



**HAL**  
open science

**Réponse à Florian Cafiero et Jean-Baptiste Camps.  
Why Molière most likely did write his plays. Science  
Advances. 5: eaax5489. 27 November 2019.**

Dominique Labbé

► **To cite this version:**

Dominique Labbé. Réponse à Florian Cafiero et Jean-Baptiste Camps. Why Molière most likely did write his plays. Science Advances. 5: eaax5489. 27 November 2019.. 2019. halshs-02383640v1

**HAL Id: halshs-02383640**

**<https://shs.hal.science/halshs-02383640v1>**

Preprint submitted on 27 Nov 2019 (v1), last revised 19 Dec 2019 (v2)

**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.

Dominique Labbé

[Dominique.labbe@umrpacte.fr](mailto:Dominique.labbe@umrpacte.fr)

<https://www.pacte-grenoble.fr/membres/dominique-labbe>

## Réponse à

Florian Cafiero et Jean-Baptiste Camps. Why Molière most likely did write his plays. *Science Advances*. 5: eaax5489. 27 November 2019.

Grenoble le 27 novembre 2019

### Résumé

Dans cet article, MM. Cafiero et Camps prétendent apporter la preuve que P. Corneille n'a écrit aucune des pièces présentées par Molière. Ils utilisent pour cela 6 "caractéristiques" (lemmes, formes, mots outils, rimes, affixes, n-grams) couplées avec des classifications automatiques. En fait, les auteurs fournissent peu d'informations précises sur ces méthodes et aucune donnée chiffrée. Les quelques informations, notamment dans les annexes en ligne, suffisent pour soulever beaucoup de doutes. Par exemple, la liste des "mots outils" comporte de nombreuses étrangetés qui ne peuvent s'expliquer simplement par des maladroites. De même, ils ont opéré un tri dans les pièces de Molière, retirant de l'expérience 24 des 33 pièces. Parmi ces pièces écartées : *Psyché* qu'il ne fallait surtout pas retirer ! Enfin, le détail des classifications (publié dans une annexe en ligne séparée de l'article) montre un échec total. Leurs méthodes se révèlent incapables de reconnaître : Boursault, Chevalier, Dancourt, Donneau de Visé, Gillet de la Tessonnerie, Pierre Corneille, Thomas Corneille, La Fontaine, Ouville, Quinault, Régnard, Rotrou... et Molière.

Deux remarques préalables.

Premièrement, une série de données est disponible sur le site :

["https://advances.sciencemag.org/content/suppl/2019/11/21/5.11.eaax5489.DC1](https://advances.sciencemag.org/content/suppl/2019/11/21/5.11.eaax5489.DC1)

Ces données, par ailleurs très intéressantes pour les chercheurs, prétendent répondre à l'exigence de transparence indispensable étant donnée la portée du débat. Nous conseillons au lecteur d'y jeter un coup d'œil et de se faire sa propre opinion. En particulier : trouvera-t-il les données chiffrées qui ont servi à tracer les graphiques et qui conduisent aux conclusions si fermes des auteurs ?

Deuxièmement, dès le premier paragraphe de cet article, il est clair que nos travaux sont la principale cible. Alors pourquoi n'avoir pas pris contact avec nous ? Pourquoi ne pas nous avoir donné la possibilité de travailler ensemble et d'arriver à une position commune ou à une liste raisonnée de divergences ? C'est ainsi que la science progresse. L'échange des données et la possibilité de répondre pour les chercheurs mis en cause sont des règles basiques dans la communauté scientifique. MM. Cafiero et Camps ignorent ces règles ?

Voici quelques observations que MM. Cafiero et Camps nous contraignent à rendre publiques, alors que nous les leur aurions volontiers communiquées de manière privée s'ils avaient respecté les usages de la communauté scientifique.

Dès les premières lignes, une certitude : ils ne nous ont pas lus. Par exemple,

- dans la première colonne de la page 1, ils affirment que notre méthode donne plus de poids aux mots de fortes fréquences. C'est faux ! S'ils s'étaient donnés la peine de lire nos articles, ils auraient vu que la distance intertextuelle donne à chaque mot exactement son poids dans les textes sans aucune déformation ;

- dans la première colonne de la page 2, ils nous font dire que les *Précieuses ridicules* seraient de P. Corneille (c'est la seule pièce qu'ils citent à notre propos). Pas de chance ! Dans notre article de 2001, cette pièce n'est pas attribuée à P. Corneille et nous n'avons pas varié depuis.

A vouloir trop "prouver"...

Sur le fond, le 27 novembre au soir, à sa parution, l'article et les annexes en ligne ne comportent aucun tableau de données, seulement des graphiques difficiles à comprendre. Dans l'état actuel de ce qui a été mis en ligne, il est impossible de contrôler sans refaire les expériences, ce qui demandera de longues semaines.

En attendant, il faut donc faire confiance ?

Malheureusement dans ce travail, beaucoup de choses soulèvent la défiance.

## 1. Qualité des données et des analyses.

Nous nous appuyons sur la table 1 (page 10) et la table S4, dans l'annexe mise en ligne avec l'article à l'adresse mentionnée ci-dessus (*Supplementary Materials for Why Molière most likely did write his plays*). Ces deux tables montrent que :

- MM. Cafiero et Camps ne travaillent pas sur les lemmes – contrairement à ce qu'ils affirment. Par exemple, la table 10 contient "ta", "ton", "tes" qui sont des flexions ("word forms") d'un même article possessif. Ou encore, la table des "mots fonctionnels" (S4) comporte : c', jusqu', l', etc. qui sont des flexions des lemmes : 'ce', 'jusque', 'le', etc.

- dans cette liste des mots "fonctionnels", on trouve les conjonctions de coordinations sauf 'or' ; on trouve 'après' mais pas 'avant' ; 'jamais' mais pas 'toujours' ; 'ici' et 'là' mais pas 'ailleurs' ; 'voilà' mais pas 'voici' ; 'se' mais pas 'il', 'elle' ou 'me', 'te' ; on trouve 'tous' et 'toute' mais pas 'toutes', etc. Toutes les prépositions, conjonctions, articles, pronoms, adverbes qui sont absents de cette liste ne seraient donc pas des "outils" ? Certaines flexions d'un même mot le seraient et d'autres pas ?

Autre surprise : figurent dans cette liste des outils : les verbes "être" et "avoir". Ce seraient donc des "outils" et pas des "mots lexicaux" ? Il faudrait alors trouver dans la table S4 toutes les flexions de ces deux verbes, ce qui est très loin d'être le cas. Dès lors, certaines de ces flexions seraient des mots "lexicaux" – comme "suis" ou "sommes" qui ne sont pas dans la liste alors qu'ils sont très utilisés dans les comédies - et d'autres des "outils", comme "est" ou "êtes", également très utilisés, qui sont, eux, dans la liste de la table S4 ?

Les auteurs ne connaissent pas le français ?

Ces curiosités sont nécessaires pour amener les pièces de Molière à la "bonne place". Effectivement, sans surprise, on lit p. 3 : "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).

Nous verrons aussi que "all" n'est pas exact puisque cette affirmation ne concerne que 9 pièces attribuées à Molière sur les 33 qu'il a présentées.

La même question peut être posée à propos de toutes les "features" (lemmes, formes, rimes, affixes, n-grams, mots outils) sélectionnées ou enlevées – selon qu'elles réussissent ou non à isoler certaines pièces attribuées à Molière - sans qu'aucune explication claire, ni chiffres précis soient donnés au lecteur.

- A la fin de cette même annexe, on trouve les "mots outils" suivants : qu / c / n / l / d / s / jusqu. Naturellement, ces mots auraient dû avoir une apostrophe terminale. Autrement dit, les textes utilisés par Cafiero et Camps comportent de nombreuses fautes de frappe qui n'ont pas été corrigées.

Rappelons la règle de base en statistique : la qualité des conclusions dépend de celle des observations du phénomène. Ici, le peu d'informations disponibles suffit largement pour démontrer que la qualité n'est pas au rendez-vous.

A cela s'ajoute une grande désinvolture dans le choix des pièces.

## 2. Une curieuse sélection

Avant d'être appliquée aux cas douteux, une méthode d'attribution d'auteur doit être mise à l'épreuve sur des cas qui ne le sont pas. Dans l'article de Cafiero et Camps, c'est le rôle des corpus "exploratoire" et de "contrôle" qu'on peut découvrir en détail dans les tableaux S2 et S3 (du document annexe). En gros, le corpus exploratoire contient, outre quelques pièces de Molière, celles d'auteurs contemporains ou antérieurs. Le corpus de "contrôle" est composé d'une trentaine de pièces postérieures.

Ces listes réservent de nombreuses surprises. Par exemple :

- La Fontaine *Ragotin*. Les auteurs ignorent que cette pièce a été jouée et publiée sous le nom de... Champmeslé. Cet acteur de la troupe de l'Hôtel de Bourgogne était comédien poète exactement comme Molière et son histoire est très semblable à celle de Molière. Après la mort de Champmeslé (et de La Fontaine), un libraire d'Amsterdam a republié cette pièce sous le nom de La Fontaine sans expliquer les raisons de cette attribution posthume. Cette pièce n'avait donc rien à faire dans un corpus de textes de paternité indiscutable. Ou alors, puisque, au début de leur papier, MM. Cafiero et Camps mettent en doute l'existence même du système du comédien poète, il fallait l'enregistrer sous le nom de Champmeslé...

- Molière : *Dom Garcie*, *Mélicerte*, *Sganarelle* figurent dans le corpus des pièces utilisées pour étalonner la méthode de manière "exploratoire". Ces pièces seraient donc "sûres" ? L'article n'apporte aucune explication concernant cette étrange décision. On ne comprend pas (ou plutôt on comprend trop bien) pourquoi ces pièces ne figurent pas dans l'expérience "finale" censée démontrer l'existence de Molière "grand auteur".

## 3. Des "oublis" fâcheux

La table S1 de l'annexe liste les 30 pièces sur lesquelles a porté l' "expérience" finale d'attribution d'auteur :

- 9 de Molière alors que celui-ci en a présenté 33 : la paternité des autres est donc certaine ?

- 8 de Pierre Corneille alors qu'il en a présenté 33 (ou 34 avec *Psyché*). Les autres ne seraient pas de lui ?

- 10 de Thomas Corneille alors que 37 pièces de lui sont disponibles.

etc., etc.

Pour Molière, Cafiero et Camps ont écarté toutes les pièces en prose (*Les précieuses ridicules, Dom Juan, l'Avare, le Bourgeois gentilhomme, le Malade imaginaire*, etc.). Elles ne posent donc aucun problème d'attribution ? En effet, leur conclusion affirme bien imprudemment que Molière est l'auteur de *toutes* ses pièces !

Dans les pièces en vers, ils ont "oublié" :

*Sganarelle* (1660)

*Dom Garcie* (1661)

*Mélicerte* (1666)

*Les amants magnifiques* (1670)

*Psyché* (1671)

Cette dernière pièce est le plus grand succès de Molière (plus de 70 représentations d'affilée, sous le seul nom de Molière, et une recette supérieure au chiffre d'affaire annuel de la troupe). Six mois après ce triomphe, la pièce est publiée – avec en couverture et en page de garde, le seul nom de Molière – mais avec un avertissement du "libraire" (l'éditeur) :

"Le libraire au lecteur

Cet ouvrage n'est pas tout d'une main. M. Quinault a fait les paroles qui s'y chantent en musique, à la réserve de la plainte italienne. M. de Molière a dressé le plan de la pièce, et réglé la disposition, où il s'est plus attaché aux beautés et à la pompe du spectacle qu'à l'exacte régularité. Quant à la versification, il n'a pas eu le loisir de la faire entière. Le carnaval approchait, et les ordres pressants du Roi, qui se voulait donner ce magnifique divertissement plusieurs fois avant le carême, l'ont mis dans la nécessité de souffrir un peu de secours. Ainsi, il n'y a que le prologue, le premier acte, la première scène du second et la première du troisième dont les vers soient de lui. M. Corneille a employé une quinzaine au reste ; et, par ce moyen, Sa Majesté s'est trouvée servie dans le temps qu'elle avait ordonné."

*Psyché* a donc été présentée au public sous le seul nom de Molière (comme toutes les autres). Cependant, dans ce cas, grâce à une indiscretion (il y en a eu d'autres du vivant de Molière), la collaboration entre Corneille et Molière n'est pas discutable. Les passages, que chacun d'eux est censé avoir écrits, sont clairement identifiés. Dès lors, ces passages fournissaient "l'épreuve décisive" qui aurait permis de juger de l'efficacité de la méthode de MM. Cafiero et Camps : est-elle capable de distinguer ces passages et de les attribuer correctement ?

Dans leur article, page 6, au bas de la deuxième colonne, les auteurs admettent implicitement que la réponse est négative et que, pour cette raison, ils ont retiré *Psyché* de toutes leurs expériences.

Faute d'avoir réalisé cette expérience décisive sur *Psyché*, leurs prétentions s'écroulent.

#### 4. La dernière table (S5) révèle un échec complet

Cette table donne – pour trois seulement des 6 séries de graphiques de l'article (graphiques difficilement compréhensibles et incontrôlables en l'absence des données chiffrées) - la composition des "clusters" découpés dans les corpus (mais sans les chiffres). On s'attend à ce que, dans l'expérience préliminaire et cruciale (décrite par les six diagrammes de la figure 1), chaque auteur "sûr" trouve sa place dans l'un des groupes. Il faut aussi que ces groupes soient homogènes et que le même classement se retrouve quelle que soit la "feature" utilisée. Sinon, l'expérience doit s'arrêter puisque le contrôle préalable a échoué.

Le ratage commence avec "Molière" dont *Amphitryon*, *Dom Garcie de Navarre*, *le Dépit amoureux*, les *Fâcheux*, *Mélicerte*, *Sganarelle* ne sont jamais groupées de manière stable et complète... Le titre de l'article de Cafiero et Camps est donc trompeur sinon mensonger.

D'autres pièces d'un même auteur, comme celles de La Fontaine, Donneau de Visé ou Boursault, se trouvent dans beaucoup de groupes différents et jamais toutes ensemble ! Les pièces de Thomas Corneille ou de Quinault se dispersent presque partout, groupées avec pratiquement tous les autres auteurs mais pas toujours les mêmes selon les "features", etc., etc.

En effet, non seulement les classements sont aberrants mais ils varient en fonction des "features" utilisées. C'est le cas notamment pour La Fontaine dont les pièces se trouvent un peu partout, rarement à la même place et jamais ensemble !

Cette table prouve que **la méthode de MM. Cafiero et Camps est incapable de reconnaître : Boursault, Chevalier, Donneau de Visé, Gillet de la Tessonnerie, Pierre Corneille, Thomas Corneille, La Fontaine, Ouville, Quinault, Rotrou... et Molière !** Pas un n'échappe au naufrage.

Et que font ces curieux expérimentateurs ?

Ils retirent toutes les pièces qui posent problème ! A la trappe : Boursault, Chevalier, Donneau de Visé, Gillet, La Fontaine, Ouville, Quinault. Ce ne sont donc pas de vrais auteurs ? Ceux qui restent, au tour final, se trouvent amputés de la plupart de leurs œuvres. Molière en garde 9 (moins d'une sur trois).

Et... miracle ! Cela marche. Ou, du moins, on arrive enfin là où on voulait en venir depuis le début : 9 des 33 pièces assignées à Molière seraient isolées (mais comment contrôler

puisque nous n'avons aucun chiffre). Et peu importe que les autres auteurs continuent à se mélanger.

On lit dans l'article que le "corpus de contrôle" ne poserait pas de problème de ce genre... sauf Dancourt et Régnard. Dancourt (Florent Carton, comédien poète lui aussi) a présenté plus de 40 pièces et Régnard plus d'une dizaine. Ils ont dominé la comédie à la fin du XVIIe et au début du XVIIIe. Et la méthode échoue sur ce cas emblématique ! Cela ne pose aucun problème à MM. Cafiero et Camps. Ils ont toujours une hypothèse "had hoc" sous la main pour expliquer l'inexplicable : ici, la femme de l'un était actrice, devait certainement connaître l'autre et influencer les deux.

Les explications embrouillées de la page 3 et de la deuxième colonne de la page 6 sont de la même eau.

Les méthodes de MM. Cafiero et Camps sont incapables de reconnaître les auteurs des pièces du XVIIe, y compris celles attribuées à Molière qui aurait mérité de meilleurs avocats !

Nos travaux sont en ligne (Archives ouvertes du CNRS et ResearchGate).

Une présentation de la méthode d'attribution d'auteur :

<https://images.math.cnrs.fr/La-classification-des-textes.html>

Le système du comédien poète :

<https://www.researchgate.net/publication/293334890>

<https://www.researchgate.net/publication/333236909>

Nos corpus sont en ligne sur le site du Centre de Linguistique de Corpus (Université de Neuchâtel) : <http://www.unine.ch/clc/home/Corpus.html>