDIE	1663	21708
Molière	LES FÂCHEUX, COMÉDIE	1662	9607
Molière	LES FEMMES SAVANTES, COMÉDIE	1672	19135
Molière	MÉLICERTE, COMÉDIE PASTORALE HÉROÏQUE	1682	6328
Molière	LE MISANTHROPE ou L'ATRABILAIRE AMOUREUX, COMÉDIE	1667	19590
Molière	LA PRINCESSE D'ÉLIDE, COMÉDIE GALANTE	1664	10951
Molière	SGANARELLE ou Le COCU IMAGINAIRE, COMÉDIE	1660	6877
Molière	LE TARTUFFE ou L'IMPOSTEUR, COMÉDIE	1669	21088
Quinault, Philippe	AMADIS, TRAGÉDIE	1684	4733
Quinault, Philippe	ARMIDE, TRAGÉDIE.	1686	6811
Quinault, Philippe	ATYS, TRAGÉDIE	1676	9181
Quinault, Philippe	CADMUS et HERMIONE, TRAGÉDIE	1673	6840

author	title	date	n. words
Quinault, Philippe	LA COMÉDIE SANS COMÉDIE, COMÉDIE	1667	17813
Quinault, Philippe	LES COUPS DE L'AMOUR ET DE LA FORTUNE, TRAGI-COMÉDIE	1655	16101
Quinault, Philippe	LE DOCTEUR DE VERRE, COMÉDIE	1689	4324
Quinault, Philippe	LE FANTOME AMOUREUX, TRAGI-COMÉDIE	1657	18411
Quinault, Philippe	LES FÊTES DE L'AMOUR ET DE BACCHUS, PASTORALE	1672	4180
Quinault, Philippe	LA GÉNÉREUSE INGRATITUDE, TRAGI-COMÉDIE PASTORALE	1656	16516
Quinault, Philippe	ISIS, TRAGÉDIE en MUSIQUE	1687	7221
Quinault, Philippe	LA MÈRE COQUETTE ou LES AMANTS BROUILLÉS, COMÉDIE	1665	19045
Quinault, Philippe	PERSÉE, TRAGÉDIE	1682	8040
Quinault, Philippe	PROSERPINE, TRAGÉDIE	1680	8342
Quinault, Philippe	ROLAND, TRAGÉDIE EN MUSIQUE	1685	8640
Quinault, Philippe	STRATONICE, TRAGI-COMÉDIE	1660	18813
Quinault, Philippe	LE TEMPLE DE LA PAIX, BALLET	1685	3254
Quinault, Philippe	THÉSÉE, TRAGÉDIE	1675	9200

A.2.2. Test

author	title	date	n. words
Boyer, Claude	FÉDÉRIC, TRAGI-COMÉDIE	1660	18625
Corneille, Pierre	PULCHÉRIE, COMÉDIE HÉROÏQUE	1673	18884
Corneille, Thomas	LES ILLUSTRES ENNEMIS, COMÉDIE	1657	20996
Donneau de Visé, Jean	DÉLIE, PASTORALE.	1668	17723
Molière	DON GARCIE DE NAVARRE, COMÉDIE	1682	19181
Quinault, Philippe	AMALASONTE, TRAGI-COMÉDIE	1661	17932