Answer to Florian Cafiero and Jean-Baptiste Camps. Why Molière most likely did write his plays.

Dominique Labbé

Dominique Labbé
Pacte – Université Grenoble-Alpes
dominique.labbe@umrpacte.fr

Answer to

Florian Cafiero and Jean-Baptiste Camps. Why Molière most likely did write his plays.

Published in *Science Advances.* 5. 27 November 2019.
(https://advances.sciencemag.org/content/5/11/eaax5489**)**

2019 December 10

Abstract

In this article, Messrs. Cafiero and Camps claim to provide definitive proofs that P. Corneille did not write any of the plays presented by Molière. They use six "features" (lemmas, word forms, function words, rhymes, affixes, n-grams) coupled with two metrics (Burrows' distance and MinMax distance) and automatic classifications. In fact, the authors provide little precise information on these methods as to no figures like the distance matrices. The limited information, particularly in the online appendices, is sufficient to raise many doubts. For example, the list of "function words" contains many oddities that cannot be explained simply by clumsiness. Similarly, they sorted Molière's plays, drawing from experience 24 of the 33 plays. Among the discarded plays: *Psyché,* that should not have been removed because it's a collaboration between Corneille and Molière that has always been so acknowledged. Finally, the details of the classifications (published in a separate online appendix to the article) show that their methods are unable to recognize any of these authors: Boursault, Chevalier, Dancourt, Donneau de Visé, Gillet de la Tessonnerie, Pierre Corneille, Thomas Corneille, La Fontaine, Ouville, Quinault, Régnard, Rotrou... and Molière.

1. INTRODUCTION

The article of Messrs Cafiero and Camps requires three preliminary remarks.

Firstly, it appeared in a generalist open access magazine which has a poor reputation in the scientific community. Despite the management statements, papers are not peer-reviewed by experts in the field. Authors pay to be published and it is very expensive (4500$)[1]. In other words, this fake scientific journal, like many other predators, lives on the credulity of academics in need of publications, especially in the Third World. No experienced researcher submits to this kind of journal and, in the scientific community, no one wastes time reading this literature.

However, the CNRS, the Ecole des Chartes, the University of Paris-Sorbonne, the AFP, etc. announced this publication through triumphalist communiqués[2]. The mainstream press and the media have covered it extensively. Some have even made this junk magazine a "prestigious scientific journal".

It was therefore necessary to read....

Secondly, from the very beginning of Cafiero and Camps' article, it is quite clear that their targets are our work on Corneille and Molière (see Appendices 1, 2 ,3) and our authorship attribution method (Appendix 4). As attested from their acknowledgements, they did not contact us and did not provided us with their article before publication as is customary when some researchers are seriously implicated by the conclusions. If they had done so, we would have privately provided them with our critical observations. Their attitude requires us to make these comments public in order to defend our own research. In addition, our observations may be of interest to *Science Advance* readers.

Thirdly, the Cafiero and Camps' article contains little information about the metrics used and provides even less quantified results. The data posted online claim to meet the requirement of transparency that is essential in the light of the scope of the debate. The reader should have a look at them and make his own opinion. In particular: will he find the figures that were used to draw the graphs and that lead to the authors' firm conclusions?

In fact, the "proofs" provided in the article and the online appendices include very few data and not any distance matrices: in the article, only small graphs difficult to decrypt; in the appendices online an annex (Supplementary Materials) and some lists, a script and a software package. The dimensions of these missing matrices are modest: for example, for the main experiment are missing 6 tables of 37 rows and 37 columns – one for each of the 37 plays included in this experiment - that

---

[1] https://advances.sciencemag.org/content/licensing-and-charges

And about the editor-in-chief (Holden Thorp) : https://www.documentcloud.org/documents/1344054-full-wainstein-report.html

[2] For example : http://www.cnrs.fr/fr/corneille-na-pas-ecrit-les-pieces-de-moliere ; http://www.chartes.psl.eu/fr/actualite/jean-baptiste-camps-florian-cafiero-confirment-paternite-oeuvres-moliere ; https://u-paris.fr/moliere-a-bien-signe-toutes-ses-pieces-corneille-ny-est-pour-rien/

could have been easily presented online. Because of this lack of information, it seems impossible to check anything without redoing the experiments, which is also difficult due to the confusing explanations about them (as shown below).

As a result, must we trust them?

Unfortunately, in this work, there are many things that arouse mistrust.

2. A CURIOUS SELECTION

Before being applied to disputed cases, an authorship attribution method must be tested on undisputed texts. In the article by Cafiero and Camps, it is the role of the "exploratory" and "control" corpuses that can be discovered in detail in Tables S2 and S3 of the online document (Supplementary Materials). Basically, the exploratory corpus contains, in addition to some plays by Molière and P. Corneille, those by contemporary or former authors. The "control" corpus is composed of thirty plays performed after the deaths of Molière and Corneille. The crucial experiment is carried on a "main corpus" containing thirty seven plays mainly by Molière (9) and the two brothers Corneille (the Annexes 1, 2 and 3 present these corpora).

These lists have many surprises in store.

A strange "exploratory corpus"

Common sense makes it clear that the method must first be tested on texts whose authors are not in doubt. Therefore, Molière and Corneille must not be in this first test…

According to the Table S2, the "exploratory corpus" contains 34 plays (the title of the Table mistakenly indicates 30):

- Molière's *Dom Garcie*, *Mélicerte*, *Sganarelle*. P. Corneille's *Don Sanche*, *Pulchérie*, *Tite and Bérénice*. These plays would seem to be of "sure authorship"? The article provides no explanation for this strange decision. We do not understand (or rather we understand quite well) why these plays do not appear in the main experiment supposed to demonstrate Molière's existence as a "great author" unrelated with P. Corneille.

- Other example, La Fontaine's *Ragotin*. Canfiero and Camps are unaware that this play was performed and published under the name of.... Champmeslé. This actor in the Hôtel de Bourgogne troupe was a 'comédien poète' just like Molière and his life is very similar to that of Molière. After the death of Champmeslé (and La Fontaine), an Amsterdam bookseller republished this play under La Fontaine's name without explaining the reasons for this posthumous attribution. Therefore, this play had nothing to do within a corpus of texts of indisputable authorship, either that, or it had to be registered under the name of Champmeslé...

But the composition of this exploratory corpus appears to be very different as shown in Table S5 (also in Supplementary Materials) which provides the legend of some of the graphs in the article. Let us examine the first part: legend for the dendrograms of Figure 1, drawn in order to present the results of the exploratory analysis. Surprisingly, an attentive reader can discover that the calibration of the method was made not using the 3 plays alleged by Moliere –as indicated in Table S2 - but with 12 of them (In alphabetic order: *Amphitryon*, *le Dépit amoureux*, *Dom Garcie*, *l'Ecole des femmes*, *l'Ecole des maris*, *l'Etourdi*, *les Fâcheux*, *les Femmes savantes*, *Mélicerte*, *le Misanthrope*, *Sganarelle*, *le Tartuffe*). Similarly, 11 plays by Corneille were used (instead of 3 listed in Table S2). In other words, the calibration of the method was done using 12 Molière's plays and 11 by P. Corneille to find out which ones "worked" and which ones had to be eliminated in order to present only successes.

Even more surprising: the 12 clusters in Table S5 contain 72 different plays: Boursault has 7 plays (instead of 6 listed in table S2); Corneille P.: 11 (instead of 3); Corneille T.: 11 (vs 1); Molière: 12 (3); Ouville: 2 (3); Rotrou: 4 (0); Scarron: 7 (1). The article and the title of the Table S2 say "30 plays"; Table S2 lists 34 plays; the experiment have been carried on 72. The authors are very cavalier. How to check something in this chaos?

Unfortunate "omissions" in the crucial experiment

Table S1 in the "Supplementary Materials" online lists the 37 plays that were the subject of the main authorship attribution "experiment" (the table title and the article – materials and design presentation - mistakenly indicate 30 instead of 37):

- 9 plays by Molière whereas his name is associated with 33 of them: the paternity of the others is therefore certain?
- 8 by Pierre Corneille when his name is associated with 33 (or 34 with *Psyché*). Are the others not his?
- 10 by Thomas Corneille while 37 plays by this author are available.

etc., etc.

For Molière, Cafiero and Camps rejected all the prose plays even the most famous ones (*Les précieuses ridicules*, *Dom Juan*, *l'Avare*, *le Bourgeois gentilhomme*, *le Malade imaginaire*, etc.), as if they pose no attribution problems. However, their conclusion states quite rashly that Molière is the author of *all* his plays! In the plays in verse, the final experiment omitted: *Sganarelle* (1660), *Dom Garcie* (1661), *Mélicerte* (1666), *les Amants magnifiques* (1670), *Psyché* (1671).

The disappearance of Psyché

This play is Molière's greatest success (more than 70 performances without interruption, under the only name of Molière, and a recipe higher than the company's annual sales revenue). Six months after this triumph, the play was published – with only Molière's name on the cover - but with a warning from the "librarian" (the publisher):

"*The librarian to the reader*

This book is not all from one hand. Mr. Quinault made the lyrics that are sung at the beginning, with the exception of the Italian complaint. M. de Molière drew up the plan of the play, and adjusted the layout, where he focused more on the beauty and pump of the show than on the exact compliance with the rules. As for the versification, he did not have the time to do it all. The carnival was approaching, and the urgent orders of the King, who wanted to give this magnificent entertainment several times before Lent, made it necessary for him to call for a little help. Thus, there is only the prologue, the first act, the first scene of the second and the first scene of the third whose verses are his. Mr. Corneille wrote the rest during about fifteen days; and by this means, His Majesty was served within the time she had ordered."

Psyché was therefore presented to the public under the sole name of Molière (like all the others plays). However, in this case, thanks to an indiscretion (there were other indiscretions during Molière's lifetime), the collaboration between Corneille and Molière is not debatable. The passages, which each of them is supposed to have written, are clearly identified (Appendix 2). Therefore, these passages provide the litmus test made it possible to judge the effectiveness of the Cafiero and Camps method: is it able to distinguish these passages and attribute them correctly?

In their article, page 6, at the bottom of the second column, Cafiero and Camps wrote: "We excluded Psyché, a very rare case of declared collaborative authorship—this play being written by P. Corneille, Molière, and Quinault". Without having carried out this decisive test on *Psyché*, how can they claim to provide "proof" that Molière wrote all his plays and that Corneille had nothing to do with it? In addition, the reader has certainly noticed that Cafiero and Camps already knew that *Psyché* is a very rare case... before their experiment.

3. DATA AND ANALYSIS QUALITY.

The limited information available on the materials and calculations raises many questions and doubts about the quality of Cafiero and Camp's work.

Strange experiments

Table 2 (p 13) summarizes the results of the experiments, including the volume of features used and the "success rate". The reader has three major surprises.

First, Table 2 gives information only about the main and control phases. The preliminary phase supposed to demonstrate the effectiveness of the method is missing! Is this absence an unfortunate omission? If so, the authors are not serious. In fact, this absence may well be deliberate, considering the confusion about the composition of this supposed previous experiment as shown above. The fact is that the reader is deprived of a crucial element of information concerning the quality of the preparatory step.

Secondly, the authors did not perform their calculations using all the material. For example, for the main experiment, they selected only 1789 lemmas out of the 8781 constituting the whole vocabulary (about 1 out of 5). Is it acceptable to claim to study texts considering such a small proportion of their vocabulary? More importantly, Table 2 indicates that, for each feature, Cafiero and Camps selected the proportion of materials by seeking maximum efficiency from the point of view of... the conformity of the classification with the conventionally "alleged authors"! This approach has been systematic and reveals the curious statisticians they are. For example, Table 2 shows that one would only have to select the analysis with 75% of the word forms to "demonstrate" that the Cafiero and Camps method fails in more than 4/10 (43%) of the cases.

Thirdly, except for the "function words", the "success rates" are very low (often less than 90%). The reader can infer from this that, even when the most favourable proportion is chosen, the error rate is very significant. Therefore, the firmness of the conclusions is unjustified. It should have been stated, that they are given with an average error rate around 10%... which is unacceptable for such an important authorship attribution.

But are they really studying the "vocabulary" of the plays?

Strange materials

Let us consider Table S4 (in the same appendix). This table shows that they drew their own "feature". The list of the 110 – in fact 112 - function words contains all the conjunctions of coordination except "or" (because). Why this omission? In the same list, we find "après" (after) but not "avant" (before) which is much more frequent than "depuis" (since) which is in the list; we also find "jamais" (never) but not "toujours" (always) which is more frequent than "assez" (enough), "là" (here), "mieux" (better) which are present in the list. The list contains "voici" (this is) but not "voilà" (that is it) while the second is more frequent than the first, etc.

In the notice above the Table, it is stated that "personal pronouns" are excluded but the list contains "s'", and "se" (him) which are personal pronouns and much less used than; "je" (I), "tu" (you), "il" (he), "elle" (she), "nous" (we), "vous" (you), "ils" and "elles" (they), "moi" (myself) which are all absent from the list.

Another surprise is that this list of the "function words" includes the verbs "être" (to be), "avoir" (to have) and "fait" (he makes, made), which, in French, are usually considered as "content words" (supposed to be excluded from the list). Moreover, it should then be necessary to find in table S4 other conjugations of these three verbs, which is far from being the case: four-fifths of the conjugations are missing, including "faire" (to make) - more frequent than "fait" - "été" (been) and "fut" (he was) which are more frequent than "avait" or "depuis", etc.

Why all these contradictions and obvious mistakes? The authors do not know the French language? Of course not! These acrobatics were needed to bring Molière's plays to the "right place".

The sceptical reader should first consult Table 2 (p 13). The last two frames show that the authors' calculations were made using all the 112 function words (as opposed to the way they did with lemmas, word forms, rhymes, etc.). The selection of the function words was therefore made in advance according to the authors' habit: to adjust the parameters of the analysis until the "good" result is reached.

As a result, the reader should not be surprised by the following statement: "The highest agglomerative coefficient is obtained for the analysis of function words. In this analysis, all plays signed by Molière are clustered together" (p. 3). It must be observed that this statement does not apply to all the plays (as the authors fallaciously write it) but only the 9 that they attribute to Molière out of the 33 plays he presented.

The same doubts hang over the five other "features" selected or removed - depending on whether or not they succeed in isolating certain plays attributed to Molière - without any clear explanation or precise figures being given the reader.

Above all, it must be noted that this short list of function words "selected among the 250 most frequent words" contains ten typographical errors: "-ce", "-là", "-même", "qu", "c", "n", "l", "d", "s", "jusqu", instead of: ce, là, même, qu', c', n', l', d', s' jusqu' (which are also in the Cafiero and Camps' list). Since these 10 anomalies are among the 250 most frequent words, it is obvious that the texts used by Cafiero and Camps contain many typographical errors.

There is little chance that these mistakes would be distributed evenly throughout the texts. Therefore, it is likely that the classifying power claimed by Cafiero and Camps about the "function words" can be explained in a large part by these typographical errors.

Let us recall the basic rule of statistics: the quality of the conclusions depends on the quality of the observations of the phenomenon. Here, the little information shown is insufficient to show that the minimum quality is there.

The alleged results fail to reach the minimum level as well.

4. THE LAST TABLE (S5) REVEALS A COMPLETE FAILURE

The table S5 gives - for only two of the six series of graphs in the article (graphs that are too small, difficult to understand and cannot be checked in the absence of figures) - the composition of the "clusters" cut out in the corpuses (but without the figures).

What should be the results of the preliminary experiment (Fig 1) in order to confirm the method's ability to classify together texts by the same authors? It is expected that each sure author will find his or her place in one of the clusters. These groups must also be homogeneous (they should not mix two or several different authors) and the same ranking must be obtained regardless of which one out of the six features is used.

With respect to these standards, Cafiero and Camps fail completely (The reader must keep in mind that, as shown in Table2, the authors have chosen the composition of the features the most favourable for their hypotheses).

More failures than success

The Table S5 makes it possible to assess the efficiency of the Cafiero and Camps' method and to correct the omissions in their Table 2 (p. 13) in which the results of the exploratory test are missing. In the absence of the distance matrices, let:

- Success be achieved when the classification results in an authorially homogeneous cluster (a single author). It is accepted that an author may be in several clusters as long as he is alone in each cluster. Each cluster containing a single author is given a score equal to the number of plays in it.

- Conversely, if two or several authors are mixed within a cluster, the score of this cluster is 0. For example, the first cluster in Table S5 groups 3 different authors (score = 0): Boursault (*la Comédie sans titre*, *les Mots à la mode*, *le Portrait du peintre, la Satire des satires);* Donneau de Visé (*les Embarras de Godard*, *le Gentilhomme Guépin*, *les Intrigues de la loterie)*; La Fontaine (*Climène).*

The second cluster is also given a 0 score because it mixes three plays presented by Molière *(les Fâcheux, Mélicerte and Sganarelle)* with T. Corneille's *Festin de Pierre,* La Fontaine *(Ragotin)* and Donneau de Visé *(la Cocue imaginaire).*

The third one receive a score equal to 4 because it contains 4 plays by the same author (Chevalier : *les amours de Calotin, l'Intrigue des carrosses, les Barbons amoureux, le Pédagogue amoureux)*, etc.

The Table below summarizes the scores obtained by Cafiero and Camps' preliminary experiment (which was intended to measure the effectiveness of the method), for the 6 features (column) and each cluster (lines).

Table 1. Cluster scores in the Cafiero and Camps' preliminary test (Table S5). Success (score = number of plays correctly clustered) and errors (several authors are mixed in the cluster, score = 0)

Clusters	Lemmas	Rhymes	Word forms	Affixes	N-Grams	Function words	Total
1	0	0	0	0	0	0	0
2	0	0	4	5	0	0	9
3	4	4	0	5	4	0	17
4	0	0	6	0	6	0	12
5	0	0	0	0	0	0	0
6	6	0	4	6	6	3	25
7	6	0	6	0	0	0	12
8	2	0	2	4	5	6	19
9	0	0	0	0	0	0	0
10	6	0	6	5	0	4	21
11	0	0	0	0	0	6	6
12	0	6	4	4	6	4	24
Total score	24	10	32	29	27	23	145
Success rate %	33	14	44	40	38	32	34

The experiment involved 72 plays: it is the maximum total score for each column. For example, using lemmas, 24 plays were correctly clustered. The success rate is 24/72 = 33%, etc. Not a feature reaches the 50% threshold.

Given that there are 6 features, the score that an authorship attribution method must achieve is 432 (72*6). To accept the method provisionally, the score must be at least 410 (error rate of about 5%). Cafiero and Camps come very far from these figures. The total success rate is 34%: twice more failures (287) than successes (145).

The order of the clusters is also to be considered because most classifiers start considering the closest pairs. Thus the totals at the ends of the lines give an idea of the relative homogeneity of the groups on which success or failure occurs… and shows that the classifier always starts by complete mistakes – the total of the first line is 0 - and reaches some success only in the least safe areas.

Finally, it must be noted that the number of clusters obtained by Cafiero and Camps is always 12 for each feature. These kinds of events cannot occur by chance. Cafiero and Camps had limited the classifier to 12 clusters. In fact, this preliminary experiment included 11 or 12 (assumed) authors: if

Corneille was the ghost-writer of Molière, there are 11 authors; if not, they are 12. This is another example of Cafiero and Camps' way of thinking: they knew that they are 12 authors before they even started the experiment.

Untraceable authorship

The failure begins with "Molière" whose *Amphitryon*, *Dom Garcie de Navarre*, *Le Dépit amoureux, les Fâcheux, Mélicerte, Sganarelle* are never grouped in a stable and complete way except for a single one of the features (the reader can guess which one). The most amusing case is *Dom Garcie* (presented by Molière). Using five of the six features (lemmas, word forms, rhymes, n-grams, affixes), this play is clustered 13 times with P. Corneille's plays, 12 times with T. Corneille's ones, 4 times with Rotrou, 1 time with Ouville and Quinault… and never with any of the other Molière's plays… But, strangely, using function words, *Dom Garcie* appears in a cluster containing 7 other plays by Molière and none by the others. In fact, it is not all the "function words" but, as seen above, a short list of words selected in order to achieve this happy issue. Therefore, the title of Cafiero et Camps' article is misleading if not false.

Other plays by the same author, such as those by La Fontaine, Donneau de Visé or Boursault, can be found in many different clusters and never all together. T. Corneille's or Quinault's plays are scattered almost everywhere, grouped with practically all the other authors (particularly Molière) but not always the same according to the "features", etc. In fact, not only are the rankings aberrant, but furthermore they vary according to the "features" used. This is the case in particular for La Fontaine, whose plays are found everywhere, rarely in the same cluster and never grouped together.

Table S5 shows that the method of Cafiero and Camps is unable to recognize: Boursault, Chevalier, Donneau de Visé, Gillet de la Tessonnerie, Pierre Corneille, Thomas Corneille, La Fontaine, Ouville, Quinault, Rotrou... and Molière! Not one of them escaped the shipwreck.

Conclusion: the experiment should stop here because this prior check has failed.

And what are the experimenters doing?

First, they erase the bad news from Table 2.

Secondly, they remove all the plays that cause problems. The following are put out of the box without any explanation: Boursault, Chevalier, Donneau de Visé, Gillet, La Fontaine, Ouville, Quinault. They were wrongly selected? They were not real authors?

The very few who remain in the main round (Molière, P. and T. Corneille, Rotrou, Scarron), are left with most of their works amputated. As seen above, for Molière, nine plays remain (less than one in three). In addition, it should be noticed that Scarron (1610-1660) and Rotrou (1609-1650) cannot have written the Molière plays (presented between 1659 and 1673)... since they were dead.

Final miracle

With these drastic cuts… it works. Or, at least, they finally arrive where they wanted to arrive from the very beginning: 9 of the 33 plays assigned to Molière would be isolated (but how to check since we have no figures). And it doesn't matter that the other authors continue to mix.

The article states that the "control corpus" would not pose such a problem... except for Dancourt and Régnard. Dancourt (Florent Carton, also a "comédien poète" of the Comédie française) presented more than 40 plays and Régnard more than a dozen. They dominated the stages at the end of the 17th century and during the beginning of the 18th. The method fails on this representative case: this is not a problem for Cafiero and Camps. They always have an "ad hoc" hypothesis at hand to explain the inexplicable. Here, Dancourt played a leading role in a play by Régnard; Dancourt's wife was an actress, certainly had to know Régnard and influenced them both. The confusing explanations on page 3 and the second column on page 6 comes from the same source. Without Cafiero and Camps being aware of it, the multiplication of ad hoc explanations destroy one of their central assumptions: each author has particular "unconscious" stylistic characteristics.

5. CONCLUSIONS

The few factual elements, given that the experiments are difficult to check, lead all to the rejection of Cafiero and Camps' claims.

The selection of the plays shows a willingness to rule out all problematic cases – especially the omission of *Dom Garcie* (from the final test) and *Psyché* (everywhere). The absence of many authors and plays raises doubts about the good faith of the authors. This doubt is reinforced by manipulations of the list of function words in order to distance nine plays, presented by Molière, from P. Corneille. Finally, what is revealed from the calibration of the method on a corpus of 'sure' texts shows an irremediable failure. This is probably the main mistake of Cafiero and Camps. They did not test their method on anything except the "ancien régime" plays, without realizing that the essential prerequisite for any authorship attribution tool is to be able to survive a large number of tests under very diverse conditions. On the contrary, here is an "ad hoc" method designed only to show that Molière is Molière. But despite the concealment of most of the precise information (including the absence of the distance matrices used for automatic classifications), the failure is clearly visible.

The methods of Cafiero and Camps are unable to recognize the authors of the 17th century plays, including those presented by Molière.

The conclusions of our 2001 study were never overturned. On the contrary, the little information given by Cafiero and Camps (such as the strange proximity between *Dom Garcie* and the

contemporary plays of the Corneille brothers), everything they clumsily hide (such as *Psyché*) and all their manipulations only reinforce our 2001 double conclusions. First, *Dom Garcie* (1661) and *Psyché* (1671) are the two sisters – presented by Molière – of the ten tragedies by P. Corneille between 1659 and 1674. Second, Corneille's *Menteur* (1643) and *Suite du menteur* (1643) are the eldest sisters of 17 out of the 31 other comedies presented by Molière: *l'Etourdi* (1659), *le Dépit amoureux* (1659), *Sganarelle* (1660), *l'Ecole des maris* (1661), *les Fâcheux* (1661), *l'Ecole des femmes* (1662), *la Princesse d'Elide* (1664), *le Tartuffe* (1664), *Dom Juan* (1665), *le Misanthrope* (1666), *Mélicerte* (1666), *Amphitryon* (1668), *l'Avare* (1668), *les Amants magnifiques* (1670), *le Bourgeois gentilhomme* (1670), *les Femmes savantes* (1672), *le Malade imaginaire* (1673). All of them are from the same pen: Pierre Corneille's.

ACKNOWLEDGMENTS

John L. Klause, Cyril Labbé, Thomas Merriam and Jacques Savoy carefully read a first version of this article and provided us useful remarks and comments. Thomas Merriam and John Klause kindly helped to translate the French version into English.

Appendix 1

The Molière plays

		Création	Genre	Length (tokens)
01	La jalousie du barbouillé*	Before 1659	Comedy prose	3 501
02	Médecin volant*	Before 1659	Comedy prose	3 876
03	L'étourdi	1659	Comedy verse	18 671
04	Dépit amoureux	1659	Comedy verse	16 242
05	Précieuses ridicules*	1660	Comedy prose	6 648
06	Sganarelle*	1660	Comedy verse	6 042
07	Dom Garcie*	1661	Heroic Comedy vers	17 049
08	L'école des maris	1661	Comedy verse	10 536
09	Les fâcheux	1661	Comedy verse	7 922
10	L'école des femmes	1662	Comedy verse	16 625
11	Critique de l'école des f.*	1663	Comedy prose	8 610
12	L'impromptu*-	1663	Comedy prose	7 168
13	Mariage forcé*	1664	Comedy prose	6 058
14	Princesse d'Elide*	1664	Comedy verse	11 333
15	Le Tartuffe	1664	Comedy verse	18 271
16	Dom Juan*	1665	Comedy prose	17 452
17	L'amour médecin*	1665	Comedy prose	6 147
18	Le Misanthrope	1666	Comedy verse	17 180
19	Médecin malgré lui*-	1666	Comedy prose	9 317
20	Mélicerte*	1666	Comedy verse	5 540
21	Coedy pastorale*	1667	Comedy veres	732
22	Le sicilien*	1667	Comedy prose	5 375
23	Amphytrion	1668	Comedy verse	15 117
24	Georges Dandin*	1668	Comedy prose	11 009
25	L'avare*	1668	Comedy prose	21 033
26	M. de Pourceaugnac*	1669	Comedy prose	11 803
27	Amants magnifiques*	1670	Comedy verse & prose	11 983
28	Bourgeois gentilhomme*	1670	Comedy prose	17 132
29	*Psyché* (see below appendix 2)*	*1671*	*Heroic Comedy verse*	*16 282*
30	Fourberies de Scapin*	1671	Comedy prose	14 245
31	Comtesse d'Escarbagnas*	1671	Comedy prose	5 564
32	Femmes savantes	1672	Comedy verse	16 863
33	Malade imaginaire*	1673	Comedy prose	19 919

* Cafiero and Camps withdrew from their experiment 24 plays of the 33 presented by Molière

Appendix 2

29. *Psyché* (1671) presented and published by Molière

	Authors	Genre	Length (tokens)
29.1	Corneille	verse	10 067
29.2	Molière	Verse	4 816
29.3	Quinault	verse	1 399

Appendix3

The P. Corneille plays

		Creation	Genre	Length (tokens)
01	Mélite	1630 ?	Comedy	16 690
02	Clitandre*	1631	Tragi-Comedy	14 402
03	La Veuve	1631	Comedy	17 661
04	La Galerie du Palais	1632	Comedy	16 140
05	La Suivante	1633	Comedy	15 160
06	Comédie des Tuileries*	1634	Comedy	3 627
07	Médée*	1635	Tragedy	14 269
08	La Place Royale	1634	Comedy	13 801
09	L'illusion comique	1636	Comedy	15 428
10	Le Cid*	1636	Tragi-Comedy	16 677
11	Cinna*	1639	Tragedy	16 126
12	Horace*	1640	Tragedy	16 482
13	Polyeucte*	1641	Tragedy	16 472
14	Pompée*	1642	Tragedy	16 492
15	Le Menteur	1642	Comedy	16 653
16	Le Menteur (suite)	1643	Comedy	17 675
17	Rodogune*	1644	Tragedy	16 842
18	Théodore*	1645	Tragedy	17 121
19	Héraclius*	1647	Tragedy	17 433
20	Andromède*	1650	Tragedy	15 514
21	Don Sanche*	1650	Heroïc Comedy	16 947
22	Nicomède*	1651	Tragedy	16 923
23	Pertharite*	1651	Tragedy	17 121
24	Œdipe*	1659	Tragedy	18 618
25	Toison d'Or*	1661	Tragedy	20 343
26	Sertorius*	1662	Tragedy	17 675
27	Sophonisbe*	1663	Tragedy	16 858
28	Othon*	1664	Tragedy	16 971
29	Agésilas*	1666	Tragedy	18 227
30	Atilla*	1667	Tragedy	16 788
31	Tite et Bérénice*	1670	Heroïc Comedy	16 697
32	Pulchérie*	1672	Tragedy	16 630
33	Suréna*	1674	Tragédie	16 545

* Off the 33 plays presented by Molière, 25 were withdrawn from the Cafiero and Camps' experiment.

Appendix 4

Our authorship attribution method

From the first lines, one thing is certain: Cafiero and Camps did not read us. For example, in the first column on page 1, they claim that our method gives more weight to words with high frequencies. If they had taken the time to read our first article (Labbé, Labbé 2001), they would have seen that the inter-textual distance takes into account the whole vocabulary and gives each word exactly its weight in the texts without distortion. Another example, in the first column of page 2, the authors have us say that the *Précieuses ridicules* (presented by Molière) is by P. Corneille (it is the only play they quote about us). No luck! In our 2001 article, this play is not attributed to P. Corneille and we have not changed our conclusions since. As can be seen, Cafiero and Camps want to "prove" too much...

If they had read us – or if they had agreed to discuss with us as is customary in the scientific community - they would have seen that, since 1999, our method has been successful and has passed the most difficult tests. For example, a blind experiment with two English researchers (Labbé 2007). After another blind test and a thorough examination, the CNRS mathematicians published our method in their online journal (*Images des mathématiques:* Labbé, Labbé 2011). The latest success was the identification of the gosh-writer behind the pseudonym of E. Ferrante (Savoy 2018). In the latter case, this conclusion has been validated by the results obtained by seven other teams of specialists around the world (Tuzzi 2018; Tuzzi, Cortelazo 2018).

Most importantly, this method is used to fight against frauds in scientific publications (Labbé, Labbé 2012) with successes that have been praised three times by the journal *Nature* (Van-Norden 2014; Phillips 2017; Byrne 2019). These tools have been integrated into the decision-making process by the Springer-Macmillan group and, when now placed in the public domain, it is used by main international scientific publishers (Minh Tien 2018). Recently, it detected a massive fraud in cancer research (Byrne, Labbé 2016), which earned one of the members (J. Byrne) of our research network the title of "Scientist of the Year 2017" by the journal *Nature*.

In other words, our tools work well and world-wide… all but Molière?

References

J. Byrne, We need to talk about systematic fraud, *Nature*, 06 February 2019.
https://www.nature.com/articles/d41586-019-00439-9

J. Byrne, C. Labbé, Striking similarities between publications from China describing single gene knockdown experiments in human cancer cell lines, *Scientometrics*, 28 December 2016.
http://membres-lig.imag.fr/labbe/Publi/ByrneLabbe2016.pdf

C. Labbé, D. Labbé, Inter-Textual Distance and Authorship Attribution Corneille and Molière, *Journal of Quantitative Linguistics*, 8-3, p. 213-231, December 2001 / https://www.researchgate.net/publication/32222053

C. Labbé, D. Labbé, La classification des textes. Comment trouver le meilleur classement possible au sein d'une collection de textes ? *Images des mathématiques. La recherche mathématique en mots et en image,* 28 March 2011. http://images.math.cnrs.fr/La-classification-des-textes.html

C. Labbé, D. Labbé, Duplicate and fake publications in the scientific literature: how many SCIgen papers in computer science? *Scientometrics*, 22 June 2012. https://www.researchgate.net/publication/257663143

D. Labbé, Experiments on Authorship Attribution by Intertextual Distance in English, *Journal of Quantitative Linguistics*, 14(1), p. 33-80, April 2007 https://www.researchgate.net/publication/32222131

N. Minh Tien. *Detection of automatically generated texts*. Ph-D Thesis. Grenoble-Alpes University. 3-04-2018. https://tel.archives-ouvertes.fr/tel-01919207

N. Philipps. Online software spots genetic errors in cancer papers. *Nature*. 20 November 2017. https://www.nature.com/news/online-software-spots-genetic-errors-in-cancer-papers-1.23003

J. Savoy, *Elena Ferrante Unmasked*. Neuchatel University, September 2017. https://www.researchgate.net/publication/320131096_Elena_Ferrante_Unmasked

A Tuzzi, *It Takes Many Hands to Draw Elena Ferrante's Profile*. Padova University Press, 2018. https://www.researchgate.net/publication/326723646_It_Takes_Many_Hands_to_Draw_Elena_Ferrante's_Profile

A. Tuzzi, M. Cortelazo, What is Elena Ferrante? A comparative analysis of a secretive bestselling Italian writer, *Digital Scholarship in the Humanities*, Volume 33, 3, p 685–702, September 2018

R. Van Noorden, Publishers withdraw more than 120 gibberish papers, *Nature*, 24 February 2014. https://www.nature.com/news/publishers-withdraw-more-than-120-gibberish-papers-1.14763